2012-07-12

Stoneflies of Unusual Size: Population Genetics and Systematics Within Pteronarcyidae (Plecoptera)

John S. Sproul
Brigham Young University - Provo

Stoneflies of Unusual Size: Population Genetics and Systematics Within Pteronarcyidae (Plecoptera)

John S. Sproul

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Dennis K. Shiozawa, Chair
C. Riley Nelson
R. Paul Evans

Department of Biology

Brigham Young University

August 2012

ABSTRACT

Stoneflies of Unusual Size: Population Genetics and Systematics
Within Pteronarcyidae (Plecoptera)

John S. Sproul
Department of Biology, BYU
Master of Science

Chapter 1. The family Pteronarcyidae (Plecoptera) is a highly studied group of stoneflies and very important to a wide variety of aquatic studies. Several phylogenies have been proposed for this group recent decades, however there is little congruence between the various topologies. The present study revises the phylogeny of the group by combining molecular data from mitochondrial cytochrome oxidase subunit II, ribosomal subunit 12S, ribosomal subunit 16S, and nuclear loci ribosomal subinit 18S and Histone H3, with published morphological data in a parsimony-based total evidence analysis. The analysis produced a well-supported phylogeny with novel relationships within the genus *Pteronarcys*. Maximum Likelihood and Bayesian analyses produced topologies congruent with parsimony analysis. Character mapping revealed several homoplasious morphological characters that were previously thought to be homologous.

Chapter 2. Phylogeographic studies in aquatic insects provide valuable insights into mechanisms that shape the genetic structure of aquatic communities. Yet studies that include broad geographic areas are uncommon for this group. We conducted a broad scale phylogeographic analysis of *P. badia* across western North America. In order to allow us to generate a larger mitochondrial data set, we used 454 seqeuncing to reconstruct the complete mitochondrial genome in the early stages of the project. Our analysis reveals what appears to be a complex history of isolation and multiple invasions among some lineages. The study provides evidence of multiple glacial refugia and suggests that historical climactic isolations have been important mechanisms in determining genetic structure of insects in western North America. Our ability to generate a large mitochondrial data set through mitochondrial genome reconstruction greatly improved nodal support of our mitochondrial gene tree, and allowed us to make stronger inference of relationships between lineages and timing of divergence events.

Keywords: insect systematics, total evidence analysis (TEA), Plecoptera, next-generation sequencing, phylogeography, Last Glacial Maximum (LGM)

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER 1

INTRODUCTION

Pteronarcyidae is a family of stoneflies (Plecoptera) with twelve described species that group into two genera: *Pteronarcys* (10 species), and *Pteronarcella* (2 species). Ten of the twelve species in the family (*Pteronarcella badia, P. regularis*, *Pteronarcys californica, P. princeps, P. dorsata, P. proteus, P. scotti, P. pictetii, P. biloba,* and *P. comstocki*) are distributed in North America, while two species (*P. sachalina* and *P. reticulata*) occur only in eastern Asia. Nymphs generally live in cold, fast-flowing rivers and emerge from late spring to mid-summer (Dewalt and Stewart, 1995). *Pteronarcys* are the largest of all stoneflies with females of some species reaching lengths over 7 cm, while *Pteronarcella* are much smaller in size, generally less than 2 cm (Branham and Hathaway, 1975).

Pteronarcyids have been the highly studied in the past century. This is due in part to their having many features that make them a convenient organism of study: they are large, easily identifiable, readily collected throughout the year (due to their mero-voltine life cycle), and where they occur, they are often abundant. They are featured in investigations on toxicity in streams (Anderson and Shubat, 1984; Clubb et al., 1975; Elder and Lords, 1974), ecology (Freilich, 1991; McDiffett, 1970; Moore and Williams, 1990; Schultheis et al., 2008), life history (Dewalt and Stewart, 1995; Holdsworth, 1941; Miller, 1939, 1940), river regulation (Angradi, 1993; Pesacreta, 1997), and biogeography and evolution (Kauwe et al., 2004; Nelson and Hanson, 1968; Stewart and Beckenbach, 2006; Stewart et al., 1988).

With only twelve species, this family also represents a very tractable group for phylogenetic studies. Investigators have presented at least five different phylogenies in the last

four decades to describe the relationships within the family. Nelson and Hanson (1971) presented an early phylogeny with *Pteronarcella* and *Pteronarcys* both as monophyletic clades (Figure 1). Subsequently, Zwick (1973, 1980) proposed an alternative phylogeny that divided the family into three major groups: *Pteronarcella, Pteronarcys,* and *Allonarcys*. Stark and Szczytko (1982) added egg morphology to some previously used morphological characters to revise Zwick (1973), and they, like Nelson and Hanson (1971), found two monophyletic genera, and therefore created a synonym of *Allonarcys* with *Pteronarcys*. While topology presented by Stark and Szczytko (1982) was congruent with Nelson and Hanson (1971) at the genus level, relationships within *Pteronarcys* were not congruent. Stark and Szczytko (1982) only included some of the characters used by previous authors. They also did not include the two Asian species in their reconstruction.

Nelson (1988) attempted to synthesize data from many previous studies in a single analysis. He combined characters used by Zwick (1973) and Nelson and Hanson (1971), with the egg morphology of Stark and Szczytko (1982), as well as the egg morphology of the Asian species (Table 2). This analysis produced two trees (through PENNY and CLIQUE methods) that closely resembled the phylogeny of Nelson and Hanson (1971), however the relationships between *Pteronarcy biloba, P. scotti, P. comstocki,* and *P. proteus* all differed slightly in each tree.

Finally, Wright and White (1992) used early molecular techniques (allozyme data) to address the relationships within the genus *Pteronarcys*. Their study included eight of the ten taxa in the genus. Different methods of tree reconstruction yielded very different relationships; some analyses agreed with certain existing morphology based phylogenies, while others did not. They concluded that the age of the family makes phylogeny reconstruction difficult. In summary,

despite being so well studied even in the systematic literature, a well-supported phylogeny remains lacking in this group.

The present study generates a robust phylogeny of Pteronarcyidae by combining the morphological character data from Nelson (1988) with DNA sequence data from all twelve species in the family. Despite Nelson (1988) including more character data than any other analysis to date, many nodes were weakly supported. By adding molecular data to the analysis, and employing newer, more effective tree searching algorithms, this study addresses two primary questiaons. First, what are the species-level sister relationships within *Pteronarcys?* Previous studies showed very little congruence with respect to these relationships. Second, in light of the new topology, do phylogeographic patterns remain consistent with those suggested by Nelson (1988)? Within the family, both *Pteronarcella* species, and two *Pteronarcys* species (*P. californica, P. princeps*) are confined to western North America in their distribution, *P. pictetii,* and *P. dorsata* occurs in the Midwest, five *Pteronarcys* species (*P. biloba, P. scotti, P. comstocki, P. scotti* and *P. proteus*) occur only in eastern North America, and the other two *Pteronarcys* species occur in eastern Asia. Finally, *P. dorsata* is the only species with confirmed collection localities spanning across all of North America.

METHODS

Sampling and tissue preparation

Tissues from 27 pteronarcyid individuals (22 adults and 5 nymphs) and eight outgroup taxa were obtained from the arthropod collection at the Bean Life Science Museum at Brigham Young University, and also through personal collecting efforts. All twelve species in the family are represented in this study. For nine of the twelve species, multiple individuals (n=2-4) from

geographically distinct populations were included. For the Asian species (*Pteronarcys sachalina*, and *P. reticulata*) and one North American species (*Pteronarcys dorsata*), fDNA-grade tissue was only available for single individuals. Within Plecoptera, the Pteronarcyidae fall within the infraorder Systellognatha (Zwick, 2000). Outgroup taxa were sampled from four of the five other families within Systellognatha including Peltoperlidae, Chloroperlidae, Perlidae, and Perlodidae. Inasmuch as molecular analysis of the order Plecoptera places Peltoperlidae as the family immediately basal to Pteronarcyidae with 100% bootstrap support (Terry and Whiting unpublished data), Peltoperlidae was used to root the topology in the present study. All specimens are vouchered in the Insect Genomics Collection and Dennis Shiozawa's personal collection at Brigham Young University. DNA was dissected from leg or thoracic muscle tissue and extracted using the Qiagen® DNeasy™ protocol.

PCR amplification and sequencing

Five genes were targeted for amplification via Polymerase Chain Reaction (PCR), mitochondrial ribosomal 12S and 16S, nuclear ribosomal 18S, mitochondrial protein coding Cytochrome Oxidase II, and nuclear protein coding H3. All targets were amplified using previously published primers (Colgan et al., 1998; Whiting, 2002; Whiting et al., 1997). An initial comparison of these genes for three pteronarcyids suggested that they contained sufficient variability (~1.5-7% sequence divergence between closely related species depending on the locus) to be informative in a genus/species level study and therefore were deemed suitable markers for the present study. PCR was preformed with a Peltier PTC-225 DNA Engine Tetrad Thermal Cycler (MJ Research, Inc., Waltham, MA) in 12.5 μL reactions containing 3 μL of DNA template, 2.25 μL sterile distilled water, 0.5 μL dNTP''s, 0.5 μL each primer, and 6.25 μL

GoTaq mastermix using the following thermal profile: 2 min. @ 95°C, followed by 35 cycles of 30 sec @ 95°C, 30 sec annealing @50°C, and 2 min. extension @72°C, with a final elongation step of 4 min at 72°C.

Successful amplification was verified using ultraviolet visualization following gel electrophoresis with a 1% agarose gel. The PCR product was purified using Gene Clean III DNA purification kits. (Bio101, Inc., Vista, CA). The purified DNA was sequenced in 10 μL reactions using ABI Big Dye terminator protocol (Applied Biosystems, Inc., Palo Alto, CA) and nucleotide sequences were generated by the BYU DNA Sequencing Center using an ABI 3730XL automated sequencer (Applied Biosystems).

Phylogenetic analysis

PCR and sequencing yielded an unaligned data set of approximately 3260 base pairs per taxon (12S = ~410bp, 16S =~561bp, 18S = ~1124bp, COII = ~792bp, H3 = ~379bp). Missing data included the H3 gene for one taxon and 18S for five taxa. Missing data were due to unsuccessful PCR amplification after repeated attempts, and perhaps related to the marginal quality DNA obtained from some of the older museum specimens. DNA sequences were edited using Sequencher 4.8 software (Gene Codes Corp., Ann Arbor, MI), and aligned using MAFFT (Katoh et al., 2005). MAFFT was chosen for its ability to consider secondary RNA structure (Katoh and Toh, 2008) and its use of an iterative process to quickly obtain optimal alignments. Protein coding genes were aligned with the L-INS-i algorithm with the scoring matrix set to 1PAM/k=2, gap penalty = 1.53, and offset value=0.1. For ribosomal RNA genes, the Q-INS-i algorithm (which considers secondary RNA structure) was used with the scoring matrix set to 1PAM/k=2, gap penalty = 1.53, and offset value=0.1. To test the sensitivity of the alignments to

changes in parameters, alignments for each gene were repeated using gap penalties ranging from 1.0 to 2.53, offset values ranging from 0.0 to 0.5, and the 200PAM/k=2 scoring matrix.

Aligned sequences were visualized and concatenated using Mesquite v2.74 (Maddison and Maddison, 2011) and then combined with the 75 character morphology matrix (see Table 1) from Nelson (1988) for a "total evidence analysis" (TEA) under a parsimony criterion. Morphology and molecular data sets were also analyzed separately, however, TEA was preferred over partitioned analysis because it has been shown to best recover hidden signal, and otherwise out-perform partitioned approaches (Bybee et al., 2008; Pilgrim et al., 2002; Wheeler et al., 2001). A parsimony-based approach allowed us to utilize current phylogenetic software that facilitates the combination of molecular and morphological data sets in a simultaneous analysis (Goloboff et al., 2008). Further, since we incorporated morphological data gleaned from previous studies that used parsimony-based methods (Nebeker, 1971; Nelson, 1988; Stark and Szczytko, 1982; Zwick, 1973), we could directly compare the additive effect of molecular data on the topology against previous results. Inasmuch as several studies have identified long branch attraction (LBA) as a weakness of parsimony methods (Bergsten, 2005), the TEA was re-run excluding the outgroups and sets of ingroup taxa in order to test whether LBA could be affecting relationships within the topology. The data set was also analyzed with and without 18S sequence to test whether the missing data within 18S was affecting relationships or nodal support within the topology.

Parsimony analysis was performed using TNT v1.1 (Goloboff et al., 2008) using the ratchet, tree drifting, and tree fusing searching methods for 100 tree bisections across 1000 replicates, followed by a bootstrap analysis of 2000 replicates. In addition to the bootstrap analysis (Felsenstein, 1985), nodal support was also estimated using Bremer methods (Bremer,

1994). To calculate Bremer support, an additional (congruent) topology was generated using a TBR tree searching strategy; support was estimated from ~56,000 trees saved during the TBR search. Partitioned Bremer analysis was performed using PAUP v4.0 (Swofford, 2003) and TreeRot v3.0 (Sorenson and Sorenson, 2007) with the data set divided into seven partitions, one partition for each of the five genes, and two partitions in the morphology matrix (characters related to egg morphology were partitioned separate from the remaining morphological characters). Consistency and retention indices were calculated in Mesquite v2.74 (Maddison and Maddison, 2011). Finally, the TEA parsimony topology was compared to topologies produced using Bayesian Inference (BI) and Maximum Likelihood (ML) strategies, both of which included the morphological data, to test for congruence between different methods of tree reconstruction.

Appropriate models of evolution for implementation in BI and ML approaches were determined using jModelTest v0.1.1 (Posada, 2008) under the Akaike information criterion. BI analysis was performed using MrBayes v3.2 (Huelsenbeck and Ronquist, 2001). The data were partitioned by data type and by gene and analyzed using GTR+gamma (12S and 18S), GTR+gamma+I (16S and COII), and HKY+gamma (H3) models of evolution and analyzed with a chain length of 5,000,000 generations, with subsamples every 1000 generations for three chains. The first 5% of trees were discarded as "burn in"; this quantity was determined to be adequate based on visualization of tracing in the program Geneious v5.5.5 (Drummond et al., 2010). Maximum Likelihood analysis was performed using RAxML v7.2.6 (Stamatakis, 2006). The data set was partitioned by data type and gene, and run with the GTR+gamma+I model of evolution for 100 replicates followed by 1000 bootstrap replicates using the rapid bootstrap algorithm (Stamatakis et al., 2008).

Morphological character and geographical distribution mapping

Mesquite v2.74 was used to reproduce the morphology-based topology from Nelson (1988) and the 75 morphological characters (Table 1) were mapped on the topology. The morphology-based tree was then constrained to match the topology obtained from the TEA to determine which morphological characters are useful in characterizing groups in the newly presented topology (Figure 3). We also mapped geographical location on the revised topology for comparison to patterns in previous studies. Geographical distribution information was obtained via *Stoneflies of the United States* (Kondratieff and Baumann, 2000), an online database that synthesizes published stonefly distributional data into a digital map based resource. Distributional categories were assigned as follows: Eastern Asia, western North America, mid-western North America, eastern North America, and widespread North American.

RESULTS

Sequence alignment in MAFFT produced an alignment of 3355 base pairs for phylogenetic analysis. Changing alignment parameters did not affect the alignment outcome for 12S, 16S, COII, or H3 alignments based on careful visual comparison in Mesquite. Visual comparison showed that 18S alignments were somewhat sensitive to alignment parameters; however, the use of differing 18S alignments in tree reconstruction had no affect on topology outcome. For the purpose of the parsimony analysis, we will refer to nodes that have bootstrap values >95 % as being "very strongly supported", >75% as "strongly supported", >50% as "weakly supported" and <50% as "poorly supported".

Partitioned analysis

Partitioned analysis on the morphology matrix yielded the same topology as Nelson (1988), despite using newer searching methods. Five of the eight nodes within *Pteronarcys* were poorly supported with bootstrap values less than 50% however a monophyletic clade containing *P. dorsata, P. pictetii, P. californica,* and *P. princeps* was strongly supported. The molecular partition produced a topology with a pectinate shape similar to that of Nelson (1988), however the sister branching order and sister relationships within *Pteronarcys* were substantially different from the morphology-based topology.

Total evidence analysis (all taxa included)

The parsimony TEA identified two most parsimonious trees that were combined using a strict consensus (score=2900, CI = 0.59, RI = 0.73) (Figure 4). Novel relationships were present in the TEA tree, which were not present in either the morphology or molecular partitioned analysis. Monophyly for both genera (*Pteronarcys* and *Pteronarcella*) was very strongly supported (100% bootstrap support). Within *Pteronarcys,* three major monophyletic groups were present: the "reticulata clade" (*P. reticulata* and *P. sachalina*) (100% bootstrap support), the "comstocki clade" (*P. comstocki*, *P. scotti*, *P. proteus*, and *P. biloba*) (92% bootstrap support) and the "dorsata clade" (*P. dorsata*, *P. pictetii*, *P. princeps,* and *P. californica*) (75% bootstrap support). The node separating the *dorsata clade* from the *comstocki clade* was weakly supported (60% bootstrap support). Bremer supports showed patterns consistent with bootstrap values, as nodes with less than 75% bootstrap support have the lowest Bremer values (2-6) while all nodes with strong or very strong bootstrap support values also had high Bremer support (≥12) (Figure 4). Species-level sister relationships were very strongly supported in eleven of the twelve taxa

(exception: the sister relationship between *P. scotti* and the *P. proteus*/*P. biloba* sister group). The Partitioned Bremer analysis showed that COII contributed the most support across the topology (40.5%) followed by 12S (16.8%), 16S (13.7%), H3 (11.9%), morphology characters 1-62 (Table 2) (8.8%), 18S (8.3%), and morphology characters 63-75 (Table 2)(-0.003%).

Removal of 18S from the TEA analysis only resulted in a slight overall decrease in nodal support and did not change relationships in the topology. All replicate individuals for each species grouped as expected based on the initial species identifications.

Total evidence analysis (excluding long branches)

Removal of the outgroup and long-branch *Pteronarcella* taxa (Asian species were left in as the single putative long branch) did not change relationships within *Pteronarcys*, however it heavily affected nodal support within the genus. Whereas there were two poorly supported nodes (above the species level) within *Pteronarcys* in the initial TEA (36, 54% bootstrap support respectively), after removal of the long-branch taxa, support for these nodes increased to 99% and 99% respectively.

Maximum Likelihood and Bayesian Inference

ML analysis produced a strongly supported topology with all but three nodes having bootstrap support of 100%. Nodes with lower support included the node separating the *dorsata clade* from the *comstocki clade* (84% bs), the node separating the *Pteronarcys dorsata*/*P. pictetii* sister group from the *P. princeps*/*P. californica* sister group, and the node separating *P. scotti* from the *P. proteus*/*P. biloba* sister group. With the exception of the level of nodal support, the

ML topology was entirely congruent with the Parsimony topology. The ML topology is shown in Figure 5.

Relationships in the BI topology were also congruent both to the parsimony and ML trees (posterior probabilities were greater than 0.99 for between species relationships in the ingroup).

Morphological character and geographical distribution mapping

Constraining the historical morphology-based tree (90 steps) to match the shape of the TEA topology increased the treelength by six steps. The previously non-homoplasious characters 24, 25, 41, 42, all required at least two steps to map onto the TEA topology. Characters 70 and 74 increased in homoplasy from two steps to three steps. Character 72 decreased in homoplasy from three steps to two steps. Character 50 went from being two steps, to a non-homoplasious character. Within each genus, species geographical distributions mapped onto the phylogeny support the relationships presented in this study. An eastern Asian distribution defines the *reticulata clade*, an eastern North American distribution defines the *comstocki clade*, a Midwestern North American distribution *P. pictetii* while its sister species *P. dorsata* is defined by a widespread North American distribution. A western North American distribution defines the *P. californica, P. princeps* sister group, as well as the entire genus *Pteronarcella*.

DISCUSSION

Similarities and differences among tree reconstruction methods

Since all three methods of tree reconstruction produced congruent topologies, the primary difference between methods was in nodal support. The parsimony topology had three poorly supported nodes within the *Pteronarcys* clade, whereas no nodes in the ML or BI topologies has less than 72% bootstrap support or 0.99 posterior probability. The three most poorly supported parsimony nodes correspond with the three ML nodes with the lowest support, suggesting that these nodes may be areas in the topology that are challenging to predict based on either limitations in the data set, or due to the evolutionary history of species within the *comstocki clade*. The lack of morphological characters supporting distinctions among species within the *comstocki clade* seen in previous studies (Nelson, 1988) (Figure 2) seems to suggest that this lack of nodal support across methods could reasonably be explained by a rapid radiation of species within the clade.

We also supposed that the very low support for these nodes in the parsimony topology could be an artifact of LBA if long-branched taxa were grouping together in some percentage of the bootstrap replicates, however removal of the long branches did not change relationships within *Pteronarcys*, and while it substantially increased bootstrap support for the weakly supported nodes, it is expected that bootstrap support would increase with fewer terminals in the analysis; we thus do not have strong evidence that LBA is affecting our parsimony results. Overall, congruence between topologies produced through parsimony, ML, and BI provide strong evidence supporting the relationships proposed in this study and collectively they overcome much of the ambiguity seen in nodes limited to the parsimony topology.

Congruence with previous phylogenies

Compared to previous studies, the presently proposed topology most closely resembles topologies presented by Zwick (1973, 1980) and Nelson (1988). In terms of topology shape, our tree shows subdivisions within *Pteronarcys* similar to those proposed by Zwick (1973, 1980); however our topology did not place the two Asian species within *Allonarcys*. Our topology agrees with Nelson (1988) in terms of placement of the Asian clade as basal to the rest of *Pteronarcys*, however subsequent branching order is not congruent. Our topology differed from both Nelson (1988) and Zwick (1973, 1980) in species level relationships among four species within *Pteronarcys*.

Species level relationships within Pteronarcys

Relationships between *Pteronarcys biloba, P. scotti, P. comstocki,* and *P. proteus* were inconsistent in previous topologies, likely due to a limited availability of morphological characters. The addition of molecular data to the analysis provides substantial clarification to these relationships. Regardless of tree reconstruction method, all topologies in the present study recovered the same relationship between these four species, with basal *P. comstocki* being sister to the other three species, followed by *P. scotti* being sister to the more derived *P. biloba/P. proteus* sister group (although support for these relationships was poor in the parsimony analysis, it was strong or very strong in the ML and BI topologies). This relationship among the eastern distributed taxa is incongruent with the previously proposed phylogenies. For example, no historical phylogeny suggests a sister relationship between *P. proteus* and *P. biloba*, however, this relationship has very high nodal support (100% bs) regardless of tree reconstruction method in the present study.

Previous phylogenies all agree in the placement of *Pteronarcella* as the basal group in the family, as well as the sister relationships between *Pteronarcys reticulata*, and *P. sachalina*, *P. californica* and *P. princeps,* and *P. dorsata* and *P. pictetii*. Our topology supports all of these relationships, while clarifying how sister groups relate to one another across deeper nodes. The mapping of geographical distribution on the phylogeny suggests that geography has been important to speciation across the family as unique distributions define monophyletic clades across the topology (Figure 3); these patterns offers further evidence to the validity of the relationships presented in the current topology.

Validity of the historical genus Allonarcys

Zwick (1973) proposed a division of the genus *Pteronarcys* into two monophyletic genera, *Pteronarcys* and *Allonarcys*. This distinction was later reversed (Stark and Szczytko, 1982). The *dorsata clade* recovered by the present study is congruent with Zwick's *Pteronarcys*, and the *comstocki clade* is comprised of the same species as Zwick's *Allonarcys* except for the absence of the Asian species in the present topology (although the remaining species level relationships within *Allonarcys* differ from those in the *comstocki clade*). Thus, excluding the Asian species, our study suggests that the species that comprised *Allonarcys*, *P. comstocki, P. scotti, P. proteus,* and *P. biloba*, do form a monophyletic clade within the genus. It is likely that this monophyletic clade was not recovered in revisions prior to, and following Zwick's study due to the use of several homoplasious characters that were not identified as such. This explanation is supported in the findings in mapping morphological characters on the TEA topology. In the previously accepted phylogeny of the family, there are a total of four non-homoplasious

characters (24, 25, 41, 42; Table 2)that support the branches which break up the monophyly of *P. comstocki, P. scotti, P. proteus,* and *P. biloba* (Figure 2). All four of these characters became homoplasious when mapped on the TEA topology - they were the only previously non-homoplasious characters to do so. Our findings also suggest that the use of egg morphology-based characters (characters 63-75; Table 2) may have been problematic to previous studies as we found eight of the thirteen egg morphology characters requiring two or more steps to map onto our phylogeny. The lack of phylogenetic signal in egg morphology for this group is further shown by the partitioned Bremer analysis, which showed that the egg morphology partition contributed slightly negative support overall to the topology. While some studies did not include egg morphology characters in the analysis, they comprised as much as 25% if the character matrix in other studies. It is likely that the presence or absence of these characters contributed significantly to the lack of congruence between previous topologies.

Although we found substantial substructure within the presently accepted genus of *Pteronarcys,* revision of the genera for such a small group would be of limited value to taxonomists. Thus, we present only new "group" distinctions within *Pteronarcys* as seen below. Morphological characters supporting each group are mapped on Figure 3.

Group classification

Genus: *Pteronarcys* Newman

Group reticulata

P. reticulata Burmeister

P. sachalina Klapalek

Group dorsata

P. dorsata Say (type)

P. pictetii Hagen

P. californica Newport

P . princeps Banks

Group comstocki

P. comstocki Smith

P. scotti Ricker

P. biloba Newman

P. proteus Newman

CONCLUSIONS

The addition of molecular data to existing morphological data produced a robust phylogeny of Pteronarcyidae. The congruence of topologies across multiple methods of tree reconstruction, high nodal support, and geographically meaningful clades all suggest that the present study is valuable in clarifying the long-standing controversial relationships within *Pteronarcys*. A well-supported phylogeny of the group can prove to be a valuable tool for future studies, including those that explore questions related to cryptic speciation and character evolution in aquatic insects.

REFERENCES

Anderson, R.L., Shubat, P., 1984. Toxicity of flucythrinate to Gammarus lacustris (Amphipoda), Pteronarcys dorsata (Plecoptera) and Brachycentrus americanus

(Trichoptera): Importance of exposure duration. Environmental Pollution Series A, Ecological and Biological 35, 353-365.

Angradi, T.R., 1993. Stable carbon and nitrogen isotope analysis of seston in a regulated rocky mountain river, USA. Regul. Rivers-Res. Manage. 8, 251-270.

Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163-193.

Branham, J.M., Hathaway, R.R., 1975. Sexual differences in growth of Pteronarcys-californica (Newport) and Pteronarcella-badia (Hagen) (Plecoptera). Can. J. Zool.-Rev. Can. Zool. 53, 501-506.

Bremer, K., 1994. Branch support and tree stability. Cladistics-Int. J. Willi Hennig Soc. 10, 295-304.

Bybee, S.M., Ogden, T.H., Branham, M.A., Whiting, M.F., 2008. Molecules, morphology and fossils: a comprehensive approach to odonate phylogeny and the evolution of the odonate wing. Cladistics 24, 477-514.

Clubb, R.W., Lords, J.L., Gaufin, A.R., 1975. Isolation and characterization of a glycoprotein from the stonefly, Pteronarcys californica, which binds cadmium. J. Insect Physiol. 21, 53-60.

Colgan, D.J., McLauchlan, A., Wilson, G.D.F., Livingston, S.P., Edgecombe, G.D., Macaranas, J., Cassis, G., Gray, M.R., 1998. Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. Aust. J. Zool. 46, 419-437.

Dewalt, R.E., Stewart, K.W., 1995. Life-histories of stoneflies (Plecoptera) in the Rio Conejos of southern Colorado. Gt. Basin Nat. 55, 1-18.

Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., Wilson, A., 2010. Geneious v5.1, Available from http://www.geneious.com.

Elder, J.A., Lords, J.L., 1974. Effects of selected mercurials on glyceraldehyde-3-phosphate-dehydrogenase in Pteronarcys-californica. J. Insect Physiol. 20, 41-48.

Felsenstein, J., 1985. CONFIDENCE-LIMITS ON PHYLOGENIES - AN APPROACH USING THE BOOTSTRAP. Evolution 39, 783-791.

Freilich, J.E., 1991. Movement patterns and ecology of Pteroarcys nymphs (Plecoptera) - Observations of marked individuals in a Rocky Mountain stream. Freshwater Biology 25, 379-394.

Goloboff, P.A., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. Cladistics 24, 774-786.

Holdsworth, R.P., 1941. The life history and growth of *Pteronarcys proteus* Newman (Pteronarcidae: Plecoptera). Annals of the Entomological Society of America 34, 495-502.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754-755.

Katoh, K., Kuma, K.-i., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Research 33, 511-518.

Katoh, K., Toh, H., 2008. Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. BMC Bioinformatics 9.

Kauwe, J.S.K., Shiozawa, D.K., Evans, R.P., 2004. Phylogeographic and nested clade analysis of the stonefly Pteronarcys californica (Plecoptera : Pteronarcyidae) in the western USA. Journal of the North American Benthological Society 23, 824-838.

Kondratieff, B.C., Baumann, R.W., 2000. Stoneflies of the United States. Jamestown, ND: Northern Prarie Wildlife Research Center Online. http://npwrc.usgs.gov/resource/distr/insects/sfly/index.htm (Version 12DEC2003).

Maddison, W.P., Maddison, D.R., 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75 http://mesquiteproject.org.

McDiffett, W.F., 1970. The Transformation of Energy by a Stream Detritivore, *Pteronarcys Scotti* (Plecoptera). Ecology 51, 975-988.

Miller, A., 1939. The egg and early development of the stonefly, *Pteronarcys proteus* Newman (Plecoptera). Journal of Morphology 64, 555-609.

Miller, A., 1940. Embryonic membranes, yolk cells, and morphogenesis of the stonefly Pteronarcys proteus Newman (Plecoptera: Pteronarcidae). Annals of the Entomological Society of America 33, 437-477.

Moore, K.A., Williams, D.D., 1990. Novel Strategies in the Complex Defense Repertoire of a Stonefly (Pteronarcys dorsata) nymph. Oikos 57, 49-56.

Nebeker, A.V., 1971. Effect of Water Temperature on Nymphal Feeding Rate, Emergence, and Adult Longevity of the Stonefly Pteronarcys dorsata. Journal of the Kansas Entomological Society 44, 21-26.

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.

Nelson, C.H., Hanson, J.F., 1968. The External Anatomy of Pteronarcys (Allonarcys) proteus

Newman and Pteronarcys (Allonarcys) biloba Newman (Plecoptera: Pteronarcidae). Transactions of the American Entomological Society (1890-) 94, 429-472.

Pesacreta, G.J., 1997. Response of the stonefly Pteronarcys dorsata in enclosures from an urban North Carolina stream. Bull. Environ. Contam. Toxicol. 59, 948-955.

Pilgrim, E.M., Roush, S.A., Krane, D.E., 2002. Combining DNA sequences and morphology in systematics: testing the validity of the dragonfly species Cordulegaster bilineata. Heredity 89, 184-190.

Posada, D., 2008. jModelTest: Phylogenetic Model Averaging. Mol. Biol. Evol. 25, 1253-1256.

Schultheis, A.S., Booth, J.Y., Vinson, M.R., Miller, M.P., 2008. Genetic evidence for cohort splitting in the merovoltine stonefly Pteronarcys californica (Newport) in Blacksmith Fork, Utah. Aquat. Insects 30, 187-195.

Sorenson, M.D., Sorenson, E.A., 2007. TreeRot, version 3. Boston University, Boston, MA.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688-2690.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers. Syst. Biol. 57, 758-771.

Stark, B.P., Szczytko, S.W., 1982. Egg Morphology and Phylogeny in Pteronarcyidae (Plecoptera). Annals of the Entomological Society of America 75, 519-529.

Stewart, J.B., Beckenbach, A.T., 2006. Insect mitochondrial genomics 2: the complete mitochondrial genome sequence of a giant stonefly, Pteronarcys princeps, asymmetric directional mutation bias, and conserved plecopteran A+T-region elements. Genome 49, 815-824.

Stewart, K.W., Szczytko, S.W., Maketon, M., 1988. Drumming as a Behavioral Line of Evidence for Delineating Species in the Genera Isoperla, Pteronarcys, and Taeniopteryx (Plecoptera). Annals of the Entomological Society of America 81, 689-699.

Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Wheeler, W.C., Whiting, M., Wheeler, Q.D., Carpenter, J.M., 2001. The phylogeny of the extant hexapod orders. Cladistics 17, 113-169.

Whiting, M.F., 2002. Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. Zoologica Scripta 31, 93-104.

Whiting, M.F., Carpenter, J.C., Wheeler, Q.D., Wheeler, W.C., 1997. The strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst. Biol. 46, 1-68.

Wright, M., White, M.M., 1992. Biochemical systematics of the North American *Pteronarcys* (Pteronarcyidae, Plecoptera) Biochem. Syst. Ecol. 20, 515-521.

Zwick, P., 1973. Insecta: Plecoptera; phylogenetisches System und Katalog. W. de Gruyter, Berlin; New York.

Zwick, P., 2000. Phylogenetic system and zoogeography of the Plecoptera. Annu. Rev. Entomol.

Table 1. Summary of collection information for each individual used in the study. Outgroup taxa are indicated by an asterisk, taxa used to root the topology are indicated by two asterisk.

Species	Collection Location	Geographic Division	ST/Prov	Country	Date
Pteronarcella badia	Red River	Taos	NM	USA	06/05/01
P. badia	Red River	Taos	NM	USA	06/06/01
P. badia	Blodgett Creek	Ravalli	MT	USA	07/16/11
P. badia	Kwethluk River	Bethel	AK	USA	06/25/11
P. regularis	Big Springs, Mt. Shasta City Park	Siskiyou	CA	USA	06/22/85
P. regularis	Whiskey Springs, Whiskey Springs Campground	Jackson	OR	USA	06/09/04
Pteronarcys californica	Yellowstone River	Park	MT	USA	07/15/11
P. californica	Pecos River	San Miguel	NM	USA	06/04/11
P. princeps	Big Springs, Mt. Shasta City Park	Siskiyou	CA	USA	05/20/98
P. princeps	Big Springs, Mt. Shasta City Park	Siskiyou	CA	USA	06/22/85
P. princeps	Wahkeena Creek	Multnomah	OR	USA	05/10/87
P. dorsata	Upper Three Runs (lab reared)	Aiken	SC	USA	03/10/84
P. proteus	Laurison River	Beauce	QC	Canada	06/12/93
P. proteus	Johnson Hollow Brook	Delaware	NY	USA	05/07/91
P. pictetii	Platte River	Hall	NE	USA	04/17/96
P. pictetii	Middle Loup River	Thomas	NE	USA	04/09/96
P. comstocki	Nepisiguit River	Gloucester	NB	Canada	06/17/93
P. comstocki	Kennebec River	Somerset	ME	USA	06/11/93
P. biloba	Big Eskedelloc River	Northumberland	NB	Canada	06/17/93
P. biloba	Mill River	Prince	PE	Canada	06/19/93
P. reticulata	Gol River	Teshig	BG	Mongolia	07/11/05
P. scotti	Howard Creek (lab reared)	Oconee	SC	USA	04/07/82
P. scotti	Davidson River	Transylvania	NC	USA	07/05/02
P. scotti	~	McDowell	NC	USA	~
P. sachalina	~	Primorskii Krai	FE	Russia	~
**Chloroperla tripunctata*	Fulda River	~	HE	Germany	~
**Neaviperla forcipata*	Skagway River	Skagway	AK	USA	~
**Claassenia sabulosa*	~	~	~	~	~
**Hesperoperla pacifica*	Provo River	Utah	UT	USA	~
**Isogenoides columbrinus*	Ross River	~	YK	Canada	~
**Isoperla davisi*	Big Juniper Creek	Santa Rosa	FL	USA	~
***Soliperla campanula*	~	Multnomah	OR	USA	~
***Yoraperla nigrisoma*	~	Jackson	OR	USA	~

Table 2. Contains descriptions of the 75 morphological characters used in the analysis. Intended to be congruent with the table of morphological characters in Nelson (1988).

	Character	Plesio-morphic (0)	Apo-morphic (1)
1	Adult body size large	No	Yes
2	Frontal lateral margins depressed ventrally	Yes	No
3	Meso- and metathoracic transcostal suture present	No	Yes
4	Meso- and metascutellar pits with well-developed apodemes	Yes	No
5	Metanotal posterior notal wing process partially or completely detached from metanotum	No	Yes
6	Metanotal posterior notal wing process completely detached from mesonotum	No	Yes
7	Male abdominal segments 2-8 with dorsolateral margins upraised into rounded protuberances	No	Yes
8	Female sternite II with darkly colored medial longitudinal region	No	Yes
9	Female sternite II posterior margin with median square-shaped depression bearing two posterior projections	No	Yes
10	Male sternite II posterior margin with medial notch	No	Yes
11	Female sternite II posterior margin bearing anterior pointed setae capable of being unfolded	No	Yes
12	Female sternite II 'vaginal valves' separated a wide notch	Yes	No
13	Male tergite 9 with anteriorly projecting tergal process arising from posterior tergal area	No	Yes
14	Male tergite 9 with anteriorly projecting tergal process arising from anterior tergal area	No	Yes
15	Male tergite 9 divided medially by broad longitudinal membranous band	No	Yes
16	Hammer present	Yes	No
17	Male sternite 9 posterior border barry produced	No	Yes
18	Male sternite 9 posterior border notched medially and greatly produced	No	Yes
19	Male tergite 9 with spinule patch covering entire area	No	Yes
20	Male tergite 9 with spinule patch restricted to anteromesal area	No	Yes
21	Male sternite 9 densely clothed with setae	Yes	No
22	Male hemitergite 10 bearing 3 transverse lobes	No	Yes
23	Median hemitergal lobes well-differentiated from anterior lobes	No	Yes
24	Median hemitergal lobes slightly expanded and rounded	No	Yes
25	Median hemitergal lobes moderately expanded and rounded	No	Yes
39	Epiproct lateral braces reduced or absent	No	Yes
40	Epiproct lateral braces absent	No	Yes
41	Epiproct inner part 3/4 length of dorsal section or greater	No	Yes
42	Epiproct inner part equal in length to dorsal section or greater	No	Yes
43	Epiproct inner part approximately twice length of dorsal section	No	Yes
44	Muscle c attached to apex of inner part of epiproct	No	Yes
45	Muscle c strongly developed	No	Yes
46	Clavi with fork-shaped structure present in inner membranous surface	No	Yes
47	Clavi with lateral arms of basal sclerotic plate greatly expanded	No	Yes
48	Subanal lobes boot-shaped and moderately or well-developed	No	Yes
49	Subanal lobes entirely membranous apically	No	Yes
50	Subanal lobes bootshaped and well-developed	No	Yes
51	Subanal lobes boot-shaped, well-developed, and sharply pointed anteriorly	No	Yes
52	Seminal receptacle and receptacular duct slender; no distinction between duct and receptacle	No	Yes
53	Seminal vesicles and testes remain separate; not fused to form anchor	No	Yes
54	Gene ASC1 absent	No	Yes
55	GlnRT2-3 absent	No	Yes
56	Gene A3 present	No	Yes
57	Larval prothoracic angles produced	No	Yes
58	Larval prothoracic and wing margins fringed by small setae not separated by space	No	Yes
59	Larval paired lateral abdominal spines present	No	Yes
60	Larvae with distinct subocular and subantennal setal fringe	No	Yes
61	Larvae with accessory intersegmental trachea connecting to gill present	No	Yes
62	Larval mesothoracosternum with narrow 'Y' arms	No	Yes
63	Egg aeropyles on slender arms	No	Yes

Table 2. Cont.

Character	Plesio-morphic (0)	Apo-morphic (1)
26 Median hemitergal lobes greatly expanded, extending dorsolaterally from surface	No	Yes
27 Median hemitergal lobes greatly expanded, extending posteriorly from surface	No	Yes
28 Posterior hemitergal lobes weakly demarked	No	Yes
29 Male sternite 10 ventrally complete	Yes	No
30 Epiproct free part with ventral section greatly reduced and not curved dorsally	Yes	No
31 Epiproct free part with ventral section longer than dorsal section	No	Yes
32 Epiproct free part with ventral section grooved ventrally	No	Yes
33 Epiproct free part with superior surface of dorsal section folded into two 'arms'	No	Yes
34 Epiproct free part with dorsal section superior surface 'arms' strongly projecting from surface	No	Yes
35 Epiproct free part with apex of dorsal section modified into 'cup'	No	Yes
36 Epiproct free part with apex of dorsal section with eversible sac in 'cup'	No	Yes
37 Epiproct free part with apex of dorsal section exhibiting 'spatula'	No	Yes
38 Epiproct lateral braces moderately developed, or reduced, or absent	No	Yes

Character	Plesio-morphic (0)	Apo-morph (1)
64 Egg micropylar orifice elevated	No	Yes
65 Eggoutline oval	No	Yes
66 Egg outline hemispherical, micropylar neck collar	No	Yes
67 Egg collar a flattened disc surrounded by a serrated border	No	Yes
68 Egg with plastrons	No	Yes
69 Egg plastrons widely spaced	No	Yes
70 Egg plastrons moderately packed	No	Yes
71 Egg plastrons each with narrow edge offset from median field	No	Yes
72 Egg plastrons rounded or triangular	No	Yes
73 Egg plastrons with surface appearing convoluted	No	Yes
74 Egg plastrons rounded	No	Yes
75 Egg plastrons individually large	No	Yes

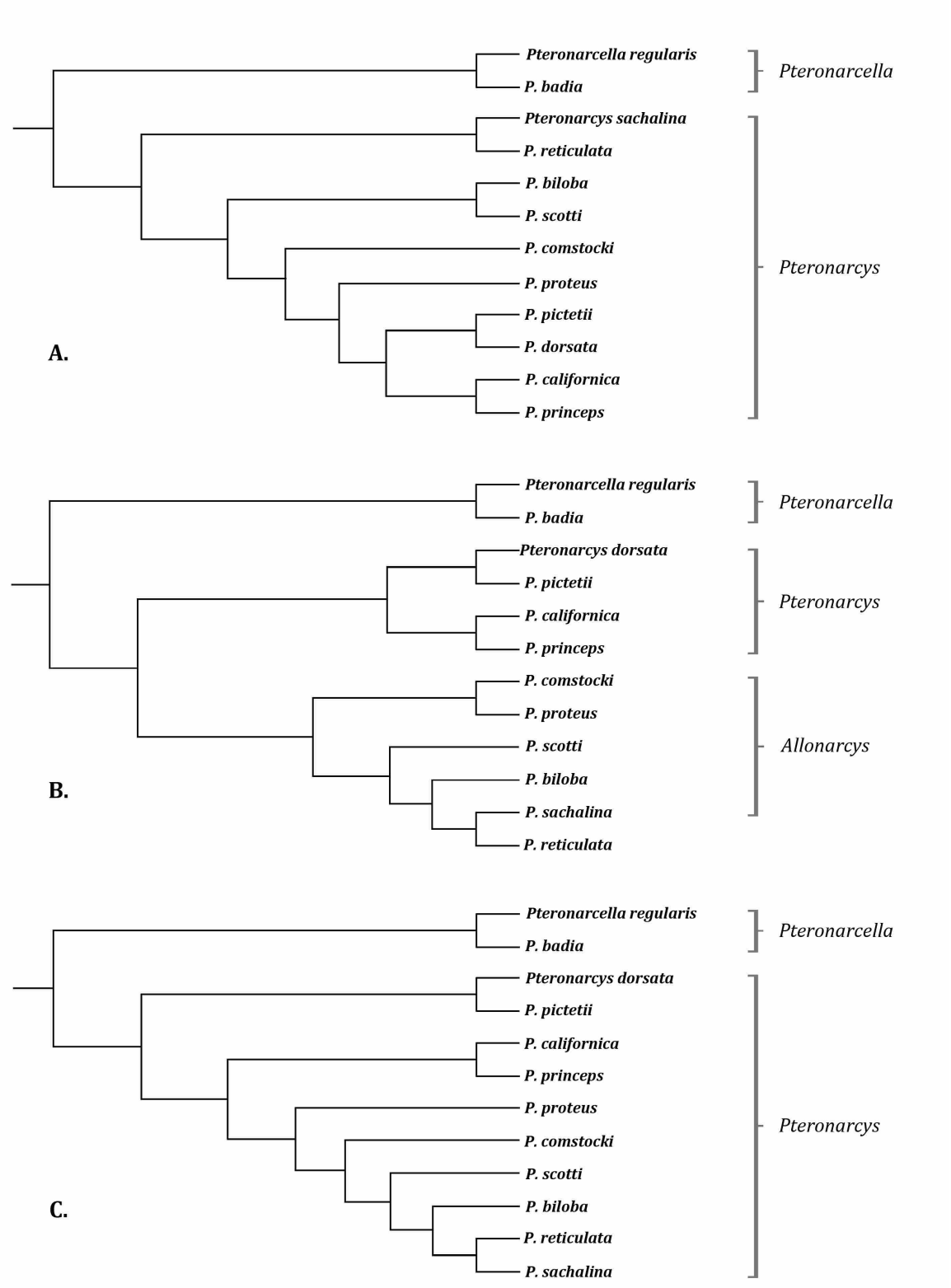

Figure 1. Three differing topologies proposed by Nelson (1971) (A), Zwick (1973) (B), and Stark (1982) (C), indicating the genera each investigator recognized in their respective studies.

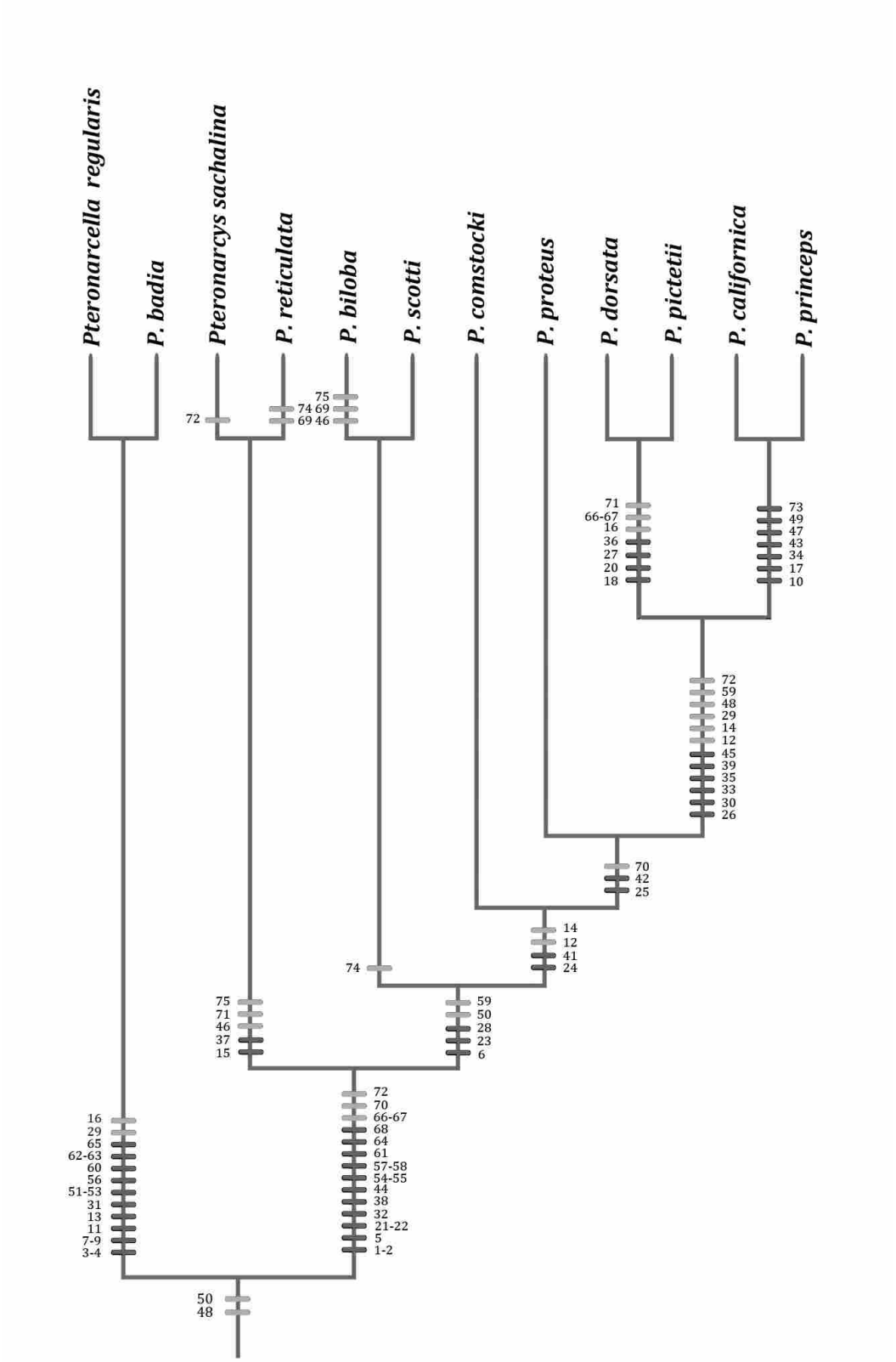

Figure 2. Intended to represent the topology presented in Nelson (1988). Tick marks on the topology represent morphological characters supporting each node. Homoplasious characters are represented in light gray, non-homoplasious characters are mapped in dark gray.

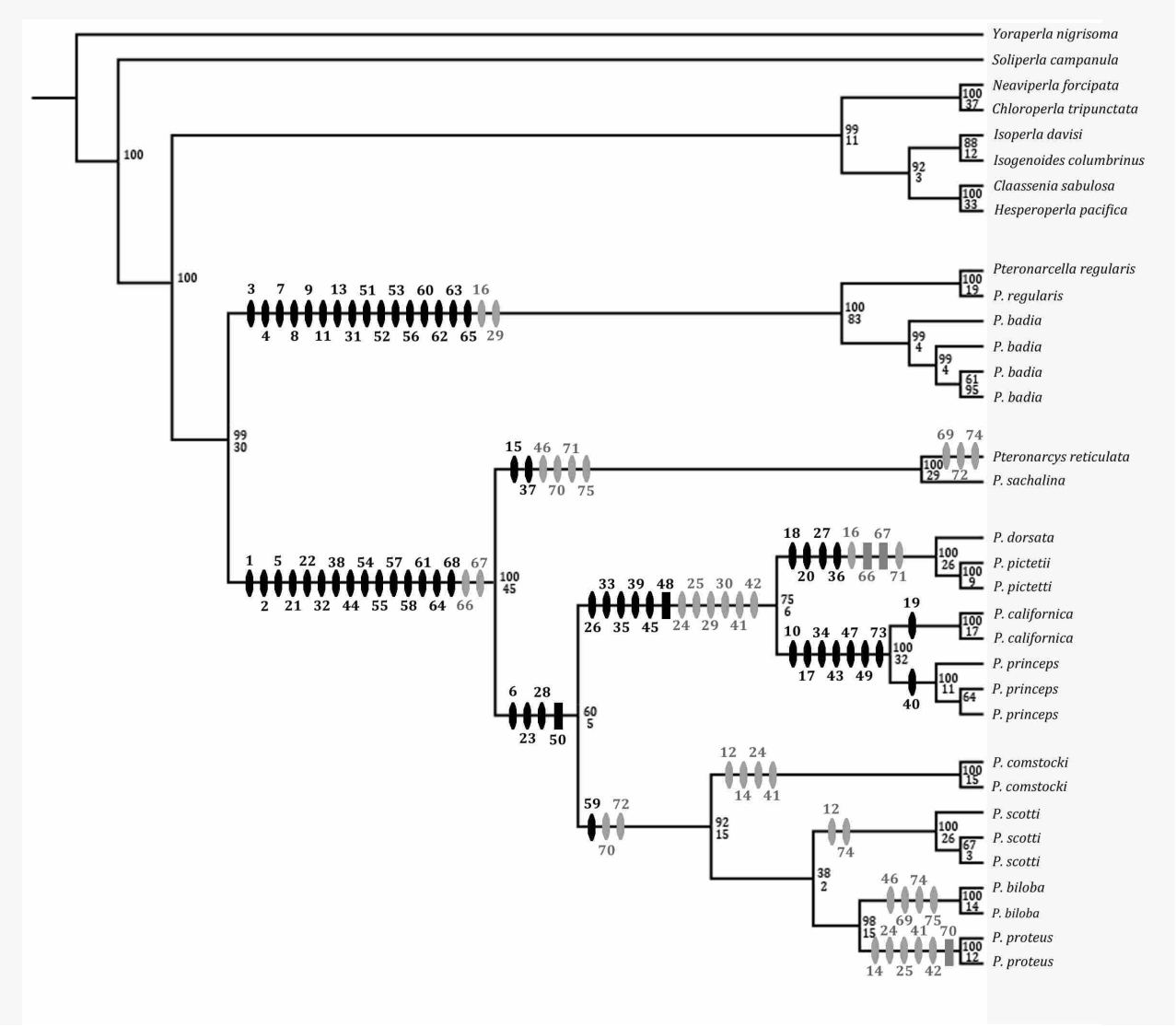

Figure 3. Toplogy resulting from our parsimony TEA including branch support and mapped morphological characters. Upper values indicate bootstrap support, and lower values indicate Bremer support. Morphological characters are mapped on branches with non-homoplasious represented by black ovals (derived character) and rectangles (lost sympleisomorphic character) and homoplasious characters represented by gray ovals and rectangles.

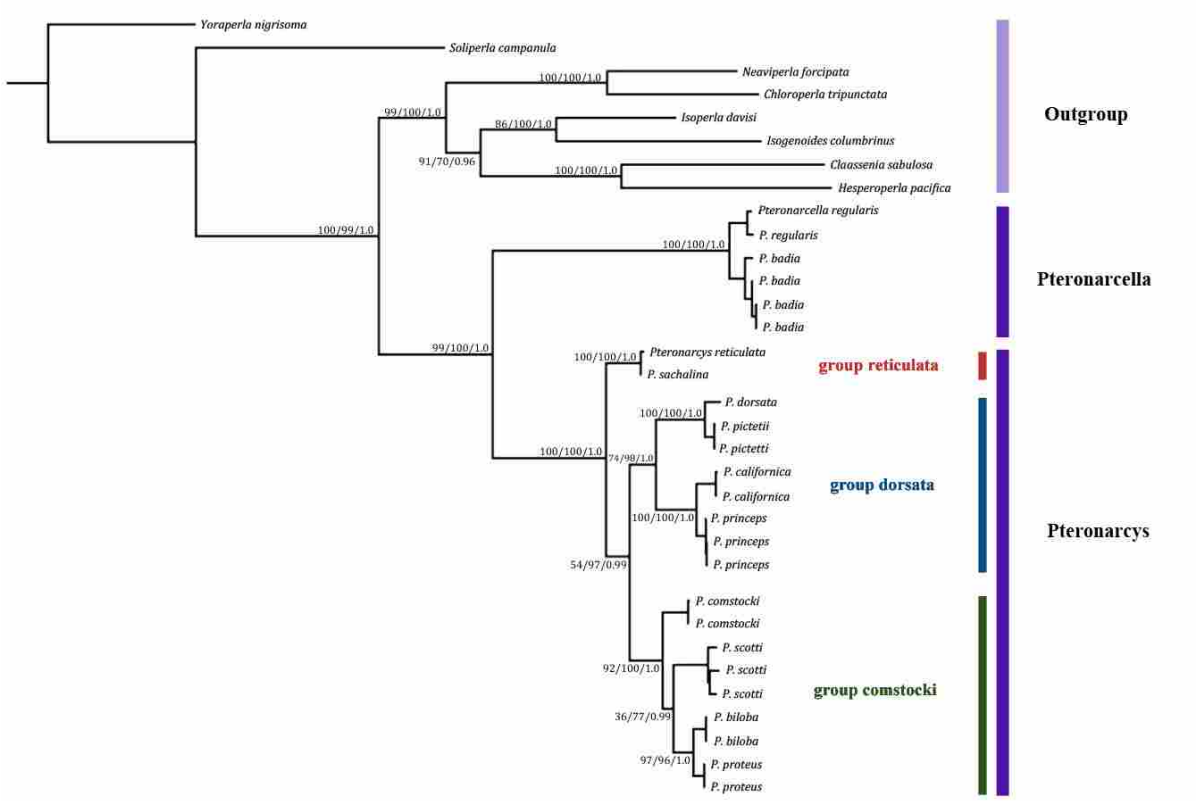

Figure 4. Our preferred tree which simultaneously represents the parsimony, ML, and BI topologies. Nodal support values represent parsimony bootstraps, ML bootstraps, and BI posterior probabilities respectively. Distinctions between genera and outgroup taxa are indicated by purple bars. Red, blue, and green bars indicated the newly revised groups within *Pteronarcys*.

Appendix 1. List of primers used in the analysis. See text for primer sources.

Gene	Primer Name	Sequence (5'-3')	Length	Direction
12S	12Sai	AAACTACGATTAGATACCCTATTAT	25	Forward
	12Sbi	AAGAGCGACGGGCGATGTGT	20	Reverse
16S	16S A	CGCCTGTTTATCAAAAACAT	20	Forward
	16S B	CTCCGGTTTGAACTCAGATCA	21	Reverse
18S	18S ai	CCTGAGAAACGGCTACCACATC	22	Forward
	18S bi	GAGTCTCGTTCGTTATCGGA	20	Reverse
COII	COII F-Leu	TCTAATATGGCAGATTAGTGC	21	Forward
	COII R-Lys	GAGACCAGTACTTGCTTTCAGTCATC	26	Reverse
H3	H3 AF	ATGGCTCGTACCAAGCAGACVGC	23	Forward
	H3 AR	ATATCCTTGGGCATGATGGTGAC	23	Reverse

Appendix 2. Accession numbers for all genes downloaded from Genbank arranged by

	Gene				
Taxon	12S	16S	18S	COII	H3
Pteronarcys scotti	EF623429.1	EF623263.1	EF622816.1	EF623112.1	EF622659.1
Pteronarcys sachalina	EF623430.1	EF623264.1	EF622817.1	EF623113.1	EF622660.1
Peltoperla arcuata	EF623424.1	EF623258.1	EF622813.1	EF623107.1	EF622656.1
Soliperla campanula	EF623422.1	EF623256.1	EF622811.1	EF623105.1	EF622654.1
Yoraperla nigrisoma	EF623423.1	EF623257.1	EF622812.1	EF623106.1	EF622655.1

CHAPTER 2

INTRODUCTION

Molecular studies in a phylogeographic context not only provide insights into the evolutionary history of the taxon of interest, but as studies across taxa accumulate, inference of broader deterministic processes is possible. Such studies are particularly valuable in understanding factors that impact complex, diverse communities as seen in freshwater aquatic systems. Insects account for much of the diversity present in freshwater communities (Dudgeon et al., 2006). With a wide array of dispersal abilities and habitat tolerance (Bilton et al., 2001), aquatic insects provide researchers a host of candidate taxa for testing phylogeographic hypotheses at many scales (Schmitt, 2009).

While aquatic insects have been reasonably well studied for many decades, relatively few molecular studies are conducted given the number of aquatic taxa. Studies that do exist are often limited in geographic scale (see review Hughes et al., 2009). Yet studies that have considered large geographic areas have given powerful insight as to the effect of historical climatic processes on genetic structure at the species and community level (Heilveil and Berlocher, 2006; Lehrian et al., 2010; Lehrian et al., 2009; Theissinger et al., 2011), as well as the importance of long-distance dispersal (Kauwe et al., 2004).

A significant obstacle to conducting molecular studies in aquatic insects is the lack of genomic information for many taxa of interest. Overwhelmingly researchers are confined to using markers available through universal or degenerate primers such as the "barcode" region of the mtDNA cytochrome oxidase I gene (Folmer et al., 1994; Simon et al., 1994). While the fast mutation rate of mtDNA makes it a desirable locus for population/species level studies, universal markers may limit the size of a mitochondrial dataset to several hundred base pairs (bp). For

phylogenetic purposes, this may not be sufficient to produce a well-supported gene tree. The ability to generate large amounts of genomic data at ever decreasing costs through next-generation sequencing approaches makes it feasible for investigators to move past constraints on genomic information in the early stages of a research project enabling them to conduct more effective molecular studies in non-model groups.

Here we present a multi-locus study on the phylogeography of *Pteronarcella badia*. This herbivorous stonefly is moderately sized, and occurs in mid-elevation mountain streams across western North America. It is one of two members of the genus *Pteronarcella* within the family Pteronarcyidae (Plecoptera). The species is readily identifiable in the field as an immature or adult (except where its range overlaps with its sister species *P. regularis* (Hagen) in the Pacific Northwest). When it is present it often occurs abundantly. It's broad western North American distribution makes it an ideal organism to study phylogeographic patterns of aquatic insects in this region, which has few related studies to date (but see (Kauwe et al., 2004; Stutz et al., 2010). Degenerate and barcode primers worked very poorly for this species in preliminary experiments with Polymerase Chain-Reaction (PCR). In order to give us flexibility in selecting a larger mitochondrial data set than the universal markers would allow, we used 454 pyrosequencing to sequence the complete mitochondrial genome (mt genome) in the early stages of the research. From the mt genome we designed new PCR primer pairs at three locations spaced across the genome. We combined sequence from these regions of the mt genome with nuclear rRNA 28S to form a dataset designed to address the following research questions: (1) *What is the population structure of P. badia?* (2) *What are the dominant modes of dispersal in P. badia?* We specifically test whether dispersal through hydrologic connectivity or overland movement (putatively through flight during the winged adult stage) appears to be more important in

determining genetic structure. (3) *How have historical climate oscillations influenced population structure?*

METHODS

Sampling and tissue preparation

We determined the approximate range of *P. badia* through species checklists (Ricker and Scudder, 1975; Stewart and Ricker, 1997) online data bases (Kondratieff and Baumann, 2000), personal communications with stonefly experts, and museum records. We collected *P. badia* nymphs and adults from 30 localities throughout its known distribution in the western United States. Where densities were sufficiently high, we collected at least 10 individuals per locality. Additional samples for eight localities in Canada and Alaska were sent to us by remote collectors or obtained from the stonefly collection at the Monte L. Bean Life Science Museum at Brigham Young University, for a total of 38 sample localities. Although the species is known to occur in both Alaska and southern regions of Canada, we were unable to find confirmed localities for the species in northern British Columbia, northern Alberta or the Yukon, indicating an apparent gap in the known species distribution.

For outgroup sampling, we obtained 10 individuals (from two localities) of the sister species *P. regularis* from both personal collecting efforts and the Monte L. Bean Life Science Museum. All samples collected through personal efforts were preserved in 100% EtOH and frozen until DNA could be extracted. Ten individuals per locality were included in the phylogenetic analysis. We determined that 10 individuals per locality would adequately sample haplotypes based on a pilot study data, which showed low haplotypic diversity within localities but high haplotypic diversity between localities. For localities where fewer than 10 individuals

were obtained, we used all available individuals for that locality. DNA was dissected from leg or thoracic muscle tissue and extracted using the Qiagen® DNeasy™ protocol. All DNA sequences were uploaded to GenBank as accession numbers XXXXXXX through XXXXXXX. Collection information is summarized in Table 1.

Mitochondrial Genome Reconstruction

Pyrosequencing, assembly, and annotation

We pooled whole genomic DNA extracted from a single *P. badia* individual with DNA from three other taxa on a 454 half plate. Library preparation and sequencing were performed on a 454 Life Sciences Genome Sequencer FLX at the Brigham Young University DNA Sequencing Center. Following pyrosequencing, reads were assembled using Newbler v.2.6 (454 Life Sciences 2006-2011). We identified mitochondrial DNA contigs using Basic Local Alignment Search Tool (BLAST), and assembled the mitochondrial contigs using the closely related (sister genus) *Pteronarcys princeps* (Hagen) mt genome (Stewart and Beckenbach, 2006) as a reference. To fill in small gaps between the *P. badia* contigs, we designed PCR primers flanking gap regions and amplified the gaps using DNA extracted from the same individual that was sequenced on the 454 run. To amplify part of the A+T rich control region not recovered in the Newbler assembly, we used Phusion High-fidelity DNA Polymerase (New England BioLabs, Ipswich, MA) on an Eppendorf Mastercycler® pro (Hamburg, Germany) in a 12.5 µL reactions containing 3 µL of DNA template, 2.25 µL sterile distilled water, 0.5 µL dNTP''s, 0.5 µL each primer, and 6.25 µL Phusion polymerase using the following thermal profile: 2 min. @ 95^{o}C, followed by 35 cycles of 30 sec @ 95^{o}C, 30 sec annealing @50^{o}C, and 2 min. extension @72^{o}C, with a final elongation step of 4 min at 72^{o}C. We used MOSAS (Sheffield et al., 2010)

to identify protein coding, ribosomal RNA, and tRNAscan-SE v1.21 (Lowe and Eddy, 1997) as implemented in MOSAS, as well as alignment to other insect mt genomes to identify t-RNA regions. We identified open reading frames and annotated the genome in Geneious v5.5.5 (Drummond et al., 2010). The complete geome is available on GenBank as accession number XXXXXXX.

Phylogenetic Analysis and Hypothesis Testing

PCR amplification and sequencing

We targeted portions of six, mitochondrial protein coding genes for PCR amplification: COI (844 bp), ATP6 (506 bp), CYTB (795 bp), COIII (711 bp), CYTB (796 bp), and ND6 (493 bp); as well as a portion of nuclear ribosomal 28S (~1,050 bp) for 275 individuals (265 ingroup, 10 outgroup) from 40 sample localities (38 ingroup, 2 outgroup) across western North America. Genes from the mitochondrial locus were amplified via PCR using primer pairs designed from the annotated mt genome which we generated through pyrosequencing. We chose primer locations flanking regions that showed moderate to high variation in an alignment between the mt genomes of *P. badia* and the closely related *P. princeps* in an attempt to select markers that would give maximum resolution to our species-level data set. We chose 28S as the second locus because it showed variation at six sites in a pilot study that considered a sub-sample of individuals from geographically distant localities. All primers used for data set generation are listed in Appendix 3.

We performed PCR with a Peltier PTC-225 DNA Engine Tetrad Thermal Cycler (MJ Research, Inc., Waltham, MA) in 12.5 μL reactions containing 3 μL of DNA template, 2.25 μL

sterile distilled water, 0.5 µL dNTP's, 0.5 µL each primer, and 6.25 µL GoTaq mastermix using the following thermal profile: 2 min. @ 95°C, followed by 35 cycles of 30 sec @ 95°C, 30 sec annealing @50°C, and 2 min. extension @72°C, with a final elongation step of 4 min at 72°C. We verified successful amplification using ultraviolet visualization following gel electrophoresis with a 1% agarose gel. We purified PCR product using Gene Clean III DNA purification kits. (Bio101, Inc., Vista, CA). The purified DNA was sequenced in 10 µL reactions using ABI Big Dye terminator protocol (Applied Biosystems, Inc., Palo Alto, CA) and sequenced at the BYU DNA Sequencing Center using an ABI 3730XL automated sequencer (Applied Biosystems). We edited DNA sequences using Geneious v5.5.5 and aligned using MAFFT v6 (Katoh et al., 2005). We aligned in MAFFT for its ability to consider secondary RNA structure (Katoh and Toh, 2008) and its use of an iterative process to quickly obtain optimal alignments. Protein coding genes were aligned with the G-INS-i algorithm with the scoring matrix set to 1PAM/k = 2, gap penalty = 1.53, and offset value = 0.25. For ribosomal RNA genes, the Q-INS-i algorithm (which considers secondary RNA structure) was used with the scoring matrix set to 1PAM/k = 2, gap penalty = 1.53, and offset value = 0.1. We translated alignments into amino acid sequences in Geneious v5.5.5 to detect unexpected stop codons, which can indicate the presence of nuclear pseudogenes (Zhang and Hewitt, 1996). We concatenated mitochondrial alignments into a single file using Mesquite v2.75 (Maddison and Maddison, 2011).

What is the population structure?

We calculated genetic diversity, and haplotype and nucleotide diversities in DnaSP v9. To estimate overall population stucture, we calculated Φ_{ST} values in Arlequin v3.5 (Excoffier, 2010). We used TCS v1.21 (Clement et al., 2000) to generate haplotype networks for both loci in

order to visualize the geographic distribution of haplotypes, and to determine if the data fit the assumption of being "tree-like".

We performed tree reconstruction using a Maximum Likelihood (ML) approach in RAxML v7.2.6 (Stamatakis, 2006). The data set was partitioned by locus, and run with the GTR+gamma+I model of evolution for 100 replicates followed by 500 bootstrap replicates using the rapid bootstrap algorithm (Stamatakis et al., 2008). We identified appropriate models of evolution for implementation in ML analysis, and all subsequent approaches with jModelTest v0.1.1 (Posada, 2008) under the Akaike information criterion.

Importance of hydrologic connectivity vs. overland dispersal

To test the importance of hydrologic connectivity in determining population structure, we performed AMOVA in Arlequin v3.5 (Excoffier, 2010) and defined our *a priori* groups according to major drainage basin as follows: Columbia, Colorado River, Great Basin, Rio Grande, Yukon, and Hudson Bay, and Kenai. We assumed for our null hypothesis that if hydrologic connectivity (as a mechanism for dispersal) is important in determining population structure, as it is with many freshwater obligates such as fish, the AMOVA should show that a large percentage of the genetic diversity is explained by differences among drainage basin (Billman et al., 2010; Houston et al., 2010; Loxterman and Keeley, 2012; Meffe and Vrijenhoek, 1988). To further test the importance of waterway connections, we performed divergence dating in BEAST v1.6.1 (Drummond and Rambaut, 2007) to determine whether divergence from the outgroup (sister species) or between lineages was coincident with known river transfer events in North America (Minckley et al., 2001). We used the relative rate test (Tajima, 1993) implemented in MEGA v5.04 (Tamura et al., 2007) to test whether our data fit the assumptions

of a molecular clock. We calibrated the clock using a mean mutation rate of 2.3% per lineage per million years, an estimated rate for arthropod mitochondrial DNA (Brower, 1994). To account for uncertainty in the actual mtDNA mutation rate, we specified a standard deviation of 10% of the mean rate. We specified a relaxed uncorrelated lognormal clock and ran two chains of 75 million generations each, sampling every 2,500 generations with a GTR+gamma+I model of evolution. Based on visualization of the tracings in the program Tracer v1.5 (Rambaut and Drummond, 2007), we discarded the first 10% of the trees as burn-in. We combined results from each run using LogCombiner v1.6.1.

To compare the importance of hydrologic connectivity as a mechanism for dispersal to overland flight, we estimated gene flow between localities within populations that can only be achieved overland.

Importance of climatic oscillations

To test the effect of past climate oscillations on historical demography we estimated changes in effective population size through time using a Bayesian skyline plot (BSP) (Drummond et al., 2005) implemented in BEAST v1.6.1 (Drummond and Rambaut, 2007). We removed outgroup taxa and ran two chains of 50 million generations under a relaxed lognormal clock prior and a mutation rate of 2.3% per lineage per million years and standard deviation of 10% of the mean. We combined multiple runs in LogCombiner v1.6.1 and visualized tracing, confirmed convergence across multiple runs, and generated the skyline plot using Tracer v1.5 (Rambaut and Drummond, 2007). To further explore the impact of known climatic events on species demography, we mapped major climate transitions (Pliocene onset, Pleistocene onset, last glacial maximum) onto the tree generated in our divergence dating analysis (described

above). As an additional test for recent patterns in population dynamics we calculated Tajima's D (Tajima, 1993) in DnaSP v5 (Librado and Rozas, 2009) and Fu's F in Arlequin v3.5 (Excoffier, 2010).

RESULTS

Mitochondrial Genome Reconstruction

Pyrosequencing produced six mtDNA contigs that represented 96.3% (15,017 of 15,586 bp) of the total genome. PCR-based sequencing of the gaps yielded the remaining 569 bps with the majority (>500 bps) of the missing sequence coming from the A+T rich control region. Nucleotide composition showed overall A+T richness that is typical in insect mt genomes (Stewart and Beckenbach, 2005) with total A+T composition = 67.4% (A = 10,119 bp, T = 9,308 bp, C = 5,786 bp, G = 3,598). Sequence analysis in MOSAS identified 36 of the 37 genes expected to be present in the mt genome. Alignment with the mt genome of *Pteronarcys princeps* confirmed the location of $tRNA^{Arg}$, the only gene unidentified by MOSAS. All protein-coding, rRNA, and tRNA genes occurred in the same relative genomic position as the "ancestral" arthropod genome based on comparison to *Drosophila yakuba* (Clary and Wolstenholme, 1985). The complete annotated sequence is available on GenBank (Benson et al., 2000), accession # XXXXXXX. An annotated visualization of the complete genome is shown in Figure 1.

Phylogenetic Analysis and Hypothesis Testing

Following alignment and trimming, our data set consisted of 2518 bp of mitochondrial and 880 bp of nuclear DNA sequence. PCR amplification for one or more genes was unsuccessful in 13 of the 265 in-group individuals included in the analysis (mostly museum

specimens with putatively degraded DNA), so they were eliminated from the population genetic analyses. Translation of mtDNA sequence into amino acids did not show any unexpected stop codons which would be indicative of contamination by nuclear pseudogenes.

What is the population strucutre?

Mitochondrial DNA sequences showed high levels of variation with 383 polymorphic sites (302 = parsimony informative, 81 = singleton) and an overall nucleotide diversity of 0.025. Of the 252 individuals sampled, 151 unique haplotypes were present with a haplotypic diversity of 0.991±0.002. TCS haplotype networks revealed very high levels of population structure with multiple sub-networks failing to connect with the connection limit set to 50 steps. High levels of population structure were also evidenced by an overall Φ_{ST} of 0.91 ($p = >0.001$). The nuclear locus showed far less variability, with only fifteen haplotypes present, and all individuals from 31 of the 38 sample localities shared the same haplotype. Thirteen of the remaining haplotypes were only present in six sample localities along the western edge of the species range from northern Nevada to southern Washington. The remaining haplotype was from a single individual in Wyoming.

Tree reconstruction in RAxML and MrBayes produced highly congruent topologies with good to excellent nodal support for all major clades (see Figure 3). Six deeply divergent clades were present within the in-group which we labeled: Old Colorado Plateau (OCP), Old Rio Grande (ORG), Western Great Basin (WGB), Pacific Northwest (PNW), Northern Rockies (NR), and Widespread (WS) (see Figures 2 and 3). Comparison of to the total evidence tree (3398 bp) to the topology produced using only COI (745 bp after trimming) showed that while COI recovered the major clades present in the total evidence topology, relationships between

clades were not identical, and three of the six clades had less than 70% bootstrap support in the COI topology. Because of the high number of inferred mutational steps between these six clades, we chose to summarize the data for visual display with phylogenetic trees and not haplotype networks.

The two basal clades, ORG and OCP, are comprised of haplotypes sampled from seven localities in the southeastern portion of the species' range (central Utah and Colorado, and northern New Mexico). Despite being highly divergent at the mitochondrial locus (up to 5.2% pairwise distance), no changes separated these clades at the nuclear locus. The PNW clade includes all individuals sampled from the western edge of the species' range, from northern California to southern Washington. Thomas Creek, CA (excluded from the full data set due to missing data) was added to this clade in Figure 3 through a secondary analysis of COI and CYTB. This clade is monophyletic with the WGB clade, which contained all individuals sampled from a single locality in northern Nevada. The PNW and WGB are the only clades supported by variation at the nuclear locus.

The NR clade includes all individuals from all localities in Montana and southern Canada. Three additional localities in Canada obtained from museum specimens (Spray River, Columbia River, and Lee Creek) were excluded from the analysis due to missing sequence data, however a separate analysis that included only sequence from COI and CYTB showed that they also fall in the NR clade. The most derived clade, the WS clade, contains haplotypes from all remaining sample localities, ranging from Alaska to New Mexico. Four sample localities (Coal Creek, Monroe Creek, Parowan Creek, and Salina Creek), all from the Colorado Plateau, contained haplotypes from two different clades. All other sample localities fell entirely within one of the six clades.

Importance of hydrologic connectivity vs. overland dispersal

AMOVA results showed that 26.15% of genetic variation was explained by differences between drainage basin (Φ_{SC} = 0.88, p = >0.001), differences between sample localities explained 64.87% of the genetic variation (Φ_{ST} = 0.91, p = >0.001), and differences within localities explained 8.98% of the variation (Φ_{CT} = 0.26, p = 0.007). Each clade, and several sub-clades, contained localities from multiple drainage basins (with the exception of NGB which only contained individuals from a single locality) providing evidence that overland dispersal across drainage basin boundaries does occur for this taxon.

Historical climate oscillations

Divergence dating suggested that *P. badia* diverged from its sister species *P. regularis* 4.95 million ybp, soon after the onset of the Pliocene Epoch. Divergence of the oldest clades within *P. badia* (OCP and ORG) dated to 2.5 million ybp, or the approximate onset of the Pleistocene, with divergence of the most derived WS clade beginning 640,000 ybp. The WS sub-clade containing all localities from Alaska is well-supported (100% bs) within the WS clade and diverged over 160,000 ybp, prior to the last two glacial maximums.

BSP detected a change in effective population size in the late Pleistocene, showing a slight decrease beginning ~100,000 ybp and continuing until ~ 60,000 ybp, followed by a rapid increase from ~60,000 ybp to present (Figure 4). Fu's F_S was highly negative and statistically significant (F_S = -29.73, p = >0.001), also indicative of a rapid demographic expansion. A negative Tajima's D value further indicated recent expansion; however it was not statistically significant (D = -0.486, p = 0.10).

DISCUSSION

Hypothesis Testing

Hydrologic connectivity vs. overland dispersal

Despite having a winged adult stage, some aquatic insects (Wishart and Hughes, 2003) have been shown to exhibit patterns of population structure consistent with a stream hierarchy model (Meffe and Vrijenhoek, 1988) where the majority of genetic variation can be explained by differences between drainage basins. This pattern is expected in aquatic insects that have limited flight ability, or that exhibit high habitat fidelity to stream corridors (Hughes et al., 2009). Our AMOVA results suggest that hydrologic connectivity is not a major determinant of genetic structure across broad geographic scales in *P. badia*.

This conclusion is scale dependent, as it is testing for dispersal across tens to hundreds of kilometers. At local levels, contemporary dispersal among networks of stream corridors likely results in high levels of connectivity, but connectivity across broad distances within drainages is not necessarily expected. This can be due in part to high order, low-elevation streams acting as barriers to the dispersal (due to temperature, gradient, dissolved O_2, etc.) of the mid-elevation distributed *P. badia*; a hypothesis that is supported by analysis of genetic structure in other other aquatic groups (Finn et al., 2007; Hughes et al., 2009). Thus, it may have been more biologically realistic to designate *a priori* groups according to sub-drainage. We found difficulty, however, in designating groups according to sub-basin in an unbiased way given our sample localities. While it is possible that grouping according to sub-drainage basin would explain a greater percentage of genetic diversity than our AMOVA as presently organized, it is clear from the geographic

distribution of the various clades (Figure 2) that dispersal events for this taxon are not confined to hydrologic connections.

The presently observed distribution of haplotypes in the WS clade suggests that this lineage achieved overland connectivity from the southern Rocky Mountains to Alaska in the relatively recent history of the species, the last 200,000 years according to our dating analysis. This pattern would only be observed if overland dispersal across drainage boundaries occurred in the group. The NR clade provides what appears to be a recent example of inter-basin transfer with very closely related haplotypes being present in both the upper Columbia (Kootenay River), and Hudson Bay drainages (Bow River, Lee Creek, and Spray River). While post-glacial colonization into the upper Columbia drainage (Kootenay River) could have proceeded from populations residing at lower elevations in the same drainage, *P. badia* in the upper Husdon Bay drainage (Bow River, Lee Creek, and Spray River) either must have colonized by way of overland dispersal in the last ~20,000 since the whole of the Hudson Bay drainage was glaciated during the LGM, or closely tracked the recession of periglacial lakes formed by the melting Laurentide ice sheet. Beyond the NR clade however, all other clades except the NGB (which consists of a single collection locality) also show closely related haplotypes that span drainage boundaries. These data suggest that, while genetic diversity is not shared exclusively along drainage basin boundaries as predicted for organisms that are extreme headwater specialists (Finn et al., 2007), dispersal across headwater reaches is still an important mechanism in shaping the population structure of *P. badia*. Finally, although overland dispersal of the species is clearly evident, the very high overall Φ_{ST} of 0.91 ($p = >0.001$) indicates that these events are rare between geographically distinct localities.

Historical climatic oscillations

Several studies have shown Pleistocene glacial cycles to be one of the main drivers in shaping genetic structure of various aquatic insect species across Europe (Kubow et al., 2010; Lehrian et al., 2010; Lehrian et al., 2009; Schmitt, 2009). Our data suggest that historical climate oscillations have been an important factor in shaping the current and past distribution of *P. badia* and may have given rise to many of the presently observed patterns of genetic structure. The presence of WS clade haplotypes in Alaska with divergence dating between 100,000 and 200,000 ybp suggests that at least intermittent connectivity existed between Alaska and the lower latitudes of the species' distribution. We hypothesize that during the most recent glacial periods, and certainly the LGM, the Cordilleran Ice Sheet pushed *P. badia* into northern (Alaskan) and southern refugia. Given that all haplotypes from the southern Canadian sample localities fall in the NR clade, it is clear that post-glacial re-colonization is expanding from refugial Montana localities and not localities having the WS clade that was previously connected with Alaskan WS subclade.

This raises the question: Is there presently connectivity between the Alaskan and southern groups? As far as our sampling can detect, no evidence of present genetic connectivity exists. The Canadian sample localities that are closest geographically to Alaska contain haplotypes of genetically distant NR clade (Figures 2 and 3). Further, we found no confirmed records of *P. badia* in northern British Columbia and the Yukon through extensive personal contacts with aquatic biologists in southeastern Alaska, British Columbia, Yukon territories, or through a careful search of Canadian stonefly checklists (Ricker and G.G.E, 1975; Stewart and Ricker, 1997). We thus conclude that it is likely that post-glacial expansion has not yet resulted in connectivity between the northern WS subclade and southern refugial groups (NR or WS clades), resulting in an actual gap in the present species distribution.

Although the presence of continental ice sheets would have caused the distribution of *P. badia* to contract northward, in general our analysis of past population demographics with Fu's F_s and BSP data show evidence of recent demographic expansion roughly coincident with the LGM. We hypothesize that the rapid increase in effective population size in the last 100,000 years shown by our BSP is driven by fluctuating climate associated with the LGM. Cooling global temperatures preceding the LGM would have likely resulted in southward expansion into warmer latitudes. In addition to a latitudinal shift, distributions would have also shifted to lower elevations within un-glaciated drainages. This could result in longer tracts of habitat connectivity and *P. badia* being able to inhabit higher order streams and rivers than would be possible during interglacial periods. Our sampling of haplotypes from the Colorado Plateau and southern Rockies indeed show patterns consistent with a recent southward range expansion. The most basal clades in our topology (OCP and ORG) occur in the southern most limits of the present species distribution. According to our divergence dating, these lineages invaded the southern Great Basin and Rio Grande drainages and began differentiating near the onset of the Pleistocene ~2.5 million ybp. However, haplotypes from the most derived WS clade are also abundant in the southern limits of the species distribution; and in three instances occur at the same localities as OCP and ORG haplotypes (Coal Creek, Parowan Creek, and Salina Creek). This suggests that recent southward shifts in the WS clade have resulted in secondary contact with the basal OCP and ORG lineages. Finally, in addition to cooling patterns allowing for southern expansion, the retreat of the continental ice sheet following the LGM has resulted in a subsequent expansion of the NR clade into formerly glaciated latitudes in Canada. Thus, both warming and cooling cycles associated with recent glacial maxima are likely drivers in the recent range expansion of the group.

Presence of cryptic species

While our mitochondrial data reveal the presence of several highly divergent clades (3-5% pairwise distance) that may be cryptic species candidates, the nuclear locus only confirms the distinctiveness of the PNW and NGB clades. This is interesting considering the OCP and ORG clades have earlier divergence dates, but do not show any variation at the nuclear locus. While this is unexpected, it is possible that the OCP and ORG have simply not yet accumulated changes at the locus we examined, but would show nuclear variation if other loci were examined. Alternatively, it may be that a gene flow event occurred sometime more recently than their divergence date that eliminated any nuclear signal that may have been accumulating. In contrast, 35 of the 36 individuals sampled from the PNW and NGB sample localities showed distinctiveness in at least one of six sites (sites here refers to nucleotide positions) in 28S, with many individuals showing substitutions at all six sites (outgroup taxa had fixed variation in at least 8 sites in 28S, all different than the variable sites in the ingroup). Since we did not detect any haplotypes from the other five clades in the PNW and NGB clades, we suspect that the lack of fixation of these six sites in 28S is due to insufficient time passage to allow the fixation rather than a result of secondary contact with individuals from other clades. The lack of haplotypes from other clades being present in the PNW also provides evidence that the PNW clade remained in isolation through most of the Pleistocene, despite the range expansions and contractions, which resulted in secondary contact between the OCP, ORG and WS clades. In addition, stonefly experts Richard Baumann and Andrew Sheldon (personal communication) note that some morphological characters in specimens of *P. badia* from this region appear ambiguous during identification. Future projects will include an in depth analysis of morphological characters from the most divergent genetic lineages identified in the present study.

CONCLUSION

Our phylogeographic analysis of *P. badia* reveals what appears to be a complex history of isolation and multiple invasions among some lineages. The study provides evidence of multiple glacial refugia and suggests that historical climactic isolations have been important mechanisms in determining genetic structure of insects in western North America. Our ability to generate a large mitochondrial data set through mitochondrial genome reconstruction greatly improved nodal support of our mitochondrial gene tree, and allowed us to make stronger inference of relationships between lineages and timing of divergence events. This highlights potential of next-generation DNA sequencing in a phylogenetic context to improve molecular data sets in understudied groups.

REFERENCES

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A., Wheeler, D.L., 2000. GenBank. Nucleic Acids Research 28, 15-18.

Billman, E.J., Lee, J.B., Young, D.O., McKell, M.D., Evans, R.P., Shiozawa, D.K., 2010. Phylogenetic divergence in a desert fish: differentiation of speckled dace within the Bonneville, Lahontan, and upper Snake River basins. Western North American Naturalist 70, 39-47.

Bilton, D.T., Freeland, J.R., Okamura, B., 2001. Dispersal in freshwater invertebrates. Annual Review of Ecology and Systematics 32, 159-181.

Brower, A.V.Z., 1994. Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution. Proceedings of the National Academy of Sciences of the United States of America 91, 6491-6495.

Clary, D.O., Wolstenholme, D.R., 1985. The mitochondrial DNA molecule of Drosophila yakuba: Nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. 22, 252-271.

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657-1659.

Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., Wilson, A., 2010. Geneious v5.1, Available from http://www.geneious.com.

Drummond, A., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology. BioMed Central Ltd, p. 214.

Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. Mol. Biol. Evol. 22, 1185-1192.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.-I., Knowler, D.J., Lévêque, C., Naiman, R.J., Prieur-Richard, A.-H., Soto, D., Stiassny, M.L.J., Sullivan, C.A., 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews 81, 163-182.

Excoffier, L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows - EXCOFFIER - 2010 - Molecular Ecology Resources - Wiley Online Library. Molecular Ecology Resources.

Finn, D.S., Blouin, M.S., Lytle, D.A., 2007. Population genetic structure reveals terrestrial affinities for a headwater stream insect. Freshwater Biology 52, 1881-1897.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3, 294-299.

Heilveil, J.S., Berlocher, S.H., 2006. Phylogeography of postglacial range expansion in Nigronia serricornis Say (Megaloptera : Corydalidae). Mol. Ecol. 15, 1627-1641.

Houston, D.D., Shiozawa, D.K., Riddle, B.R., 2010. Phylogenetic relationships of the western North American cyprinid genus Richardsonius, with an overview of phylogeographic structure. Mol. Phylogenet. Evol. 55, 259-273.

Hughes, J.M., Schmidt, D.J., Finn, D.S., 2009. Genes in Streams: Using DNA to Understand the Movement of Freshwater Fauna and Their Riverine Habitat. Bioscience 59, 573-583.

Katoh, K., Kuma, K.-i., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Research 33, 511-518.

Katoh, K., Toh, H., 2008. Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. BMC Bioinformatics 9.

Kauwe, J.S.K., Shiozawa, D.K., Evans, R.P., 2004. Phylogeographic and nested clade analysis of the stonefly Pteronarcys californica (Plecoptera : Pteronarcyidae) in the western USA. Journal of the North American Benthological Society 23, 824-838.

Kondratieff, B.C., Baumann, R.W., 2000. Stoneflies of the United States. Jamestown, ND: Northern Prarie Wildlife Research Center Online. http://npwrc.usgs.gov/resource/distr/insects/sfly/index.htm (Version 12DEC2003).

Kubow, K.B., Robinson, C.T., Shama, L.N.S., Jokela, J., 2010. Spatial scaling in the phylogeography of an alpine caddisfly, Allogamus uncatus, within the central European Alps. Journal of the North American Benthological Society 29, 1089-1099.

Lehrian, S., Balint, M., Haase, P., Pauls, S.U., 2010. Genetic population structure of an autumn-emerging caddisfly with inherently low dispersal capacity and insights into its phylogeography. Journal of the North American Benthological Society 29, 1100-1118.

Lehrian, S., Pauls, S.U., Haase, P., 2009. Contrasting patterns of population structure in the montane caddisflies Hydropsyche tenuis and Drusus discolor in the Central European highlands. Freshwater Biology 54, 283-295.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451-1452.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Research 25, 0955-0964.

Loxterman, J.L., Keeley, E.R., 2012. Watershed boundaries and geographic isolation: patterns of diversification in cutthroat trout from western North America. BMC Evolutionary Biology 12.

Maddison, W.P., Maddison, D.R., 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75 http://mesquiteproject.org.

Meffe, G.K., Vrijenhoek, R.C., 1988. Conservation Genetics in the Management of Desert Fishes. Conservation Biology 2, 157-169.

Minckley, W., Hendrickson, D., Bond, C., 2001. Geography of western North American freshwater fishes: description and relationships to transcotinental tectonism. In The zoogeography of North American freshwater fishes. Edited by: Hocutt CH, Wiley EO. New York, NY: Wiley Interscience; 1986:519-613.

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.

Posada, D., 2008. jModelTest: Phylogenetic Model Averaging. Mol. Biol. Evol. 25, 1253-1256.

Rambaut, A., Drummond, A., 2007. Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer.

Ricker, W.E., G.G.E, S., 1975. An annotated checklist of the Plecoptera (Insecta) of British Columbia. Syesis 8, 333-348.

Schmitt, T., 2009. Biogeographical and evolutionary importance of the European high mountain systems. Front. Zool. 6.

Sheffield, N.C., Hiatt, K.D., Valentine, M.C., Song, H.J., Whiting, M.F., 2010. Mitochondrial genomics in Orthoptera using MOSAS. Mitochondrial DNA 21, 87-104.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. Annals of the Entomological Society of America 87, 651-701.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688-2690.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers. Syst. Biol. 57, 758-771.

Stewart, J.B., Beckenbach, A.T., 2005. Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug Philaenus spumarius (Hemiptera: Auchenorrhyncha: Cercopoidae). Genome 48, 46-54.

Stewart, J.B., Beckenbach, A.T., 2006. Insect mitochondrial genomics 2: the complete mitochondrial genome sequence of a giant stonefly, Pteronarcys princeps, asymmetric directional mutation bias, and conserved plecopteran A+T-region elements. Genome 49, 815-824.

Stewart, K.W., Ricker, W.E., 1997. Stoneflies (Plecoptera) of the Yukon. *in* H.V. Danks and J.A. Downes (Eds.), Insects of the Yukon. Biological Survey of Canada (Terrerstial Arthropods), Ottowa., pp.201-222

Stutz, H.L., Shiozawa, D.K., Evans, R.P., 2010. Inferring dispersal of aquatic invertebrates from genetic variation: a comparative study of an amphipod and mayfly in Great Basin springs. Journal of the North American Benthological Society 29, 1132-1147.

Tajima, F., 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135, 599-607.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596-1599.

Theissinger, K., Balint, M., Haase, P., Johannesen, J., Laube, I., Pauls, S.U., 2011. Molecular data and species distribution models reveal the Pleistocene history of the mayfly Ameletus inopinatus (Ephemeroptera: Siphlonuridae). Freshwater Biology 56, 2554-2566.

Wishart, M.J., Hughes, J.M., 2003. Genetic population structure of the net-winged midge, Elporia barnardi (Diptera : Blephariceridae) in streams of the south-western Cape, South Africa: implications for dispersal. Freshwater Biology 48, 28-38.

Zhang, D.-X., Hewitt, G.M., 1996. Nuclear integrations: challenges for mitochondrial DNA markers. Trends in Ecology & Evolution 11, 247-251.

Table 1. Sample locality data. Outgroup taxa are indicated with an asterisk.

Stream/River	N=	State/Prov	County	Country	Lat	Lon	Ele
Asotin Creek	2	WA	Asotin	USA	46.33174°	-117.08283°	350 m
Beaver Creek	10	CO	Gunnison	USA	38.49465°	-107.03188°	2289 m
*Big Springs	1	CA	Siskiyou	USA	–	–	–
Blodgett Creek	10	MT	Ravalli	USA	46.28908°	-114.16093°	1084 m
Bow River	4	AB	Wheatland	USA	50.77491°	-113.04139°	948 m
Canyon Creek	10	OR	Grant	USA	44.29128°	-118.95702°	1167 m
Chama River	10	NM	Rio Arriba	USA	36.91238°	-106.57317°	2410 m
Clearwater River	1	AB	Clearwater	Canada	51.98999°	-115.39132°	1516 m
Coal Creek	8	UT	Iron	USA	37.67167°	-113.04125°	1824 m
*Cold Stream (Creek)	9	CA	Sierra	USA	39.52390°	-120.28835°	1851 m
Conejos River	10	CO	Conejos	USA	37.05743°	-106.19032°	2539 m
Crooked Creek	5	AK	Kenai Penn.	USA	60.56419°	-149.95447°	600 m
Deer Creek	9	UT	Garfield	USA	38.01253°	-111.97403°	2110 m
Diamond Fork River	10	UT	Utah	USA	40.07865°	-111.37079°	1653 m
Green River	9	WY	Sublette	USA	43.01808°	-110.11820°	2291 m
Hitman Creek	10	CO	Routt	USA	40.75120°	-106.83826°	2339 m
Hoback River	5	WY	Sublette	USA	43.24552°	-110.47662°	1967 m
Kisaralik River	8	AK	Bethel	USA	–	–	–
Kootenay River	7	BC	Kootenay	Canada	50.88342°	-116.34326°	1729 m
La Plata River	6	CO	La Plata	USA	37.27518°	-108.03335°	2444 m
Lee Creek	1	AB	Cardston	Canada	–	–	–
Leeds Creek	1	UT	Washington	USA	37.26553°	-113.36881°	1233 m
Mammoth Creek	10	UT	Garfield	USA	37.62803°	-112.45667°	2129 m
Middle Fk South Platte River	10	CO	Park	USA	39.21892°	-105.99405°	3007 m
Mill Creek	12	UT	San Juan	USA	38.48357°	-109.40855°	1640 m
Monroe Creek	10	UT	Sevier	USA	38.61050°	-112.10577°	1733 m
N. Fk Humboldt River	4	NV	Elko	USA	41.57825°	-115.93085°	2066 m
Parowan Creek	10	UT	Iron	USA	37.80847°	-112.80937°	1952 m
Pecos River	1	NM	San Miguel	USA	35.69339°	-105.69415°	2265 m
Red River	10	NM	Taos	USA	36.69828°	-105.47640°	2473 m
Rock Creek	10	ID	Cassia	USA	42.33495°	-114.28223°	1368 m
Rock Creek	10	MT	Granite	USA	46.71092°	-113.67357°	1085 m
Salina Creek	10	UT	Sevier	USA	38.87907°	-111.55307°	2077 m
Saugache Creek	1	CO	Saguache	USA	38.12948°	-106.45683°	2562 m
Soldier Creek	10	UT	Utah	USA	39.96474°	-111.30976°	1767 m
Spray River	1	AB	Banff Park	Canada	–	–	–
Thomas Creek	1	CA	Modoc	USA	–	–	–
Umatilla River	10	OR	Umatilla	USA	45.67448°	-118.75873°	335 m
Walla Walla River	8	OR	Umatilla	USA	45.92142°	-118.37373°	340 m
Wind River	1	WY	Fremont	USA	43.52537°	-109.61702°	2097 m

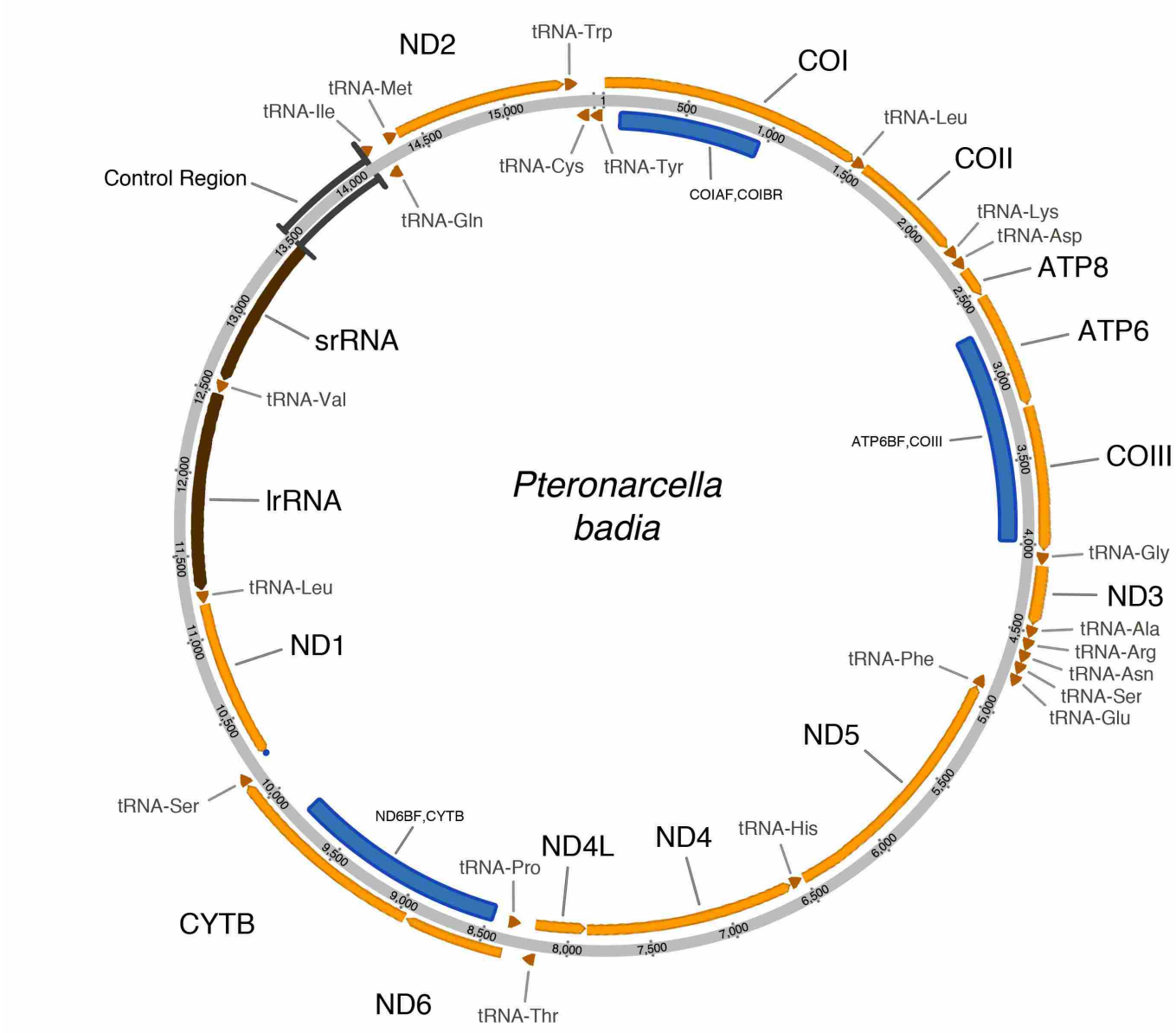

Figure 1. Annotated visualization of the *P. badia* mt genome. Genes located outside the gray circle occur on the majority-coding strand (J-strand), and genes inside the gray circle occur on the minority-coding strand (N-strand). Blue bars represent regions that were amplified in the present study for phylogenetic inference, accompanying labels indicate the primer pairs used for PCR amplification (see Appendix 3)

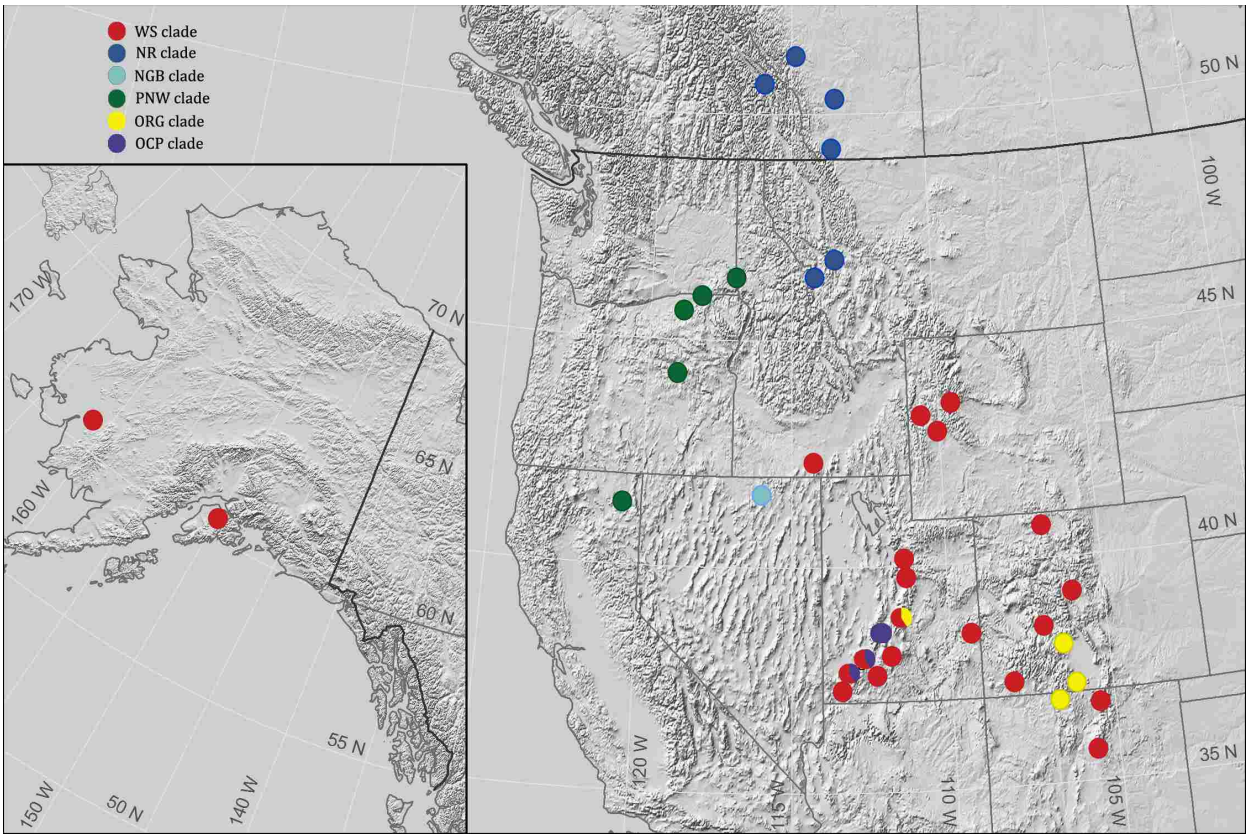

Figure 2. A map showing the distribution of sample localities (I am adding a color scheme to this to reflect the distribution of haplotypes).

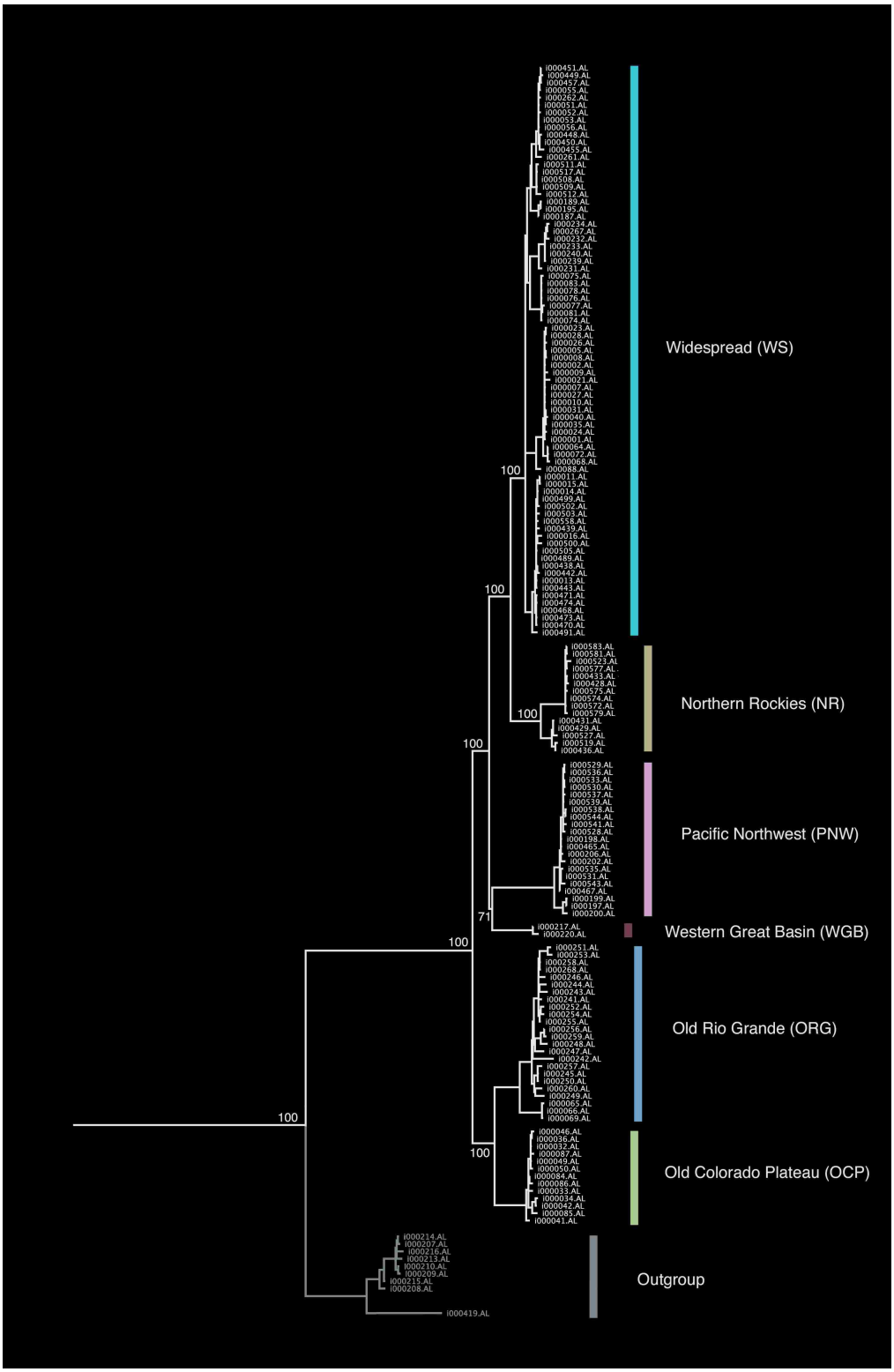

Figure 3. A TEA ML topology showing the relationships among *P. badia.* Deeply divergent clades are labeled and identified by color.

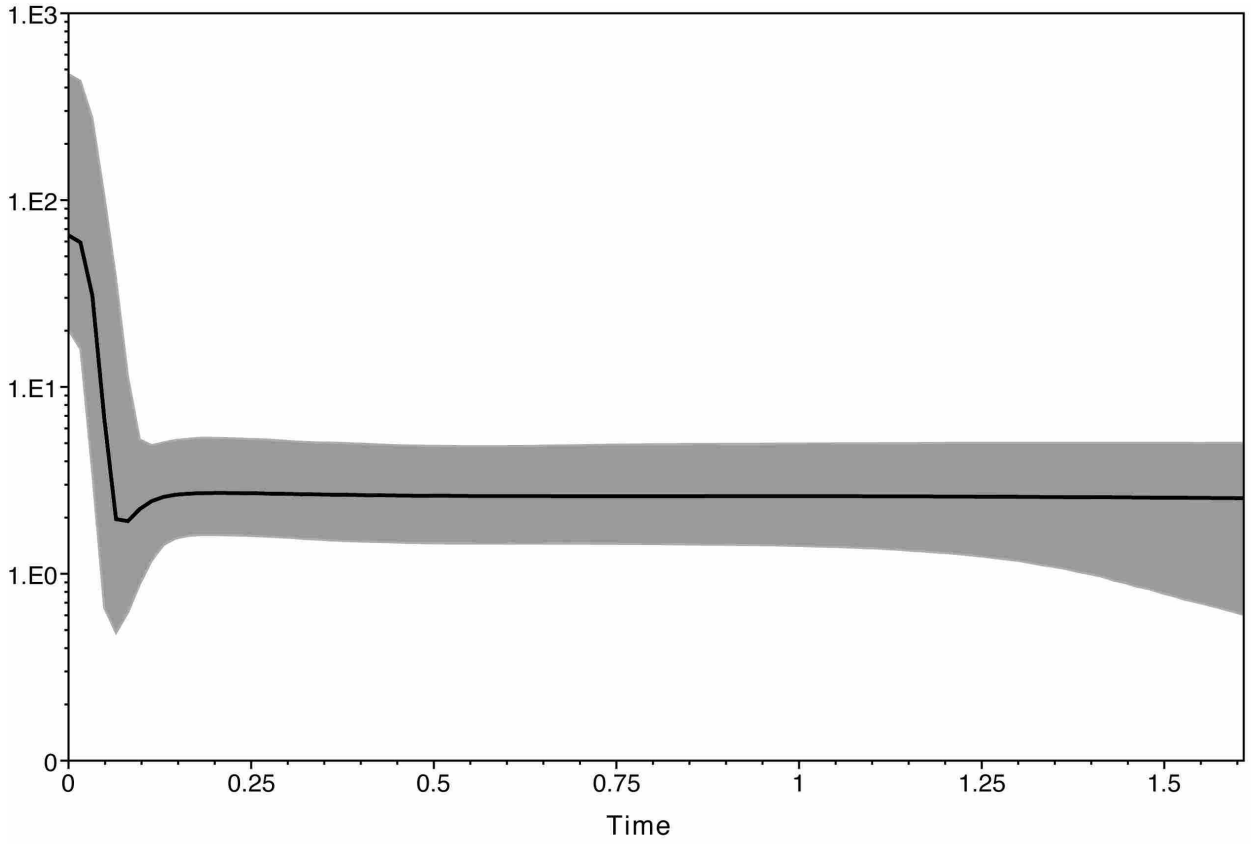

Figure 4. A Bayesian skyline plot showing the recent demographic history of *P. badia*.

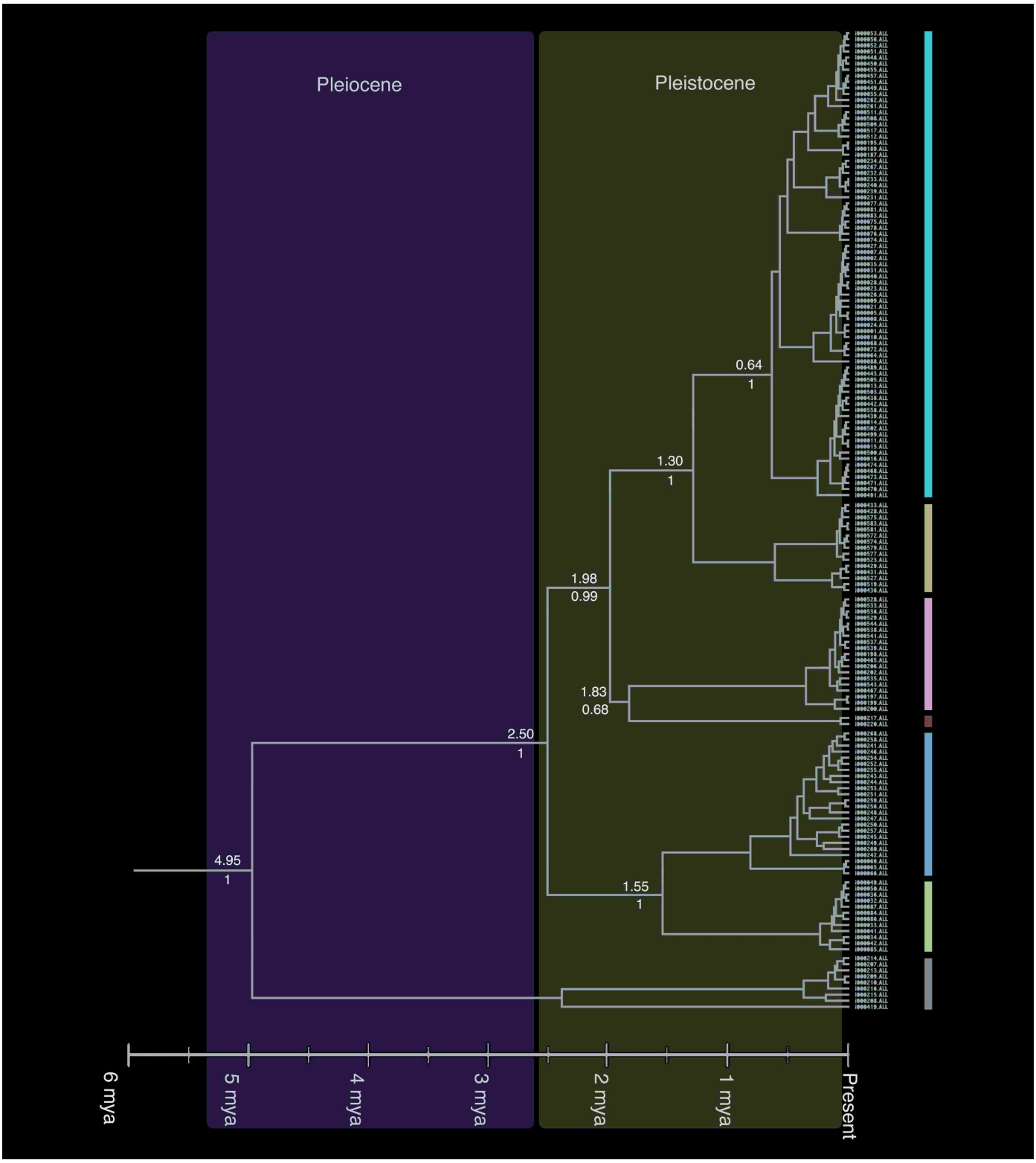

Figure 5. A BI tree reconstruction with Pliocene and Pleistocene epochs shown by the colored boxes. Major clades are color coded in the same color scheme as Figure 3. Date of divergence is shown above node lines with posterior probabilities shown below node lines.

Appendix 3. List of primers used in the analysis. See text for primer sources.

Genes	Primer Name	Sequence (5'-3')	Length	Direction
COI	Ptero.COIAF	ACTTGGCCAACCTGGTTCTCTT	22	Forward
	Ptero.COIBR	GTGGAGGGTTGCTAGTCAGCTA	22	Reverse
ATP6, COIII	Ptero.ATP6BF	CACAGGACACGCTGGTAGAACT	22	Forward
	Ptero.COIIIR	AGTGTCAATATCAGGCTGCTGCT	23	Reverse
ND6, CYTb	Ptero.ND6BF	CCCAAATAAGTCACCCCTTAGCCA	24	Forward
	Ptero.CYTBR	CATTCTGGTTGAATGTGGACGGG	23	Reverse
28S	Ptero.28SBF	ACACGTTGGGACCCGAAAGA	20	Forward
	Ptero.28SBR	TTCCAGGGAACTCGAACGCTT	21	Reverse