2010-04-16

Gene Flow and Dispersal of the Caddisfly, *Neothremma alicia*, in the Rocky Mountains of Utah: A Multiscale Analysis

Xioben Jiang
Brigham Young University - Provo

Gene Flow and Dispersal of the Caddisfly, *Neothremma alicia*, in the Rocky Mountains of Utah:

A Multiscale Analysis

Xiaoben Jiang

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

Master of Science

Russell B. Rader
Keith A. Crandall
Jerald B. Johnson

Department of Biology

Brigham Young University

April 2010

ABSTRACT

Gene Flow and Dispersal of the Caddisfly, *Neothremma alicia*, in the Rocky Mountains of Utah:

A Multiscale Analysis

Xiaoben Jiang

Department of Biology

Master of Science

We determined genetic variance and gene flow across multiple scales (reaches, tributaries, and catchments) to examine the dispersal ability of the caddisfly, *Neothremma alicia* in streams along the Wasatch Range in the Rocky Mountains of Utah. *Neothremma alicia* is one of the most abundant caddisflies in this region. We generated DNA sequence data (mitochondrial COI) from 34 reaches, nested in 15 tributaries distributed across 3 adjacent catchments. We identified 47 haplotypes from a total of 486 individuals. The most abundant haplotype (H1) was found at all sites/reaches and comprised 44% of the total number of individuals sequenced. The remaining rare haplotypes (46) were recently derived from the dominant, H1 haplotype. All of the rare haplotypes were restricted to a single catchment with 81 % restricted to either a single tributary or to two adjacent tributaries. We found the largest F_{ST} values among tributaries and the smallest F_{ST} values between reaches within tributaries suggesting that dispersal and gene flow is largely confined to within tributaries. This result supports the observation that aerial adults commonly crawl and fly along the stream corridor, especially in deeply incised valleys of mountainous regions. Our analyses show that this population has experienced a bottleneck that may have reduced population genetic variance from many haplotypes to one single dominant haplotype, H1. The rare haplotypes may have diverged since the bottleneck from the H1 haplotype and thus, have not had time to disperse outside their catchment and in most cases outside their specific tributary. Our analyses indicated that the bottleneck took place between 1,000 and 10,000 years ago. Thus, it appears that most rare haplotypes have been unable to colonize outside of the tributary they originated in for around 1,000 years.

Keywords: caddisfly, mtDNA, COI, SHM, PRH

ACKNOWLEDGEMENTS

I would first of all like to thank my advisor Dr. Russell Rader for all the time and effort he has put into helping me as a graduate student. I would also like to thank the rest of my graduate committee, Dr. Keith Crandall and Dr. Jerald Johnson for their help and advice throughout various stages of my project. I offer a special thanks to Dr. Peter Unmack for the genetic lab training. I really appreciate Jeffrey Moore's help with the field collections. Last and most importantly, I would like to thank my mother Manman Song who has give me unending support and has been very patient throughout the course of my studies. I would like to thank my father Jianzhong Jiang for his love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Population genetic theory predicts a lack of genetic differentiation and homogenous allele frequencies among local populations across a species range if there are high levels of dispersal and thus, gene flow. Alternatively, when dispersal among populations is limited, allele frequencies will diverge as a result of genetic drift, natural selection or a combination of both (Slatkin 1985). Dispersal in aquatic insects is difficult to measure directly because they are difficult to mark and because their long range dispersal may be a rare event making detection almost impossible (Bohonak 1999; Bohonak& Jenkins 2003; Feral 2002). Thus, most studies of aquatic insects have attempted to examine dispersal using indirect methods, such as genetic markers. By measuring allele frequencies and genetic divergence among populations, it is possible to estimate the relative levels of dispersal (Hughes *et al.* 2008; Slatkin 1985).

According to the "isolation by distance" model of gene flow, the highest level of genetic variance should be found between sites separated by the greatest distance. Thus, maximum genetic divergence between local populations should occur at the largest spatial scales (Hughes *et al.* 1999; Miller *et al.* 2002). Poorly dispersing aquatic insects, stream invertebrates without an aerial adult stage (e.g. crustaceans), and stream fish have shown a hierarchical pattern of genetic variance consistent with the "isolation by distance" model. That is, genetic variation was minimal within a reach, increased between tributaries within a catchment, and reached a maximum among catchments. This is called the Stream Hierarchy Model (SH model) of genetic variation (Meffe& Vrijenhoek 1988).

Interestingly, stream insects with an aerial adult stage often show the opposite pattern (Hughes *et al.* 2008). That is, there is no correlation between genetic distance and geographic distance. The maximum genetic divergence is frequently detected at the smallest scales (e.g.

sites nested in reaches) with little divergence between catchments at large scales. That is, the number of haplotypes at a local site is a small fraction of the total number of haplotypes across all sites. This pattern was first detected using three aquatic insects in south-east Queensland, Australia (Schmidt et al. 1995, Bunn and Hughes 1997, and Hughes et al. 1998). Since then, it has been detected in a variety of stream insects with an aerial adult stage (Hughes *et al.* 2008).

Bunn and Hughes (1997) developed the "Patchy Recruitment Hypothesis" (PR hypothesis) as a potential explanation for this pattern. They suggested that adult flight across catchment boundaries should result in most haplotypes occurring in all catchments and thus, little genetic differentiation at large scales. They invoked a "recruitment effect" to explain why only a small fraction of the total number of haplotypes across a region occurred within single reaches. In the recruitment effect adult dispersal is widespread across the study area but a stream reach is re-populated each generation by the offspring of only a few females from a small subset of the total number of haplotypes (Bunn& Hughes 1997). Over the course of several generations, all haplotypes should produce some adults that could successfully colonize reaches in each catchment. How can adult flight account for most haplotypes being represented in all catchments at large scales but not at small scales? If they can fly across catchment boundaries why can't they fly up and down stream corridors within catchment boundaries? We must assume that they could but don't. Most adults must oviposit in the same local vicinity from which they emerged and there must be little movement of haplotypes between local sites by immature stages (e.g. larval drift and crawling; Bunn & Hughes 1997). Even then, all haplotypes would gradually accumulate at local scales if long range dispersal was usually successful. However, if most new long range colonization attempts failed, then only a subset of haplotypes that by chance did not fail, would persist in any given reach even if adult long range dispersal

was sufficiently common for all haplotypes to eventually become established in some reaches in each catchment.

This explanation also assumes that adults are capable of flying over catchment boundaries. Hughes et al. (2008) suggested that the PR hypothesis may not apply in mountainous terrain with deeply incised valleys that would prevent adult dispersal over catchment boundaries. In mountainous terrain we might expect the geographic distribution of haplotypes to more closely follow the SH model.

Our objective was to test the SH model versus the PR hypothesis using the caddisfly, *Neothremma alicia*, in three deeply incised drainages along the Wasatch Range in the Rocky Mountains of Utah. Patterns of genetic variation in other caddisflies have often conformed to the PR hypothesis (Hughes *et al.* 1998; Schultheis& Hughes 2005) partly because caddisflies in general are considered good long range dispersers and have shown little genetic divergence across catchment boundaries at large scales (Hughes *et al.* 2008). However, such studies have rarely been conducted in mountainous regions with deeply incised valleys that might restrict adult dispersal among catchments (Bunn& Hughes 1997; Hughes *et al.* 1998; Schultheis& Hughes 2005). We hypothesized that *N. alicia* would show little genetic variation at small scales (sites nested in tributaries) with increasing genetic variation at intermediate scales (tributaries nested in catchments), and maximum genetic differentiation at large scales (among catchments). Thus, we expected that genetic variation of *N. alicia* in the deeply incised valleys of the Rocky Mountains would conform to the Stream Hierarchy Model.

METHODS

Focal Species

Neothremma alicia (Trichoptera, Uenoidae) (Flint& Wiggins 1961; Wiggins *et al.* 1985) is a stone-cased caddisfly common in western North American. Larvae of *N. alicia* are sedentary grazers that feed mainly on fine organic particles, with small proportions of diatoms and other algae (Wiggins *et al.* 1985). The altitudinal distribution of *N. alicia* in Utah typically ranges from 1600 m asl to 3400 m asl (Walker, 2008 and personal observations).

This species has a two-year life cycle in Utah similar to populations in the Canadian Rockies (Ogilvie& Clifford 1986). Also, *N. alicia* adults emerge synchronously in Utah over two months during July and August at lower elevations (≈ 1600 m asl to 2500 m asl), and during August and September at higher elevations (Walker 2008).

Sampling Design

We used a nested hierarchical sampling design based on the spatial organization of streams (Frissell *et al.* 1986). That is, we collected samples from 34 sites nested in 15 tributaries, nested in three adjacent catchments along the Wasatch Range (Figure 1, Appendix 1). This design allowed us to examine patterns of genetic differentiation across all of the spatial scales recognized as relevant in stream ecosystems. Most previous studies have only examined genetic variation at one or two spatial scales (Hughes et al. 2008).

Provo River is a fourth order catchment, whereas American Fork and Little Cotton Wood Canyon are smaller third-order basins. A site was a stream reach of 50 m to 100 m in length. Two and sometimes three sites were sampled within each tributary separated by at least 500 m in order to examine small scale population genetic structure. We would expect significant genetic

variation among sites within tributaries according to the PR hypothesis but not according to the SH model. Specimens were collected by hand from the underside of the rocks and stored in 95% ethanol at -80° C prior to analysis. At most sites, approximately 30% of the boulders contained clusters of *N. alicia*. We intensively searched for *N. alicia* in all tributaries within each drainage. *Neothremma alicia* was very rare or absent from the lowest sections of each drainage and from five sub-basins in the Provo River drainage (Figure 1).

DNA extraction, PCR amplification and automated DNA sequencing

Our results are based on 15 individuals analyzed from each site. Fifteen or fewer individuals have been sufficient to accurately characterized genetic variation at a site in previous studies involving caddisflies (Baker *et al.* 2003; Hughes *et al.* 1998). Genomic DNA was extracted using the Qiagen DNeasy Tissue Kit (Qiagen, Hilden, Germany), and the cytochrome *c* oxidase I mitochondrial gene (COI) was amplified and sequenced. We chose the COI gene and as our genetic marker and sequenced 15 individuals for each site because mitochondrial DNA is a maternally inherited, haploid marker with a fourfold smaller effective population size relative to nuclear DNA. So, mitochondrial DNA is much more sensitive to restricted gene flow than nuclear DNA (Birky *et al.* 1989).

We amplified and sequenced two fragments of COI gene and aligned these into one 800bp fragment. We designed our own primers, 715F (5'-GAAGTTTATATTCTCATTTTACCTG -3'), 1186R (5'- GGATTTATAGTTAAACCTGTA -3') and 1061F (5' –GCTAATTCTTCTATTGATATTATACTTC -3'), and Leu 25R (5'-CTTTATAAATGGGGTTTAAATCCAT -3') to avoid amplifying pseudogenes.

Amplifications contained 2.5 µl of each primer, 2 µl of template, 2.5 µl buffer, 4 µl dNTPs, 0.15 µl *Taq* polymerase, 4 µl $MgCl_2$ and 9.85 µl of sterile water. The program for polymerase chain reactions (PCR) consisted of a 3 minute denaturation step at 94ºC, 40 cycles of 30s at 94ºC, 45s at 53ºC, 60s at 72ºC, and a 10 minute extension step at 72ºC. Amplified DNA was checked by running on a 1.5% agarose gel stained with ethidium bromide.

PCR products were purified with the Qiagen QIAquick PCR Purification Kit and then sequenced. Each individual DNA fragment was sequenced from both directions. DNA was cycle sequenced using ABI Big Dye terminator protocol. The reactions were done in 10 µl total volumes containing 2 µl templates, 1 µl primer, 0.5 µl Big Dye, and 6.5 µl sterile water. Big Dye products were cleaned over Sephadex columns and dehydrated in the appropriate well of the sample plate. Sequences were obtained using an Applied Biosystems 3730 XL automated DNA sequencer at the Brigham Young University DNA Sequencing Center. Chromatograms were edited using SEQUENCHER™ 4.7 (Gene Codes, Ann Arbor, MI, USA) and aligned manually. Sequences were checked for unexpected frame shift errors or stop codons in Mega 4.0 (Tamura *et al.* 2007)

Data Analysis

We used several analytical approaches to compare the SH model versus the PR hypothesis. First, we compared genetic differentiation based on F_{ST} values across scales using an analysis of molecular variance (AMOVA) implemented in *Arlequin* version 3.11 (Excoffier *et al.* 2005). According to the SH model, F_{ST} values should be largest at the largest scales but decrease at intermediate and small scales. By contrast, the smallest scales should show the largest F_{ST} values according to the PR hypothesis.

We created a haplotype network, a phylogenetic program based on maximum likelihood, and used the software program *Migrate* to show the geographical distribution of haplotypes across catchments, to identify ancestral and recently diverged haplotypes, and to evaluate the direction of gene flow among catchments. According to the PR hypothesis all haplotypes should occur in each catchment, whereas haplotypes should be restricted to a single catchment in the SH model. The haplotype network was created using TCS software v1.21 (Clement *et al.* 2000). TCS software builds networks by using population level events, like recombination and the presence of ancestral haplotypes, to create a genealogical network showing population level divergences, where a bifurcating tree is a poor representation of the mutation process (Clement *et al.* 2002). In addition to accurate haplotype networks, TCS is also faster than similar network construction programs. We also employed a traditional set of phylogenetic analysis such as, maximum likelihood (ML).

We used TCS (v1.21) to create a minimum spanning phylogram (tree) of COI haplotypes using statistical parsimony with a 95% probability that no multiple substitutions had occurred. We estimated phylogenetic relationships among haplotypes using maximum likelihood criteria in the software program PAUP* 4.0b10 (Swofford 2003). Maximum likelihood analysis requires selecting an appropriate model of molecular evolution. We used the Akaike Information Criterion (AIC) in ModelTest v3.06 (Posada& Crandall 1998) to select the TIM+I model based on: Lset Base = 0.3265 0.1634 0.1144; Nst = 6; Rmat = 1.0000 12.2651 0.1050 0.1050 4.4545; Rates = equal; Pinvar = 0.6340. Our ML analysis used a heuristic search with 100,000 random replicates and TBR branch swapping.

We evaluated gene flow among the three catchments by estimating migration rates using a Markov Chain Monte Carlo (MCMC) method implemented in the software package *Migrate*

V3.1.2 (Beerli& Felsenstein 1999). We conducted four independent replicate runs. In each run, we used 20 short chains with 1,000 genealogies followed by four long chains with 10,000 genealogies and a burn-in of 10,000.

We also used *Arlequin* (version 3.11) to compare genetic diversity among scales (reaches within tributaries, tributaries within drainages, and among drainages) using the average of Θ_S (an estimate of genetic diversity in the distant past) and the average of Θ_π (an estimate of genetic diversity in the recent past). We used this analysis to test for the effects of historical events in addition to isolation by distance (SH model) and PR hypothesis, on genetic variation and gene flow. For example, when Θ_π is less than Θ_S we have evidence that the effective population size has decreased at some point in the recent past (Buhay and Crandall 2005). Thus, this is would be evidence of a bottleneck effect. In a separate analysis, we also used *Beast* version 1.5.3 (Drummond& Rambaut 2007) to calculate data used to create a Bayesian skyline plot to estimate changes in the effective population size over time (decrease, increase or stable). A skyline plot and an analysis of Θ_S and Θ_π provide corroborating evidence of a change in the effective population size. Because *Beast* requires a large sample size, we combined data from all three catchments. This analysis also requires a molecular clock, which is an estimate of the population mutation rate. We used the same rate of molecular divergence (2.2% per Myr) used in previous analyses for caddisflies (Baker *et al.* 2003; Gaunt& Miles 2002). We ran five initial short runs of 10^6 generations to optimize the settings in *Beast* analysis that were then used to conduct multiple longer runs with a chain length of 5×10^7. The Bayesian skyline plot was created in the software program, *Tracer* version 1.4 (Rambaut& Drummond 2007).

RESULTS

The pattern in this study was the geographical distribution of haplotypes across scales combined with the pattern of relatedness among haplotypes. Subsequent analyses help to explain this pattern providing evidence to support or refute our competing hypotheses (the SH model versus the PR hypothesis).

We identified 47 haplotypes from a total of 486 individuals (Appendix 2). The nucleotide base frequencies were A = 0.318, T = 0.400, C = 0.149 and G = 0.133. The most abundant and the ancestral haplotype (H1) was found at all sites across the three catchments and comprised 44% of the total number of individuals sequenced which was between 13% and 73% of the individuals at any particular site (Appendix 2). In addition to H1, we also found between two and six rare haplotypes at each site (Appendix 2). However, all rare haplotypes were restricted to single drainages (46 in total). Thirty-one haplotypes (H2 – H21 and H37 – H47) were found in the Provo River catchment, five (H22 – H26) in American Fork, and ten (H27 – H36) in Little Cotton Wood Canyon (Appendix 2). Also, 58.7% of the rare hyplotypes were restricted to a single tributary with an additional 21.7% found only in two adjacent tributaries. Thus, 81% of the rare hyplotypes were found in only one, or sometimes, two tributaries indicating a very restricted distribution.

The haplotype network showed a well-resolved pattern with only a few missing haplotypes (Figure 2). H1, the most common haplotype, had a high root probability and was the only potential ancestral haplotype identified in this analysis (Figure 2). Most clades in the network were only one-step removed from the H1 haplotype indicating that most haplotypes had recently diverged from H1. Individuals in clades with the greatest number of steps removed from H1 (6 or 7 steps) clustered in 3 isolated sub-catchments.

Maximum likelihood methods produced a single tree with a –lnL score = 1271.45771 (Figure 3). Relationships among haplotypes in the ML tree topology were nearly completely congruent with the haplotype network. There were only a few minor differences; H5 was closer to H41 and H7 was closer to H9 in the ML tree than in the haplotype network.

The ubiquitous geographic distribution of the dominant, ancestral haplotype shows that it has been able to disperse and colonize among catchments at the largest scale consistent with the PR hypothesis. However, the restricted distribution of the rare haplotypes is most consistent with isolation by distance and the SH model.

The AMOVA F_{ST} analysis showed that *N. alicia* populations did not conform to either the SH model or the PR hypothesis (Table 1). According to the SH model and the PR hypothesis, we expected the greatest average F_{ST} values at the largest and smallest scales, respectively. F_{ST} was only significant at the intermediate scale (among tributaries within catchments), whereas F_{ST} at the largest (among catchments) and smallest (among reaches within tributaries) scales were not significantly different from zero.

Estimates of migration rates averaged over the recent and distant past for the three catchments showed that Provo River had high levels of movement into both American Fork (1,120 individuals per generation scaled by the mutation rate) and Little Cotton Wood Canyon (893 individuals per generation) with little migration to Provo River from American Fork River (2.49×10^{-10} individuals per generation) or Little Cotton Wood Canyon (56 individuals per generation; Table 3). There were also high levels of movement (653 individuals per generation) from Little Cotton Wood Canyon into the American Fork River, but a much smaller amount (1.97×10^{-8} individuals per generation) in the opposite direction. This analysis suggests that *N. alicia* originated in Provo River and migrated from Provo River into Little Cotton Wood Canyon

and The American Fork River. It also suggests that there was and is considerable gene flow at the largest scales, across catchments, contrary to the SH model and consistent with the PR hypothesis. However, it is not entirely consistent with the predictions of the PR hypothesis because dispersal is not equal between drainages but flows primarily from the largest catchment to the smallest catchments.

As expected, the largest scale (catchments) and the largest drainage had the greatest effective population sizes. That is, Θ_S and Θ_π were greatest at the catchment scale (Table 1) and in Provo River (Table 2). Interestingly, Θ_S, a measure of population genetic diversity in the distant past, was greater at each scale than Θ_π, a measure of population genetic diversity in the recent past. A decline in genetic diversity suggests the occurrence of a bottleneck.

Our Bayesian skyline analysis is also consistent with a bottleneck event in the recent past. It revealed a stable population size through time for the overall population of *N. alicia* in this study with a dramatic decline in population size between 1,000 and 10,000 years ago (Figure 4). It also shows that the effective population size of *N. alicia* rapidly increased following the decline. A bottleneck effect is a historical explanation for the geographical distribution and relatedness of haplotypes of *N. alicia* that is independent of and neither supports nor refutes the validity of the SH model versus the PR hypothesis.

DISCUSSION

Our results suggest that a bottleneck event, differences in time for dispersal between ubiquitous versus rare haplotypes, and the effects of mountainous terrain on adult dispersal provide a plausible explanation for one ubiquitous, dominant, ancestral haplotype and numerous rare haplotypes with a restricted distribution. It appears that a single haplotype (H1) was able to expand and disperse following a bottleneck event between 10,000 and 1,000 years ago. Either it

had colonized all three drainages before the bottleneck or it dispersed from the Provo River drainage during the rapid expansion of the population following the bottleneck. By contrast, the rare haplotypes were recently derived from the dominant haplotype and have been unable to disperse beyond the tributary within which they originated. Deeply incised valleys separated by mountainous terrain may account for the restricted distribution of rare haplotypes. Mountainous terrain may increase the time needed for long range dispersal of *N. alicia* in the Wasatch Mountains.

Alternatively, the rare haplotypes may be primarily restricted to single tributaries because they are rare. Population size is often correlated with an increased capacity to disperse (Hughes *et al.* 2008). Thus, H1 might occur across all three basins because abundant populations produce more adults thus, increasing the probability that some will successfully colonize new tributaries and catchments. Propagule pressure is an important aspect in the successful colonization of introduced species (Lockwood *et al.* 2005).

This bottleneck explanation includes aspects of both the SH model and the PR hypothesis. That is, the presence of a ubiquitous haplotype that has migrated between drainages is consistent with long range adult dispersal in the PR hypothesis. However, the restricted distribution of rare haplotypes to single tributaries is most consistent with isolation by distance and the SH model. Thus, our data does not completely support nor refute either hypothesis.

An additional explanation for the pattern of one ubiquitous and many rare haplotypes invokes natural selection. That is, the H1 haplotype may be a generalist adapted to a variety of conditions with superior dispersal abilities and thus able to colonize and persist in all reaches of this study. By contrast, the rare haplotypes may be specifically adapted to the conditions present in the tributary they originated in. Rare haplotypes may or may not be capable of long range

dispersal but they can only persist in their resident tributary. However, the mitochondrial CO1 gene is neutral and does not indicate functional or adaptive differences between haplotypes. Plus, most of the rare haplotypes were only one base pair removed from the H1 haplotype. It seems unlikely that a single base pair difference in a neutral marker would indicate sufficient time for specialization. Clearly, these explanations (bottleneck and selectionist) are not mutually exclusive. For example, rare haplotypes may have a restricted distribution relative to the H1 haplotype because they have a low propagule pressure and because they are adapted to conditions within specific tributaries and cannot colonize other basins when they do disperse.

The nested hierarchical design of our study allowed us to examine patterns not just across small and large scales (Hughes *et al.* 2008), but also at intermediate scales. Thus, we can make a more refined analysis of movement and dispersal, which appears to primarily occur within and not between tributaries for rare haplotypes. The greatest genetic variation (F_{ST} values) was detected among tributaries rather than at the smallest scale (reaches nested in tributaries) which is contrary to the PR hypothesis. It appears that the dispersal of rare haplotypes is largely confined to within tributaries. Aerial adults commonly fly along the stream corridor (Hughes *et al.* 2008; Müller 1982). For example, (Müller 1982) proposed that adults of aquatic insects generally fly upstream to compensate for the downstream drift of immature stages.

How do the rare haplotypes coexist with the dominant haplotype in the same reach? There are two potentially related explanations: 1) asynchrony in the timing of adult emergence and 2) mating preferences within haplotypes. The rare haplotypes may coexist with the dominant haplotype if there is little overlap in the period of adult emergence. For example, the H1 haplotype may be adapted to a variety of temperature regimes found in various catchments and thus, show an extended period of emergence compared to the rare haplotypes that show a

much synchronized emergence adapted to the specific conditions within a single catchment. Also, individuals of rare haplotypes may have mating preferences for individuals of the same haplotype. Mating preferences could reinforce a highly synchronized emergence period, especially for rare populations where there is a low probability of finding another rare haplotype. Finally, it is possible that they don't coexist. The rate at which the rare haplotypes go extinct is somewhat balanced by the rate at which they are produced. Our data, and most previous studies, are a single snap-shot in time. A single snap-shot may produce an illusion of coexistence.

Future research might explore the relationship between population size and the dispersal ability of stream insects, including observations on gene flow and movements of rare taxa and haplotypes at intermediate scales of tributaries within catchments in mountainous regions. Future studies might also compare gene flow and patterns of dispersal using a variety of gene markers (nuclear, mitochondrial, and microsatellite). Hughes (2008) provides an excellent discussion on the advantages and disadvantages of each as it relates to gene flow in aquatic insects.

LITERATURE CITED

Baker AM, Williams A, Hughes J (2003) Patterns of spatial genetic structuring in a hydropsychid caddisfly (Cheunmatopsyche sp. AV1) from southeastern Australia.

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* **152**, 763-773.

Birky WC, Fuerst P, Maruyam T (1989) Organelle gene diversity under migration, mutation and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells and comparison to nuclear genes. *Genetics* **121**, 613-627.

Bohonak AJ (1999) Dispersal, Gene Flow, and Population Structure *The Quarterly Review of Biology* **74**, 21-45.

Bohonak AJ, Jenkins DG (2003) Ecological and evolutionary significance of dispersal by freshwater invertebrates. *Ecology Letters* **6**, 783–796.

Bunn SE, Hughes JM (1997) Dispersal and recruitment in streams: evidence from genetic studies. *Journal of the North American Benthological Society* **16**, 338-346.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-1659.

Clement M, Snell Q, Walker P, Posada D, Crandall KA (2002) TCS: Estimating Gene Genealogies.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214.

Excoffier L, Laval LG, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis *Evolutionary Bioinformatics Online* **1**, 47-50.

Feral J-P (2002) How useful are the genetic markers in attempts to understand and manage marine biodiversity? *Journal of Experimental Marine Biology and Ecology* **268**, 121-145.

Flint OS, Wiggins GB (1961) Records and descriptions of North American species in the genus Lepidostoma, with a revision of the Vernalis group (Trichoptera: Lepidostomatidae). *Canadian Entomologist* **93(4)**, 279-297.

Frissell CA, Liss WJ, Warren CE, Hurley MD (1986) A hierarchical framework for stream habitat classification: Viewing streams in a watershed context *Environmental Management* **10**, 199-214.

Gaunt MW, Miles MA (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Molecular Biology and Evolution* **19**, 748-761.

Hughes JM, Mather PB, Sheldon AL, Allendorf FW (1998) Dispersal and recruitment of *Tasiagma ciliata* (Trichoptera : Tasimiidae) in rainforest streams, south-eastern Australia. *Freshwater Biology* **41**, 63-72.

Hughes JM, Mather PB, Sheldon AL, Allendorf FW (1999) Genetic structure of the stonefly, Yoraperla brevis, populations: the extent of gene flow among adjacent montane streams. *Freshwater Biology* **41**, 63-72.

Hughes JM, Schmidt DJ, Mclean A, Wheatley A (2008) Population genetic structure in stream insects: What have we learned? In: *Aquatic Insects*, p. 284.

Lockwood JL, Cassey P, Blackburn T (2005) The role of propagule pressure in explaining species invasions. *Trends in Ecology & Evolution* **20**, 223-228.

Meffe GK, Vrijenhoek RC (1988) Conservation genetics and the management of desert fishes. *Conservation Biology* **2**, 157-169.

Miller MP, Blinn DW, Keim P (2002) Correlations between observed dispersal capabilities and patterns of genetic differentiation in populations of four aquatic insect species from the Arizona White Mountains, U.S.A. *Freshwater Biology* **41**, 63-72.

Müller K (1982) The Colonization Cycle of Freshwater Insects. *Oecologia* **52**, 202-207.

Ogilvie GA, Clifford HF (1986) Life histories, production, and microdistribution of two caddisflies (Trichoptera) in a Rocky Mountain stream. *Canadian Journal of Zoology/Revue Canadienne de Zoologie* **64**, 2706.

Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **14**, 817-818.

Rambaut A, Drummond AJ (2007) Tracer V1.4, Available from http://beast.bio.ed.ac.uk/Tracer.

Schultheis AS, Hughes JM (2005) Spatial patterns of genetic structure among populations of a stone-cased caddis (Trichoptera: Tasimiidae) in south-east Queensland, Australia. *Freshwater Biology* **50**, 2002-2010.

Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics* **16**, 393-430.

Swofford DL (2003) *PAUP*. Phylogenetic Analysis Using Parsimony (* and other methods), Version 4.0b10.*, Sinauer, Sunderland.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology and Evolution* **24**, 1596-1599.

Walker S (2008) *Altitudinal and seasonal survey of Ephemeroptera, Plecoptera, and Trichoptera in American Fork Canyon, Utah* Honored Thesis, Brigham Young University.

Wiggins GB, Weaver JS, Unzicker JD (1985) Revision of the caddisfly family Uenodiae (Trichoptera). *Canadian Entomologist* **117**, 763-800.

Table 1. Results of hierarchical AMOVA analysis performed with 16,000 permutations of mtDNA haplotype frequency and Tajima-Nei estimates of sequence divergence (F_{ST}). Values in the body of the table are average theta estimates (Θ_S and Θ_π) at each spatial scale.

Hierarchical level	$\mathbf{F_{ST}}$	Θ_S	Θ_π
Among catchments	0.001	3.06	1.72
Among tributaries	0.15**	1.66	1.50
Among reaches within tributaries	0.02	1.61	1.48

**$P < 0.01$.

Table 2. Theta estimates (Θ_S and Θ_π) for each of the three catchments.

Catchment	Θ_S	Θ_π
Provo River	4.98	1.78
American Fork	1.45	1.60
Little Cotton Wood	2.75	1.77

Table 3. Migration rates (m/mμ; m is gene flow rate and μ is mutation rate) for each of the three catchments calculated in Migrate. For the migration rates, catchments in the first column represent the destination of the migration and catchments on the top row represent the source of the migration.

	Provo	**American Fork**	**Little Cotton Wood**
Provo		2.49×10^{-10}	56
American Fork	1,120		653
Little Cotton Wood	893	1.97×10^{-8}	

FIGURE LEGENDS

Figure 1. Map showing the sites (reaches) within catchments of this study (Provo River, American Fork, and Little Cotton Wood).

Figure 2. A minimum spanning haplotype network of the COI gene using statistical parsimony with a 95% probability that no multiple substitutions have occurred. Three colors represent the different catchments (Provo River, American Fork River, and Little Cottonwood River). Haplotypes are identified by numbers and the numbers correspond to Appendix 2.

Figure 3. Maximum likelilhood phylogram of 47 COI haplotypes of *Neothremma alicia*. Different colors represent the three catchments (Provo River, American Fork River, and Little Cottonwood River). Haplotypes are identified by numbers and the numbers correspond to Appendix 2.

Figure 4. Bayesian skyline plot for COI individuals showing change in effective population size over time in *N.alicia* in all three catchments. The black line represents the mean effective population size and blue lines represent the upper and lower 95% confidence intervals.

Figure 1.

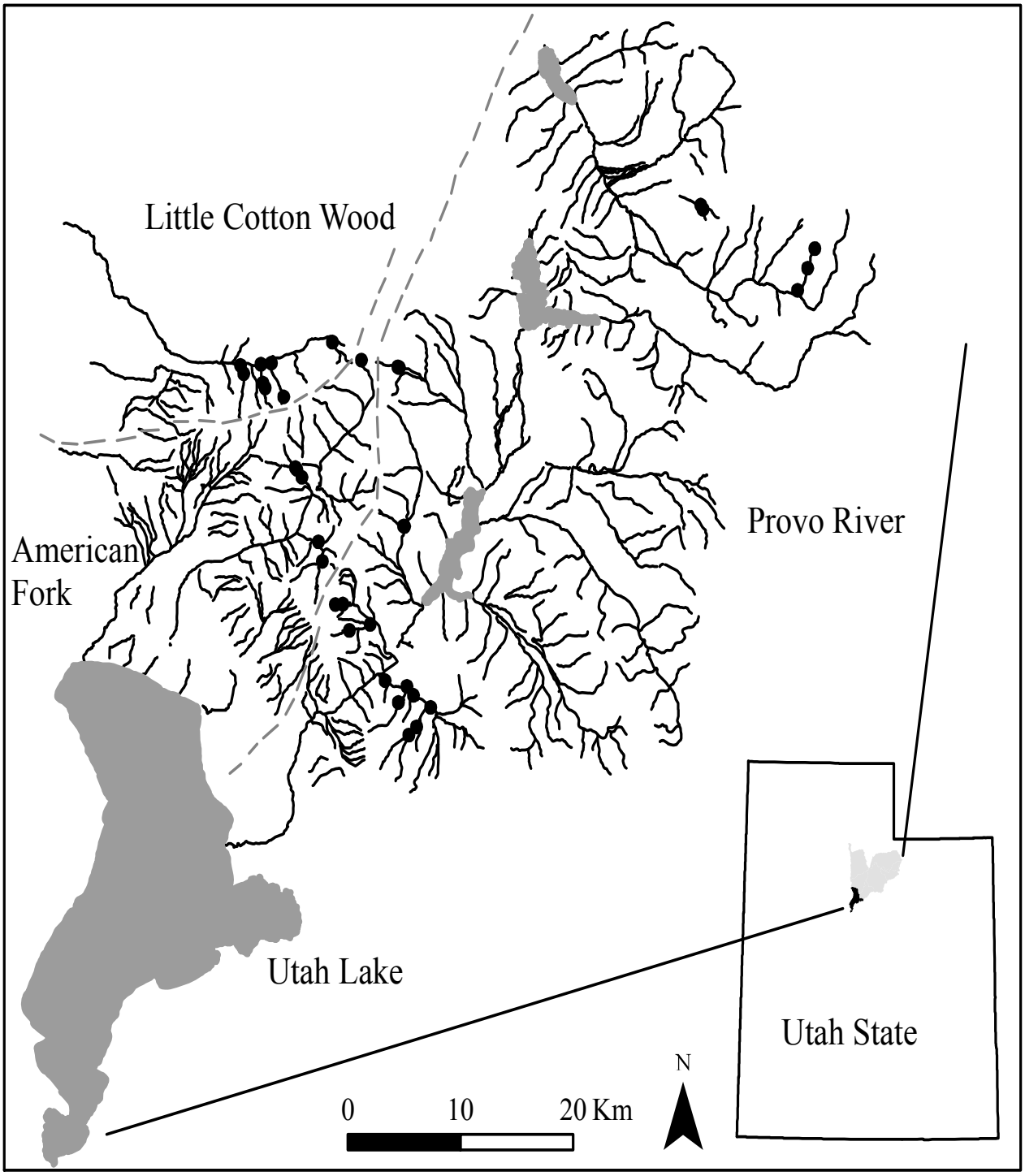

Little Cotton Wood
American Fork
Provo River
Utah Lake
Utah State
N
0
10
20 Km

Figure 2

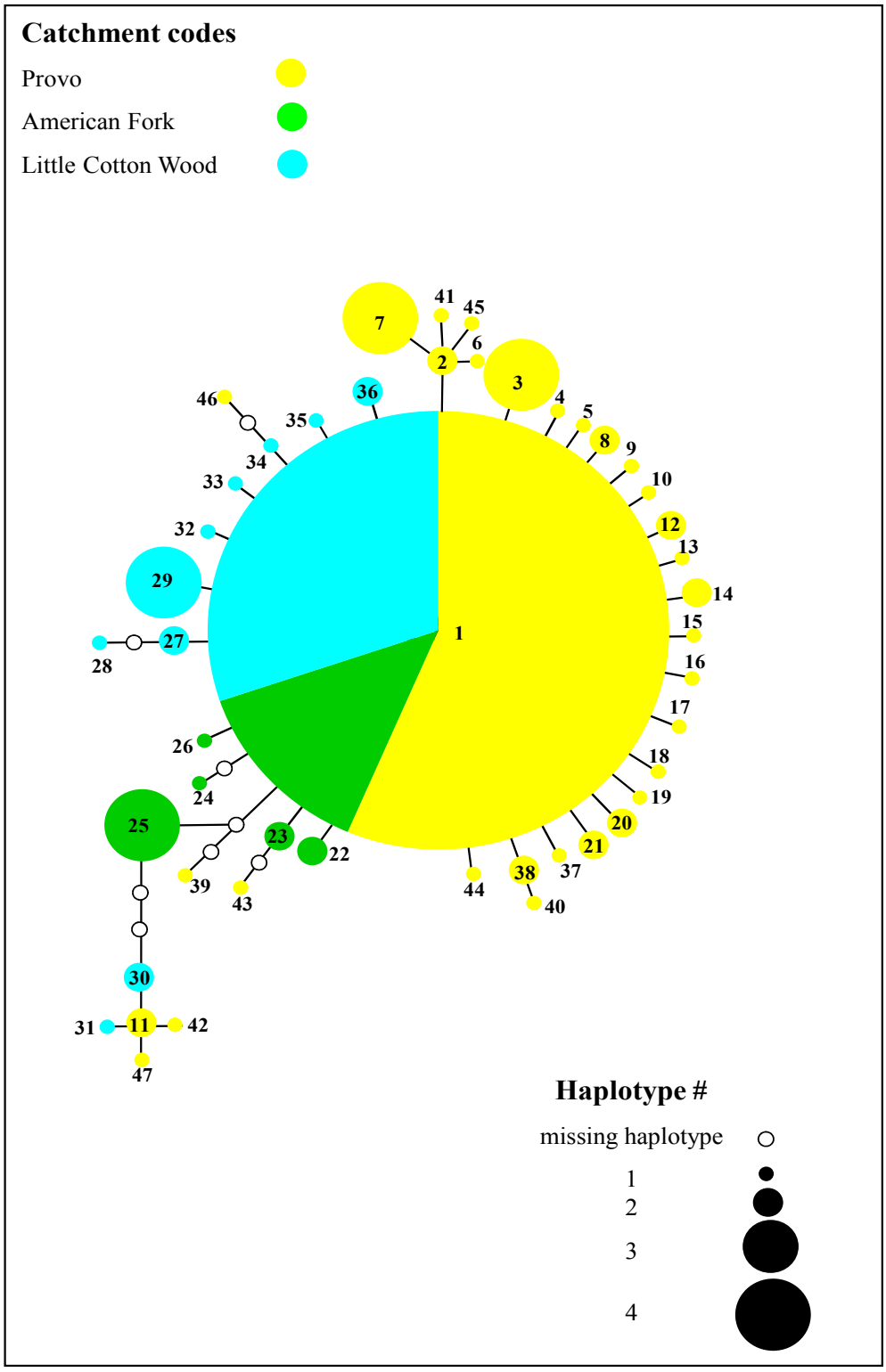

Catchment codes
Provo
American Fork
Little Cotton Wood
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
Haplotype #
missing haplotype
1
2
3
4

Figure 3.

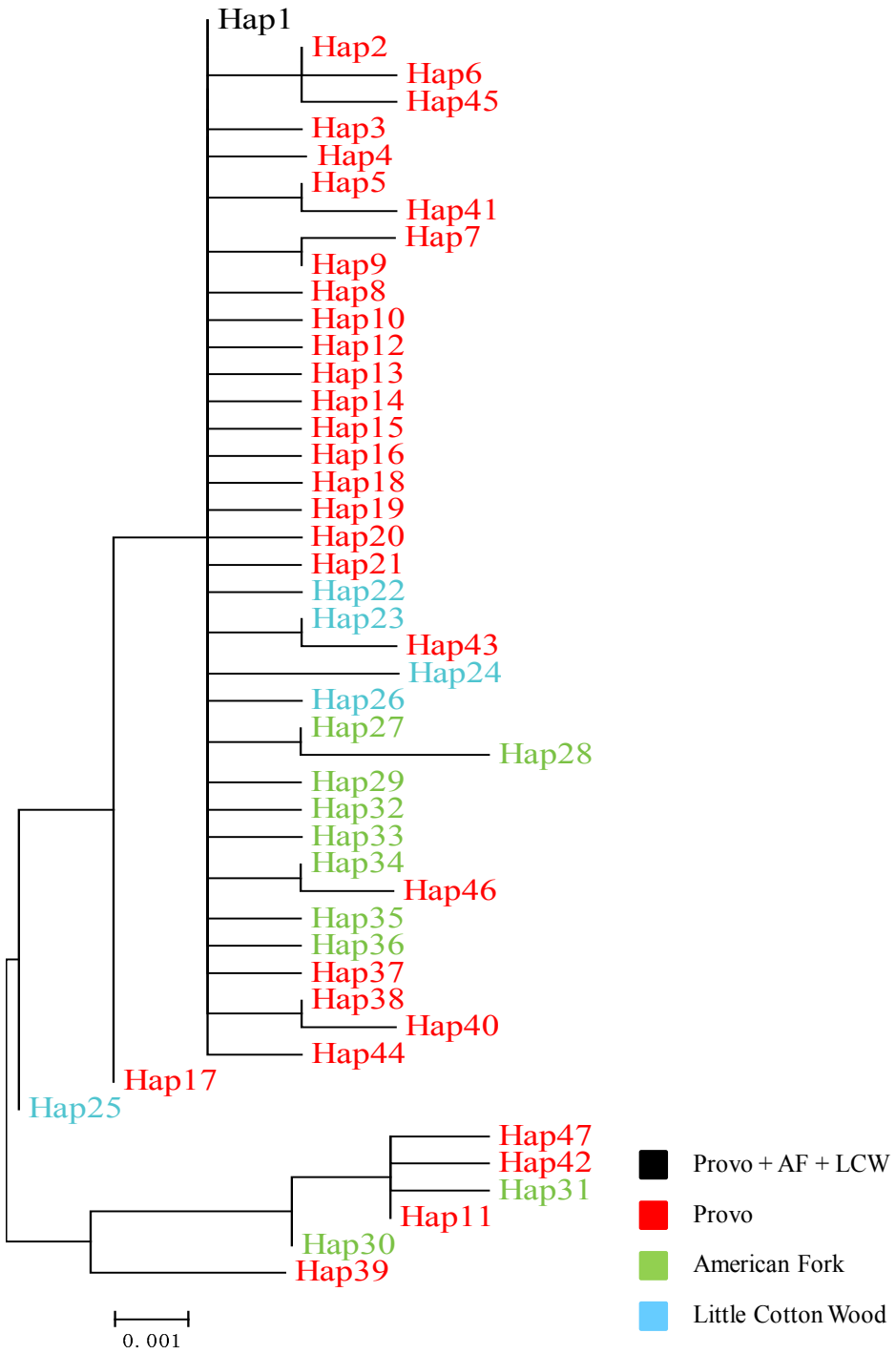
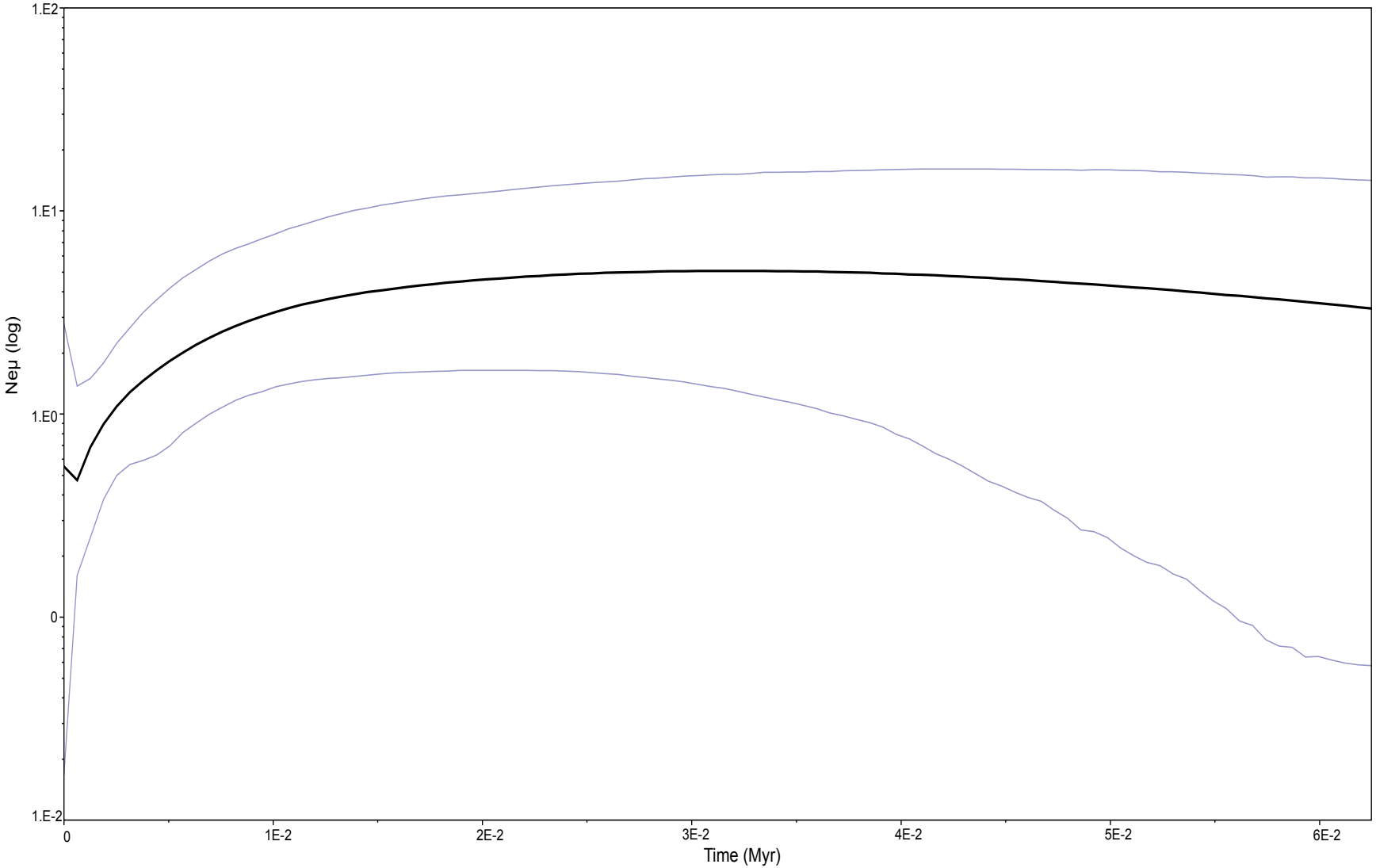

Hap1
Hap2
Hap6
Hap45
Hap3
Hap4
Hap5
Hap41
Hap7
Hap9
Hap8
Hap10
Hap12
Hap13
Hap14
Hap15
Hap16
Hap18
Hap19
Hap20
Hap21
Hap22
Hap23
Hap43
Hap24
Hap26
Hap27
Hap28
Hap29
Hap32
Hap33
Hap34
Hap46
Hap35
Hap36
Hap37
Hap38
Hap40
Hap44
Hap17
Hap25
Hap47
Hap42
Hap31
Hap11
Hap30
Hap39
0. 001
Provo + AF + LCW
Provo
American Fork
Little Cotton Wood

Figure 4.



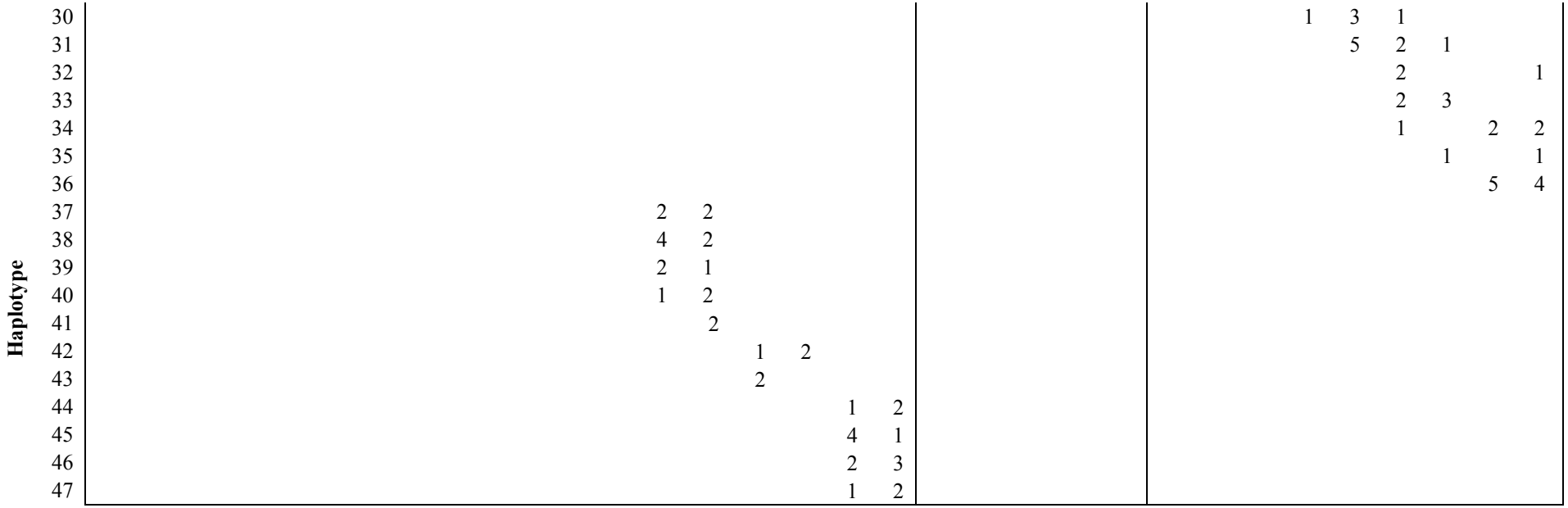
Appendix 1. GPS coordinates and elevation (m asl) for sites sampled in this study. Site numbers correspond to those used in Appendix 2.

Sites	Catchment	GPS Coordinates	Elevation
1	Provo	N 40°18.820' W111°32.368'	2013
2	Provo	N 40°19.183' W111°32.103’	1914
3	Provo	N 40°20.004' W111°31.241'	1737
4	Provo	N 40°20.858' W 111°32.728'	1683
5	Provo	N 40°20.842' W111°32.582'	1656
6	Provo	N 40°20.192' W111°33.284'	1831
7	Provo	N 40°21.111' W111°34.057'	1596
8	Provo	N 40°24.240' W111°37.245'	2315
9	Provo	N 40°24.268' W111°36.780'	2181
10	Provo	N 40°23.424' W111°35.055'	1907
11	Provo	N 40°39.129' W111°07.208'	2783
12	Provo	N 40°38.488' W111°08.318'	2618
13	Provo	N 40°37.372' W111°08.221'	2274
14	Provo	N 40°23.167' W111°36.370'	2241
15	Provo	N 40°27.521' W111°32.992'	1885
16	Provo	N 40°27.509' W111°32.928'	1851
17	Provo	N 40°34.135' W111°33.414'	2386
18	Provo	N 40°34.099' W111°33.311'	2377
19	Provo	N 40°40.749' W111°14.226'	2224

20	Provo	N 40°40.917' W111°14.369'	2202
21	American Fork	N 40°26.025' W111°38.083'	2230
22	American Fork	N 40°26.835' W111°38.335'	2023
23	American Fork	N 40°29.841' W111°39.737'	2182
24	American Fork	N 40°29.928' W111°39.805'	2198
25	American Fork	N 40°29.524' W111°39.398'	2085
26	Little Cotton Wood	N 40°34.590" W111°36.009"	2853
27	Little Cotton Wood	N 40°34.739' W111°38.195'	2805
28	Little Cotton Wood	N 40°32.839' W111°40.394'	2892
29	Little Cotton Wood	N 40°34.272' W111°41.098'	2405
30	Little Cotton Wood	N 40°33.211' W111°41.709'	2663
31	Little Cotton Wood	N 40°33.419' W111°41.840'	2587
32	Little Cotton Wood	N 40°34.202' W111°42.019'	2184
33	Little Cotton Wood	N 40°34.172' W111°43.282'	1997
34	Little Cotton Wood	N 40°34.121' W111°43.029'	2014

Appendix 2. Haplotypes generated from COI sequences presented by sites within catchments. The number of individuals from each site with a particular haplotype are shown in the body of the table. Empty cells indicate that no individuals representing the specified haplotype occurred at that site.

	Sites																																
	Provo																				American Fork					Little Cotton Wood								
Haplotype	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	7	8	6	6	5	8	6	10	6	5	7	5	10	8	6	3	4	2	5	5	6	5	5	3	6	10	10	5	11	6	4	8	7	7
2	1	1		1			2				1									1														
3	3	2															6	5																
4	4	1																																
5		3																2																
6			1		2																													
7			2		6		3																											
8				4		5																												
9				1	2	1	1																											
10				3																														
11						1											2	4																
12								2	6					1																				
13								2	1																									
14								1	2	4																								
15										2				1																				
16										3				2																				
17										1				3																				
18											2	2																						
19											2	2	1																					
20											3	3																						
21												3	4																					
22																					3	4		2	1									
23																					4	4	1											
24																					2	2	1											
25																							3	7	6									
26																							3	2										
27																										3		1	2		1		1	
28																										2	2	1						
29																											3	8	1	1				


Haplotype
30 1 3 1
31 5 2 1
32 2 1
33 2 3
34 1 2 2
35 1 1
36 5 4
37 2 2
38 4 2
39 2 1
40 1 2
41 2
42 1 2
43 2
44 1 2
45 4 1
46 2 3
47 1 2