Genetic models reveal historical patterns of sea lamprey population fluctuations within Lake Champlain

Cassidy C D'Aloia, Christina B Azodi, Sallie P Sheldon, Stephen C Trombulak, Willam R Ardren

The origin of sea lamprey (*Petromyzon marinus*) in Lake Champlain has been heavily debated over the past decade. Given the lack of historical documentation, two competing hypotheses have emerged in the literature. First, it has been argued that the relatively recent population size increase and concomitant rise in wounding rates on prey populations are indicative of an invasive population that entered the lake through the Champlain Canal. Second, recent genetic evidence suggests a post-glacial colonization at the end of the Pleistocene, approximately 11,000 years ago. One limitation to resolving the origin of sea lamprey in Lake Champlain is a lack of historical and current measures of population size. In this study, the issue of population size was explicitly addressed using nuclear (nDNA) and mitochondrial DNA (mtDNA) markers to estimate historical demography with genetic models. Haplotype network analysis, mismatch analysis, and summary statistics based on mtDNA noncoding sequences for NCI (479 bp) and NCII (173 bp) all indicate a recent population expansion. Coalescent models based on mtDNA and nDNA identified two potential demographic events: a population decline followed by a very recent population expansion. The decline in effective population size may correlate with land-use and fishing pressure changes post-European settlement, while the recent expansion may be associated with the implementation of the salmonid stocking program in the 1970s. These results are most consistent with the hypothesis that sea lamprey are native to Lake Champlain; however, the credibility intervals around parameter estimates demonstrate that there is uncertainty regarding the magnitude and timing of past demographic events.

TITLE:

Genetic models reveal historical patterns of sea lamprey population fluctuations within Lake Champlain

Authors:

1. Cassidy C. D'Aloia*; Middlebury College, Department of Biology, Middlebury, VT, USA
Current Affiliation: University of Toronto, Department of Ecology & Evolutionary Biology, Toronto, ON, CA

2. Christina B. Azodi; Middlebury College, Department of Molecular Biology and Biochemistry, Middlebury, VT, USA
Current Affiliation: Michigan State University, Department of Plant Biology, East Lansing, MI, USA

3. Sallie P. Sheldon; Middlebury College, Department of Biology, Middlebury, VT, USA

4. Stephen C. Trombulak; Middlebury College, Department of Biology and Program in Environmental Studies, Middlebury, VT, USA

5. William R. Ardren; U.S. Fish and Wildlife Service, Western New England Complex, Essex Junction, VT, USA

*Corresponding Author:
Current contact information:
Cassidy C. D'Aloia; University of Toronto, Department of Ecology & Evolutionary Biology, 25 Harbord St, Toronto, ON M5S 3G5, Canada; Phone: (647) 939-7722;
cassidy.daloia@utoronto.ca

ABSTRACT

The origin of sea lamprey (*Petromyzon marinus*) in Lake Champlain has been heavily debated over the past decade. Given the lack of historical documentation, two competing hypotheses have emerged in the literature. First, it has been argued that the relatively recent population size increase and concomitant rise in wounding rates on prey populations are indicative of an invasive population that entered the lake through the Champlain Canal. Second, recent genetic evidence suggests a post-glacial colonization at the end of the Pleistocene, approximately 11,000 years ago. One limitation to resolving the origin of sea lamprey in Lake Champlain is a lack of historical and current measures of population size. In this study, the issue of population size was explicitly addressed using nuclear (nDNA) and mitochondrial DNA (mtDNA) markers to estimate historical demography with genetic models. Haplotype network analysis, mismatch analysis, and summary statistics based on mtDNA noncoding sequences for NCI (479 bp) and NCII (173 bp) all indicate a recent population expansion. Coalescent models based on mtDNA and nDNA identified two potential demographic events: a population decline followed by a very recent population expansion. The decline in effective population size may correlate with land-use and fishing pressure changes post-European settlement, while the recent expansion may be associated with the implementation of the salmonid stocking program in the 1970s. These results are most consistent with the hypothesis that sea lamprey are native to Lake Champlain; however, the credibility intervals around parameter estimates demonstrate that there is uncertainty regarding the magnitude and timing of past demographic events.

INTRODUCTION

The origin of the landlocked population of sea lamprey (*Petromyzon marinus*) in Lake Champlain has been the subject of an ongoing debate in recent years (Bryan et al., 2005; Waldman, Grunwald & Wirgin, 2006; Waldman et al., 2009; Eshenroder, 2009, 2014). The sea lamprey is an anadromous fish that has a parasitic juvenile phase during which it feeds on the bodily fluids of a variety of prey fishes, including large salmonids such as lake trout (*Salvelinus namaycush*), Atlantic salmon (*Salmo salar*), and lake whitefish (*Coregonus clupeaformis*), as well as lake sturgeon (*Acipenser fulvescens*). Recent research has suggested that native coastal populations of sea lamprey may have positive environmental impacts on freshwater streams. For example, post-spawning sea lamprey carcasses may be important sources of marine-derived nutrients and materials in oligotrophic streams (Guyette et al., 2014). However, the overall effect of landlocked sea lamprey populations in the Great Lakes and Lake Champlain has been detrimental. Lamprey-induced collapses of native fish populations have been well documented in the region since the 1970s (Smith, 1971; Smith & Tibbles, 1980).

Consequently, control efforts were developed in the Great Lakes to suppress sea lamprey populations and facilitate restoration of native species. Control methods are varied and include widespread biocide use, physical migration barriers, and spawning-phase traps in tributaries. Although these methods have been criticized for the potential negative effects that non-target species may experience (McLaughlin, Marsden & Hayes, 2003), the effort has been largely successful in terms of lamprey control. Using the methods developed in the Great Lakes, an experimental lamprey control program was implemented in Lake Champlain in 1990, followed by a long-term program beginning in 2001. Although the control methods in Lake Champlain are similar to those in the Great Lakes, a major difference is the consensus regarding the fish's status

77 as native versus invasive; in the Great Lakes, sea lamprey are known to be invasive (with the
78 exception of Lake Ontario), while much debate surrounds the population in Lake Champlain.
79 To date, two alternative historical scenarios have dominated the discussion regarding the
80 origin of sea lamprey in Lake Champlain. First, it has been argued that individuals from the
81 Atlantic coast population invaded Lake Champlain via the canal system sometime between the
82 1840s and the 1920s (Eshenroder, 2009, 2014). This hypothesis is based on the fact that the first
83 documentation of the species in the lake was in 1929 (Eshenroder, 2014), 13 years after the
84 construction of the Champlain Barge Canal (the third version of the Champlain Canal), which
85 directly connected Lake Champlain to the Hudson River. Second, recent genetic analyses based
86 on both nuclear (nDNA) and mitochondrial DNA (mtDNA) markers suggest that sea lamprey are
87 native to Lake Champlain (Bryan et al., 2005; Waldman, Grunwald & Wirgin, 2006; Waldman et
88 al., 2009). Bryan et al. (2005) tested alternative coalescent-based colonization models using
89 microsatellite markers and found evidence of long-term vicariance in the Lake Champlain
90 population, concluding that the most probable route of entry was via the St. Lawrence River
91 upon the initial formation of Lake Champlain about 12,500 years ago. Waldman et al. (2006)
92 compared haplotype frequencies of the Lake Champlain population to Atlantic Coast and Great
93 Lakes populations using mitochondrial non-coding DNA and concluded that the data were most
94 consistent with a post-glacial colonization during a period when modern-day Lake Champlain
95 was an arm of the Atlantic Ocean called the Champlain Sea. However, neither hypothesis has
96 been widely accepted throughout the scientific and management communities. The “nonnative”
97 hypothesis has been criticized because it is based on the absence of species documentation data
98 during the 1800s — a time when systematic biological censuses were not conducted in the lake
99 — and because it lacks due consideration of the post-glacial geological history of the region.

Likewise, the "native" hypothesis has recently been called into question after an extensive review of historical documentation suggested that the time of origin used in genetic models (1841) was based on an erroneous species identification and, subsequently, may have biased the results (Eshenroder, 2014).

As a result of this ongoing debate, prior research has focused exclusively on the timing of the origin of *P. marinus* in Lake Champlain, while little is known about the rest of the population's history. The two previous genetic studies of *P. marinus* in the lake were regional studies that compared the Lake Champlain sea lamprey population to other populations from the Great Lakes and the Atlantic Ocean (Bryan et al., 2005; Waldman, Grunwald & Wirgin, 2006). While both studies found evidence for long-term vicariance from anadramous and other freshwater populations, it has proven challenging to reconcile the results with the complete lack of historical documentation of sea lamprey in the Lake Champlain (Bryan et al., 2005). Given that these regional studies have already shown the Lake Champlain population to be differentiated from all other populations, additional analyses can further probe the genetic data by conducting rigorous intra-population analyses to model historical population dynamics *within* Lake Champlain. An objective assessment of the timing and magnitude of fluctuations in population size over time may therefore provide a more complete understanding of the history of *P. marinus* in Lake Champlain, while simultaneously shedding new light on the contentious topic of whether or not the species is invasive.

In cases such as this one, where historical census data are unavailable, genetic markers can be powerful tools for inferring demographic fluctuations. These inferences are possible because the census population size (N_c) is generally proportional to the effective population size (N_e). Effective population size is the size of an idealized population (i.e., with binomial variance

123 in reproductive success, an equal sex ratio, and discrete generations), that is subject to the same
124 level of genetic drift and inbreeding as the census population (Wright, 1938). Drawing on the
125 relationship between effective and census population size, and the fact that rapid demographic
126 fluctuations can be detected with genetic markers, we can investigate how effective population
127 size has changed over time and, in turn, infer proportional changes in the overall population
128 (Waples, 1989; Frankham, 1995).

129 Traditionally, genetic approaches to estimating historical demography have used
130 summary statistics to test whether extant population-level data deviate from theoretical
131 expectations under alternative models of population stasis, contractions, and expansions (Cornuet
132 & Luikart, 1996; Harpending et al., 1998; Schneider & Excoffier, 1999; Garza & Williamson,
133 2001). For example, mismatch analysis uses sequence data to compare the distribution of
134 observed pairwise differences between all haplotypes in a population to the distribution expected
135 under a specified population change. Expansion, contractions, and equilibrium each generate a
136 particular pattern of the distribution of pairwise differences among sequences. Contractions or
137 equilibrium lead to multimodal, ragged distributions while expansions result in a smooth
138 unimodal Poisson distribution of pairwise differences (Harpending et al., 1998; Schneider &
139 Excoffier, 1999). These moment-based metrics are widely used because they are easy to obtain
140 with sequence and/or allele frequency data. However, they provide crude approximations of
141 population changes, and their precision is linked to the timing and magnitude of the demographic
142 change in question.

143 Arguably, a more powerful approach to inferring past demographic change is coalescent
144 modeling (Storz & Beaumont, 2002; Beaumont & Rannala, 2004). Coalescent theory seeks to
145 describe the ancestral relationship of a particular gene or set of genes by recognizing that the

146 probability of two lineages coalescing during a particular generation is inversely proportional to
147 effective population size at that time (Beaumont & Rannala, 2004; Kuhner, 2008;). Thus, these
148 models trace separate genetic lineages back to their most recent common ancestor (Kuhner,
149 2008) and connect these genealogies to changes in effective population size (Storz & Beaumont,
150 2002). Models such as BEAST and Msvar adopt Bayesian Markov chain Monte Carlo (MCMC)
151 methods to explore parameter space and sample the posterior distributions of the demographic
152 parameters of interest (Beaumont, 1999; Drummond & Rambaut, 2007). A key advantage to
153 these methods is the ability to include time as one of the estimated parameters, as opposed to
154 using fixed time points as assumptions in the model. Directly estimating time is particularly
155 important for models of historical sea lamprey demography, as previous genetic studies that
156 focused on the population's origin have been criticized for using fixed, potentially-incorrect
157 dates (Eshenroder, 2014).
158 The sea lamprey is a tractable study species for coalescent modeling of effective
159 population size fluctuations because several genetic resources are available. The entire
160 mitochondrial genome is sequenced (Lee & Kocher, 1995), enabling researchers to sequence
161 hypervariable regions of mitochondrial DNA (mtDNA). For decades, mtDNA has been widely
162 used to infer demographic processes because of its maternal inheritance, small effective
163 population size, and relatively fast rate of evolution (Avise, 1994; White et al., 2008). These
164 characteristics are particularly useful for demographic studies because uniparental inheritance
165 can be modeled without the complications of recombination, and signatures of relatively recent
166 demographic events are more readily detectable when effective population size is small and
167 mutation rate is elevated. Also, *P. marinus* has two non-coding regions in the mitochondrial
168 genome: Non-coding region one (NCI) is 491-bp long and non-coding region two (NCII) is 199-

bp long (Lee & Kocher, 1995). Finally, a suite of microsatellite markers has already been developed for this species (Bryan et al., 2005); thus, markers from the mitochondrial and nuclear genomes can be used concurrently to study the population's history. The concurrent use of two genomic marker types enables intra-population replication for inferring demographic history (Eytan & Hellberg, 2010).

The purpose of this study was to use multiple analytical approaches and two sets of genetic markers to investigate historical population fluctuations in Lake Champlain sea lamprey. First, moment-based methods — including a mismatch distribution and Fu's F_S statistic — were used to generate coarse estimates of historical population expansions and/or contractions based on mtDNA sequence data (NCI and NCII). Second, two coalescent MCMC models were used to explicitly estimate changes in effective population size over time. Mitochondrial sequence data (NCII) were used to generate a Bayesian Skyline Plot (BSP) in the program BEAST to model effective population size history while taking into account coalescent and phylogenetic uncertainty. Previously-published allele frequency data for eight nuclear microsatellite loci (Bryan et al., 2005) were also used in the program Msvar to estimate four demographic parameters: historical effective population size, current effective population size, mutation rate, and time. In total, we used 10 loci (n = 2 non-coding mtdna; n=8 nuclear microsatellites), and employed both moment-based methods as well as two classes of coalescent models to explore historical demography. Taken together, these data can provide insight into signatures of demographic events within two separate genomes.

METHODS

Sample collection and mitochondrial DNA sequencing

To estimate historical population fluctuations within Lake Champlain, fin tissue samples from spawning-phase sea lamprey were obtained from the US Fish and Wildlife Service (USFWS) in May-June 2009. All tissue collection was conducted as part of routine USFWS sampling; the authors of this study received tissue, but did not handle any vertebrate specimens. First, to estimate fluctuations based on mtDNA sequence data, samples were collected at three Lake Champlain tributaries: Great Chazy River (n = 33), Malletts Creek (n = 33), and Beaver Brook (n = 28), representing the northern, central, and southern regions of Lake Champlain, respectively. A broad geographic sampling regime was used in order to test for population structure. All tissue samples were stored in 95% non-denatured ethanol and genomic DNA was extracted using DNeasy Blood and Tissue Kits (Qiagen).

The two non-coding regions of the mtDNA genome, NCI and NCII, were amplified in all Lake Champlain samples using PCR with the lamprey-specific primers CR1 (Waldman et al., 2004) and LampR (5'- AATAGACGGTTGGTGGGACA - 3'). PCR reactions were performed in 25 µl volumes with the following reagents: 0.2 µl Qiagen *taq* polymerase (5 units/ µl), 2.5 µl 10X PCR buffer (with 1.5 mM $MgCl_2$), 10 µM each primer, 10 µM dNTP, and 50-100 ng template DNA. Thermal cycler settings were set at an initial denaturation at 95°C for 5 minutes followed by 40 cycles of 95° C for 45 seconds, 56.5° C for 45 seconds, and 72° C for 1 minute; and a final extension at 72° C for 10 minutes. PCR products were cleaned using the Wizard SV Gel and PCR Clean-Up Kit (Promega) and eluted in 25 µl H_2O.

To sequence the Lake Champlain samples at both non-coding regions, forward extension reactions were run for each individual using the forward primers CR1 (for NCI) and CASSFN (for NCII) (5'- GACCCCTAAGTTCATTGC - 3'). All primers designed specifically for this study, including LampR, CR1, and CASSFN, were designed using Primer3 (Untergrasser et al.,

2012). Extension reactions were prepared using ¼ reactions from the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). For each 20 µl reaction, the following reagents were combined: 2 µl Ready Reaction Mix, 3 µl 5x sequencing buffer, 3.2 µl primer (1 µM), 1 µl DNA template (20-30 ng), and 10.8 µl water. Thermal cycler settings were set at an initial denaturation at 96° C for 1 minute followed by 30 cycles of 96° C for 20 seconds, 50° C for 20 seconds, and 60° C for 4 minutes. Reactions were cleaned with 2 µl of 2.2% sodium dodecyl sulfate and returned to the thermal cycler at 98° C for 5 minutes followed by 25° C for 10 minutes. Reactions were then purified in Sephadex size-exclusion columns (GE Healthcare Life Sciences) and 12 µl of purified products were loaded into an Applied Biosystems 3130 Genetic Analyzer for sequencing.

Sequences from individuals that successfully amplified at both non-coding regions (n=54) were concatenated into a single sequence denoted by NC_{total}. Concatenated sequences were then trimmed to a 652-bp region to exclude repetitive portions of reads prone to slippage (NCI: excluded 14 bp before position 15382; NCII: excluded 28 bp after position 16173). Regions prone to slippage were identified by superimposed sequences on the chromatograms. All trimmed sequences were aligned with the MUSCLE algorithm (Edgar, 2004). Finally, a haplotype network was generated for all NC_{total} haplotypes to visualize their relationship using TCS v. 1.18 (Clement, Posada & Crandall, 2000)

Second, to estimate population fluctuations based on nuclear microsatellite data, we used previously-published allele frequency data (Bryan et al., 2005). Bryan et al. (2005) genotyped individuals from Great Chazy River (n=40) and Lewis Creek (n=40) at 8 microsatellite loci. There was no evidence for linkage disequilibrium or significant deviations from Hardy-Weinberg equilibrium (see Bryan et al., 2005 for more details).

238

239 *Moment-based analyses of historical population changes*

240 Historical population fluctuations were inferred using three moment-based methods. First, the
241 mismatch distribution of pairwise differences between NC_{total} haplotypes was plotted. The
242 observed distribution of pairwise differences was compared to the expected number of pairwise
243 differences under a model of population expansion with 1,000 bootstrap replicates in Arlequin v.
244 3.1 (Excoffier, Laval & Schneider, 2005). Second, Harpending's raggedness index (r) was used
245 to test whether the observed distribution was significantly different from the expected theoretical
246 distribution under a model of expansion (Harpending, 1994). Third, an alternative metric, Fu's F_s
247 statistic, was used to test the selective neutrality of mutations (Fu, 1997). Fu's F_s can detect an
248 excess (or deficiency) of haplotypes, given the observed haplotype diversity, thereby indicating a
249 population expansion (or contraction).

250

251 *Coalescent analyses of historical population changes*

252 Two Bayesian coalescent MCMC models were used to estimate historical demographic
253 fluctuations of *P. marinus* in Lake Champlain over time. First, the program BEAST v.1.6.2
254 (Drummond & Rambaut, 2007) was used to make inferences based on NCII sequences. This
255 region was selected because of its high concentration of polymorphic sites. BEAST applies a
256 Bayesian coalescent-based procedure, using MCMC to sample the posterior distribution of
257 genealogical trees, demographic parameters over time, and coalescent events given sequence
258 information and a set of priors. For our demographic model, we applied the Bayesian Skyline
259 Plot (BSP). BSP is a change-point model that, assuming a single panmictic population, estimates
260 fluxes in population size through time and uses a smoothing procedure to visualize these changes

(Drummond et al., 2005). To determine which nucleotide substitution model fit the data, Akaike information criteria (AIC) values were calculated in jMODELTEST v.0.1.1 (Guindon & Gascuel, 2003). The substitution model selected by AIC was then used as a prior in BEAST. For our baseline study, we assumed a strict molecular clock of 3.6×10^{-8} substitutions per base per year (i.e., 3.6% substitutions per million years), based on previous estimates of divergence in mtDNA noncoding regions in fishes (Donaldson & Wilson, 1999). The BSP group number (*m*) was set to 15. The parameter *m* allows adjacent coalescent intervals to be grouped so that they can have the same N_e—it serves to smooth the resulting BSP. The maximum effective population size was set at 10,000, a high estimate based on preliminary coalescent model runs. MCMC chains were run for 50^6 iterations in triplicate, sampling the posterior distribution every 1,000 iterations. The sampling distribution of the model was evaluated in TRACER v.1.5, with the first 10% discarded as burn-in (Rambaut et al., 2014). Quality of the MCMC convergence was assessed by the effective sample sizes (ESS): if the ESS value was less than 100, it was assumed that the MCMC chain had not been run long enough to get an accurate representation of the posterior distribution and the trace was discarded (Drummond et al., 2007).

A sensitivity analysis for the BEAST modeling was carried out on two parameters—clock rate and maximum effective population size—to assess whether priors biased parameter estimates. These parameters were selected because the priors were based on our preliminary estimates. For each parameter change, three additional MCMC chains were run keeping all other baseline input values constant. We used two alternative maximum population sizes (20,000 and 100,000) and at one alternate clock rate (2.0×10^{-8} substitutions per base per year, i.e. 2.0% substitutions per million years). To determine how sensitive model results were to the priors, runs with altered priors were compared.

284 Historical demographic changes were also inferred from coalescent modeling of nuclear
285 microsatellite data. Allele frequencies at eight microsatellite loci, previously published by Bryan
286 et al. (2005), were used as input for the Bayesian MCMC model Msvar v.1.3 (Beaumont, 1999).
287 Msvar uses probable genealogies of allele frequency data to generate posterior probability
288 distributions of four demographic parameters: current effective population size (N_0), historical
289 effective population size (N_1), mutation rate (μ), and time since the demographic change began
290 (t) (Storz & Beaumont, 2002; Beaumont & Rannala, 2004). Broad prior distributions were
291 defined for each parameter to test whether the model could detect true population fluctuations
292 (Table S1). Five independent chains were run for a panmictic sea lamprey population (Bryan et
293 al., 2005; Waldman, Grunwald & Wirgin, 2006), under a model of exponential growth, with an
294 average generation time of six years (Hardisty & Potter, 1971), and a N_e/N_c ratio of 0.2
295 (Frankham, 1995). Each chain consisted of 8×10^8 iterations, with sample points taken once
296 every 10,000 iterations. The burn-in portion of the chain was excluded by truncating runs to
297 consider only the second half. To estimate each demographic parameter, the five truncated
298 chains were combined into one posterior distribution and peak density values were recorded for
299 each parameter, along with 95% highest probability density (HPD) credibility intervals. A Bayes
300 factor was also calculated to test whether there was more support for a population expansion or
301 contraction using the method described by Storz and Beaumont (2002). Under the Bayesian
302 statistical framework, a Bayes factor is an appropriate metric for comparing alternative models.
303 Sensitivity analyses were not conducted for Msvar because the model is computationally
304 demanding; however, simulation modeling has revealed that Msvar has the power to detect true
305 population contractions even with broad priors (Girod et al., 2011). These simulation models
306 have also shown that the precision of estimates increases when demographic parameters are

scaled using coalescent theory. Thus the marginal posterior distributions of the following scaled parameters were also plotted: $\theta_0 = 4N_0\mu$ and $\theta_1 = 4N_1\mu$ (effective population sizes scaled by mutation rate); $t_f = t/(2N_0)$ (time scaled by current effective population size).

RESULTS

Haplotype diversity & population structure

We observed 14 unique haplotypes among the 54 NC_{Total} mitochondrial sequences (Table 1). The haplotype network of all 14 NC_{total} haplotypes visualizes their relationships as well as their relative abundance (Fig. 1). Haplotype 1 is predominant in Lake Champlain, with five other rare haplotypes (2, 3, 6, 7, and 9) having only one base pair difference from haplotype 1. Most of the mutations within these five haplotypes are point mutations within NCI or the 5′ end of NCII (Table 1). In contrast, most of the remaining rare haplotypes branch off of haplotype 7 and are characterized by mutations within the A/T-rich, repetitive 3′ region of NCII. In general, these mutations in the repetitive region were retained in each subsequent repeat of the sequence, leading to a larger number of overall mutations in NCII.

An exact test of population differentiation revealed that haplotypes were randomly distributed across the three sampling locations (n=54; p=0.265; # dememorization steps = 10,000; # steps Markov chain = 100,000). Thus, the Lake Champlain population is considered to be panmictic, consistent with results from previous studies (Bryan et al., 2005; Waldman, Grunwald & Wirgin, 2006).

Moment-based analyses of historical population changes

329 The mismatch distribution of NC_{total} was smooth and unimodal, suggesting that the sea lamprey
330 population has undergone an expansion (Fig. 2). The distribution of observed pairwise
331 differences closely matches the expected distribution of pairwise differences under a model of
332 population expansion, with a raggedness index that was positive, but not significant (r = 0.04; n
333 = 54; p = 0.86). This indicates that there was no significant deviation from the theoretical model
334 of expansion. Additionally, Fu's F_s statistic was significantly negative (F_s = -6.61; n = 54; p =
335 0.02), indicating an excess of rare haplotypes, which would be predicted under a scenario of a
336 recent population expansion. Overall, these moment-based methods of studying historical
337 demography with mtDNA data strongly support a population expansion, without explicit
338 estimates of the timing and magnitude of the event.
339
340 *Coalescent analyses of historical population changes*
341 For the BEAST analysis of mtDNA NCII data, the best-fit model of sequence evolution
342 determined by jMODELTEST was Hasegawa, Kishino and Yano (HKY) + I, where I means
343 there is a significant proportion of invariable sites. The BSP derived from the NCII data shows a
344 decrease in effective population size starting around 400 years ago that continued until 50 to 100
345 years ago, at which point effective population size slowly began to increase (Fig. 3). The
346 sensitivity analysis showed that the maximum population size prior was directly proportional to
347 the current N_e estimate; when doubled, the N_e estimates also doubled. However, changes in the
348 maximum population size and clock rate priors did not affect the overall pattern of the BSP
349 (Table S2).
350 In contrast to the BEAST analysis, coalescent modeling of microsatellite alleles estimated
351 one historical population contraction. The five Msvar chains converged below a critical potential

scale reduction factor of 1.2 (Brooks & Gelman, 1998) for all four demographic parameters, indicating that there was good convergence of parameter estimates. The Bayes factor (BF) for a population contraction indicated very strong support for a population contraction over a population expansion (2*ln(BF) = 10.3) (Kass & Raftery, 1995), with the vast majority of MCMC iterations estimating a population contraction (Fig. 4a). The posterior density distributions reveal the estimates for each demographic parameter with 95% HPD intervals (Fig. 4b-d). There was a decrease in effective population size from approximately 2,660 ($HPD_{95\%}$ = 153, 65750) down to approximately 50 currently ($HPD_{95\%}$ = 0.06, 1440) (Fig. 4b). The density peak was higher, with narrower credibility intervals, for historical effective population size compared to current effective population size. This population contraction was estimated to have begun approximately 820 years ago, although the peak density was only 40% for this time parameter, with wide credibility intervals ($HPD_{95\%}$ = 0.81, 71558) (Fig. 4c). Finally, there was a high density peak for the mutation rate estimate, with over 80% of the estimates approaching a modal value of 3.26×10^{-4} mutations/site/generation ($HPD_{95\%} = 3.52 \times 10^{-5}$, 2.78×10^{-3}) (Fig. 4d).

The posterior distributions of the scaled demographic parameters also indicate a population contraction (Fig. 5). As predicted by simulation analyses, scaling the parameters increased the precision of parameter estimates. The magnitude of the scaled population size estimates were very similar to the unscaled estimates. Modal θ estimates corresponded with a decline from approximately 2,335 individuals historically ($HPD_{95\%}$ = 359, 29614) to 60 individuals currently ($HPD_{95\%}$ = 0.08, 772). Despite increased precision, there was still some overlap in the 95% credibility intervals for θ (Fig. 5A). In contrast, there was a substantial increase in precision of the time estimate, suggesting a more ancient demographic event: the

375 modal value of the scaled time distribution corresponded with an event beginning approximately
376 1,230 years ago ($HPD_{95\%}$ = 128, 7522) (Fig. 5B).
377

378 **DISCUSSION**

379 Genetic-based methods can be useful tools for studying demographic changes in the absence of
380 historical population records, though their precision is linked to the timing and magnitude of the
381 events themselves (Beaumont & Rannala, 2004; Lawton-Rauh, 2008). In this study, we applied
382 multiple analytical approaches and used two sets of genetic markers to investigate the population
383 history of *P. marinus* in Lake Champlain. Synthesizing the results to draw conclusions about the
384 most likely historic scenario of demographic change requires a critical analysis of the results
385 generated by each method, an exploration of sea lamprey ecology, and consideration of the
386 history of land-use and management practices in the region.
387

388 *Moment-based analyses of historical population changes*

389 Mismatch distribution analysis is based on the assumption that demographic changes leave
390 corresponding genetic signatures in neutral sequence data and gene trees (Rogers & Harpending,
391 1992; Harpending et al., 1998). The mismatch distribution of this study fits a model of
392 population expansion well, being both smooth and unimodal. However, a major weakness of this
393 method is its poor ability to make specific inferences about the timing of demographic changes.
394 In general, when using moment-based methods, only a general inference as to whether the
395 change was relatively recent or ancient can be made. The presence of many low-frequency
396 mutations is one indication of a 'recent' expansion (Schneider & Exoffier, 1999). This pattern
397 was evident in the haplotype network, in which the majority of haplotypes in Lake Champlain

398 were present in only one to three individuals (Fig. 1). Additionally, the mean number of pairwise
399 differences in the mismatch distribution can be used for a crude estimate of timing: a 'low' mean
400 can indicate a 'recent' expansion while a 'high' mean can indicate a more 'ancient' expansion
401 (Okello et al., 2005). The haplotypes in the Lake Champlain lamprey population had a mean of
402 1.74 mismatches, which supports a 'recent' expansion. Yet, without explicit guidelines for
403 differentiating between 'low' and 'high', or inferring what 'recent' and 'ancient' indicate on an
404 evolutionary timescale, caution should be used in interpreting these results. Instead, these results
405 are more appropriately used as supplemental evidence to be considered alongside inferences
406 made by coalescent-based analyses.
407
408 *Coalescent analyses of historical population changes*
409 BEAST and Msvar analyses both detected an initial decline in effective population size. These
410 congruent results between two separate genomes strengthen the evidence for a decline; however,
411 the magnitude and timing of the event(s) remain uncertain. The BSP generated using BEAST
412 provides credibility intervals at every time point that reflect both coalescent and phylogenetic
413 uncertainty. This proved to be important, as the BSP shows that the credibility intervals for
414 effective population size ranged from 51 to 9,500. These wide credibility intervals around N_e
415 could be due to the limited number of segregating sites within the NCII sequence data. The
416 sensitivity analysis also showed that the maximum population size prior affected the BSP results,
417 with the estimated N_e being proportional to the prior. However, the overall pattern of
418 demographic change was consistent regardless of the priors (see Table S2), indicating that the
419 estimated mean N_e always showed a contraction followed by an expansion.

420 In contrast to BEAST, Msvar estimates four parameters simultaneously and can only
421 detect the single most likely demographic trend over time based on the strongest genetic signal.
422 Bayesian models, such as Msvar, require prior distributions for each parameter (Storz &
423 Beaumont, 2002; Beaumont & Rannala, 2004). Generating priors can be difficult when no
424 accurate estimates of demographic parameters are available, as is the case for sea lamprey in
425 Lake Champlain. Estimates of current sea lamprey population size are poor because census data
426 are inferred from wounding rates on prey species. Moreover, no historical documentation of sea
427 lamprey population size in Lake Champlain exists. In the absence of reliable data, broad and
428 equivalent priors were used for current and historical effective population size (Goossens et al.,
429 2006). Furthermore, through simulation analyses, Girod et al. (2011) showed that Msvar is
430 powerful at detecting population fluctuations even without informed priors if the demographic
431 event was sufficiently large and ancient. The contraction detected in this analysis indicates that
432 effective population size has declined from roughly 2,660 to 50 (a low but plausible estimate
433 given the aggressive population control program and the fish's high fecundity). It is important to
434 note that these values represent best estimates based on the mode of the posterior density
435 distributions, but that the 95% credibility intervals of historical and current population size
436 overlap due to the uncertainty associated with the current effective population size parameter.
437 Scaling the population size parameters by the mutation rate did increase precision, but there was
438 still overlap in the credibility intervals. Thus, for a comprehensive population history, the Msvar
439 results must be interpreted alongside results from the BEAST and moment-based analyses.
440
441 *Scenario 1: Evidence for fluctuations within a native population*

Capitalizing on multiple analytical approaches and two unique genetic data sets used in this study, it is possible to infer multiple demographic fluctuations. While the mismatch distribution detected a population expansion and coalescent modeling using Msvar detected a decline, the results may not be in conflict. Instead, they may be detecting the two different demographic changes that are both evident in the BSP generated by BEAST. By considering known events in the ecological history of Lake Champlain and the approximate time scale estimated by each method, we present the most parsimonious demographic history of a native sea lamprey population in Lake Champlain.

The decline in effective population size detected by both coalescent models may correlate with land-use changes and fishing pressures that began after the arrival of European settlers in the mid-18th century. The BSP from BEAST indicated that the decline began approximately 400 years ago. Given that there is uncertainty around this time estimate, one hypothesis is that a gradual decline in population size began around 1753 when a large number of European settlers arrived in the area. At this time, human impact on the landscape increased with the onset of large-scale agricultural and clear-cutting practices and, later, the construction of mills and dams that would have limited upstream lamprey spawning migrations (Klyza & Trombulak, 1999). Moreover, the acceleration of the decline shown in the BSP correlates with the extirpation of salmonids in Lake Champlain—the primary food source of parasitic-phase sea lamprey. The last documentation of native Atlantic salmon in the basin was 1830, with lake trout extirpation following shortly thereafter in the 1890s (Fisheries Technical Committee, 2009). Thus, it is plausible that the population contraction detected by both coalescent models can be explained by known ecological changes in the region, though some caution is warranted given that the scaled Msvar distributions suggest that the decline was more ancient (ca. 1230 years ago).

The mismatch distribution analysis may be detecting a more recent population expansion associated with salmonid stocking in Lake Champlain. This recent expansion is supported by the upward trend in effective population detected in the BSP in the last 50 years. A stocking program began for both Atlantic salmon and lake trout in 1972 and now focuses on stocking the main part of the lake with yearlings (Marsden et al., 2003; Fisheries Technical Committee, 2009). The reintroduction of their primary food source could have allowed the sea lamprey population to expand.

While this demographic history is plausible given the ecological context of Lake Champlain over the last 300 years, it is important to consider whether the data represent two real population changes, or whether these changes are artifacts of the models. To address this issue, an important consideration is the difference in mutation rates between the nuclear and mitochondrial markers used in the study, as faster-evolving markers generally have a greater power to detect more recent events. While the mitochondrial genome tends to have a higher mutation rate than the nuclear genome due to the mutagenic properties of respiration by-products and the limited DNA repair mechanisms of the mitochondrial genome (White et al., 2008), nuclear microsatellites have an even higher average mutation rate due to strand slippage. Indeed, Msvar estimated a fast mutation rate for the microsatellites used in the study, with the narrow posterior distribution peaking at 3.24×10^{-4} mutations/site/generation. Nevertheless, Msvar, a model that uses microsatellite markers exclusively, has been shown to have a bias towards detecting ancient declines in effective population size (Beaumont, 1999). Most recently, this tendency has been rigorously confirmed through simulation studies (Girod et al., 2011). Therefore, even if there were signatures of two demographic fluctuations in the population's history, Msvar, which can only identify a single event, is predicted to detect the more ancient

488 decline. As to whether or not the mtDNA sequence data show a true signature of expansion
489 within the last 40 years, some insight can be gained from a comparison between BEAST model
490 inputs and outputs. Even though an extremely strict (i.e., slow) molecular clock was used as an
491 input in this analysis, the posterior distribution of effective population size still shows an upward
492 trend beginning approximately 50 years ago. Posterior estimates that deviate from prior
493 predictions generally indicate a strong genetic signal. Taken together, these lines of evidence
494 suggest that the population fluctuations detected are likely to be real.
495
496 *Scenario 2: Evidence for a founder event in the early 20th century*
497 Differentiating between founder events resulting from an invasion versus bottlenecks within
498 native populations can be challenging, as both events are predicted to lead to a decline in
499 effective population size and a loss of genetic diversity (Nei, Maruyama & Chakraborty, 1975).
500 Thus, an alternative interpretation of the data is that the population contraction detected in Msvar
501 represents a bottleneck immediately following an invasion from the anadromous Atlantic Coast
502 population sometime after the completion of the Champlain Barge Canal in 1916, while the
503 expansion detected in the moment-based and BEAST analyses represents the population boom in
504 the 1970s. While the timing of such an invasion falls within the bounds of the BEAST and
505 unscaled Msvar credibility intervals, both coalescent models assume a single panmictic
506 population. Therefore, if the present-day Lake Champlain population is significantly
507 differentiated from the anadromous population, the coalescent models could not accurately
508 estimate historical changes beginning in the differentiated population of origin. Indeed, Bryan et
509 al. (2005) found evidence for significant genetic structure between Lake Champlain and

510 anadromous sea lamprey populations, and long-term vicariance for the Lake Champlain
511 population.
512 The mitochondrial haplotype network can provide additional insight into founder events:
513 the random sampling associated with a *recent* founder event should lead to the loss of rare
514 haplotypes. Therefore, a recent founder event is unlikely given the observed network, which
515 features a number of rare haplotypes branching off of a predominant ancestral haplotype (Fig. 1).
516 In sum, the majority of the evidence from the genetic models suggests that the population is
517 native to Lake Champlain, but uncertainty remains and a founder event associated with an
518 invasion in the early 1900s cannot be completely excluded given the genetic data that are
519 presently available.
520
521 **CONCLUSIONS AND FUTURE DIRECTIONS**
522 This study builds upon previous research by modeling historical sea lamprey population
523 fluctuations in Lake Champlain. Considering both coalescent models and moment-based genetic
524 approaches, we conclude that multiple demographic events are likely to have occurred over the
525 past 300 years. Importantly, however, there is a large amount of uncertainty around these
526 estimates. While we argue that the data largely align with prior genetic studies and are most
527 consistent with the native hypothesis, the wide credibility intervals around our estimates cannot
528 exclude an alternative interpretation that a founder event occurred in the early 20th century.
529 As such, we propose two potential lines of future research aimed at resolving the
530 residency debate. First, expanded genomic sampling could provide more accurate estimates of
531 historical population sizes within Lake Champlain, as well as the timing of divergence between
532 the Lake Champlain and Atlantic populations. Second, given the uncertain results from neutral

533 genetic data, it would be useful to investigate genes that may be under selection, e.g., those
534 regulating Na/K-ATPase pumps in the gills**.** Population genetic analyses, gene expression
535 analyses, and physiological saltwater challenges could be undertaken in tandem. Together, these
536 lines of research may provide further clarity to the history of sea lamprey in Lake Champlain.
537

538 **ACKNOWLEDGEMENTS**

539 The authors thank U.S. Fish and Wildlife Service professionals in Essex Junction, VT for
540 providing sea lamprey tissue. We also thank Vicenta Hudziak and Livingston Burgess for
541 assistance in DNA sequencing and Jeremy Ward, Wayne Bouffard, Steve Smith, and Brad
542 Young for providing feedback. We are grateful to John Waldman, Amy Russell, and one
543 anonymous reviewer for helpful comments on the manuscript. The findings and conclusions in
544 the article are those of the authors and do not necessarily represent the views of the USFWS.

545 **REFERENCES**

546 Avise JC. 1994. *Molecular Markers, Natural History and Evolution*. Norwell, MA: Chapman &
547 Hall.

548 Beaumont MA. 1999. Detecting population expansion and decline using microsatellites.
549 *Genetics* 153:2013-2029.

550 Beaumont MA, Rannala B. 2004. The Bayesian revolution in genetics. *Nature Reviews Genetics*
551 5:251-261.

552 Brooks SP, Gelman A. 1998. General methods for monitoring convergence of iterative
553 simulations. *Journal of Computational and Graphical Statistics* 7:434-455.
554
555 Bryan MB, Zalinksi D, Filcek B, Libants S, Li W, Scribner KT. 2005. Patterns of invasion and
556 colonization of the sea lamprey (*Petromyzon marinus*) in North America as revealed by
557 microsatellite genotypes. *Molecular Ecology* 14:757-3773.

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

Cornuet JM, Luikart G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144:2001-2014.

Donaldson KA, Wilson Jr RR. 1999. Amphi-panamic geminates of snook (Percoidei: Centropomidae) provide a calibration of the divergence rate in the mitochondrial DNA control region of fishes. *Molecular Phylogenetics and Evolution* 13:208-213.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* 22:1185-1192.

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

Drummond AJ, Ho SYW, Rawlence N, Rambaut A. 2007. A Rough Guide to BEAST 1.4. University of Auckland, New Zealand.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792-1797.

Eshenroder RL. 2009. Comment: Mitochondrial DNA analysis indicates sea lampreys are indigenous to Lake Ontario. *Transactions of the American Fisheries Society* 138:1178-1189.

Eshenroder RL. 2014. The role of the Champlain Canal and Erie Canal as putative corridors for colonization of Lake Champlain and Lake Ontario by sea lampreys. *Transactions of the American Fisheries Society* 143:634-649.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.1: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50.

Eytan RI, Hellberg ME. 2010. Nuclear and mitochondrial sequence data reveal and conceal different demographic histories and population genetic processes in Caribbean reef fishes. *Evolution* 64:3380-3397.

Fisheries Technical Committee of the Lake Champlain Fish and Wildlife Management Cooperative. 2009. Strategic Plan for Lake Champlain Fisheries. Lake Champlain Fish and Wildlife Management Cooperative, USFWS, Essex Junction, Vermont.

Frankham R. 1995. Effective population size/adult population size ratios in wildlife: a review. *Genetics Research* 2:91-107.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915-925.

Garza JC, Williamson EG. 2001. Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* 10:305-318.

Girod C, Vitalis R, Leblois R, Fréville H. 2011. Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the Msvar method. *Genetics* 188:165-179.

Goossens B, Chikhi L, Ancrenaz M, Lackman-Ancrenaz I, Andau P, Bruford MW. 2006. Genetic signature of anthropogenic population collapse in orangutans. *PLoS Biology* 4:e25.

Guindon S, Gascuel O. 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* 52:6965-704.

Guyette MQ, Loftin CS, Zydlewski J, Cunjak R. 2014. Carcass analogues provide marine subsidies for macroinvertebrates and juvenile Atlantic salmon in temperate oligotrophic streams. *Freshwater Biology* 59:392-406.

Hardisty MW, Potter IC, eds. 1971. *The biology of lampreys,* vol. 1. New York: Academic Press.

Harpending RC. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* 66:591-600.

Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST. 1998. Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences USA* 95:1961-1967.

Kass RE, Raftery AE. 1995. Bayes Factors. *Journal of the American Statistical Association* 90:773-795.

Klyza CM, Trombulak SC. 1999. *The Story of Vermont: A Natural and Cultural History*. Hanover: University Press of New England.

Kuhner MK. 2008. Coalescent genealogy samplers: windows into population history. *Trends in Ecology & Evolution* 24:86-93.

Lawton-Rauh A. 2008. Demographic processes shaping genetic variation. *Current Opinion in Plant Biology* 11:103-109.

Lee WJ, Kocher TD. 1995. Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. *Genetics* 139:873-887.

Marsden JE, Chipman BD, Nashett LJ, Anderson JK, Bouffard W, Durfey L, Gersmehl JE, Schoch WF, Staats NR, Zerrenner A., 2003. Sea lamprey control in Lake Champlain. *Journal of Great Lakes Research* 29:655-676.

McLaughlin RL, Marsden JE, Hayes DB. 2003. Achieving the benefits of sea lamprey control while minimizing effects on nontarget species: conceptual synthesis and proposed policy. *Journal of Great Lakes Research* 29:755-765.

Nei M, Maruyama T, Chakraborty R. 1975. The bottleneck effect and genetic variability in populations. *Evolution* 29:1-10.

Okello JB, Nyakaana S, Masembe C, Siegismund HR, Arctander P. 2005. Mitochondrial DNA variation of the common hippopotamus: evidence for a recent population expansion. *Heredity* 95:206-215.

632 Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6. Available at
633 http://beast.bio.ed.ac.uk/Tracer (accessed 29 July 2015).

634 Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of pairwise
635 genetic differences. *Molecular Biology and Evolution* 9:552-569.

636 Schneider S, Excoffier L. 1999. Estimation of past demographic parameters from the distribution
637 of pairwise differences when the mutation rates vary among sites: application to human
638 mitochondrial DNA. *Genetics* 152:1079-1089.

639 Smith BR. 1971. Sea lampreys in the Great Lakes of North America. In: Hardisty MW, Potter IC
640 meds. *The Biology of Lampreys*, vol. 1. New York: Academic Press, 207-247.

641 Smith BR, Tibbles JJ. 1980. Sea Lamprey (*Petromyzon marinus*) in Lakes Huron, Michigan, and
642 Superior: history of invasion and control, 1936– 1978. *Canadian Journal of Fisheries*
643 *and Aquatic Sciences* 37:1780-1801.

644 Storz JF, Beaumont MA. 2002. Testing for genetic evidence of population expansion and
645 contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian
646 model. *Evolution* 56:154-166.

647 Untergrasser A. Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012.
648 Primer3 – new capabilities and interfaces. *Nucleic Acids Research* 40:e115.

649 Waldman JR, Grunwald C, Roy NK, Wirgin I. 2004. Mitochondrial DNA analysis indicates sea
650 lampreys are indigenous to Lake Ontario. *Transactions of the American Fisheries Society*
651 133:950-960.

652 Waldman JR, Grunwald C, Wirgin, I. 2006. Evaluation of the native status of sea lampreys in
653 Lake Champlain based on mitochondrial DNA sequencing analysis. *Transactions of the*
654 *American Fisheries Society* 135:1076-1085.

655 Waldman JR, Daniels R, Hickerson M, Wirgin I. 2009. Mitochondrial DNA analysis indicates
656 sea lampreys are indigenous to Lake Ontario: response to comment. *Transactions of the*
657 *American Fisheries Society* 138:1190-1197.

658 Waples R. 1989. A generalized approach for estimating effective population size from temporal
659 changes in allele frequencies. *Genetics* 121:379-391.

660 White DJ, Wolff JN, Pierson M, Gemmell NJ. 2008. Revealing the hidden complexities of
661 mtDNA inheritance. *Molecular Ecology* 17:4925-4942.

662 Wright S. 1938. Size of population and breeding structure in relation to
663 evolution. *Science* 87:430-431.

Table 1 (on next page)

Polymorphic sites among 14 mtDNA haplotypes from the concatenated non-coding region sequences (NC_{total}), relative to the most common haplotype (first entry).

Base pair positions relative to the reference mitochondrial genome are provided (Lee and Kocher, 1995). Dots indicate no change, dashes indicate a deletion, and G/C/A/T represent point mutations.

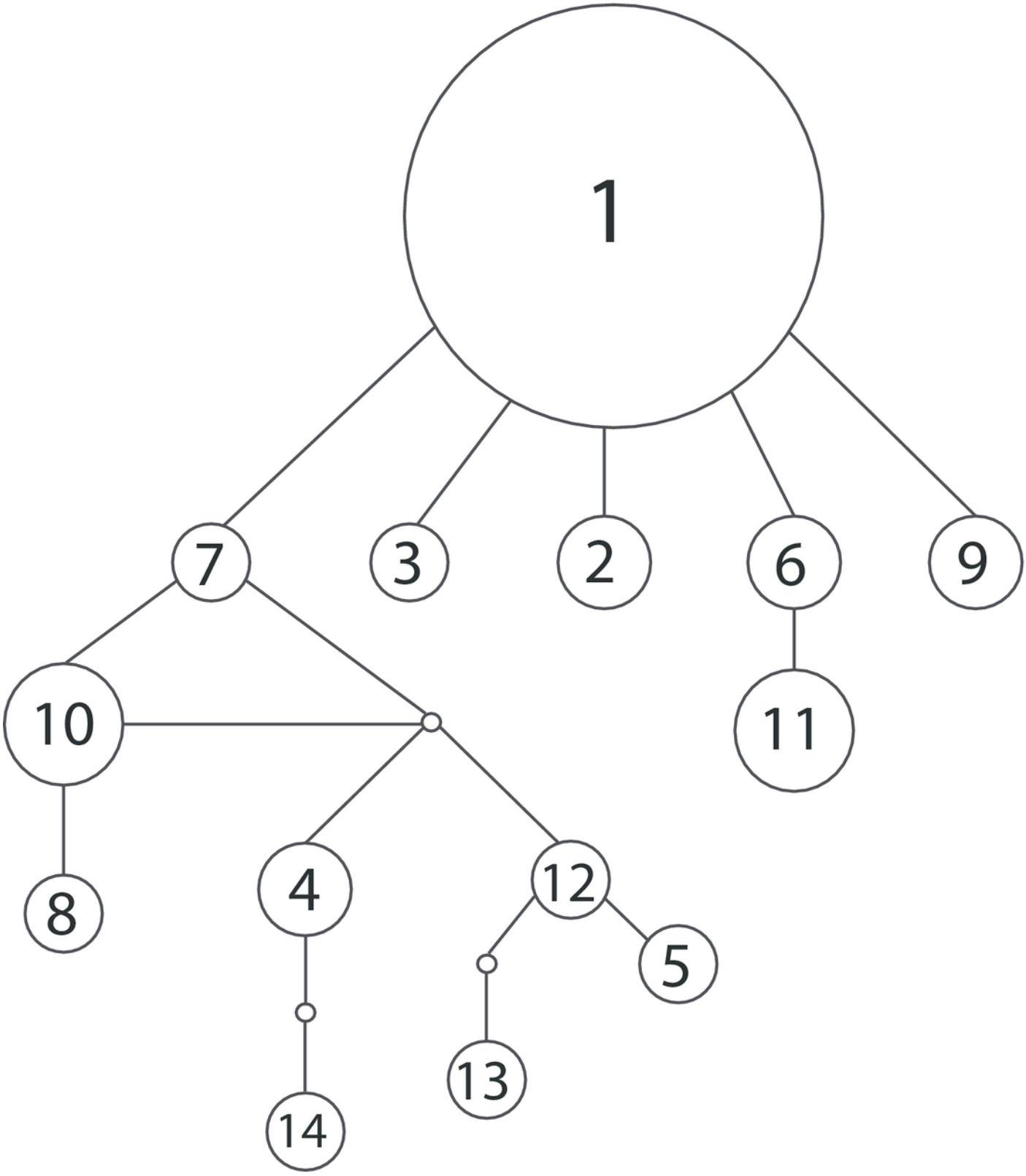
		NCI			NCII								
ID	**Genbank Accession #**	**15398**	**15402-15403**	**15403**	**16049-16050**	**16062**	**16112**	**16114**	**16115**	**16134**	**16139**	**16142**	**16169**
1	GU459340	G	–	C	T	T	T	T	C	T	T	C	C
2	GU459341	.	C	.	.	.	.	.	.	.	.	.	.
3	GU459342	.	–	.	.	–	.	.	.	.	.	.	.
4	GU459343	.	–	.	.	.	.	.	T	.	.	T	T
5	GU459344	.	–	.	.	.	A	.	–	.	.	T	T
6	GU459345	.	–	.	–	.	.	.	.	.	.	.	.
7	GU459346	.	–	.	.	.	.	.	.	.	.	.	T
8	GU459347	.	–	.	.	.	.	.	.	A	.	–	T
9	GU459348	A	–	.	.	.	.	.	.	.	.	.	.
10	GU459349	.	–	.	.	.	.	.	.	.	.	–	T
11	GU459350	.	–	–	–	.	.	.	.	.	.	.	.
12	GU459351	.	–	.	.	.	.	.	–	.	.	T	T
13	GU459352	.	–	.	.	.	.	–	–	.	A	T	T
14	GU459353	.	–	.	.	.	.	.	T	.	A	T	.

1

1

Haplotype network for 14 concatenated mtDNA haplotypes found in Lake Champlain, constructed using TCS v1.2.1 with 95% parsimony.

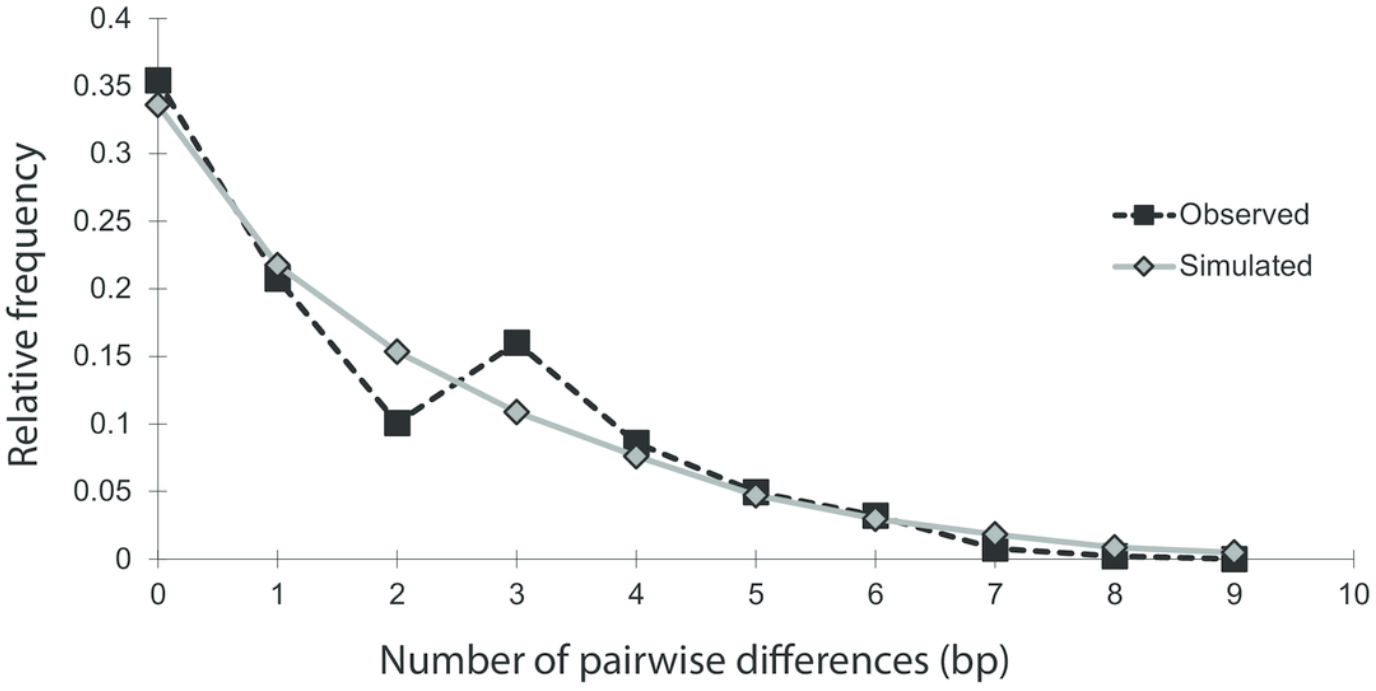
The size of the circle is proportional to the relative abundance of the haplotype. Solid lines represent one point mutation and small, unfilled circles represent inferred haplotypes. The mutations resulting in branching off of haplotype 7 are concentrated in the repetitive region of the NCII 3’ region, which has an elevated mutation rate due to strand slippage.


1
7
3
2
6
9
10
11
8
4
12
5
13
14

2

Mismatch distribution of the 14 concatenated mtDNA haplotypes, conducted in Arlequin v.3.1.

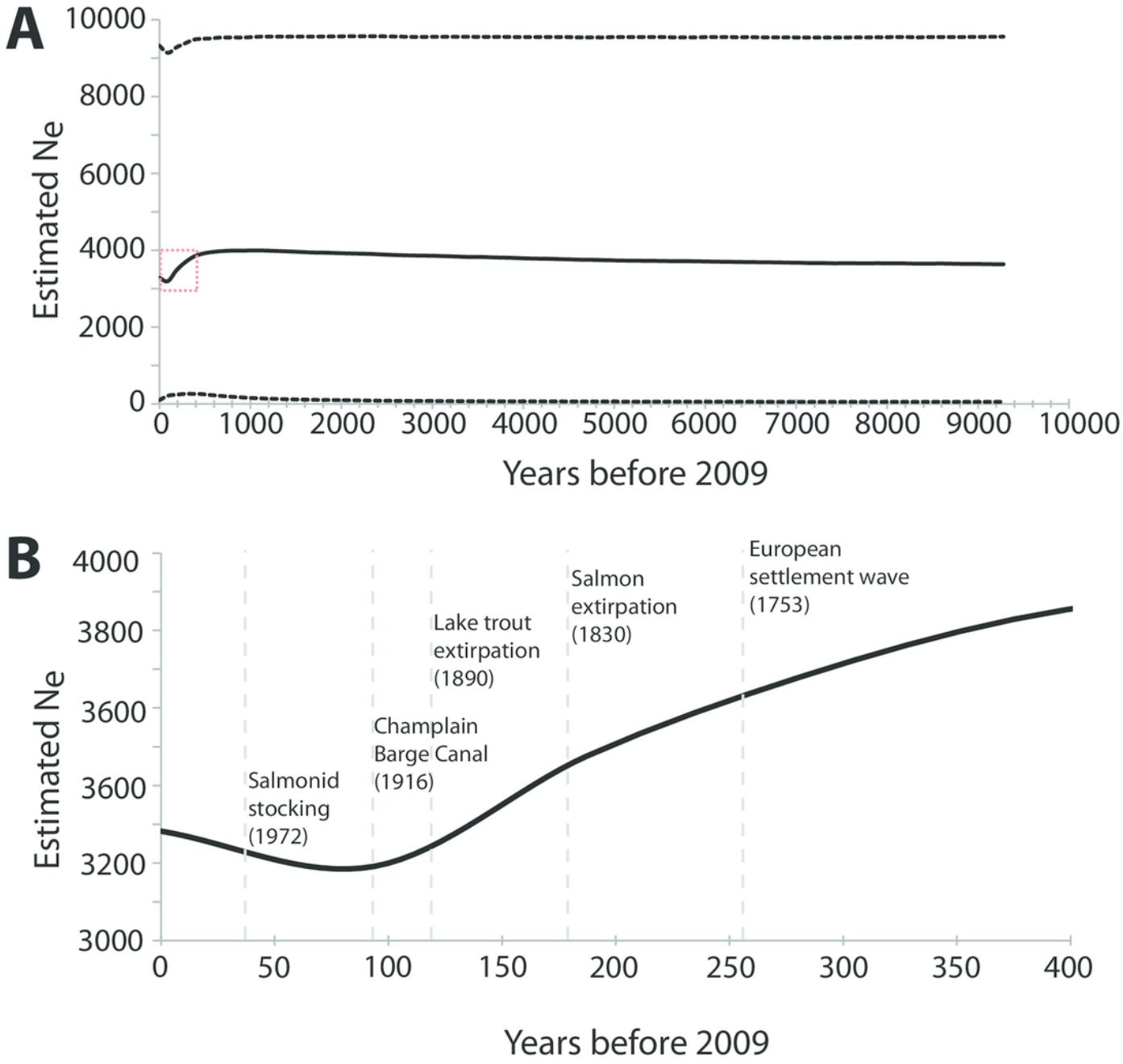
The black squares represent the observed number of pairwise bp differences between haplotypes, while the gray diamonds represent the expected number of pairwise bp differences between haplotypes, based on a model of population expansion (bootstrap replicates = 1,000). The relatively smooth and unimodal shape of the observed distribution closely matches the expected distribution for a demographic expansion.



3

Bayesian skyline plot (BSP) derived from NCII sequence alignments from Lake Champlain lamprey collected in May-June 2009.

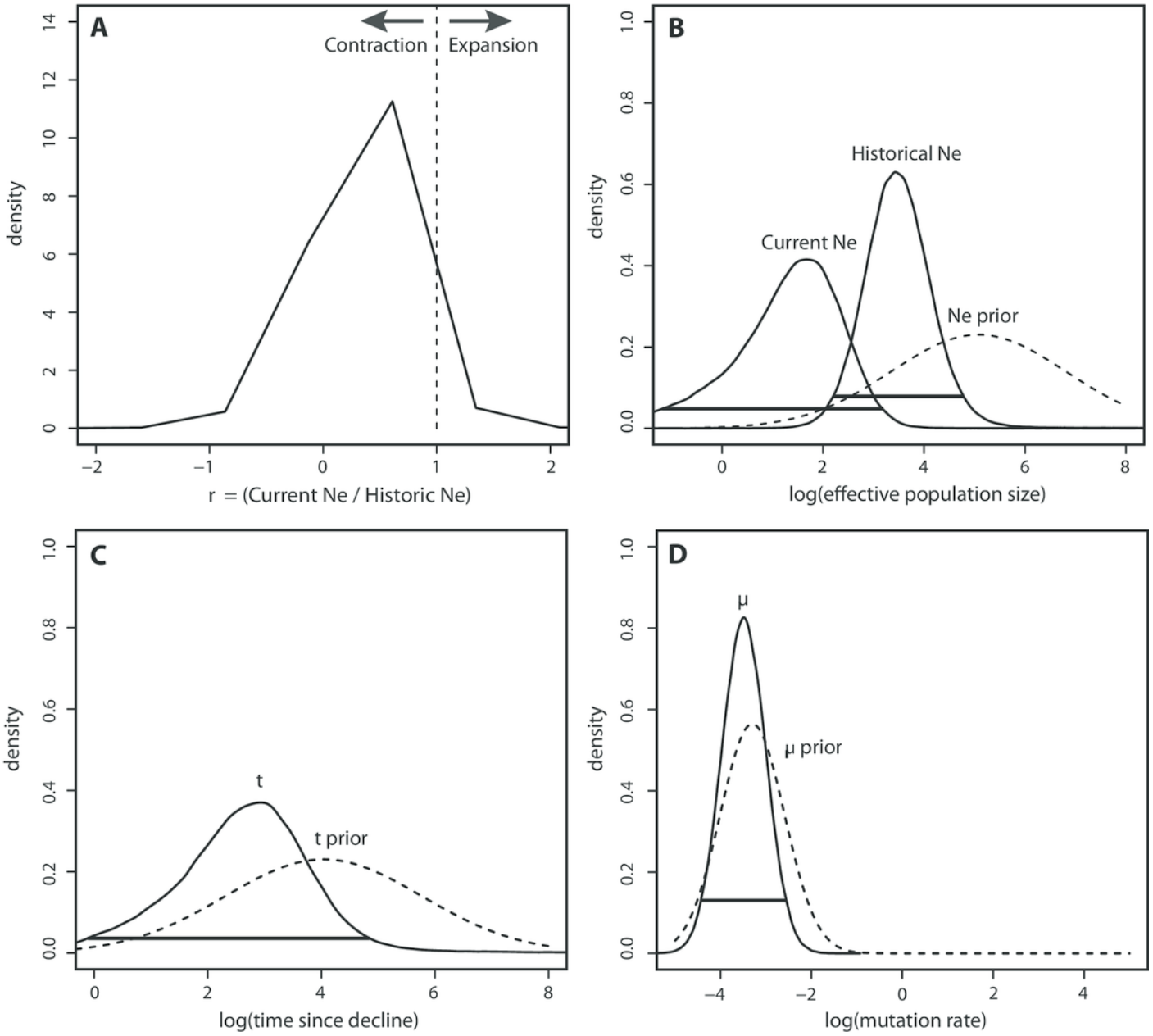
A) The mean effective population size is estimated from the Bayesian posterior distribution and is shown as the thick solid line. The horizontal dashed lines show the 95% HPD intervals around the N_e estimate. The dashed red box highlights the portion of the BSP shown in panel B; B) Zoom-in of the BSP for 400 years prior to 2009. The dashed vertical gray lines represent the timing of relevant historical events in the region. Note that the x and y axes differ between the two panels.


A
Estimated Ne
10000
8000
6000
4000
2000
0
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
Years before 2009
B
Estimated Ne
4000
3800
3600
3600
3200
3000
Salmonid stocking (1972)
Champlain Barge Canal (1916)
Lake trout extirpation (1890)
Salmon extirpation (1830)
European settlement wave (1753)
0 50 100 150 200 250 300 350 400
Years before 2009

4

Msvar results based on microsatellites.

A) Density plot across MCMC iterations for r, which represents the ratio of current N_e to historical N_e, with r > 1 indicating a population expansion and r < 1 indicating a population contraction. B—D) Posterior density distributions for each demographic parameter, plotted on a log10 scale. Solid lines represent posterior density distributions and dashed lines represent prior distributions (with solid horizontal lines representing the 95% HPD credibility intervals for the posterior distributions); B) Effective population sizes; C) Time in years since population contraction; D) Mutation rate as # mutations/site/generation.


A
Contraction
Expansion
density
r = (Current Ne / Historic Ne)
B
Historical Ne
Current Ne
Ne prior
density
log(effective population size)
C
t
t prior
density
log(time since decline)
D
μ
μ prior
density
log(mutation rate)

5

Marginal posterior distributions of scaled parameters from Msvar.

Solid lines represent posterior density distributions and dashed lines represent prior distributions (with solid horizontal lines representing the 95% HPD credibility intervals for the scaled posterior distributions). A) Scaled effective population sizes on a $\log_{10}$ scale; B) Scaled time on a $\log_{10}$ scale.