sabin METAL CORPORATION

CORPORATE OFFICE: 300 PANTIGO PLACE - SUITE 102, EAST HAMPTON, NEW YORK 11937

PHONE: 631-329-1717 • FAX: 631-329-1985 • Email: sabincorp@aol.com • WEBSITE: www.sabinmetal.com

REFINERS OF PRECIOUS METALS
GOLD • SILVER • PLATINUM • PALLADIUM • RHODIUM
RUTHENIUM • IRIDIUM

February 1, 2017

Hon. Sean Walter – Supervisor
Town of Riverhead Long Island
200 Howell Avenue
Riverhead, N.Y. 11901

Dear Hon. Supervisor Walter:

As discussed at our meeting approximately two years ago, I funded a study on the endangered Eastern Tiger Salamanders on Long Island (Focus: Calverton) Enclosed is a copy of the completed research report. We would like to make a presentation, at your convenience, and will schedule a meeting with you within the next few months.

Thank you for your consideration.

Sincerely,

Andrew E. Sabin
AES/rr

Enclosure

Cc: Councilman John Dunleavy
Councilwoman Jodi Giglio
Councilman Tim Hubbard
Councilman James Wooten

Population Genomics of Endangered Tiger Salamanders (*Ambystoma tigrinum*) on Long Island, NY Reveals a Highly Structured Species Impacted by Major Roads

Evan McCartney-Melstad[1*], Jannet Vu[1,2], H. Bradley Shaffer[1]

1: Department of Ecology and Evolutionary Biology, La Kretz Center for California Conservation Science, and Institute of Environmental Studies, University of California, Los Angeles
2: Department of Ecology and Evolutionary Biology, Stony Brook University, New York

Abstract

We used DNA sequence data from thousands of nuclear loci to characterize the population structure of endangered tiger salamanders (*Ambystoma tigrinum*) on Long Island and quantify the impacts of human development on this species. We uncovered highly genetically structured populations over an extremely small spatial scale (approximately 40 km^2) in an increasingly human-modified landscape. Geographic distance and the presence of major roads between ponds are both strong predictors of genetic divergence in this system, which suggests both natural and anthropogenic factors are responsible for the observed patterns of genetic variation. This study demonstrates the added value of genomic approaches in molecular ecology, as these patterns were not apparent in an earlier study of the same system using microsatellite loci. Ponds exhibited small effective population sizes, and there is a strong correlation between pond surface area and salamander population size. When combined with the high degree of structuring in this heavily modified landscape, our study suggests that these endangered amphibians require management at the individual pond, or pond cluster, landscape level. Particular efforts should be made to preserve large vernal pools, which harbor greater genetic diversity, and their surrounding upland habitat. Contiguous upland landscapes between ponds that encourage natural metapopulation dynamics and demographic rescue from future local extirpations should also be protected.

Introduction

Genetic, and, increasingly, genomic analyses constitute a powerful tool kit for understanding how species move through landscapes, particularly for secretive species such as reptiles and amphibians (Shaffer *et al.* 2015). When studying endangered species, we are often concerned with the degree to which human activity has impacted the size and movement of populations. This human interference often occurs at very small spatial scales compared to species range

sizes—for example, building a road between two nearby populations that exchange migrants regularly—as well as short temporal scales, given that humans often have been impacting wildlife populations for tens or hundreds of generations. As conservation and resource managers and as population biologists, we are often less interested in larger scale effects across thousands of kilometers of a species range than we are about dynamics across a few kilometers on specific landscapes. This is especially true for low-vagility species like amphibians, reptiles, small mammals, and many invertebrates that often move a kilometer or less per generation (Blaustein *et al.* 1994). For such taxa, the genetic relationships between distant populations are often a result of ancient demographic processes, but interruption of gene flow at an extremely fine spatial scale is the defining component of human impacts. For protected or endangered species, understanding the extent to which human activities at the finest spatial scales alter demographic and population processes is the key to effective management.

Discerning gene flow and differentiation at very fine spatial scales is challenging because populations located proximately to one another tend to be very closely related (Wright 1943). Furthermore, the ability to detect differentiation between genetically very closely related populations is limited by the number of samples and genetic loci assayed (Patterson *et al.* 2006). Until now, nearly all population genetic studies of amphibians have been limited to mitochondrial DNA or a small number of nuclear loci (typically microsatellites). This is due at least in part to the large, highly repetitive genomes of many amphibians that make it difficult to generate genomic resources (Licht & Lowcock 1991; Sun *et al.* 2012). While this is slowly changing as genomic technologies are beginning to be being applied to amphibians (Keinath *et al.* 2016; McCartney-Melstad *et al.* 2016; Portik *et al.* 2016; Newman & Austin 2016), most systems that could benefit from genomic scale data remain unexplored. Custom target

enrichment assays built from transcriptomic resources are promising intermediate solutions that bridge the gap between microsatellites and whole genome sequencing while allowing for flexibility in which genomic regions to study (McCartney-Melstad *et al.* 2016; Portik *et al.* 2016).

One interesting case study where the added resolution of genomic-scale datasets may make a difference is for tiger salamanders (*Ambystoma tigrinum*) on Long Island, a New York-listed endangered species (6 CRR-NY 182.5) where fine-scale population dynamics are critical for management decision making. *A. tigrinum* was historically found in scattered localities across New York at the northern limits of its range in the eastern US, including Albany County, Rockland County, and across Long Island. However, the species has experienced dramatic declines in the region, and it is currently restricted to Suffolk and Nassau Counties, primarily in central Long Island (Bishop 1941; Stewart & Rossi 1981). In recent years, surveyors have witnessed a decrease in the observed number of individuals, with approximately 90 breeding ponds remaining (New York State Department of Environmental Conservation 2015).

The species suffers a range of threats including disease, predation, pollution, invasive species, and climate change-induced sea level rise. Development is not only a source of habitat loss, but also creates direct mortality risk from road kill, degrades pond viability from pollutants, and creates barriers to migration and population fragmentation (Titus *et al.* 2014). Telemetry studies have documented individuals traveling at least 500 meters from breeding ponds, and confirmed that individuals tend to avoid paved roads, dirt roads, and grassy areas (Madison & Farrand 1998). Movements, which are often studied during the annual breeding migration, are generally oriented towards upland refugia in their preferred habitat of sandy soil, pine barren habitat (Madison & Farrand 1998; Titus *et al.* 2014).

Prior genetic work using twelve microsatellite loci recovered two distinct populations of *A. tigrinum* across 17 ponds spanning 50 km on Long Island, both of which exhibited low diversity and high relatedness among ponds (Titus *et al.* 2014). The authors attributed the low diversity and high relatedness to post-glacial colonization from North Carolina (Church *et al.* 2003) and relatively frequent migration of salamanders between ponds. Their primary conclusion was that Long Island and New Jersey tiger salamanders were genetically uniform within each state, but were differentiated between states due to geographic isolation and range fragmentation.

Most of the ponds analyzed by Titus *et al.* (2014) on Long Island were fewer than six kilometers apart, and their analyses and conclusions required genetic markers capable of discerning fine scale ecological processes. However, the microsatellite loci used showed relatively low diversity (1-13 alleles per locus across ponds and an average of 1-3 alleles per locus within ponds), and therefore were not the most imformative (Reyes-Valdés 2013). This leaves open the real possibility that these markers lacked the statistical power to detect real patterns of landscape-driven differentiation. This was not a fault of the Titus *et al.* (2014) work, but rather a reflection of the tools available when their work was undertaken.

To explore recent anthropogenic impacts on this endangered, fragmented set of populations further, we applied a genomic target capture approach with 5,237 random nuclear exons to ponds in the same system to quantify the degree to which ponds are isolated from one another and whether or not major roads act as barriers to dispersal for extant populations of *Ambystoma tigrinum* on Long Island. We sought to answer three separate questions: 1) To what degree are ponds genetically connected to or differentiated from one another?, 2) what are the effective population sizes of ponds in the system, are they related to pond area, and how do these values compare to other amphibians?, and 3) what are the effects of roads on connectivity between

ponds in the system? The increased resolution recovered from the genomic dataset collected here demonstrates the increased power and utility of genomic-scale data for population genetics of threatened species, and highlights the fundamentally different conclusions for appropriate management interventions that such data can provide.

Methods

Sampling and Data Generation

Larval tissue samples were collected in Suffolk County over three consecutive breeding seasons between 2013 and 2015 using seines and dipnets. We timed our sampling to occur in the late spring when larvae were large enough to sample non-destructively with small tail clips (Polich *et al.* 2013). Tail tips were placed in 95% ethanol within 30 seconds of clipping, larvae were immediately released at the site of capture, and tail tips were stored at -80C until use. A hand-held GPS unit was used to locate ponds in the field, and final spatial coordinates and areas of ponds were taken from tracings of Google Earth images from March 2007. We sampled larvae from multiple sites at each pond to randomly sample the genetic variation present. DNA was extracted from samples using a salt extraction protocol (Sambrook & Russell 2001), diluted to 100 ng/μL, and sheared for 28 cycles (30s on, 90s off) using the "high" setting on a Bioruptor NGS (Diagenode). After shearing, samples were dual-end size selected to approximately 300-500bp using 0.8X-1.0X SPRI beads (Rohland & Reich 2012).

Libraries were prepared with 419-2000 ng of starting input DNA using Kapa LTP library prep kit half reactions (Kapa Biosystems, Wilmington MA). Libraries were dual-indexed using the iTru system (Glenn *et al.* 2016), which adds 8bp indices to the adapters of both ends of library fragments for demultiplexing. Next, 500ng of each library were combined into pools of 8 (4,000ng total input DNA) and enriched using a MYcroarray (Ann Arbor, MI) biotinylated RNA probe set designed from 5,237 exons from unique genes from the California tiger salamander

genome (McCartney-Melstad *et al.* 2016). Given the relatively close phylogenetic relationships of all members of the tiger salamander complex (Shaffer & McKnight 1996; O'Neill *et al.* 2013), we predicted that most of the probes would also capture the eastern tiger salamander homolog. A total of 30,000 ng of c_0t-1 prepared from *Ambystoma californiense* was used for each capture reaction to block repetitive DNA from hybridizing with probes or captured fragments. Probes were hybridized for 30 hours at 60C, bound to streptavidin-coated beads, and washed four times with wash buffer 2.2 (MYcroarray). Enriched libraries were then amplified on-bead with 14 cycles of PCR, cleaned using 1.0X SPRI beads, and sequenced on three 150bp PE lanes on an Illumina HiSeq 4000.

Reference Assembly

We built a reference assembly for read mapping and SNP calling using the Assembly by Reduced Complexity (ARC) pipeline (Hunter *et al.* 2015). To do this, the reads from the 10 samples that received the greatest number of reads were pooled and mapped to the 5,237 *A. californiense* targets across which capture probes were tiled using bowtie2 v.2.2.6 (Langmead & Salzberg 2012). Pools of reads mapping to each one of these targets were independently assembled using SPAdes v.3.8.2 (Bankevich *et al.* 2012), and the contigs assembled for each target then replaced their respective targets and another round of mapping was performed to these contigs. This process was repeated for 10 iterations to extend assembled targets several hundred bp in both directions from their central probe-tiled regions. Reciprocal best blast hits (RBBHs) were then found to represent each target locus using blast+ 2.2.30 (Camacho *et al.* 2009). The set of RBBHs was then blasted against itself to find similar regions among targets, which may be indicative of chimeric assemblies. Regions within each RBBH that were found to be similar to other RBBHs were trimmed to the ends of the RBBH contigs.

SNP Calling and Genotyping

Reads for all samples were trimmed to 150bp (if the 151st base was reported by the sequencing facility) and adapters were trimmed using skewer 0.1.127 (Jiang *et al.* 2014). These trimmed reads were then mapped to the reference assembly using BWA-mem 0.7.15 (Li 2013). Read group information was added to the aligned reads and PCR duplicates were marked using picard tools v2.0.1 (https://broadinstitute.github.io/picard/).

SNP calling and genotyping was performed according to GATK best practices (DePristo *et al.* 2011; Van der Auwera *et al.* 2013). First, a set of high-quality reference SNPs was generated to assess and recalibrate base quality scores within each sample. HaplotypeCaller from GATK nightly-2016-11-21-g69e703d (McKenna *et al.* 2010) was run separately on each sample in GVCF mode followed by joint genotyping with GenotypeGVCFs. Then, any SNP that met any of the following criteria were removed from the reference set: QD < 2.0, MQ < 40.0, FS > 60.0, MQRankSum < -12.5, ReadPosRankSum < -8.0, QUAL < 100. Similarly, any indel that failed any of the following criteria were also removed from the reference set: QD < 2.0, SOR > 10.0, FS > 60.0, ReadPosRankSum < -8.0, QUAL < 100. Base quality score recalibration was then performed at the lane level (three different platform units among all of the read groups) using GATK.

HaplotypeCaller in GATK was then used with recalibrated reads to generate sample-level GVCF files that were jointly genotyped using GATK's GenotypeGVCFs function. The same hard filters outlined above were then applied to the resulting VCF files, except that all SNPs with QUAL values above 30 (instead of 100) were kept. Genotype calls with phred-scaled quality scores under 20 (1 in 100 chance of being incorrect) were set to "missing" data, and SNPs with greater than 50% missing data were removed. Samples with missing data rates greater than 30% were also removed.

Given the extremely large genomes of ambystomatid salamanders (roughly 30GB) (Licht & Lowcock 1991; Keinath *et al.* 2015), we were concerned about the possibility of including duplicated paralogous loci in our analyses. We attempted to correct for this by filtering out loci that contained excessive heterozygosity, as fixed differences between true paralogs interpreted as homologs will typically appear as variable sites that are always heterozygous. To do this, VCFtools v.0.1.15 was used to calculate p-values for heterozygote excess for every SNP (Wigginton *et al.* 2005; Danecek *et al.* 2011). Target regions that contained at least one SNP with an excess heterozygote p-value below 0.001 were removed from the analysis. A set of SNPs was then generated by randomly choosing a single SNP from each qualifying target region (those targets that did not contain any excessively heterozygous SNPs). This dataset with a single SNP taken from each target region is referred to hereafter as the "linkage-pruned" dataset.

Population Genetic Analysis

The presence of isolation by distance (IBD)—the relationship between geographic and genetic distance—was tested at both the individual and pond (population) levels. Individual genetic similarity was calculated as the percentage of SNPs that were identical-by-state using SNPRelate v1.6.4 (Zheng *et al.* 2012). These values were regressed on geographic distance and the significance of the correlation between genetic distance and geographic distance was tested using a simple Mantel test with 999,999 permutations in the R package vegan 2.4-0 (Mantel 1967; Oksanen *et al.* 2016). At the pond level, Fst/(1-Fst) (Slatkin 1995) was calculated using SNPRelate v1.6.4 and regressed on geographic distance to estimate the slope of isolation by distance. Rousset (1997) recommends regressing Fst/(1-Fst) on the logarithm of geographic distance in the case of two-dimensional habitats or non-transformed geographic distance in the case of one-dimensional habitats. Since the sampling area for this study is very narrow and is over three times longer than it is wide (approximately 15.5 km x 4.5 km), it is unclear whether it

is more appropriate to treat the study area as linear or two dimensional, and regressions and Mantel tests are reported for both raw and log-transformed geographic distances. Fst values were also calculated using Arlequin v3.5.2.2 (Excoffier & Lischer 2010) to determine significance p-values using 100,172 permutations of the data. P-values from Arlequin were adjusted for multiple testing using the Benjamini-Yekutieli correction implemented in base R (Benjamini & Yekutieli 2001). For individual-based analyses, logarithms of geographic distances were set to a minimum value of 0.

We were interested in characterizing the level of genetic diversity present in tiger salamanders on Long Island. To estimate genetic diversity we determined per-base pair Watterson's θ, an estimator that characterizes the level of genetic diversity in populations based on the number of segregating sites per base pair sequenced (Watterson 1975). We calculated θ for each pond with samples pooled across years. As a basis of comparison, a population sample of 15 California tiger salamanders (*A. californiense*) from a single pond in Great Valley Grasslands State Park, California (McCartney-Melstad and Shaffer, unpublished data) was genotyped under similar filtering parameters for the same set of loci, and θ was estimated for this group in the same way.

The linkage-pruned dataset was visualized using principal components analysis (PCA) in the R package SNPRelate v1.6.4 (Zheng *et al.* 2012). The first eight principal components were plotted with letters corresponding to the collection sites of samples. The proportion of the variance explained by each principal component was also obtained using SNPRelate v1.6.4.

To estimate the number of distinct population clusters in the data, ADMIXTURE v1.3.0 was run using the linkage-pruned dataset containing all samples from all ponds across all three years of sampling for K=1 to K=30 with ten different random number seeds (Alexander *et al.* 2009). Each replicate was subjected to 100-fold cross validation, and CV errors were used to choose a

"reasonable" set of K values. If the standard deviation of CV values for any K value overlapped with the standard deviation of the best-scoring K value, it was included as a reasonable value for K.

Effective population sizes (Ne) for each pond were estimated using the linkage disequilibrium (LD) method in NeEstimator v2.01 with a minor allele frequency cutoff of 0.05 (Hill 1981; Do *et al.* 2014). Estimates were calculated for all cohorts (a given pond in a given year), and, when more than one year of sampling was conducted for a pond, Ne was also calculated for the pooled sample of either two or three cohorts. LD-based estimates of effective population size from single cohorts represent the harmonic mean between the effective number of breeders (Nb) and the true effective population size (Ne) (Waples *et al.* 2016). Alternatively, as the number of pooled cohorts approaches the generation length (the average age of parents for a cohort), LD-based estimators should approach the true Ne (Waples & Do 2010; Waples *et al.* 2014).

Effective population size estimates using the LD method can be downwardly biased for multiple reasons. First, estimates may be biased when many loci are used due to physical linkage among loci, given that the method assumes the loci being used are unlinked (Waples *et al.* 2016). This effect is predictable, however, and can be corrected if the number of chromosomes or total linkage map length is known. Estimates of linkage map length for the closely related axolotl, *Ambystoma mexicanum*, are known, and this number (4200cm) was used to correct estimates of effective population size for dense locus sampling by dividing them by 0.9170819 (which is equal to -0.910 + 0.219 x ln(4200)) (Voss *et al.* 2011; Waples *et al.* 2016).

LD based estimates of effective population size can also be downwardly biased when analyzing mixed cohorts in iteroparous species such as *A. tigrinum*, although this bias appears to decrease as the number of sampled cohorts approaches the generation length of the species

(Waples & Do 2010; Waples *et al.* 2014). Therefore, single-cohort estimates of Ne were further corrected by dividing dense-locus adjusted estimates by 0.8781801, the product of two equations from Table 3 of Waples et al. (2014) that use the ratio of adult lifespan (estimated at 7 years for the closely related *A. californiense*) to age at maturity (4 years, also in *A. californiense*) (Trenham *et al.* 2000) to compensate for the downward bias introduced by iteroparity: (1.103-0.245 * log(7/4)) * (0.485+0.758 x log(7/4)). For ponds in which multiple years of sampling were conducted, we report both pooled-cohort estimates (corrected for dense locus sampling) and per-cohort estimates (corrected both for dense locus sampling and single-cohort sampling). We used linear regression to visualize the relationship between pond area (as traced from Google Earth images) and effective population size, using multi-year estimates of Ne when available.

Impact of Roads

We were interested in assessing to what degree human habitat modifications have restricted movement of this species, and whether or not human activity has contributed to the observed patterns of population structure. To explore this, we created a matrix that indicated whether or not pairs of ponds were separated by a major road (New York State Route 25, Suffolk CR 46, or Interstate 495, see Figure 6). This matrix was included as a predictor variable for genetic distance in linear regression and was tested for correlations to genetic distance (while controlling for geographic distance) using a partial Mantel test with vegan v2.4-0 in R (Mantel 1967; Smouse *et al.* 1986; R Core Team 2015; Oksanen *et al.* 2016).

Results

Sampling: A total of 283 salamanders were genotyped from 17 ponds spread over an approximately 40 km^2 area (Figure 6, Table 4). More than 1.9 billion 150-bp sequencing reads were generated from three Illumina HiSeq 4000 lanes across these samples (mean=6.8 million reads/sample, min=1.8 million reads, max=10.9 million reads).

Reference assembly: The ten samples that received the most sequencing reads were pooled to generate a *de novo* reference assembly, for a total of 66.9 million merged and paired-end sequencing reads (11.7 billion total bp). Assembly of target regions with the ARC assembler produced a set of 74,109 contigs (47.5 million bp) from which 5,057 reciprocal best blast hits were recovered (6.7 million bp). After blasting these contigs against themselves, trimming self-complementary regions to the ends of contigs, and re-determining reciprocal best blast hits, a 6.6 million bp assembly with 5,050 target regions (96.4% of the originally targeted regions) was recovered for mapping reads and calling SNPs.

SNP Calling and Genotyping: An average of 29.27% of raw reads mapped to the reference assembly using BWA-mem across all 283 samples (sd=2.47%, min=20.33%, max=34.30%). After removing PCR duplicates (read pairs that map to the exact same position on the reference, indicating that they may be PCR amplicons from the same molecule), an average of 17.03% unique reads mapped to the reference (sd=2.47%, min=8.51%, max=24.59%). After joint genotyping, a total of 82,005 raw SNPs were recovered across 4,400 target regions. After applying hard filters to SNP loci, setting the minimum genotype call quality to 20, discarding variants genotyped in less than 50% of all samples, and removing the one sample with a missing data rate greater than 30%, a total of 21,998 biallelic SNPs were retained across 3,631 target regions. Tests for Hardy Weinberg equilibrium revealed 533 targets contained at least one SNP with clear ($p<0.001$) heterozygote excess, which is consistent with (though not definite evidence of) the presence of an unknown paralogous copy of this gene in the genome. After removing these target regions from the analysis, a total of 12,924 biallelic SNPs remained across 3,098 target regions. The final matrix containing 282 individuals had a mean missing data rate of 7.7%

(max=27.8%, min=1.8%, sd=4.5%). The linkage-pruned dataset contained one random biallelic SNP from each final target, for a total of 3,098 variants.

Genetic variation within cohorts: Values of Watterson's θ for ponds ranged from $3.26x10^{-4}$ to $5.77x10^{-4}$ (Table 4), and was $3.19\ x10^{-4}$ after pooling the 282 samples from all ponds together for a single estimate of θ. The comparative sample of 15 *A. californiense* from a pond in Merced County, CA had a θ value of $7.09x10^{-4}$, which was higher than each of the values calculated for ponds in Long Island *A. tigrinum.* This suggests that genetic diversity is lower for *A. tigrinum* in Long Island than it is for *A. californiense* in Great Valley Grasslands State Park, CA, and is in keeping with the low estimates of variation found by Titus *et al.* (2014).

Isolation by Distance (IBD): IBD was apparent at both the individual and pond level (Figures 7 and 8, Table 5). Regressions of individual identity-by-state on both raw and log-transformed geographic distances yielded negative relationships with p-values below $2x10^{-16}$ (Figure 7). Adjusted R^2 values were higher for log-transformed distances when comparing pairwise individual genetic relationships and geographic distances (0.2861 vs. 0.1764). Similarly, regression coefficients were positive and highly significant when testing for the relationship between pairwise Fst of ponds and raw and log-transformed geographic distances (Figure 8, $p < 2.6x10^{-16}$ and $p < 4.12x10^{-11}$ for raw and log-transformed distances, respectively). Unlike the individual-based measure, the pond-based model with raw geographic distances fit the data better (R^2=0.39) than log-transformed geographic distances (R^2=0.27). Testing the significance of isolation by distance using regression coefficient p-values is inappropriate because many of the pairwise observations are not independent. Therefore, simple Mantel tests were used to test the significance of correlations between pond/individual genetic and raw/log-transformed geographic distances, all of which yielded p-values lower than 0.000011 (Table 5). This indicates

that there is a significant relationship between geographic and genetic distance, even at the extremely fine scale studied here.

Pairwise Fst values between ponds ranged from 0.005 to 0.207 (136 comparisons, median=0.064, sd=0.042, Table 6). Using Benjimini-Yekutieli (BY)-corrected p-values, 118 out of 136 of these pairwise comparisons were significantly different from 0. Of the 18 non-significant pairwise comparisons, 16 were from pond L, which contained only a single sample and therefore had extremely low power. Many of the highest Fst values are from pairwise comparisons containing ponds A or Q. These ponds are both outliers separated by greater geographic distances and by major roads from all other ponds (Figure 6).

Principal Component Analysis: The first eight principal components (PCs) are shown as pairwise plots in Figure 9. In all PC graphs, samples are coded by letters representing the ponds from which they were collected (Figure 6). PC1 groups samples from pond A to the exclusion of the other samples, while PC 2 does the same for samples from ponds E, F, and G. PC 3 separates samples from ponds B, C, and D from the other ponds (especially pond N), and PC4 appears to be an axis of variation between ponds J and Q (which is also apparent in PC5). Finally, PCs 6, 7, and 8 correspond to axes that differentiate ponds N, P, and Q, along with some samples from ponds A and J. Overall, clustering of single ponds and small groups of closely adjacent ponds is quite apparent, which indicates the presence of easily detectable population structure with the genomic data that we have collected in this study.

Population Clustering: The value of K in ADMIXTURE with the lowest mean CV error was K=12. Four other K values (9, 10, 11, and 13) had CV error standard deviations that overlapped with K=12 (Figure 10). Admixture proportions for K=9 through K=13 are shown in Figure 11, and are split by both pond and sampling year (Glasbey *et al.* 2007). Results from ADMIXTURE

analyses corroborated the qualitative patterns observed in the PCA. First, pond A generally formed one to three clusters to the exclusion of all other ponds (as recapitulated in PCs 1, 6, and 8). Ponds B, C, and D form a single cluster to the exclusion of other ponds (as also seen in PC 3). Similarly, ponds E and G form a unique cluster at K=9 (corresponding to PC 2), but are separated into their own private clusters at K=10 through K=13. Pond F, geographically separated from its closest neighbors (ponds E and G) by NY State Route 25, appears strongly admixed at K=9 through K=12, and receives its own cluster at K=13. Ponds H, I, J, K, L, and M appear to be strongly associated across all K values (though ponds I, L, and M appear highly admixed at these K values), with the exception of one year of sampling in pond J (2014) that produced a group of animals that formed their own cluster. Pond N appears quite distinct across all K values (which can also be seen on PCs 3-8). Pond O appears highly admixed across all K values, but tends to share a considerable admixture component with the cluster formed by pond P (and pond Q for K=9 through K=11). At K=12 and K=13, pond Q forms its own strong cluster to the exclusion of all other ponds, a pattern that is also quite apparent in PC5.

Effective Population Size: Estimates of effective population size ranged from 10.3 for pond N to 135.0 for pond K (Table 4). For ponds with multiple years of sampling, single-cohort estimates were generally close to those for pooled-cohort, with the exception of pond O, which had a pooled-cohort estimate of 68.8 and a 2013-cohort estimate of 17,689. This single-cohort estimate was extremely sensitive to the minor allele frequency cutoff—changing the threshold to 0.10 from 0.05 lowered the estimate to less than 600. The 95% confidence interval was also extremely wide for this cohort estimate, ranging from 953.0 to Infinite/incalculable. The surface area of ponds was strongly correlated with effective population size estimates ($p=0.00122$, $R^2=0.5619$, Figure 12). The number of samples included in the calculation of Ne was not

correlated with the resulting Ne estimate (linear regression p=0.513, adj R^2= -0.0438), suggesting that sample size *per se* was not a driver of Ne estimates.

Roads as Barriers to Dispersal: Roads appear to play a strong role in structuring among-pond genetic divergence in Long Island tiger salamanders. Specifically, linear regression supports roads as an explanatory factor in pairwise Fst values between ponds, as adding this term increased the adjusted R^2 of models including only geographic distance from 0.39 to 0.68 (with both terms highly significant). This is apparent from visualizing the distances, as a distinct upwards shift in genetic distance is apparent for pairwise comparisons separated by major roads (Figure 13). Similarly, partial Mantel tests recovered strong and highly significant correlations between genetic distance and being separated (or not) by major roads after controlling for geographic distance (p=0.000608, Mantel R^2=0.48). This suggests that dispersal may be limited across major roads, and that human activity has contributed to isolation of ponds in this relatively highly developed region.

Discussion

Population structure is difficult to detect and quantify accurately in subtly differentiated populations, and populations in close geographic proximity tend to be subtly differentiated (Wright 1943). In conservation genetics, however, we are often interested in understanding limitations in gene flow at the temporal and spatial scales at which humans impact populations. Furthermore, as the number of generations over which humans have affected most populations is usually relatively small, many cases of human-induced structure will be difficult to detect with conventional genetic datasets.

Several amphibian studies have attempted to quantify spatial genetic structure of populations at very fine spatial scales. Jehle *et al.* (2005) found evidence of pond clustering in *Triturus* newts over a 26.5 km^2 landscape using a hierarchical Bayesian clustering algorithm (Corander *et al.*

2003), although ponds did not cluster cleanly in STRUCTURE analyses (Pritchard *et al.* 2000). Hitchings and Beebee (1997) used allozyme data in common frogs in the UK and found evidence for significant structuring over a few kilometers in urbanized environments, but not in rural environments, suggesting that human development was acting to isolate ponds from one another in this system. Similarly, Lampert et al. (2003) recovered significant isolation by distance over roughly 8km between ponds in Túngara frogs (*Physalaemus pustulosus*), although 51 of 64 pairwise Fst values on the same side of the 100m-wide Chagres River were non-significant, and no population clustering methods were attempted. Conversely, Newman and Squire (2001) recovered significant differentiation and isolation by distance in wood frogs (*Rana sylvatica*) ponds separated by roughly 20km but could not genetically differentiate ponds at closer distances. Lampert *et al.* (2003) attributed the differences in discriminating power between these two studies to the low levels of diversity in microsatellite loci for wood frogs. Zamudio and Wieczorek (2007) found evidence for two genetic clusters of *Ambystoma maculatum* from 29 ponds spread over 1272km^2 in upstate New York, but little support for substructuring among ponds within each cluster. A number of other studies have found strong support for population structure among breeding ponds of amphibians in small landscapes using microsatellite loci (Wang *et al.* 2009, 2011, Wang 2009b, 2012; Savage *et al.* 2010). Conversely, several amphibian studies using microsatellites have failed to find significant genetic differentiation among ponds for pond-breeding amphibians (Coster *et al.* 2015; Furman *et al.* 2016), while others have found evidence of isolation by distance and limited clustering (Sotiropoulos *et al.* 2013; Peterman *et al.* 2015).

These studies illustrate that, in amphibians, genetic differentiation is sometimes detectable at very fine spatial scales, and sometimes it is not. This may hinge largely on the variability of the

markers studied, which itself is shaped by deeper-time demographic processes such as bottlenecks and range expansions (Watterson 1984; Slatkin 1993). While microsatellite loci have been extremely valuable for conservation genetics, a panel of 20 microsatellites (which is towards the high end employed by most studies) has been shown in one instance to be approximately as effective for estimating genetic relationships as 50 SNP loci (Santure *et al.* 2010). While it is laborious to increase the number of microsatellite loci above the 20 or so that are typically used in conservation genetics, it is very straightforward to scale the number of SNPs assayed into the thousands or tens of thousands, which greatly increases our ability to distinguish barriers to gene flow that are subtle or have only been operating for a small number of generations (Patterson *et al.* 2006; Anderson *et al.* 2010). As genomic-scale datasets become comparable with microsatellites in terms of cost and feasibility, the added resolution from thousands of loci will give a particular boost to population genetic studies in systems with low genetic diversity, and will open entire new classes of analyses to both low- and high-diversity systems.

While a lack of statistical power is one reason why population structure may not be detected in pond-breeding amphibians, another possibility is that, even in low-vagility species, ponds in some systems are truly unstructured, and that failing to recover population structure reflects a biological reality of panmixia across these ponds. Differentiating between low resolving power and true panmixia is critical for conservation and management decision makers. Multiple studies of the same systems with both conventional and genomic datasets can help clarify whether the null hypothesis of population differentiation and strong isolation by distance is a general rule for pond-breeding amphibians, or whether such rules may be habitat or lineage-specific.

The current study is among the first to use thousands of nuclear loci across hundreds of individuals in a large-genome amphibian, and represents an opportunity to compare results between the two genetic approaches in the same system. While little genetic clustering was apparent in the microsatellite loci analyzed by Titus et al. (2014), our dataset of thousands of nuclear SNPs reveals clear population genetic structuring among breeding ponds of *Ambystoma tigrinum* on Long Island. The major genetic patterns in our data are readily apparent in both ADMIXTURE and PCA results. Genetic structuring of ponds generally shows consistent results across years (Figure 11), with two exceptions. First, samples from 2013 in Pond A were classified consistently as a unique population that is admixed with the Pond A lineages sampled in 2014 and 2015. Second, some of the samples from 2014 in Pond J appear to belong to a unique lineage that was not sampled in any other ponds or years. Aside from these two results, consistency between sampling years in the different ponds suggests that the observed patterns of genetic structure are likely driven by geography and not year-to-year variation.

Species with low genetic diversity require collecting data from a greater number of genetic loci to detect population structure (Patterson *et al.* 2006). One cause of low genetic diversity is a range expansion. Church *et al.* (2003) analyzed *Ambystoma tigrinum* mitochondrial DNA and determined that New York was likely recolonized by salamanders from Pleistocene refugia in North Carolina. This was corroborated by Titus et al. (2014), who found low genetic diversity in microsatellite loci in New Jersey and Long Island tiger salamanders. To try to understand whether this low genetic diversity led to the apparent differences between microsatellite and target capture datasets, we compared estimates of genetic diversity from Long Island tiger salamanders to other amphibian systems. Crawford (2003) used a single gene (*c-myc*) to estimate θ in populations of *Eleutherodactylus* frogs in Costa Rica and Panama and obtained values

ranging from 0.00080 to 0.01148 (excluding one population that was fixed for a single haplotype across eight diploid individuals). Weisrock *et al.* (2006) estimated θ at eight nuclear loci from 217 *Ambystoma ordinarium* (a member of the *Ambystoma tigrinum* complex) larvae from across the geographic range of the species (spanning roughly 200km) and obtained an average θ of 0.00208 across loci (min=0.0006, max=0.0034). Similarly, Nadachowska and Babik (2009) sequenced eight nuclear loci for 20 different populations of smooth newt subspecies in Turkey (*Lissotriton vulgaris kosswigi* and *Lissotriton vulgaris vulgaris*). They calculated θ for each population and, after averaging across loci, recovered population estimates ranging from 0.0019 to 0.0081. Finally, we calculated θ as 0.000709 in a collection of 15 *A. californiense* from Merced County, CA. This calculation was performed for a collection of individuals across the same set of nuclear loci presented here, so it is the most direct comparison available. All of these values of θ are greater than the largest value obtained in Long Island tiger salamander ponds (0.000577, mean=0.000427), which indicates that these populations likely do have lower genetic diversity than is normally seen in amphibians.

Breeding ponds that we examined generally exhibited small effective population sizes (< 100), consistent with results found for many other amphibian species (Schmeller & Merilä 2007; Phillipsen *et al.* 2011; McCartney-Melstad & Shaffer 2015). Our estimates (mean=36.9) are larger than, but of the same magnitude as microsatellite-based estimates performed by Titus *et al.* (2014) using the sibship method (Wang 2009a), which had a mean value of 20.9. We did, however, recover several ponds with effective population sizes higher than 44, which was the maximum value recovered by Titus *et al.* (2014). These included pond H (Ne=91.0), pond K (Ne=135.0), pond M (Ne=82.9), and pond O (Ne=68.8). This may indicate that the area around

these ponds, which was not directly sampled by Titus et al. (2014), may harbor greater effective population sizes than elsewhere on Long Island.

A clear relationship between pond size and effective population size was recovered ($p=0.00122$, $R^2=0.5619$, Figure 12). This relationship has been previously observed in *A. californiense* (Wang *et al.* 2011). Interestingly, the pond for which surface area did the worst job predicting Ne, Pond H, had a much higher effective population size estimate than expected by the model (that is, it had the largest residual from the regression line). Pond H is geographically closest pond to Pond K, which has the largest effective population size estimate of any pond. The landscape between Pond H and Pond K is largely forested with no major roads or other anthropogenic barriers to gene flow, the Fst value between ponds H and K is the lowest of any pairwise comparison between ponds (Fst=0.005, Table 6), and these ponds are consistently recovered in the same cluster in ADMIXTURE analyses. Taken together, this suggests that migration has been common between Pond H and Pond K, and that the effective population size of Pond H is augmented by its close relationship with the very large Pond K.

Our approach afforded us the resolution to evaluate the contributions of human disturbance on the movement of salamanders in the form of roads limiting dispersal between ponds. Based on the y-intercepts of linear regressions, the presence of a major road between ponds raised Fst values by approximately 0.04. Pond A was quite distinct from all the other ponds, as was Pond Q (Table 6). These ponds are generally separated from other ponds by greater geographic distance, but they are also separated from all other ponds by major roads. Similarly, ponds E and G tend to separate from all other ponds (PC2 in Figure 9)—these are the only ponds besides pond A that are north of New York State Route 25, a high-traffic road that constitutes a substantial barrier to salamander movement. The combination of geographic distance and roads did an excellent job of

explaining the observed genetic distances between ponds (linear regression, adj. R^2=0.6814). These results suggest that both geographic distance and the presence of roads have affected salamander dispersal for many generations, which has important implications for conservation strategies.

Conclusion

The results of this study show that *Ambystoma tigrinum* ponds on Long Island generally have relatively small effective population sizes that are correlated with the surface area of ponds, that migration is limited among most ponds in the area, and that major roads further limit dispersal. The interrelationships between these factors are important for conservation management. Small effective population sizes imply that ponds are more likely to suffer random demographic extinction, and highly structured populations indicate that locally extirpated ponds (such as those that do not fill with water for many years in a row) may not be easily recolonized by individuals from nearby ponds. Roads and other human activities add to these natural dynamics, and emphasize the critical importance of conserving blocks of contiguous habitat with a complex of ponds that can act as semi-isolated metapopulations. Within the Long Island landscape studied here, there appear to be several clusters of interconnected ponds that periodically share migrants (ponds B, C, and D; ponds H, I, J, K, L, and M; and ponds O and P). For such clusters migrants from interconnected ponds may be expected to "rescue" nearby ponds that go locally extinct, and maintaining these dynamics is probably critical to the long-term persistence of tiger salamanders locally. However, the presence of major roads appears to disrupt this pattern, as seen by the tendency of nearby ponds separated by major roads to fall out in different genetic clusters (such as Pond A vs. ponds B, C, and D and Pond F vs. ponds E and G).

A genomic approach was critical for this experiment to detect the observed population structure at such a fine spatial scale in a post-glacially recolonized area. The distinction between

inferences made from relatively few microsatellite loci from the data generated in this study have important consequences for our understanding of ecological dynamics in the system. Titus *et al.* (2014) recovered little genetic structure among endangered populations of Long Island tiger salamanders and inferred relatively high migration rates between ponds. Conversely, our genomic approach revealed the restrictions in movement between many groups of ponds, despite low overall levels of genetic differentiation.

This study suggests that monitoring of individual ponds is necessary, especially during and following droughts. Our genetic results suggest that ponds not separated by major roads may have increased resilience to local extirpation via demographic rescue from neighboring ponds, so efforts should be made to prevent activities that separate such clusters of ponds. In the event of an observed local extirpation of a pond, the genetic results herein provide information regarding the best source of animals to use for translocations to preserve the current genetic landscape, which is a result of a combination of current and historical patterns of dispersal among ponds.

Acknowledgments

Funding was provided by the Andrew Sabin Family Foundation. EMM and HBS also received support from NSF-DEB 1457832 and NSF-DEB 1257648. We thank Alvin Breisch, Andy Sabin, Pete Davis, and Jeremy Feinberg for assistance with fieldwork, and Megan Sha and Rice Zhang for laboratory support. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. Computing resources were provided by XSEDE and the Comet supercomputer at SDSC.

Figures

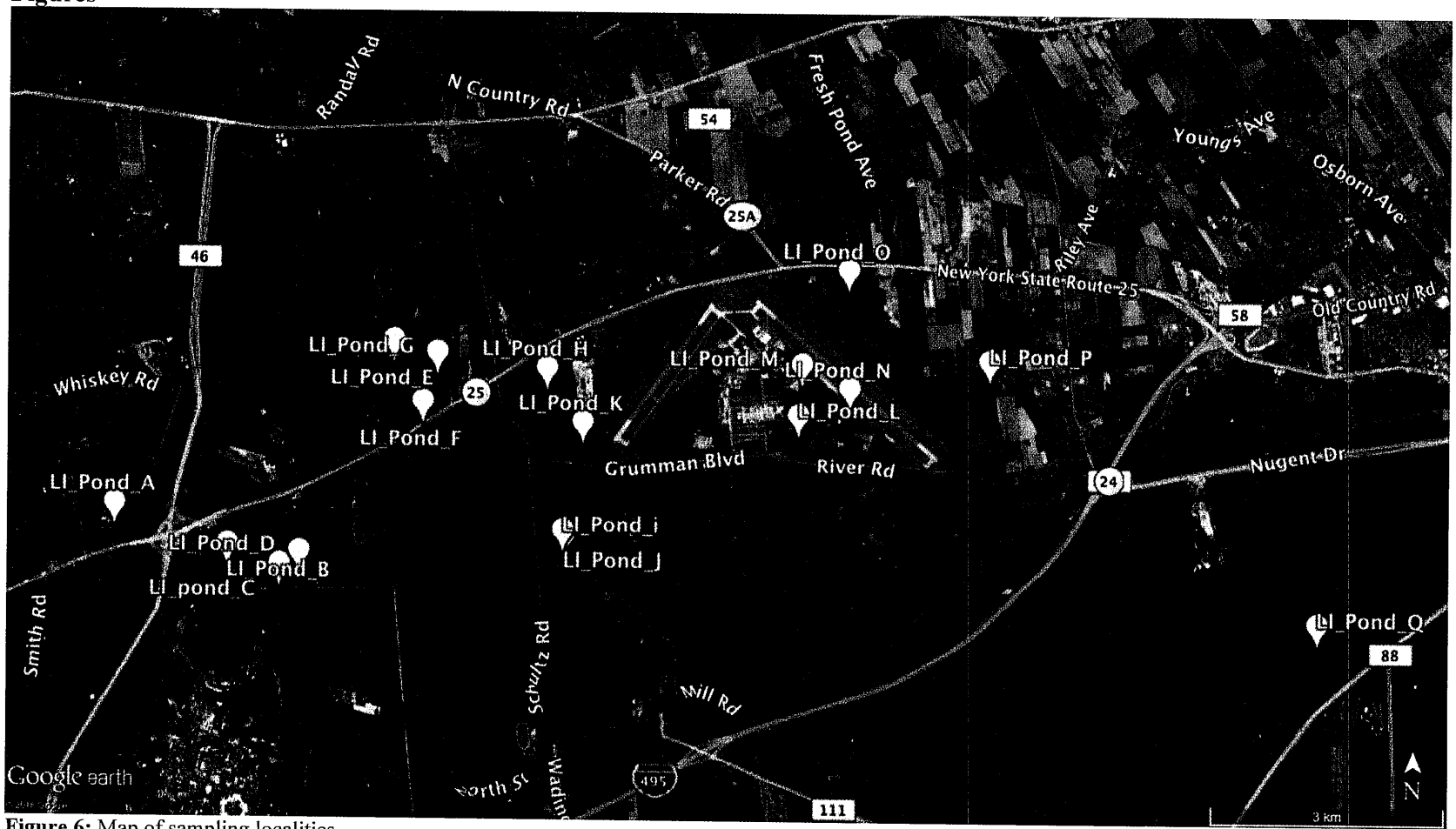

Figure 6: Map of sampling localities.

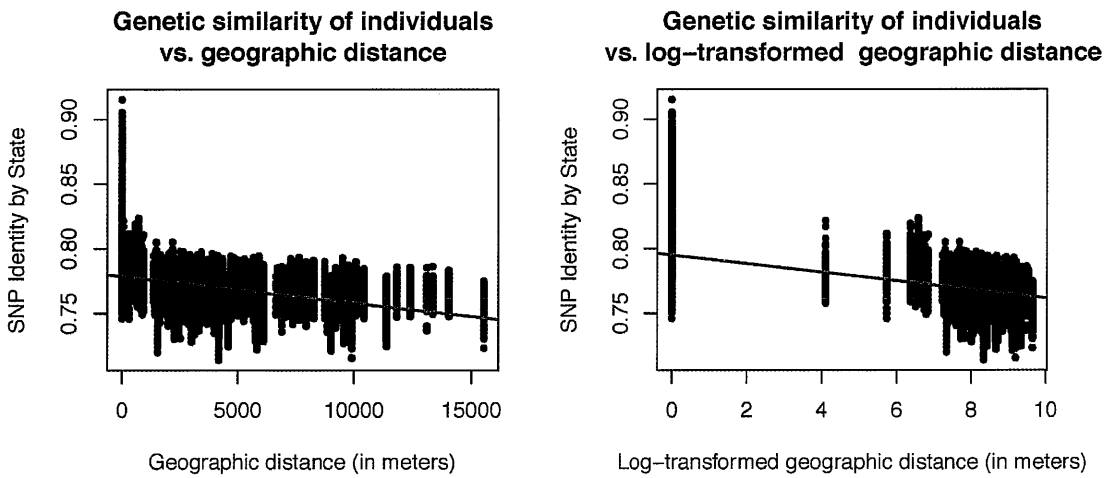

Figure 7: Relationship between genetic similarity and geographic distance between individuals. The plot on the left uses raw Euclidean distance between individuals, while the plot on the left uses log-transformed Euclidean distances.

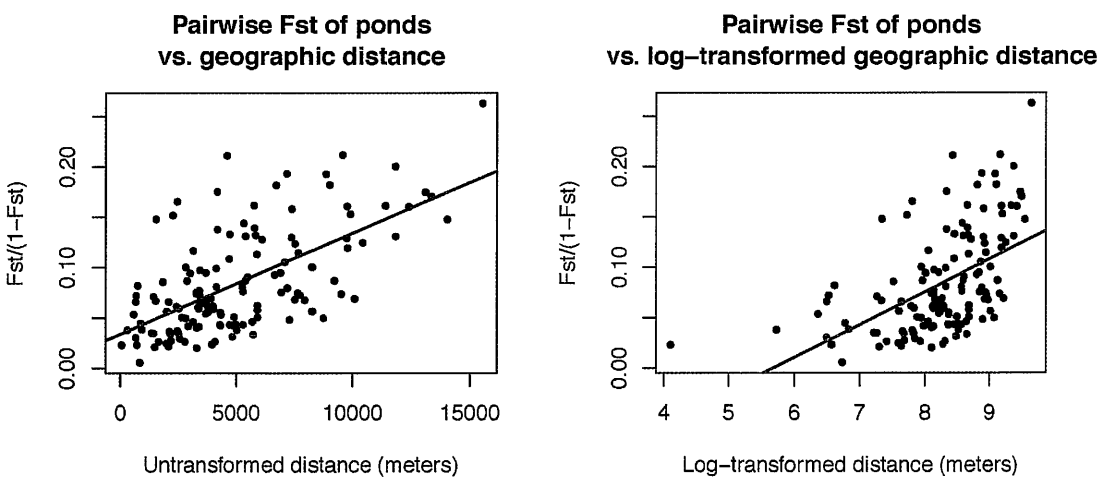

Figure 8: Relationship between genetic distance and geographic distance between ponds. The plot on the left uses raw Euclidean distance between ponds, while the plot on the left uses log-transformed Euclidean distances.

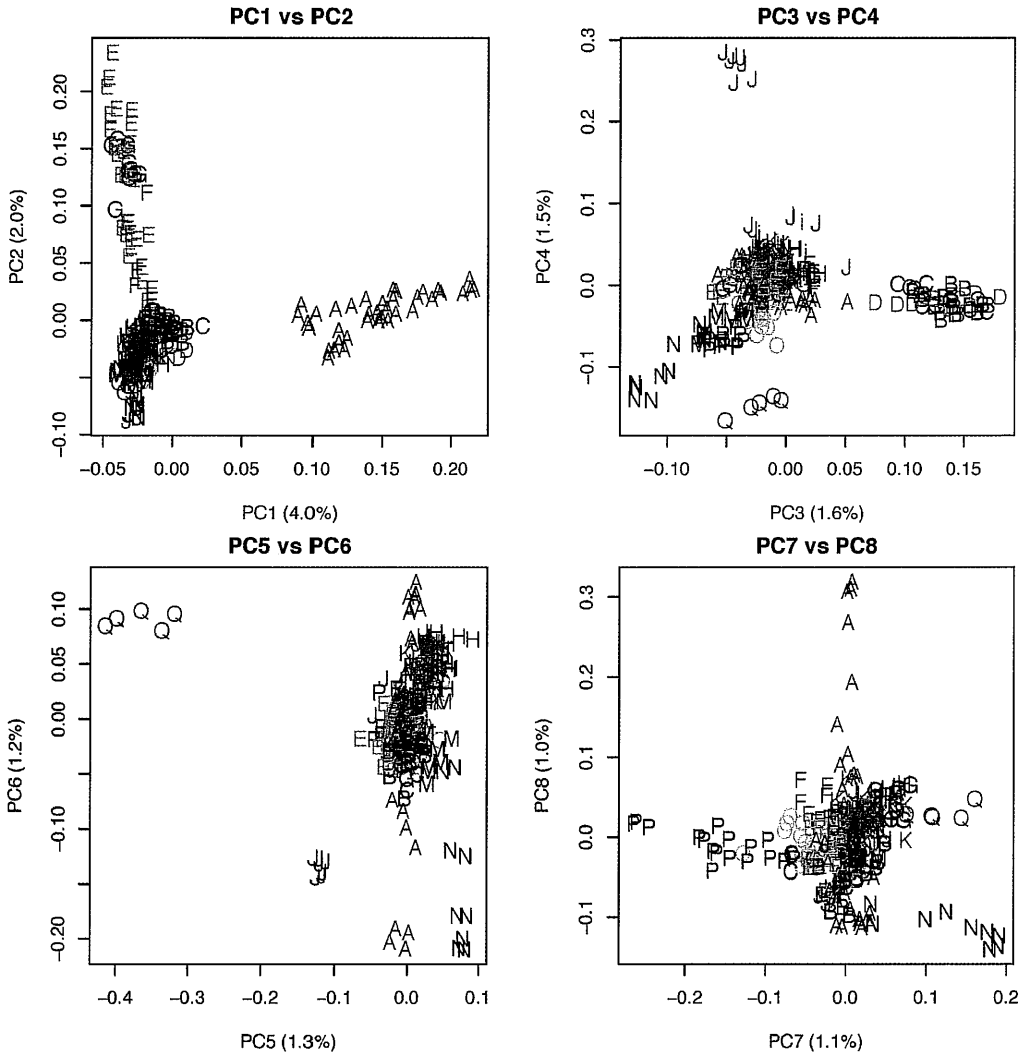

Figure 9: First eight principal components of the data. Letters on the graph correspond to samples from the same pond. Colors are used only to aid in distinguishing between letters.

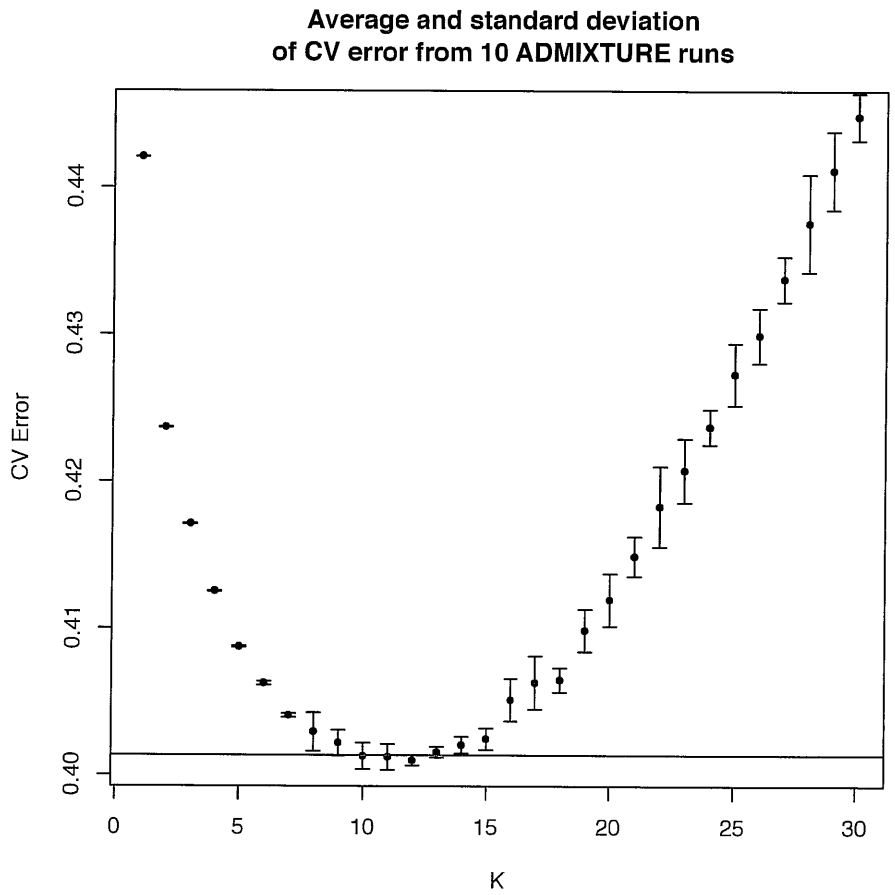

Figure 10: Cross-validation error mean and standard deviations from 10 ADMIXTURE runs using different seeds. The red line is drawn at the mean+SD of the best-performing K value (K=12). The standard deviations for K=9 through K=13 overlap this line.

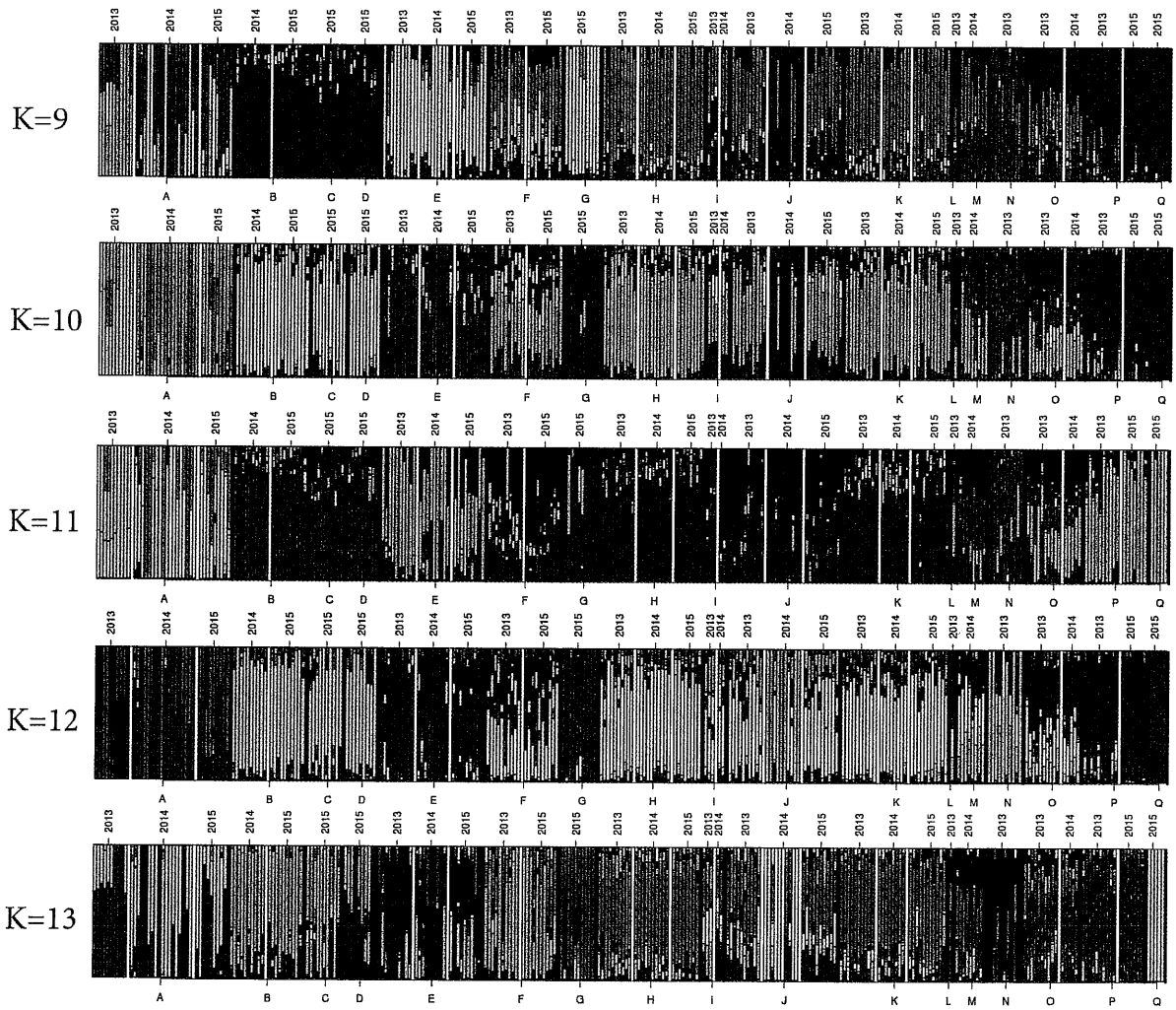

Figure 11: Admixture results from all 282 samples. Letters correspond to ponds from the sample map (Figures 6 and 9). White vertical lines separate sampling years within ponds, and black vertical lines separate ponds from one another.

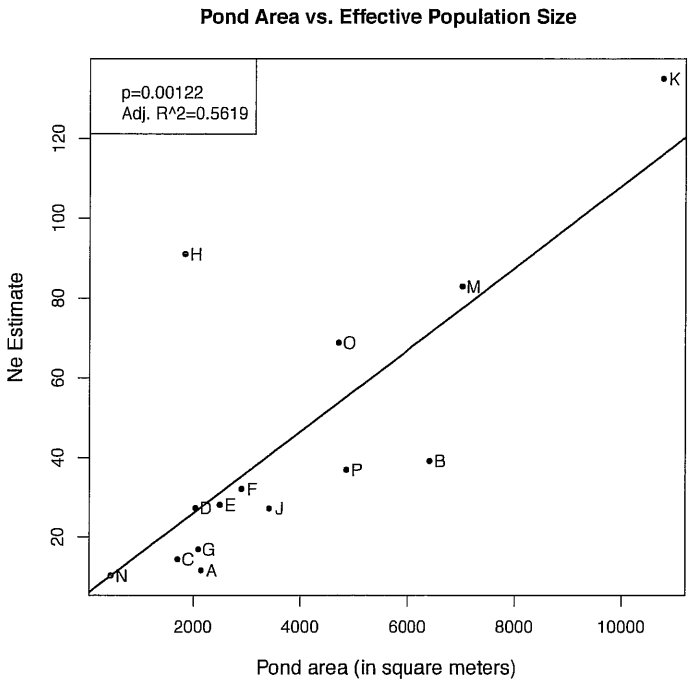

Figure 12: Relationship between pond area and effective population size estimate. Ne estimates represent multiple-cohort calculations if multiple cohorts were samples, otherwise adjusted single-year estimates were used. Ponds i, L, and Q were omitted because they did not contain enough samples to generate an estimate of Ne.

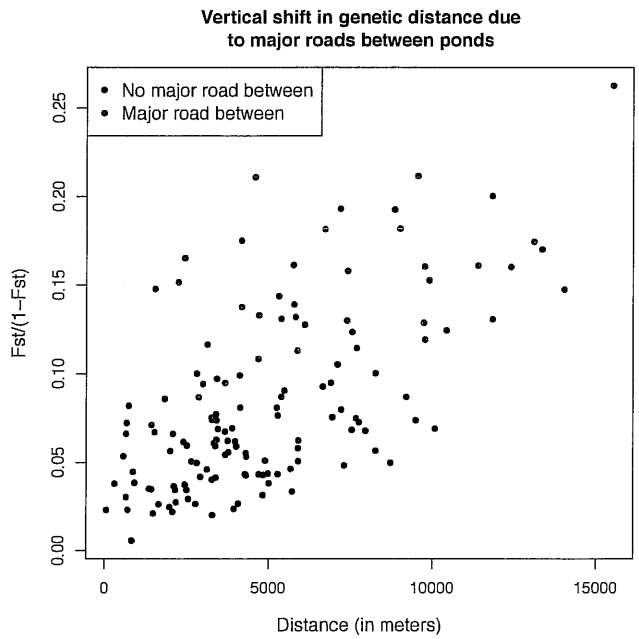

Figure 13: Visualizing the impacts of major roads on genetic differentiation between ponds. For the same geographic distance, ponds separated by major roads (indicated by red dots) tend to have higher levels of genetic differentiation.

Tables

Table 4: Pond localities, areas, Watterson's θ estimates, sampling, and effective population size estimates. Pond areas were estimated from Google Earth satellite images taken in March 2007. Single-year estimates were corrected for iteroparity-induced downward bias as explained in Methods, and both single-year and pooled-year estimates were corrected for dense locus sampling on chromosomes. Infinite values indicate that sample sizes were likely too small to estimate Ne. N=number of samples included in analyses. Ne=Effective population size estimates using LD method.

Pond	Latitude	Longitude	Pond Area (m^2)	Watterson's θ	N (2013/2014/2015)	Ne (2013/2014/2015)
A	40.896379	-72.892071	2147	$3.26x10^{-4}$	37 (10/18/9)	11.6 (6.7/7.1/20.1)
B	40.891766	-72.874854	6413	$4.05x10^{-4}$	20 (0/10/10)	39.1 (NA/37.7/47.2)
C	40.889497	-72.866932	1706	$4.45x10^{-4}$	10 (0/0/10)	14.4 (NA/NA/14.4)
D	40.891043	-72.863908	2039	$4.38x10^{-4}$	9 (0/0/9)	27.3 (NA/NA/27.3)
E	40.915705	-72.849554	2493	$3.75x10^{-4}$	28 (10/9/9)	28.1 (40.4/14.3/19.7)
F	40.908597	-72.845109	2898	$4.16x10^{-4}$	20 (10/0/10)	32.1 (30.8/NA/31.3)
G	40.914317	-72.842938	2094	$4.21x10^{-4}$	10 (0/0/10)	16.9 (NA/NA/16.9)
H	40.912580	-72.826168	1840	$4.06x10^{-4}$	28 (10/10/8)	91.0 (55.5/187.2/515.2)
I	40.893704	-72.823658	944	$4.98x10^{-4}$	5 (3/2/0)	Inf (Inf/Inf/NA)
J	40.893182	-72.823465	3418	$3.94x10^{-4}$	30 (10/10/10)	27.2 (136.2/4.0/91.3)
K	40.906296	-72.820671	10773	$4.07x10^{-4}$	29 (10/8/11)	135.0 (602.1/Inf/27.4)
L	40.907237	-72.787736	8587	$5.77x10^{-4}$	1 (1/0/0)	Inf (Inf/NA/NA)
M	40.913165	-72.787206	7020	$4.62x10^{-4}$	8 (0/8/0)	82.9 (NA/82.9/NA)
N	40.910430	-72.779946	464	$4.22x10^{-4}$	10 (10/0/0)	10.3 (10.3/NA/NA)
O	40.924112	-72.780170	4710	$4.36x10^{-4}$	15 (10/5/0)	68.8 (17689.4/Inf/NA)
P	40.913681	-72.758595	4854	$4.15x10^{-4}$	17 (10/0/7)	36.9 (Inf/NA/11.1)
Q	40.883585	-72.708374	1302	$4.08x10^{-4}$	5 (0/0/5)	Inf (NA/NA/Inf)

Test	R_M	R_M^2	p-value
Individual with log(geographic distance)	0.5349	0.2861	$1x10^{-6}$
Individual with raw geographic distance	0.4200	0.1764	$1x10^{-6}$
Ponds with log(geographic distance)	0.5276	0.2784	$1x10^{-6}$
Ponds with raw geographic distance	0.6305	0.3975	$1.1x10^{-5}$

Table 5: Mantel test results: P-values calculated using 999,999 permutations. R_M is the Mantel R statistic, and R_M^2 is the square of the Mantel R statistic.

	A	B	C	D	E	F	G	H	i	J	K	L	M	N	O	P	Q
A																	
B	0.127																
C	0.129	0.020															
D	0.129	0.030	**0.022**														
E	0.147	0.066	0.069	0.072													
F	0.119	0.043	0.044	0.048	0.040												
G	0.173	0.085	0.086	0.091	0.050	0.060											
H	0.115	0.038	0.041	0.042	0.052	0.022	0.066										
i	0.137	0.050	0.061	0.059	0.068	0.034	0.088	0.035									
J	0.120	0.048	0.050	0.045	0.056	0.031	0.080	0.026	0.022								
K	0.111	0.037	0.040	0.039	0.047	0.018	0.062	0.005	0.033	0.020							
L	**0.159**	**0.072**	**0.064**	**0.099**	**0.071**	0.039	**0.111**	0.018	0.043	0.041	0.024						
M	0.151	0.066	0.070	0.076	0.073	0.047	0.095	0.037	0.056	0.051	0.038	0.028					
N	0.173	0.090	0.100	0.102	0.101	0.082	0.123	0.064	0.088	0.074	0.064	**0.069**	0.065				
O	0.128	0.045	0.049	0.055	0.054	0.031	0.076	0.021	0.038	0.033	0.018	0.030	0.030	0.059			
P	0.136	0.063	0.065	0.071	0.066	0.043	0.092	0.040	0.052	0.045	0.037	0.033	0.056	0.077	0.031		
Q	0.207	0.131	0.142	0.144	0.139	0.118	0.164	0.111	**0.135**	0.115	0.106	**0.153**	0.135	0.151	0.110	0.116	

Table 6: Pairwise Fst values between ponds. Cells are colored by the magnitude of difference between ponds, with red being relatively low differentiation and green being relatively high differentiation. Bolded cells/values are not significantly different from 0 ($p >$ Benjamini-Yekutieli-corrected 0.05).

References

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 1655–1664.

Anderson CD, Epperson BK, Fortin M-J *et al.* (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology*, **19**, 3565–3575.

Bankevich A, Nurk S, Antipov D *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, **19**, 455–477.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, **29**, 1165–1188.

Bishop SC (1941) *The salamanders of New York.* University of the State of New York.

Blaustein AR, Wake DB, Sousa WP (1994) Amphibian Declines: Judging Stability, Persistence, and Susceptibility of Populations to Local and Global Extinctions. *Conservation Biology*, **8**, 60–71.

Camacho C, Coulouris G, Avagyan V *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

Church SA, Kraus JM, Mitchell JC, Church DR, Taylor DR (2003) Evidence for Multiple Pleistocene Refugia in the Postglacial Expansion of the Eastern Tiger Salamander, Ambystoma Tigrinum Tigrinum. *Evolution*, **57**, 372–383.

Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian Analysis of Genetic Differentiation Between Populations. *Genetics*, **163**, 367–374.

Coster SS, Babbitt KJ, Cooper A, Kovach AI (2015) Limited influence of local and landscape factors on finescale gene flow in two pond-breeding amphibians. *Molecular Ecology*, **24**, 742–758.

Crawford AJ (2003) Huge populations and old species of Costa Rican and Panamanian dirt frogs inferred from mitochondrial and nuclear gene sequences. *Molecular Ecology*, **12**, 2525–2540.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

DePristo MA, Banks E, Poplin RE *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, **43**, 491–498.

Do C, Waples RS, Peel D *et al.* (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, **14**, 209–214.

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources*, **10**, 564–567.

Furman BLS, Scheffers BR, Taylor M, Davis C, Paszkowski CA (2016) Limited genetic structure in a wood frog (Lithobates sylvaticus) population in an urban landscape inhabiting natural and constructed wetlands. *Conservation Genetics*, **17**, 19–30.

Glasbey C, van der Heijden G, Toh VFK, Gray A (2007) Colour displays for categorical images. *Color Research & Application*, **32**, 304–309.

Glenn TC, Nilsen R, Kieran TJ *et al.* (2016) Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). *bioRxiv*, 049114.

Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. *Genetics Research*, **38**, 209–216.

Hitchings SP, Beebee TJ (1997) Genetic substructuring as a result of barriers to gene flow in urban Rana temporaria (common frog) populations: implications for biodiversity conservation. *Heredity*, **79**, 117–127.

Hunter SS, Lyon RT, Sarver BAJ *et al.* (2015) Assembly by Reduced Complexity (ARC): a hybrid approach for targeted assembly of homologous sequences. *bioRxiv*.

Jehle R, Burke T, Arntzen JW (2005) Delineating fine-scale genetic units in amphibians: Probing the primacy of ponds. *Conservation Genetics*, **6**, 227–234.

Jiang H, Lei R, Ding S-W, Zhu S (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, **15**, 182.

Keinath MC, Timoshevskiy VA, Timoshevskaya NY *et al.* (2015) Initial characterization of the large genome of the salamander Ambystoma mexicanum using shotgun and laser capture chromosome sequencing. *Scientific Reports*, **5**, 16413.

Keinath MC, Voss SR, Tsonis PA, Smith JJ (2016) A linkage map for the Newt Notophthalmus viridescens: Insights in vertebrate genome and chromosome evolution. *Developmental Biology*.

Lampert KP, Rand AS, Mueller UG, Ryan MJ (2003) Fine-scale genetic pattern and evidence for sex-biased dispersal in the túngara frog, Physalaemus pustulosus. *Molecular Ecology*, **12**, 3325–3334.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-bio]*.

Licht LE, Lowcock LA (1991) Genome size and metabolic rate in salamanders. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, **100**, 83–92.

Madison DM, Farrand L (1998) Habitat Use during Breeding and Emigration in Radio-Implanted Tiger Salamanders, Ambystoma tigrinum. *Copeia*, **1998**, 402–410.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

McCartney-Melstad E, Mount GG, Shaffer HB (2016) Exon capture optimization in amphibians with large genomes. *Molecular Ecology Resources*, **16**, 1084–1094.

McCartney-Melstad E, Shaffer HB (2015) Amphibian molecular ecology and how it has informed conservation. *Molecular Ecology*, **24**, 5084–5109.

McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, **20**, 1297–1303.

Nadachowska K, Babik W (2009) Divergence in the Face of Gene Flow: The Case of Two Newts (Amphibia: Salamandridae). *Molecular Biology and Evolution*, **26**, 829–841.

New York State Department of Environmental Conservation (2015) Eastern tiger salamander fact sheet. Available http://www.dec.ny.gov/animals/7143.html. (Accessed September, 2016).

Newman CE, Austin CC (2016) Sequence capture and next-generation sequencing of ultraconserved elements in a large-genome salamander. *Molecular Ecology*, In press.

Newman RA, Squire T (2001) Microsatellite variation and fine-scale population structure in the wood frog (Rana sylvatica). *Molecular Ecology*, **10**, 1087–1100.

Oksanen J, Blanchet FG, Friendly M *et al.* (2016) The vegan package. *https://CRAN.R-project.org/package=vegan.*

O'Neill EM, Schwartz R, Bullock CT *et al.* (2013) Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (Ambystoma tigrinum) species complex. *Molecular Ecology*, **22**, 111–129.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet*, **2**, e190.

Peterman WE, Anderson TL, Ousterhout BH *et al.* (2015) Differential dispersal shapes population structure and patterns of genetic differentiation in two sympatric pond breeding salamanders. *Conservation Genetics*, **16**, 59–69.

Phillipsen IC, Funk WC, Hoffman EA, Monsen KJ, Blouin MS (2011) Comparative analyses of effective population size within and among species: Ranid frogs as a case study. *Evolution*, **65**, 2927–2945.

Polich RL, Searcy CA, Shaffer HB (2013) Effects of tail-clipping on survivorship and growth of larval salamanders. *The Journal of Wildlife Management*, 77, 1420–1425.

Portik DM, Smith LL, Bi K (2016) An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). *Molecular Ecology Resources*, **16**, 1069–1083.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

R Core Team (2015) R: A language and environment for statistical computing. https://www.R-project.org/.

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics*, **197**, 573–589.

Reyes-Valdés MH (2013) Informativeness of Microsatellite Markers. In: *Microsatellites* Methods in Molecular Biology Vol 1006. (ed Kantartzi SK), pp. 259–270. Humana Press.

Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, **22**, 939–946.

Rousset F (1997) Genetic Differentiation and Estimation of Gene Flow from F-Statistics Under Isolation by Distance. *Genetics*, **145**, 1219–1228.

Sambrook J, Russell DW (2001) *Molecular cloning: a laboratory manual (3-volume set)*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Santure AW, Stapley J, Ball AD *et al.* (2010) On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology*, **19**, 1439–1451.

Savage WK, Fremier AK, Bradley Shaffer H (2010) Landscape genetics of alpine Sierra Nevada salamanders reveal extreme population subdivision in space and time. *Molecular Ecology*, **19**, 3301–3314.

Schmeller DS, Merilä J (2007) Demographic and genetic estimates of effective population and breeding size in the amphibian Rana temporaria. *Conservation Biology*, **21**, 142–151.

Shaffer HB, Gidiş M, McCartney-Melstad E *et al.* (2015) Conservation Genetics and Genomics of Amphibians and Reptiles. *Annual Review of Animal Biosciences*, **3**.

Shaffer HB, McKnight ML (1996) The Polytypic Species Revisited: Genetic Differentiation and Molecular Phylogenetics of the Tiger Salamander Ambystoma tigrinum (Amphibia: Caudata) Complex. *Evolution*, **50**, 417–433.

Slatkin M (1993) Isolation by Distance in Equilibrium and Non-Equilibrium Populations. *Evolution*, **47**, 264–279.

Slatkin M (1995) A Measure of Population Subdivision Based on Microsatellite Allele Frequencies. *Genetics*, **139**, 457–462.

Smouse PE, Long JC, Sokal RR (1986) Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. *Systematic Zoology*, **35**, 627–632.

Sotiropoulos K, Eleftherakos K, Tsaparis D *et al.* (2013) Fine scale spatial genetic structure of two syntopic newts across a network of ponds: implications for conservation. *Conservation Genetics*, **14**, 385–400.

Stewart MM, Rossi J (1981) The Albany Pine Bush: a northern outpost for southern species of amphibians and reptiles in New York. *American Midland Naturalist*, 282–292.

Sun C, Shepard DB, Chong RA *et al.* (2012) LTR Retrotransposons Contribute to Genomic Gigantism in Plethodontid Salamanders. *Genome Biology and Evolution*, **4**, 168–183.

Titus VR, Bell RC, Becker CG, Zamudio KR (2014) Connectivity and gene flow among Eastern Tiger Salamander (Ambystoma tigrinum) populations in highly modified anthropogenic landscapes. *Conservation Genetics*, **15**, 1447–1462.

Trenham PC, Bradley Shaffer H, Koenig WD, Stromberg MR, Ross ST (2000) Life history and demographic variation in the California tiger salamander (Ambystoma californiense). *Copeia*, **2000**, 365–377.

Van der Auwera GA, Carneiro MO, Hartl C *et al.* (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, **43**, 11.10.1-33.

Voss SR, Kump DK, Putta S *et al.* (2011) Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes. *Genome Research*, **21**, 1306–1312.

Wang J (2009a) A new method for estimating effective population sizes from a single sample of multilocus genotypes. *Molecular Ecology*, **18**, 2148–2164.

Wang IJ (2009b) Fine-scale population structure in a desert amphibian: landscape genetics of the black toad (Bufo exsul). *Molecular Ecology*, **18**, 3847–3856.

Wang IJ (2012) Environmental and topographic variables shape genetic structure and effective population sizes in the endangered Yosemite toad. *Diversity and Distributions*, **18**, 1033–1041.

Wang IJ, Johnson JR, Johnson BB, Shaffer HB (2011) Effective population size is strongly correlated with breeding pond size in the endangered California tiger salamander, Ambystoma californiense. *Conservation Genetics*, **12**, 911–920.

Wang IJ, Savage WK, Bradley Shaffer H (2009) Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (Ambystoma californiense). *Molecular Ecology*, **18**, 1365–1374.

Waples RS, Antao T, Luikart G (2014) Effects of Overlapping Generations on Linkage Disequilibrium Estimates of Effective Population Size. *Genetics*, **197**, 769–780.

Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, **3**, 244–262.

Waples RK, Larson WA, Waples RS (2016) Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. *Heredity*, **117**, 233–240.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Watterson GA (1984) Allele frequencies after a bottleneck. *Theoretical Population Biology*, **26**, 387–407.

Weisrock DW, Shaffer HB, Storz BL, Storz SR, Voss SR (2006) Multiple nuclear gene sequences identify phylogenetic species boundaries in the rapidly radiating clade of Mexican ambystomatid salamanders. *Molecular Ecology*, **15**, 2489–2503.

Wigginton JE, Cutler DJ, Abecasis GR (2005) A Note on Exact Tests of Hardy-Weinberg Equilibrium. *The American Journal of Human Genetics*, **76**, 887–893.

Wright S (1943) Isolation by Distance. *Genetics*, **28**, 114–138.

Zamudio KR, Wieczorek AM (2007) Fine-scale spatial genetic structure and dispersal among spotted salamander (Ambystoma maculatum) breeding populations. *Molecular Ecology*, **16**, 257–274.

Zheng X, Levine D, Shen J *et al.* (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics (Oxford, England)*, **28**, 3326–3328.