Ghosts of a structured past: Impacts of ancestral patterns of isolation-by-distance on divergence-time estimation

Zachary B. Hancock*1,2, Heath Blackmon1,2

1Department of Biology at Texas A&M University

2Ecology & Evolutionary Biology Interdisciplinary Program at Texas A&M University

*Corresponding author: zhancock@bio.tamu.edu
Abstract

Isolation-by-distance is a widespread pattern in nature that describes the reduction of genetic correlation between subpopulations with increased geographic distance. In the population ancestral to modern sister species, this pattern may hypothetically inflate population divergence time estimation due to allele frequency differences in subpopulations at the ends of the ancestral population. In this study, we analyze the relationship between the time to the most recent common ancestor and the population divergence time when the ancestral population model is a linear stepping-stone. Using coalescent simulations, we compare the coalescent time to the population divergence time for various ratios of the divergence time over the population size. Next, we simulate whole genomes to obtain SNPs, and use the Bayesian coalescent program SNAPP to estimate divergence times. We find that as the rate of migration between neighboring demes decreases, the coalescent time becomes significantly greater than the population divergence time when sampled from end demes. Divergence-time overestimation in SNAPP becomes severe when the divergence-to-population size ratio < 10 and migration is low. Finally, we demonstrate the impact of ancestral isolation-by-distance on divergence-time estimation using an empirical dataset of squamates (*Tropidurus*) endemic to Brazil. We conclude that studies estimating divergence times should be cognizant of the potential ancestral population structure in an explicitly spatial context or risk dramatically overestimating the timing of population splits.

Subject area: Tree of Life: Population structure, phylogeography and phylogenomics

Keywords: Phylogenetics, multispecies coalescent, pairwise divergence, *Tropidurus*
Introduction

A major goal in phylogenetic and phylogeographic studies is the estimation of species divergence times. The topic has a long and contentious history largely centered around questions of how to appropriately apply fossil calibrations (e.g., Heath et al. 2014; Brown and Smith 2018), rate heterogeneity (Pond and Muse 2005), rate of morphological evolution (Lynch 1990), and selecting an adequate clock model (Douzery et al. 2004; Lepage et al. 2007).

Beyond methodological concerns are those that emerge from the nature of the data itself. Most phylogenetic models assume that fixed differences between species are the result of genetic drift, and under the neutral theory of molecular evolution (Kimura 1968; King and Jukes 1969) the rate of evolution (or substitution rate) is equal to the per generation neutral mutation rate, $\pi / 4N_e$ (Kimura 1983). For well-calibrated molecular clocks (e.g., Knowlton and Weigt 1998; Weir and Schluter 2008; Herman et al. 2018), we can estimate the time of divergence (usually in years) as $\pi_{12} / 2\pi$, where $\pi_{12}$ is the pairwise sequence divergence between species 1 and 2. However, in general we are not interested in estimating the divergence time of specific genetic variants, but rather the time of population divergence ($T_D$). For example, we might be interested in estimating the timing of a vicariant event that we suspect corresponds to a past geological upheaval.

There is a known discrepancy between the coalescent time of neutral genetic variants ($T_{MRCA}$) and $T_D$ (Nei and Li 1979; Nei and Takahata 1993). The degree of this discrepancy is determined by the ratio of $T_D / N_e$, where $N_e$ is the effective population size (Edwards and Beerli 2000; Rosenberg and Feldman 2002). This is because lineages must first be within the same population, which occurs $T_D$ generations in the past, followed by coalescence, which on average
requires $2N_e$ generations. Therefore, for a completely panmictic population: $T_{MRCA} = T_D + 2N_e$. The expected amount of pairwise sequence divergence is

$$E(\pi_{12}) = 2\pi [\pi + 2\pi]$$

(Wakeley 2000). When the ratio of $T_D / N_e$ is large, the bias in coalescent time in the ancestral population is minimal compared to $T_D$ (Edwards and Beerli 2000; Arbogast et al. 2002). However, as $T_D / N_e$ becomes small, $2N_e$ plays a major role in the overall sequence divergence between species. Rosenberg and Feldman (2002) evaluated the relationship between $T_{MRCA}$ and $T_D$ in a simple two population split model using coalescent simulations. They found that $T_{MRCA}$ converged on $T_D$ when the ratio of $T_D / N_e \approx 5$. Importantly, the $N_e$ in these models is that of the ancestral population; therefore, the extent of overestimation is the result of demographic conditions present in the ancestor. Demographic conditions that inflate $N_e$, such as ancestral population structure or a bottleneck following the split, are expected to have a major impact on divergence-time estimation (Gaggiotti and Excoffier 2000; Arbogast et al. 2002; Angelis and Dos Reis 2015).

Wakeley (2000) demonstrated that in descendant species who share an ancestor whose population dynamics are characterized by an island model (Wright 1931) with free migration between demes, overestimation of divergence-times are on the order of $2N_eD[1 + 1/(2M)]$ where $M = 2N_emD/(D - 1)$, $m$ is the migration rate and $D$ is the number of demes. The expected amount of pairwise sequence divergence is therefore

$$E(\pi_{12}) = 2\pi [\pi + 2\pi] \left(1 + \frac{1}{2M}\right)$$
where $D$ is the number of demes.

Population subdivision initially leads to shallow coalescent times where individuals within a shared deme rapidly find ancestors (the “scattering phase”; Wakeley 1998). However, since ancestral lineages must be in the same deme to coalesce, the rate in the “collecting phase” is characterized by the migration rate that shuffles ancestors around the range, reducing the probability that lineages coalesce (Wakeley 1998; 1999).

In the context of real populations, the island model of migration rarely applies (Meirmans 2012; Whitlock & McCauley 1999). Instead, population structure is the product of the spatial distribution and dispersal potential of the organism in question. Often this structure is in the form of isolation-by-distance (IBD). IBD is a widespread pattern in natural systems, characterized by a reduction in the probability of identity by descent (Wright 1943) or genetic correlation (Malécot 1968) with geographic distance. Patterns of IBD are most pronounced in stepping-stone models (Kimura 1953; Kimura and Weiss 1964) in which migration is restricted to neighboring demes. In this way, demes in close proximity share a greater proportion of migrants than they do with more distant demes. Distributions of coalescent times in stepping-stone models have been studied both in the context of one dimensional and two-dimensional models that are circular or toroidal (Maruyama 1970a; 1970b; Slatkin 1991), and in continuous models with joined ends (Maruyama 1971) or with discrete edges (Wilkins and Wakeley 2002). Slatkin (1991), using a circular stepping-stone model, showed that the probability for two genes sampled $i$ steps apart have an average coalescent time:

$$T_{MRCA} = 2N_eD + \left(\frac{\theta}{2}\right)\frac{D}{\theta}$$

Therefore, the amount of expected pairwise sequence divergence is:
The circular stepping-stone model should overestimate $T_D$ more dramatically as the number of demes becomes large and the distance between them increases. However, like the island model of free migration, circular ranges are likely rare in nature. Instead, natural populations are characterized by discrete range edges where end demes may only receive migrants from one direction (e.g., Peterson and Denno 1998; Broquet et al. 2006; Aguillon et al. 2017). Hey (1991) showed analytically in the case of a linear stepping-stone model that the distribution of coalescent times of two alleles from demes at the extremes of the range should coalesce much deeper than any two alleles chosen randomly from the population.

Vicariant speciation is considered one of the most common forms of allopatry (Coyne & Orr 2004), and results from the cessation of gene flow at some discrete barrier in a species range. This form of speciation has been invoked across many empirical systems (e.g. Riddle et al. 2000; Bocxaeler et al. 2006; Hancock et al. 2019). Vicariant speciation in organisms with low dispersal abilities may maintain strong allelic differences at the range edges ancestrally consistent with a pattern of IBD. For example, Hancock et al. (2019) found that sister species of beach amphipods in the Gulf of Mexico with large ranges showed patterns of IBD within species. In addition, they identified a distinct barrier to gene flow (the Mississippi River) and posited that this resulted in vicariant speciation. Since both IBD and vicariant speciation are presumed common in nature, biases in divergence-estimation based on $\pi_{12}$ could be widespread.

Ultimately, the degree to which $T_{MRCA}$ impacts phylogenetic inference and divergence-time estimation is dependent on its impact on $\pi_{12}$. Given that lower migration rates lead to greater $T_{MRCA}$ (Hey 1991), we expect that differentiation ($\pi_{12}$) between end demes compared to
center demes will become more pronounced at smaller $m$. If the difference between the $T_{\text{MRCA}}$ of central demes and end demes is dramatic enough, we expect that divergence dating of species that arose from ancestral end demes may significantly overestimate $T_D$.

In this study, we estimate mean $T_{\text{MRCA}}$ for two genes sampled in descendant species (either from the ends or the center of the ancestral range) in which the ancestral population is characterized by a stepping-stone model with discrete ends using a simulation approach. In particular, we are interested in what value of $T_D / ND$ we expect $T_{\text{MRCA}}$ to converge on $T_D$. We use $ND$ (the product of the census size and deme number) as our expected $N_e$ under panmixia (Wakeley 2009). Next, we examine the distribution of $\pi_{12}$ across the genome under different simulated migration conditions to compare with expectations under a panmictic model. We then test the performance of the phylogenetic inference program SNAPP (Bryant et al. 2012) on simulated single nucleotide polymorphism (SNP) data to evaluate how these trends may bias our inference of species divergence times. SNAPP is a BEAST (Bouckaert et al. 2014) package that operates under an explicit coalescent framework, inferring gene trees from individual SNPs. The program is ideal for phylogeographic studies that utilize RADseq and other genotype-by-sequencing (GBS) technologies to generate thousands of SNPs across the genome (e.g. Manthey et al. 2015; Dowle et al. 2017; Manthey et al. 2017; Leslie and Morin 2018), and has been used explicitly in divergence-time dating previously (Strange et al. 2018; Spalink et al. 2019; Fang et al. 2020). Finally, we illustrate how ancestral IBD can inflate divergence-time estimates on an empirical phylogenomic dataset of lizards (Domingos et al. 2017).
Methods

In the following methods, we use the term “deme” to represent a subpopulation of randomly mating individuals within a broader collection of demes that we refer to as the “population”. To be a part of the population, a deme must be able to share migrants with other demes within the population. We use the term “species” to represent an isolated randomly mating unit that no longer shares migration with other populations or demes. This is not meant to reflect any species definition. Finally, we use the term “end species” and “center species” to refer to a set of sister species that either descend from demes on opposite ends of the ancestral range (end species) or descend from neighboring demes in the range center (center species). Our focus below is on evaluating the coalescent times, pairwise differences, and estimated divergence times between these sister species (i.e., the outgroup – “sp3” in Fig. 1A – is only meant to root the tree).

For all simulation models, we used the product of the census population size ($N$) and the deme number ($D$) to evaluate the relationship of $T_D / N$ (Rosenberg and Feldman 2002). Since sp1 and sp2 (Fig. 1) transition to panmixia following the split at time $T_D$, $ND$ should approximate $N_e$, though there will be a period of nonequilibrium immediately following divergence. Therefore, the contribution to $\pi_{12}$ from $T_D$ is on the order of $ND$. The ancestral contribution of $\pi_{12}$ will necessarily be some value greater than $ND$ due to population structure (i.e., $N_e > ND$; Wakeley 2000). Our interest here is explicitly on how much greater this contribution is relative to $T_D$. 
Coalescent simulations

To evaluate the relationship between T_D / N and (T_{MRCA} – T_D) / T_D when the ancestral population is characterized by a stepping-stone model of migration, we used fastsimcoal2 (Excoffier et al. 2013) simulations over a wide range of T_D / N values. Specifically, starting at time 0 and going backwards, each simulation consisted of initially two species with no migration between them until time T_D in the past. At T_D, these two species merge into an ancestral population with 10 demes (D) following a linear stepping-stone migration model. For simulations of end species, the ancestral deme of each species was on opposite ends of the range (i.e., demes 1 and 10 in Fig. 1A). For the center species, the ancestral demes were neighboring and in the range center. For each of these two models, we sampled k = 2 individuals to coalesce, and each simulation terminated upon coalescence.

In the ancestral population, center demes received migrants from neighboring demes at rate 2m, whereas demes at the end of the range received migrants at rate m. This is due to the fact that end demes have only a single neighbor, whereas all center demes have two neighbors (Fig. 1A). The ancestral population was simulated for migration rates of 0.1, 0.01, and 0.001, and a range of T_D / ND values from 0.01–10. In addition, we simulated an island model of migration for comparison with the stepping-stone model. In the island model, the ancestral population consisted of 10 demes with free migration between each at rate m. This resulted in a total of 84 distinct simulation scenarios, and each were replicated 1,000 times. We did not explicitly model chromosomes; instead, replicates were treated as independent loci.

To statistically compare between the three models (end species sampled in stepping-stone, center species in stepping-stone, and the island model), we subset ratios of T_D / ND to
values of 10, 5, 2, 1, 0.5, and 0.1. Resulting $T_{MRCA}$ distributions for each population model were compared using a pairwise Wilcoxon test in the R platform (R Core Team 2019), as the resulting distributions were non-normal.

*Genome simulations*

To evaluate how ancestral IBD impacts pairwise sequence divergence ($\pi_{12}$), genome-wide coalescent times ($T_{MRCA}$), and divergence-time estimation, we performed hybrid simulations that combined the coalescent simulator *msprime* (Kelleher et al. 2016) and the forward-time simulator SLiM v3.3 (Haller and Messer 2019). Since forward-time simulators begin with individuals that are completely unrelated, often a neutral burn-in period is required to allow coalescence or mutation-drift equilibrium to occur (Haller et al. 2019). This can be computationally costly and time consuming; however, using tree-sequence recording methods in SLiM (Haller et al. 2019) we can bypass the need to equilibrate during the forward-time simulation. To generate a panmictic ancestral population with a coalescent history, we simulated 2000 individuals ($N_e = 4000$) using *msprime* with genome sizes of 10 Mb and a recombination rate of $10^{-8}$ (~0.1 recombination events per individual per generation). The resulting coalescent trees were then imported into SLiM as the basis for the starting population.

In SLiM, the initial population was split into two populations of $N = 1000$: 1) an outgroup that remained panmictic (“sp3” in Fig. 1A) and 2) the ancestral population of the sister species “sp1” and “sp2”, which was subdivided into 10 demes ($N = 100$ per deme) in a linear stepping-stone model (Fig. 1A). These dynamics persisted for 50,000 generations after which the ancestral population was split into either end species or center species (see Fig. 1A). This was done by
removing the intermediate demes and instantaneously adjusting $N$ for the two species to 1000 so that the census size remained constant. The resulting 3-species were then allowed to evolve for $T_D$ generations before the simulation was terminated. Five different $T_D$ values were simulated which correspond to $T_D / ND$ ratios of 50, 25, 10, 5, and 1 ($T_D$ values of 50,000, 25,000, 10,000, 5,000, and 1,000 generations). These values of $T_D / ND$ were chosen based on the results from the coalescent simulations in fastsimcoal2 (see Results); for values $>10$, $T_{\text{MRCA}}$ is expected to converge on $T_D$, whereas values $<10$ are expected to overestimate $T_D$ regardless of migration rates.

The resulting tree-sequences from the SLiM simulation were imported into Python 3 using pyslim, and we overlaid neutral mutations ($\mu = 10^{-7}$ per base per generation) onto the trees using msprime. Pairwise divergence ($\pi_{12}$) was then estimated across the genome in windows of 100 kb for both end demes and center demes. These values were also converted into generations using $\pi_{12} / 2\mu$, which gives a rough estimate of divergence time per window.

Equation 1 primes our expectation for the amount of sequence divergence expected given some value of $T_D$ and ancestral $N_e$. By rearranging equation 1, we can naively calculate the ancestral $N_e$ from genome-wide $\pi_{12}$ as:

$$N_e = \frac{\pi_{12} - 2T_D\mu}{4\mu}$$

(5)

From this, we plot estimated ancestral $N_e$ within 100 kb windows across the genome to compare with the known census population size ($N_c = 1000$), and to evaluate the relationship between $N_e$ and $N_c$ in the presence of IBD.
Next, we plotted the distribution of coalescent times ($T_{\text{MRCA}}$) across the genome to visualize differences between $T_{\text{MRCA}}$ of end and center species. Median $T_{\text{MRCA}}$ for each ratio and migration rate was compared via a Kruskal-Wallis test and a pairwise Wilcoxon rank test in R due to the data violating normality.

Each simulation produced >200,000 SNPs. For divergence-time analysis, we randomly sampled 3000 SNPs—a number found by Strange et al. (2018) to optimally perform in SNAPP. Each run consisted of 10 individuals from species sp1 and sp2, and 1 individual from the outgroup population, sp3 (Fig. 1). Unlike other fully coalescent models, SNAPP does not sample from gene trees directly to estimate the species tree, but instead integrates over all possible gene trees using biallelic SNPs. The method has been found previously to perform well on both simulated and empirical data (Bryant et al. 2012; Strange et al. 2018). We designated a gamma-distributed prior on $\Theta (=4N_e\theta)$ with a mean equal to the expected $\pi_{12}$ (equation 1). Forward ($u$) and backward ($v$) mutation rates were estimated within BEAUti (Bouckaert et al. 2014) from the empirical SNP matrix using the tab Calc_mutation_rates, and these values were sampled during the MCMC. The rate parameter $\theta$, which is the birth-rate on the Yule tree prior, was gamma-distributed with $\alpha = 2$ and $\beta = 200$, where the mean is $\alpha / \beta$ (Leaché and Bouckaert 2018).

SNAPP is designed to handle incomplete lineage sorting (ILS), but to minimize its effects—since we are not interested in the program’s ability to estimate topology but rather branch-lengths—we applied a fixed species tree. Branch-lengths in SNAPP do not scale to time, but instead are measured in number of substitutions. Given a fixed mutation rate, we convert the number of substitutions separating sp1 and sp2 to the number of generations as $g = s / \theta$, where $s$ is branch-lengths in units of substitutions (Bouckaert and Bryant 2015). The MCMC chain length was 10–50 million sampling every 1000 with a burn-in of 10%, ensuring that ESS values of
interest were all >200. Runs were performed on the high-performance computing cluster CIPRES (www.phylo.org; Miller et al. 2010).

MCMC log files were then downloaded and analyzed in R. The performance of SNAPP was evaluated by comparing traces of end and center species across migration rates and $T_D / ND$ values. Results were evaluated by first randomly sampling 1000 rows for each migration rate, and then performing a Kruskal-Wallis test. Trees from the MCMC were summarized in TreeAnnotator v.2.6.0 (Bouckaert et al. 2014) and visualized in R using the package ggtree (Yu et al. 2017). Branch colors were scaled by estimated median $\theta$ per branch.

To ensure the trends observed were the result of inflated $\pi_{12}$ when $T_D / ND$ and migration rate is low and not an issue unique to SNAPP, we performed pairwise $F_{ST}$ tests of end and center demes that were used in the divergence-time estimation (see Supplementary material; Fig. S10). These tests were performed using tskit. All SLiM recipes, python and R code, and xml files can be found at https://github.com/hancockzb/ancestralIBD.

**Empirical dataset**

To evaluate how ancestral patterns of IBD may impact divergence-time estimation in practice, we analyzed the phylogenomic dataset of endemic squamates from Domingos et al. (2017). The dataset consists of 12 species (including the identified cryptic lineages) sampled broadly across the geographic range of *Tropidurus itambere*, which is native to the Cerrado, a tropical savanna in Brazil that stretches across the states of Goiás, Mato Grosso do Sul, Mato Grosso, Tocantins, Minas Gerais, and the Federal District. Domingos et al. (2017), using anchored hybrid enrichment, identified five cryptic species they designated A–E within *T. itambere*. Each of these
species were sampled across multiple localities with some locations spatially nearer to their close relatives than others (see Fig. 1 in Domingos et al. 2017).

This geographically broad sample scheme is ideal to test the impact of IBD on divergence-time estimation. If ancestral IBD has influenced $\pi_{12}$, we expect species localities more distant to one another to be more deeply diverged than when two species are sampled from locations nearby. Importantly, this pattern should only hold if the range was once continuous (as in our simulations above); otherwise, there should be no difference in $\pi_{12}$ between species even if there are current patterns of IBD within species. Ongoing gene flow between the spatially close localities could also generate this pattern, but Domingos et al. (2017) found no evidence for this.

The alignments used in the coalescent species delimitation in Domingos et al. (2017) were downloaded from https://datadryad.org/stash/dataset/doi:10.5061/dryad.1hs2m. We randomly selected 10 loci conditional on them containing samples from all locations (total length = 15,303 bp). We generated two separate alignment files based on the sampled locality’s distance from their nearest relative. Preliminary analyses including the entire dataset showed that species (DE) were sister to A, and B was sister to C (which was also found from the entire concatenated dataset in Domingos et al. 2017). Therefore, the first dataset we designated as “far”, and it included only the location of E that was most distant from A (i.e., São João D’Aliança; roughly 1,000 km), and the location of C that was the most distant to B (i.e., Ribeirão Cascalheira; ~1,000 km). The second dataset was designated “near” which included localities of E geographically closest to A (Brasilia and Pirenópolis; ~160 km) and C to B (Barra do Garças; ~600 km).
Phylogenetic inference was performed using the fully coalescent software *BEAST (Heled & Drummond 2010). For each locus, we applied the HKY model (Hasegawa et al. 1985) and a strict molecular clock. We also performed runs with a relaxed lognormal clock for each locus to ensure age differences across the tree were not the result of restricting substitution rate variation. Domingos et al. (2017) only focus on topology and time is not considered; therefore, we estimate branch-lengths in units of substitutions per site. We use an arbitrary rate of mutation ($10^{-8}$) to convert substitutions to generations. This analysis is not meant to be a rigorous evaluation of the true time of divergence, but merely a demonstration of the impacts of ancestral IBD on empirical data. Each analysis was run for 100 million generations with a burn-in of 10%. Estimated divergence times between the two models were compared using a two-way ANOVA in R. Differences in the densities of estimated gene trees from the posterior were visualized using DensiTree (Bouckaert 2010).

Results

Coalescent simulation results

The coalescent simulations produced trends superficially similar to those found by Rosenberg and Feldman (2002). At the lowest $T_D / ND$, the proportion of deep coalescence was dramatically greater than at higher values with the curve producing a similar logarithmic relationship (Fig. S1). However, $T_D$ and $T_{MRCA}$ did not necessarily converge when $T_D / ND = 5$. Instead, the rate of convergence was dependent on both the deme sampled and the migration rate.

When migration was high ($m = 0.1$) and $T_D / ND$ was less than 0.5, there was no significant difference between center or end species in the stepping-stone model or the island
model. However, for values of \( T_D/ND > 0.5 \), the \( T_{MRCA} \) of end species became significantly different from both island \((p < 0.02)\) and center species \((p < 0.01; \text{see Table S1})\). When migration was reduced below 0.1, this pattern became more extreme. End species were significantly different in all pairwise comparisons of models \((p < 0.000001)\), and center species differed from the island model at \( T_D/ND \) ratios of 0.5, 2, and 10 \((p < 0.03)\) when \( m = 0.01 \). At the lowest migration rate simulated \((m = 0.001)\), all pairwise model comparisons were significantly different when \( T_D/ND > 0.5 \) \((p < 0.001; \text{see Table S1})\).

**Genome simulation results**

Results from the genome simulation approach corroborated those found with *fastsimcoal2*. Regardless of \( T_D/ND \), when \( m = 0.1 \) the difference between center and end species was less severe relative to when \( m < 0.1 \) (Table S2). Across the simulated genomes, \( T_{MRCA} \) became dramatically deeper between end than center species as migration fell below 0.01. For the genome-wide divergence estimates, the degree of overestimation depended on the ratio of \( T_D/ND \). While all scenarios where \( m = 0.001 \) overestimated the true \( T_D \), when \( T_D/ND < 10 \) end species were 5–60 times more diverged than expected (Fig. S3). This is a direct result of the deeper coalescent times between end species when \( m < 0.1 \), as these longer branches provide more time for mutations to occur and accumulate (Fig. S2).

Genome-wide coalescent times \((T_{MRCA})\) are shown in Fig. S2. When \( m = 0.1 \), only \( T_D/ND = 25 \) and 10 were significantly different between end and center species \((p < 0.005)\). Regardless of \( T_D/ND \), the variance in \( T_{MRCA} \) steadily increased with decreasing \( m \). Indeed, the increase in mean \( T_{MRCA} \) when \( m = 0.001 \) appears largely driven by an increase in the variance at
this lower rate. Due to this, we find that ancestral $N_c$ dramatically exceeds $N_c$ when $m = 0.001$ (Fig. 3).

Despite the potential for divergence-time overestimation to be extreme, SNAPP was relatively resilient when $T_D / ND > 10$ and when $m > 0.001$. When $T_D / ND = 50$, SNAPP was overly conservative and underestimated the number of substitutions expected to occur (Fig. 2). When $T_D / ND = 25$, the mean estimate of both center and end species when $m > 0.001$ either underestimated the true age or was within 5%. However, for end species where $m = 0.001$ the estimated divergence time exceeded the true age by $\sim 80\%$ (Table S3). A similar trend occurred when $T_D / ND = 10$ and 5. Here, both center and end species overestimated the true age, but the end species did so more dramatically (138% the true age versus 81% for 10; 184% versus 67% for 5). The most dramatic overestimation occurred between end species when $T_D / ND = 1$ at $\sim 700\%$ the true age. Importantly, this was not merely the result of a low $T_D / ND$ ratio, as the other migration regimes performed well. In fact, most were closer to the true $T_D$ than the expected $\pi_{12}$ accounting for $2N$ (Table S3).

Estimated $\theta$ for each branch is shown in Fig. 4 for $T_D / ND = 10$, and in Figs. S4–S7 for the remaining ratios. For all $T_D / ND$ values except 1, the median ancestral $\theta$ was higher for end species than center when $m = 0.001$, and the estimated $\theta$ for the descendant species (sp1 and sp2 in Fig. 1) was considerably lower than for the ancestor or the outgroup, sp3 (Fig. 4; Figs. S4–S7). These patterns are consistent with a population bottleneck, despite $N$ being maintained throughout the simulation.
Empirical results

The estimated divergence-time for the clades ((DE)A) and (BC) were significantly older when samples were from geographically distant localities as opposed to those nearby (p < 0.00001; Fig. 4c). For the ((DE)A) clade, the “far” dataset inferred an age of divergence of 270,000 generations, which was 40,000 generations (or ~14%) higher than the “near” estimate. The (BC) divergence was even more extreme, with the “far” being ~24% older than the “near” (Fig. 4c). Interestingly, despite the fact that all other samples were included in both analyses, the “far” dataset estimated older ages for most of the other nodes in the tree as well (Fig. 4). The total tree height of the “far” was 220,000 generations deeper in time than the “close” (or ~ 9%). The relaxed clock estimates were more extreme, with all node heights being higher in the “far” versus “near” datasets (Fig. S11).

Discussion

Macroevolutionary patterns are ultimately governed by microevolutionary processes (Li et al. 2018), an observation Lynch (2007), extending Dobzhansky’s (1973) maxim, summed up as “nothing in evolution makes sense except in light of population genetics”. In this light, we have demonstrated that the population genetic environment of the ancestor shapes the genetic landscape of descendant species. This has been known to impact tree topology when ILS is common (Kubatko and Degnan 2007) and overestimate divergence times in the presence of population structure caused by an island model of migration (Edwards and Beerli 2000; Wakeley 2000). Extensive prior work has shown that the stepping-stone model of migration reduces genetic correlation between demes (Kimura and Weiss 1964; Maruyama 1970a) and that demes
farther apart should coalesce deeper in time than those geographically closer (Slatkin 1991; Hey 1991). However, to our knowledge, the impact of ancestral IBD has not been evaluated in the context of divergence-time estimation previously.

Rosenberg & Feldman (2002) found previously that when $T_D / N = 5$, $T_{MRCA}$ and $T_D$ largely converged in a simple population split model. However, when in the presence of ancestral IBD we found that convergence was dependent on the migration rate (i.e., the strength of ancestral IBD) and whether surviving species neighbored each other or were at the range ends in the ancestral population.

When $T_D / ND > 10$, the ancestral dynamics contribute little to the divergence-time estimate differences between center and end species. However, as this ratio decreases the contribution of $2N_c$ to overall sequence divergence becomes non-trivial. The probability that genetic variants share an ancestor just prior to the population split is higher between species that were geographically closer than those more distant in the ancestral population. This is mediated by the migration rate, which, when high enough, can largely erase the differences between center and end demes. When migration is high (10%, or $m = 0.1$), individuals move well between demes and the coalescent times largely converge (though deeper in time depending on the ratio of $T_D / ND$). However, as $m$ falls below 1% ($m = 0.01$), or less than one migrant per generation being shared between demes, dispersal cannot keep up with genetic differentiation. Despite all migration regimes producing similar patterns of IBD (Fig. S9), $F_{ST}$ becomes dramatically higher as migration drops below 1%. This differentiation in the ancestor contributes to the overall sequence divergence ($\pi_{12}$) between species, which drives an overestimation of the time of the population split ($T_D$) when surviving lineages descend from demes on the opposite ends of the range.
As expected, ancestral IBD skews $\pi_{12}$ and $T_{\text{MRCA}}$ away from expected values in a panmictic population, and this caused an inflation in $N_e$ relative to $N_c$. For $T_D / ND = 50$ and $m = 0.1$, the mean $\pi_{12}$ for end species was 0.010459 and 0.010419 for center species. Using equation 5, ancestral $N_e$ was 1147.5 for end species and 1047.5 for central. However, when $m = 0.001$, $\pi_{12}$ for end species was 0.012948, an ancestral $N_e = 7370$. Center species, on the other hand, only increased to $N_e = 1255$. As with the coalescent times, at lower migration rates the variance in $N_e$ becomes exceedingly large, driving up the mean. Importantly, mean genome-wide $N_e$ always exceeds $N_c$ in the presence of ancestral IBD at a level dictated by the migration rate.

This feature of ancestral IBD has important consequences for conservation genetics. Many studies use $N_e$ as a rough biological measure of population size (Turner et al. 2002; Rieman and Allendorf, 2011; Hare et al. 2011), and therefore a metric of the health of a population. However, a common phenomenon in range contractions is fragmentation and isolation (Ceballos et al. 2017), which may result in IBD. If many of the demes once contributing to the connectivity of the population have become extinct, and $N_e$ is estimated based on the surviving demes, it will overestimate the actual number of individuals within the population (i.e., the census size, $N_c$). Thus, we might incorrectly conclude that a species has a larger population size than it actually does, which may lead to mismanagement.

Since $N_e$ is inflated in the ancestral lineage, the descendant species appear to pass through a bottleneck despite $N$ remaining constant (Figs. S4–S8). Estimated $\theta$ in SNAPP captured this dynamic with more extreme differences in $\theta$ (i.e., more dramatic bottlenecks) being inferred between end species and when $m = 0.001$. Population bottlenecks have been found to cause divergence-time overestimation due to random differential survival of ancestral alleles into the descendant species (Gaggiotti and Excoffier 2000). In the presence of IBD, this differential
allelic persistence between species is mimicking a bottleneck—when demes were far apart this pattern is more extreme as they already maintain different allelic patterns ancestrally. However, because this pattern is recognizable (Figs. S4–S8) it can be used to signal when ancestral IBD may be impacting our divergence-time estimation. Unfortunately, without prior range-size knowledge it may be impossible to differentiate between ancestral IBD and a bottleneck since these produce virtually identical genetic patterns. However, it may not be necessary to do so for simple divergence estimates.

Demes need not necessarily go extinct for ancestral IBD to still influence $\pi_{12}$, as seen from the results of the empirical dataset; however, the persistence of demes into the present allows for a geographically aware sampling scheme. Since Domingos et al. (2017) sampled broadly across the range of *T. itambere*, they would be well-positioned to identify inflated $\pi_{12}$ resulting from ancestral structure. For example, as we have done here, by subsetting the dataset by geographic proximity one can explicitly test for ancestral IBD. For ancestrally structured populations, geography should dictate the degree of $\pi_{12}$. Importantly, this also requires that species diverged via vicariance (Coyne & Orr 2004) – the splitting of a once larger range by a discrete barrier – and not some other means, such as population expansions following divergence from a more restricted habitat.

The broader impact of ancestral IBD on divergence-time estimation when in the context of large phylogenies is beyond the scope of this work, but it is conceivable that the longer than expected branches between sister species might bias rate estimation (Aris-Brosou and Excoffier 1996; Magallón 2010). In the case of ancestral IBD, the inflated $N_e$ is mimicking a pattern of substitution rate increase. Under neutrality, the rate of substitution is equal to the per generation mutation rate, $\mu$ (Kimura 1983); however, in the presence of population structure, substitutions
may occur in the ancestral lineages between demes separated by large geographic distances. If
the true age of the sister taxa is known but ancestral structure is not accounted for, the
substitution rate will be upwardly biased. We found some evidence for this in the *T. itambere*
dataset, in which we found higher estimates of $\pi_{12}$ in the “far” versus the “close” dataset even for
nodes more distantly related to the focal clades (Fig. 4).

Ancestral structured populations leave their imprint on descendent species in the form of
greater coalescent times, and therefore larger than expected pairwise divergences between
species. Further, these patterns cause inflated $N_e$ relative to census sizes. Since ancestral IBD
mimics the signature of a population bottleneck, coalescent methods that co-estimate $\theta$ along
with the topology and $\pi_{12}$, such as SNAPP and *BEAST, may be the best suited to reveal this
potential source of bias. However, fully coalescent models such as these are infamously
computationally costly and not presently used for whole-genome sequence data or for
phylogenies with large numbers of tips. Indeed, SNAPP becomes prohibitively slow when the
number of tips is ~30 (Leaché and Bouckaert, 2018).

In the context of larger phylogenies or organisms in which little is known about their
ancestral range, it may be impossible to know if extant species descend from range centers or
ends, or the level of IBD present in the ancestor. The genetic consequences of ancestral structure
therefore behave much like “ghost” populations (Slatkin 2005); despite being extinct, their
influence haunts our ability to adequately assess the phylogenetic history of their descendants.
Funding

This work was supported by National Institute of General Medical Sciences at the National Institutes of Health R35GM138098.

Acknowledgements

Thanks to Ben Haller and Wesley Brashear for coding help.

Data availability

SLiM recipes, R and python code, and .XML files have been uploaded to

https://github.com/hancockzb/ancestralBD.
References

Aguillon, S.M., Fitzpatrick, J.W., Bowman, R., Schoech, S.J., Clark, A.G., Coop, G., and N. Chen. 2017. Deconstructing isolation-by-distance: The genomic consequences of limited dispersal. PLoS Genet. 13(8): e1006911.

Angelis, K. and M. Dos Reis. 2015. The impact of ancestral population size and incomplete lineage sorting on Bayesian estimation of species divergence times. Current Zoology. 61(5): 874–885.

Arbogast, B.S., Edwards, S.V., Wakeley, J., Beerli, P., Slowinski, J.B. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Ann. Rev. Ecol. Syst. 33: 707–740.

Aris-Brosou, S., and L. Excoffier. 1996. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Mol. Biol. Evo. 13(3): 494–504.

Bocxlaer, I.V., Roelants, K., Biju, S.D., Nagaraju, J., Bossuyt, F. 2006. Late Cretaceous vicariance in Gondwanan amphibians. PloS ONE, 1(1): e74.

Bouckaert, R. 2010. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics, 26(10): 1372–1373.

Bouckaert, R., and Bryant, D. 2015. A rough guide to SNAPP. https://www.beast2.org/snapp/

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C-H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comp. Bio. 10(4): e1003537.
Broquet, T., Ray, N., Petit, E., Fryxell, J.M., and F. Burel. 2006. Genetic isolation by distance and landscape connectivity in the American marten (Martes americana). Land. Eco. 21: 877–889.

Brown, J.W., and S.A. Smith. 2018. The past sure is tense: On interpreting phylogenetic divergence time estimates. Syst. Biol. 67(2): 340–353.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A. and A. RoyChoudhury. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Bio. Evo. 29(8): 1917–1932.

Ceballos, G., Ehrlich, P.R., and R. Dirzo. 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. PNAS. 114(30): E6089-E6096.

Coyne, J. and H.A. Orr. 2004. Speciation. Sinauer Associates, Inc. Sunderland, Massachusetts, 545 pp.

Dobzhansky, T. 1973. Nothing in biology makes sense except in the light of evolution. American Biology Teacher. 35: 125–129.

Domingos, F.M.C.B, Colli, G.R., Lemmon, A., Lemmon, E.M., Beheregaray, L.B. 2017. In the shadows: Phylogenomics and coalescent species delimitation unveil cryptic diversity in a Cerrado endemic lizard (Squamata: Tropidurus). Mol. Phylo. Evo., 107: 455–465.

Douzery, E.J.P., Snell, E.A., Bapteste, E., Delsuc, F., and H. Philippe. 2004. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? PNAS. 101(43): 15386–15391.

Dowle, E.J., Bracewell, R.R., Pfrender, M.E., Mock, K.E., Bentz, B.J., Ragland, G.J. 2017. Reproductive isolation and environmental adaptation shape the phylogeography of mountain pine beetle (Dendroctonus ponderosae). Mol. Ecol. 26(21): 6071–6084.
Edwards, S.V., and P. Beerli. 2000. Gene divergence, population divergence, and the variance in coalescent time in phylogeographic studies. Evolution. 54(6): 1839–1854.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C., and M. Foll. 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9(10): e1003905.

Fang, B., Merilä, J., Matschiner, M., Momigliano, P. 2020. Estimating uncertainty in divergence times among three-spined stickleback clades using the multispecies coalescent. Mol. Phylogenet. Evol. 142: 106646.

Gaggiotti, O.E., and L. Excoffier. 2000. A simple method for removing the effect of a bottleneck and unequal population sizes on pairwise genetic distances. Proc. B. 267(1438): 81–87.

Haller, B.C., and P.W. Messer. 2019. SLiM 3: Forward genetic simulations beyond the Wright-Fisher Model. Mol. Bio. Evo. 36(3): 632–637.

Haller, B.C., Galloway, J., Kelleher, J., Messer, P.W., and P.L. Ralph. 2019. Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. Mol. Eco. Res. 19(2): 552–566.

Hancock, Z.B., Hardin, F.O., Light, J.E. 2019. Phylogeography of sand-burrowing amphipods (Haustroriidae) supports an ancient suture zone in the Gulf of Mexico. J. Biogeo. 46(11): 2532–2547.

Hare, M.P., Nunney, L., Schwartz, M.K., Ruzzante, D.E., Burford, M., Waples, R.S., Ruegg, K., and E. Palstra. 2011. Understanding and Estimating Effective Population Size for Practical Application in Marine Species Management. 25(3): 438–449.

Heath, T.A., Hueslenbeck, J.P., and T. Stadler. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. PNAS. E2957-E2966.
Maldonado, E., Gross, J.B., Cartwright, R.A., Rohner, N., Warren, W.C., and S.E. McGaugh. 2018. The role of gene flow in rapid and repeated evolution of cave-related traits in Mexican tetra, *Astyanax mexicanus*. Mol. Eco. 27(22): 4397–4416.

Hey, J. 1991. A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. Theo. Pop. Bio. 39: 30–48.

Kelleher, J., Etheridge, A.M., and G. McVean. 2016. Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. PLoS Comput. Biol. 12(5): e1004842.

Kimura, M. 1953. “Stepping Stone” model of population. Ann. Rept. Nat. Inst. Genetics, Japan. 3: 62–63.

Kimura, M. 1968. Evolutionary rate at the molecular level. Nature. 217: 624–626.

Kimura, M. 1983. The Neutral Theory of Molecular Evolution. Cambridge Univ. Pres., Cambridge, UK.

Kimura, M. and G. Weiss. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics. 49: 561–576.

King, J.L., and T.H. Jukes. 1969. Non-Darwinian evolution. Science. 164: 788–798.

Knowlton, N. and L.A. Weigt. 1998. New dates and new rates for divergence across the Isthmus of Panama. Proc. B. 265(1412): 2257–2263.

Kubatko, L.S., and J.H. Degnan. 2007. Inconsistency of Phylogenetic Estimates from Concatenated Data under Coalescence. Sys. Bio. 56(1): 17–24.

Leaché, A., and Bouckaert, R. 2018. Species Trees Estimation with SNAPP: A Tutorial and Example. Workshop on Population and Speciation Genomics, Český Krumlov.
Lepage, T., Bryant, D., Philippe, H., and N. Lartillot. 2007. A general comparison of relaxed molecular clock models. Mol. Biol. Evo. 24(12): 2669–2680.

Leslie, M.S. and P.A. Morin. 2018. Structure and phylogeography of two tropical predators, spinner (Stenella longirostris) and pantropical spotted (S. attenuata) dolphins, from SNP data. R. Soc. 5:171615.

Li, J., Huang, J-P., Sukumaran, J., and L.L. Knowles. 2018. Microevolutionary processes impact macroevolutionary patterns. BMC Evo. Bio. 18: 123.

Lynch, M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. Am. Nat. 136(6): 727–741.

Lynch, M. 2007. The Origins of Genome Architecture. Sinauer Associates, Inc. Sunderland, MA.

Magallón, S. 2010. Using Fossils to Break Long Branches in Molecular Dating: A Comparison of Relaxed Clocks Applied to the Origin of Angiosperms. Sys. Bio. 59(4): 384–399.

Malécot, G. 1968. The Mathematics of Heredity. Translated from the French edition (Paris, 1948). Ed. Yermanos, D.M. Freeman, San Francisco, 1969.

Manthey, J.D., Klicka, J., Spellman, G.M. 2015. Chromosomal patterns of diversity and differentiation in creepers: a next-gen phylogeographic investigation of Certhia americana. Hereditiy. 115: 165–172.

Manthey, J.D., Moyle, R.G., Gawin, D.F., Rahman, M.A, Ramji, M.F.S., Sheldon, F.H. 2017. Genomic phylogeography of the endemic Mountain Black-eye of Borneo (Chlorocharis emiliae): montane and lowland populations differ in patterns of Pleistocene diversification. J. Bio. 44(10): 2272–2283.

Maruyama, T. 1970a. Effective number of alleles in a subdivided population. Theo. Pop. Bio. 1: 273–306.
Maruyama, T. 1970b. The rate of decrease of heterozygosity in a population occupying a circular or linear habitat. Genetics. 67: 437–454.

Maruyama, T. 1971. Analysis of population structure: II. Two-dimensional stepping stone models of finite length and other geographically structured populations. Ann. Hum. Gen., Lon. 35: 179–196.

Meirmans, P.G. 2012. The trouble with isolation by distance. Mol. Eco. 21(12): 2839–2846.

Miller, M.A., Pfeiffer, W., and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans, LA.

Nei, M., and W.H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. PNAS. 76(10): 5269-5273.

Nei, M., and N. Takahata. 1993. Effective population size, genetic diversity, and coalescence time in subdivided populations. J. Mol. Evo. 37: 240–244.

Peterson, M.A., and R.F. Denno. 1998. The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. Am. Nat. 152(3): 428–446.

Pond, S.K., and S.V. Muse. 2005. Site-to-site variation of synonymous substitution rates. Mol. Biol. Evo. 22(12): 2375–2385.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Riddle, B.R., Hafner, D.J., Alexander, L.F., Jaegar, J.R. 2000. Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. PNAS, 97(26): 14438–14443.
Rieman, B.E., and F. W. Allendorf. 2011. Effective Population Size and Genetic Conservation Criteria for Bull Trout. North American Journal of Fisheries Management. 21(4): 756–764.

Rosenberg, N.A., and M.W. Feldman. 2002. The relationship between coalescent times and population divergence times. In: Modern Developments in Theoretical Population Genetics. Eds. Slatkin, M. and M. Veuille. Oxford Univ. Pres. New York, NY.

Slatkin, M. 1991. Inbreeding coefficients and coalescence times. Genet. Res., Camb. 58: 167–175.

Slatkin, M. 2004. Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. Mol. Eco. 14(1): 67–73.

Spalink, D., MacKay, R., Sytsma, K.J. 2019. Phylogeography, population genetics and distribution modelling reveal vulnerability of Scirpus longii (Cyperaceae) and the Atlantic Coastal Plain Flora to climate change. Mol. Ecol. 28(8): 2046–2061.

Strange, M., Sánchez-Villagra, M.R., Salzburger, W., and M. Matschiner. 2018. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfish (Ariidae) supports Miocene closure of the Panamanian Isthmus. Sys. Bio. 67(4): 681–699.

Turner, T.F., Wares, J.P., and J.R. Gold. 2002. Genetic Effective Size Is Three Orders of Magnitude Smaller Than Adult Census Size in an Abundant, Estuarine-Dependent Marine Fish (Sciaenops ocellatus). Genetics. 162(3): 1329–1339.

Wakeley, J. 1998. Segregating sites in Wright’s island model. Theoret. Pop. Biol. 53:166–175.

Wakeley, J. 1999. Non-equilibrium migration in human history. Genetics. 153:1863–1871.

Wakeley, J. 2000. The effects of subdivision on the genetic divergence of populations and species. Evolution. 54(4): 1092–1101.
Weir, J.T., and D. Schluter. 2008. Calibrating the avian molecular clock. Mol. Eco. 17: 2321–2328.

Whitlock, M.C. and D.E. McCauley. 1999. Indirect measures of gene flow and migration: FST not equal to 1/(4Nm + 1). Heredity. 82: 117–125.

Wilkins, J.F., and J. Wakeley. 2002. The coalescent in a continuous, finite, linear population. Genetics. 161: 873–888.

Wright, S. 1931. Evolution in Mendelian populations. Genetics. 16(2): 97–159.

Wright, S. 1943. Isolation by distance. Genetics. 28:114–138.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., and T.T.Y. Lam. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Meth. Eco. Evo. 8(1): 28–36.
Figure captions

**Figure 1.** Population model for SLiM simulations. A) Three-taxon species tree: 1) coalescent simulations in *msprime* with $N = 2000$; 2) ancestral stepping-stone conditions begin (see B); 3) $N = 1000$, panmictic; 4) population split, leaving end or center species surviving as sp1 and sp2. B) Ancestral population dynamics. Circles designated “1” and “10” are end species; center species are “5” and “6”.

**Figure 2.** Box plots of the estimated $T_{MRCA}$ by SNAPP; ns = “not significant”, $p < 0.05$ (*), $p < 0.001$ (**), $p < 0.0001$ (***)**, $p < 0.00001$ (****). Dashed lines represent when the estimated age converges on the true age (i.e., at 0). Note that the y-axis is different between the panels. Center species are on the left, end species on the right (see online version for full color).

**Figure 3.** Density plot of scaled ancestral $N_e$ (/1000) based on mean $\pi_{12}$ across genomic windows of 100kb. Dashed line is when $N_e / N_c = 1$.

**Figure 4.** Estimated divergence-times for the Domingos et al. (2017) dataset. A) the “near” trees from the posterior distribution; B) the “far” trees; C) boxplot of the two nodes of focus, where ADE is for clade ((DE)A) and BC is (BC); D) map of included sites from Domingos et al. (2017). Significance is as in Fig. 2. Numbers at nodes represent the median height in units of millions of generations ago (mga). (See online version for full color.)
Figure - 1

(a)...

(b)...

End species

Center species
Figure 3

Migration rate

Density

$N_e/1000$

Dem Sampled: Center, End

$T_b/ND$
