Ecological speciation under adaptation from standing genetic variation

Background

It is increasingly recognized that natural selection is an important driver of speciation (Schluter 2001). Speciation via divergent natural selection—ecological speciation (Schluter 2000)—results when reproductive isolation evolves in association with population divergence (Nosil 2012). Natural selection can also result in speciation when populations adapting to similar environments fix alternative, incompatible, adaptive mutations (Schluter 2009). Although we are beginning to understand the prevalence of speciation by natural selection in nature, we have a poor understanding of the general mechanisms that cause divergent and parallel evolution to generate reproductive isolation.

Theoretical models have made progress investigating the mechanisms linking natural selection with reproductive isolation. A model by Slatkin and Lande (1994) found that segregation variance in recombinant hybrids increases as populations evolve in allopatry. Barton (2001) confirmed that recombinant hybrids are less fit on average than either parent, and that reproductive isolation caused by reduced hybrid fitness increases with divergence time. Extending this model, Chevin et al. (2014) found that the rate at which reproductive isolation evolves does not differ between allopatric populations undergoing either parallel or divergent evolution. Thus, increasing segregation variance appears to be an important mechanism generating reproductive isolation between populations evolving in allopatry.

While these models are valuable, some of their conclusions conflict empirical results. In particular, data from natural populations often finds greater reproductive isolation between populations adapted to different environments than between populations adapted to the same environment (Schluter and Conte 2009). For example, in the threespine stickleback (*Gasterosteus aculeatus*) F1 and F2 hybrids between limnetic and benthic species have reduced performance in either parental environment (Hatfield and Schluter 1999). Meta-analyses also support the positive link between ecological divergence and reproductive isolation (Bolnick et al. 2006; Funk et al. 2006; Shafer and

Wolf 2013). Although there is an obvious conflict between theory and empirical data, the reasons for this conflict are not clear.

To date, models that examine the fitness of hybrids between divergent populations have assumed that all adaptation is from new mutations. However, a large amount of adaptive evolution—especially over short timescales—is not from *de novo* mutation but rather from standing genetic variation (Barrett and Schluter 2008). Analysis of genomic data reveal that many genetic changes that occur during parallel speciation are identical across populations (Soria-Carrasco et al. 2014). The probability of parallel evolution from standing genetic variation increases with the effect size of adaptive alleles (MacPherson and Nuismer 2016). With standing genetic variation, there is thus a high probability that identical alleles will fix under parallel selection—especially relative to divergent selection—which will reduce genomic divergence under parallel evolution. Incorporating standing genetic variation into theoretical models of hybrid fitness may thus improve their ability to predict patterns in nature.

The goal of this project is thus to evaluate whether standing genetic variation influences the evolution of postzygotic reproductive isolation between populations that are evolving in allopatry. Our main research questions are:

1. Does adaptation from standing genetic variation affect the evolution of reproductive isolation between populations evolving in allopatry?
2. Does standing genetic variation differentially effect the evolution of reproductive isolation in populations undergoing divergent vs. parallel evolution?

We predict that the fixation of identical alleles in allopatric populations will reduce segregation variance in hybrids and thus cause higher fitness than when different alleles are fixed. In addition, we expect that standing genetic variation will have this effect only when populations evolve in parallel, because the same alleles are less likely to fix in populations that are adapting to different optima. Testing these predictions will advance our understanding of general mechanisms linking ecological divergence to reproductive isolation.

The Model

We used individual-based simulations to investigate the role of standing genetic variation in speciation. As an underlying framework, we implemented Fisher's geometrical model of adaptation (hereafter 'FGM') (Fisher 1930). FGM is an increasingly popular model of evolutionary genetics (Orr 2005; Tenaillon 2014) that has been used to study speciation (Barton 2001; Chevin et al. 2014; Frïasse et al. 2016). FGM makes predictions that are testable in empirical systems, and has received support from both lab (MacLean et al. 2010) and field (Rogers et al. 2012; Stearns and Fenster 2016) studies.

All simulations were implemented in Python 3.6 and used the 'NumPy' package (Van Der Walt et al. 2011). Simulation code is accessible from the project's GitHub repository (github.com/Ken-A-Thompson/SVS), and is archived on Dryad (link TBA), along with resulting data and R script (R Core Team 2014). Figures were generated in R using the 'ggplot2' (Wickham 2011) package after data were reorganized using 'dyplr' (Wickham and Francois 2015).

Adaptive landscape model

Our models assume a life cycle that is sexual and haploid with non-overlapping generations. Viability selection occurs at the beginning of each generation, and surviving individuals mate randomly to give birth to the next generation. Following birth, each individual acquires a new mutation with probability U. Individuals all have n traits, and we assume universal pleiotropy of mutations such that each mutation affects all traits. Mutations act additively on phenotypic values, and are random numbers drawn from a normal distribution with a mean of 0 and SD = α. Thus, we assume that mutations are isotropic. The model is an extension of models by Barton (2001), Chevin et al. (2014) and Frïasse et al. (2016). While these models follow a population that is effectively monomorphic, a key difference in our model is that we model a population with extensive genetic variation.

Accumulation of standing genetic variation

Before splitting into isolated populations, we allow the ancestral population to accumulate standing genetic variance via *de novo* mutation. As in Frïasse et al. (2016), we allow the ancestral population to adapt to an initial environment before splitting into

two populations. During this 'burn-in' period, the population adapts to an optimum that fluctuates randomly around 0 for each trait independently, with each optimum being drawn from a mean of zero and SD = β. We also assume weaker selection ($\lambda = 0.5$) at this phase than during the adaptation phase ($\lambda = 1$) in order to allow for both large- and small-effect mutations to accumulate. During the burn-in phase, the ancestral population evolves for t_{burn} generations.

We quantify the amount of standing genetic variance in the ancestral population as nucleotide diversity, π, after (Nei and Li 1979) which is given by : π = …. Thus, π increases with the number of unique mutations, the among-individual variance in 'genome' similarity, and with total population size. π has a lower bound of 0 when the ancestral population at time $t_{burn} = t_0$, and an upper bound of ∞. The relationship between the duration of the burn-in phase, t_{burn}, and the amount of standing genetic variation, π, is shown in Fig. X.

Adaptation to a new environment

As in the model of Chevin et al. (2014), we consider an ancestral population, "A", that splits into two isolated populations, "I" and "J", which evolve in allopatry for an equal amount of time, t_{max}. Each population is founded by some number of individuals, $N_{c,i}$, sampled without replacement from the ancestral population at the end of the burn in period—that is when $t_A = t_{burn}$. $N_{c,i}$ has a lower bound of 1 and an upper bound equal to the size of the ancestral population, $N_{c,A}$. When $N_{c,i} = N_{c,A}$, then both derived populations are identical to the ancestral population. Thus, founder effects are increasingly likely as [limit of $N_{c,i} \rightarrow 1$].

We first consider populations that undergo either parallel or divergent evolution. In the parallel case, populations adapt to an identical multivariate optimum, such that $\theta_i = \theta_j$. In the divergent case, populations adapt to orthogonal multivariate optima, such that $\theta_i = -\theta_j$. For simplicity, we assume that the optimum for all traits During the process of adaptation, individuals in both populations acquire new mutations with probability U, as in the ancestral population. For simplicity, we assume that the optimum for all traits is identical, such that $\theta_1 = \theta_2 = \ldots \theta_n$.

Recent extensions of FGM allow adaptation to optima with varying degrees of similarity (Martin and Lenormand 2015), and we briefly consider this case… (TBA!).

Formation of hybrids and quantifying hybrid fitness

After t_{adapt} generations, derived populations 'meet' in secondary contact and produce hybrids. Hybrids are formed by randomly pairing individuals between populations, which mate to form recombinant F1 hybrids. Hybrids inherit either parental allele with an equal probability. If an allele is shared by both parents, which is only possible with ancestral standing variation, then the allele will be present in all hybrids. If an allele fixed from *de novo* mutation in the adapting population, it cannot be shared by the other population. Thus, the probability that an allele fixed from mutation is present in a hybrid is 0.5.

We assigned fitness values to individual hybrids based on their phenotypes. We use hybrid load (*sensu* Chevin et al. [2014]) as our metric of hybrid performance. Hybrid load is simply the reduction in fitness of hybrids relative to the best possible phenotype in the environment where fitness is measured. Mean hybrid fitness has an extrinsic, or environment-dependent component determined by mean distance of individuals from the phenotypic optimum. In addition, the mean fitness of hybrids is determined in part by their segregation variance, which is constant across environments and generally leads to deviations from the optimum phenotype.

Results

Fig. X. Standing genetic variation accumulation w/ burn-in.

Fig. X. Example figure of adaptation from SGV/DNM to parallel, divergent optima.

Fig. X. Parallelism with distance to optimum; for different founding AND max pop sizes.

- Mutational constraint

•

Ackowledgements

- SP Otto
- LH Riseberg

References

Barrett, R. D. H., and D. Schluter. 2008. Adaptation from standing genetic variation.

Barton, N. H. 2001. The role of hybridization in evolution. Mol. Ecol. 10:551–568.

Bolnick, D. I., T. J. Near, and P. C. Wainwright. 2006. Body size divergence promotes post-zygotic reproductive isolation in centrarchids. Evol. Ecol. Res. 8:903–913.

Chevin, L. M., G. Decorzent, and T. Lenormand. 2014. Niche dimensionality and the genetics of ecological speciation. Evolution 68:1244–1256.

Fisher, R. F. 1930. THE GENETICAL THEORY OF NATURAL SELECTION.

Frïasse, C., P. A. Gunnarsson, D. Roze, N. Bierne, and J. J. Welch. 2016. The genetics of speciation: Insights from Fisher's geometric model. Evolution 70:1450–1464.

Funk, D. J., P. Nosil, and W. J. Etges. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. Proc. Natl. Acad. Sci. U. S. A. 103:3209–3213.

Hatfield, T., and D. Schluter. 1999. Ecological speciation in sticklebacks: environment-dependent hybrid fitness. Evolution 53:866–873.

MacLean, R. C., G. G. Perron, and A. Gardner. 2010. Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in *Pseudomonas aeruginosa*. Genetics 186:1345–1354.

MacPherson, A., and S. L. Nuismer. 2016. The probability of parallel genetic evolution from standing genetic variation. J. Evol. Biol., doi: 10.1111/jeb.13006.

Martin, G., and T. Lenormand. 2015. The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima. Evolution 69:1433–1447.

Nei, M., and W.-H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. U. S. A. 76:5269–5273.

Nosil, P. 2012. Ecological speciation. Oxford University Press, Oxford, UK.
Orr, H. A. 2005. The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6:119–127.
R Core Team. 2014. R: A language and environment for statistical computing, Vienna, Austria. http://www.R-project.org/.
Rogers, S. M., P. Tamkee, B. Summers, S. Balabahadra, M. Marks, D. M. Kingsley, and D. Schluter. 2012. Genetic signature of adaptive peak shift in threespine stickleback. Evolution 66:2439–2450.
Schluter, D. 2001. Ecology and the origin of species.
Schluter, D. 2009. Evidence for ecological speciation and its alternative. Science 323:737–741.
Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, New York.
Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. Proc. Natl. Acad. Sci. USA 106 Suppl:9955–62.
Shafer, A. B. A., and J. B. W. Wolf. 2013. Widespread evidence for incipient ecological speciation: A meta-analysis of isolation-by-ecology.
Slatkin, M., and R. Lande. 1994. Segregation variance after hybridization of isolated populations. Genet. Res. 64:51–56.
Soria-Carrasco, V., Z. Gompert, A. a. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, and P. Nosil. 2014. Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation. Science 344:738–742.
Stearns, F. W., and C. B. Fenster. 2016. Fisher's geometric model predicts the effects of random mutations when tested in the wild. Evolution 70:495–501.
Tenaillon, O. 2014. The Utility of Fisher's Geometric Model in Evolutionary Genetics. Annu. Rev. Ecol. Evol. Syst. 45:179–201.
Van Der Walt, S., S. C. Colbert, and G. Varoquaux. 2011. The NumPy array: A structure for efficient numerical computation. Comput. Sci. Eng. 13:22–30.
Wickham, H. 2011. Ggplot2. Wiley Interdiscip. Rev. Comput. Stat. 3:180–185.
Wickham, H., and R. Francois. 2015. dplyr: A Grammar of Data Manipulation. R

Packag. version 0.4.2. 3.