Current Projects Summary

I am working with collaborators in Brazil, the U.S., and Mexico to complete genetic data analyses and manuscripts from two postdoctoral research fellowships.

The following slides briefly describe the two funded projects:

justinbagley.org

1. Ecological & Evolutionary Genomics of Southwestern White Pine (*Pinus strobiformis*)

Southwestern White Pine (SWWP) Genomics

This project constitutes the VCU/ Eckert Lab portion of a broader ~$4 million NSF MacroSystem Biology grant to study the ecological and evolutionary processes influencing the distribution, genetic diversity, adaptive evolution, and persistence of southwestern white pine (*Pinus strobiformis*) in the face of ongoing climate change and an encroaching fungal pathogen, white pine blister rust (*Cronartia ribicola*).

P. strobiformis

Sub-projects:

· Ecological speciation (demography, niche divergence) in SWWP

· Adaptation across a longitudinal climatic gradient

· Adaptation to challenging environments, i.e. high elevations

Southwestern White Pine (SWWP) Genomics

VCU Team

Andrew J. Eckert
PI, Team Leader

Justin C. Bagley
Postdoc

Mitra Menon
PhD Student

Southwestern White Pine (SWWP) Genomics

- Southwestern white pine is an alpine species that occurs across a range of moderate to higher elevations in disjunct population scattered across the North American desert southwest, from the southern Rockies to the Sierra Madre Occidental of northern Mexico.

- Our NAU, UNAM, and USDAFS collaborators sampled SWWP and limber pine (*P. flexilis*; LP) from throughout their ranges, and we have been tasked with genotyping samples using genome-wide ddRAD-seq data and inferring genomic signatures of selection, adaptation, and demographic processes in SWWP and LP across the range of each species sampled (Fig. 1).

Figure 1
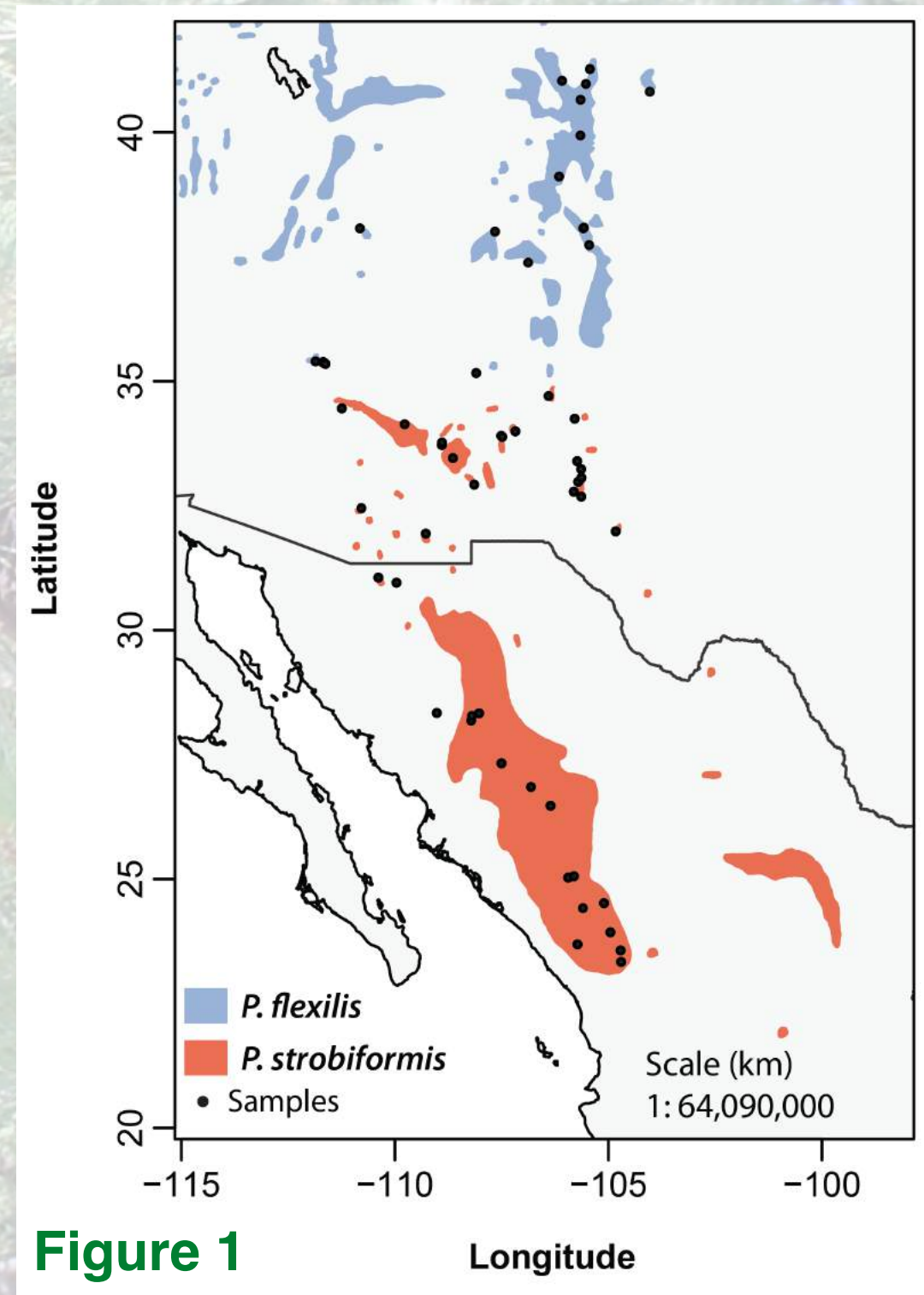
Population genomics supports speciation with gene flow, not genomic islands of differentiation, in sky-island populations of southwestern white pine

Justin C. Bagley,[1,2,*] Mitra Menon,[1,3] Christopher Friedline,[1] Amy Whipple,[4] Anna Schoettle,[5] Alejandro L. Sáenz,[6] Christian A. Wehenkel,[6], Daniel McGarvey,[7] Lluvia H. Flores-Renteria,[8] Richard Sneizko,[5] Sam Cushman,[5] Kristen Waring,[9] and Andrew J. Eckert[1]

1 Department of Biology, Virginia Commonwealth University, Richmond, VA 23284, **2** Departamento de Zoologia, Universidade de Brasília, 70910-900 Brasília, DF, Brazil, **3** Integrative Life Sciences, Virginia Commonwealth University, Richmond, VA 23284, **4** Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 36011, **5** USDA Forest Service, **6** Instituto de Silvicultura e Industria de la Madera, Universidad Juárez del Estado de Durango, 34120 Durango, México, **7** Center for Environmental Studies, Virginia Commonwealth University, Richmond, VA 23284, **8** Department of Biology, San Diego State University, San Diego, CA 92182, **9** School of Forestry, Northern Arizona University, Flagstaff, AZ 36011.
***E-mail correspondence:** jcbagley@vcu.edu.

Introduction

Understanding **speciation**, including processes leading to lineage divergence and the origin and maintenance of reproductive barriers, is a fundamental goal of evolutionary biology (Losos *et al.* 2013). As populations move across a fitness landscape, they form different ecotypes resulting in shifts in allele frequency correlated with environmental differences. Given sufficient time or strong diversifying selection, ecotypes can develop reproductive isolation, forming ecologically differentiated species via **ecological speciation** (Schluter & Conte 2009). Two models explain the maintenance of species boundaries during ecological speciation predict varying demographic scenarios, with different genomic signatures, especially patterns of gene flow (**Table 1**).

Table 1. Two models for the maintenance of species boundaries during ecological speciation.

Model	Model description	Predicted pattern of gene flow
Tension zone (Barton and Hewitt 1985)	Reduction in *hybrid fitness* due to lack of genomic cohesiveness and absence of a different niche available for hybrids	Secondary contact between divergent parental lineages (no ancient migration)
Bounded hybrid superiority (Moore 1977; Gross and Rieseberg 2005)	Restricted gene flow between diverging lineages due to *a*) positive *epistasis*, or *b*) because these loci facilitate *adaptation* to novel environmental conditions	Little to no contemporary gene flow between lineages (with or without ancient migration)

If selection is strong and remains constant, then loci contributing to initial ecological divergence may become associated with mate recognition and form coadapted gene complexes with reduced recombination, thereby generating **'genomic islands of differentiation'** (**GID**; Christe et al. 2017; Lindtke & Buerkle 2015). However, this pattern is only expected under the **tension zone model**, or where adaptation occurs from *de novo* mutations (Lackey & Boughman 2017).

Here, we test the above predictions on the prevalence of gene flow during species formation in two species of North American pine trees, *Pinus strobiformis* and *P. flexilis*, that are broadly distributed across the desert southwest, with a narrow range of sympatry in the southern Rocky Mountains (**Fig. 1**). These taxa exhibit few morphological or reproductive differences (e.g. Benkman *et al.* 1984) and are probably locally adapted to varying climate across their ranges. Moreno-Letelier
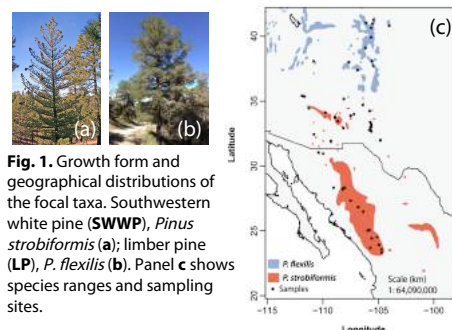

Fig. 1. Growth form and geographical distributions of the focal taxa. Southwestern white pine (**SWWP**), *Pinus strobiformis* (**a**); limber pine (**LP**), *P. flexilis* (**b**). Panel **c** shows species ranges and sampling sites.

et al. (2013) and Moreno-Letelier & Barraclough (2015) provided the first evidence of ecological divergence in these two species based on species distribution models and differentiation at climate-associated candidate genes. We use demographic modeling on genome-wide single nucleotide polymorphism (SNP) data to infer demographic changes, migration rates, and divergence times of these taxa, and to test the two models of ecological speciation discussed above.

Materials and Methods

We sampled *P. strobiformis* across its geographical range, and *P. flexilis* mainly from the southern periphery and center of its range (**Fig. 1**). We extracted whole genomic DNA then prepared five ddRAD-seq libraries (Peterson *et al.* 2012) each containing up to 96 multiplexed samples. Libraries were sequenced on an Illumina HiSeq 2500, and read processing, and SNP filtering and genotyping, were performed in dDOCENT (Puritz *et al.* 2014).

To infer the timing and influence of demographic processes shaping divergence of the focal species plus two intraspecific genetic lineages within *P. strobiformis* (geographical range **'core' and 'periphery' lineages**), we conducted demographic modeling analyses using $\partial a \partial i$ v1.7 (Gutenkunst *et al.* 2009). To avoid issues with linkage disequilibrium, we ran $\partial a \partial i$ on 1 SNP per RAD tag drawn from a reduced subset of 10,053 SNPs (out of 51,633 SNPs total). We compared a 'pure divergence' model (M_1) against 10 alternative demographic models (M_2–M_{11}) representing different speciation scenarios with varying timing and directionality of ancient versus contemporary gene flow (**Fig. 2**). Models M_8–M_{11} were similar
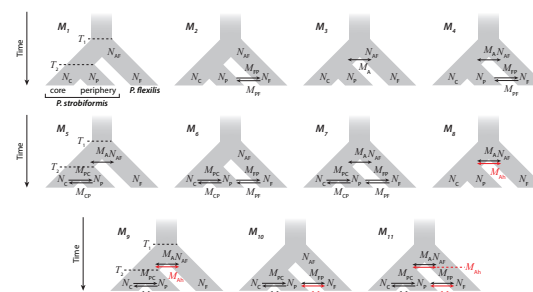

Fig. 2. Schematics and parameter details for each of the 11 demographic models of the divergence of *P. strobiformis* core and periphery lineages and *P. flexilis* run in our $\partial a \partial i$ analysis. Parameters include divergence times (T_j), population sizes (N_j), homogeneous migration rates (M_{ij}) and heterogeneous migration rates (M_{ijh}).

to the others, except they modeled ancient migration or *P. strobiformis* periphery–*P. flexilis* migration as 'heterogeneous migration', with neutrally evolving loci experiencing differential migration rates relative to those in **GIDs** (**Fig. 2**). We ran 10 replicate runs of each model in $\partial a \partial i$, using a 200 × 220 × 240 grid space and the nonlinear BFGS optimization routine. We specified heterogeneous migration parameters using Python code from Tine *et al.* (2014). We conducted model selection using Akaike information criterion (AIC) and ΔAIC$_i$ (AIC$_{model\ i}$ − AIC$_{best\ model}$) scores (Burnham and Anderson 2002) calculated using results from the best replicate (highest composite likelihood) for each model. We converted parameter estimates

from the single best-supported model (minimum AIC) using equations in Gutenkunst *et al.* (2009), a per-site mutation rate (μ) calculated from the 7.28 × 10^{-10} substitutions/site/year rate estimated for Pinacea by De La Torre *et al.* (2014) using 42 single-copy nuclear loci, and a generation time (g) of 50 years.

Results

The best-supported demographic model identified during AIC model selection (i.e. with highest information content) was M_4, a model of symmetric ancient migration between ancestral *P. strobiformis* and *P. flexilis* lineages, followed by contemporary gene flow only between the *P. strobiformis* periphery lineage and *P. flexilis* (**Table 2; Figs 2 and 3**). This model was supported by a very distinct minimum AIC score that was better than that of all other $\partial a \partial i$ models by a margin of at least 44.8 information units (ΔAIC$_i$= 44.8), indicating other models, including all island of differentiation models, were unlikely. Models with ΔAIC$_i$ > 10 have no support and fail to explain any substantial variation in the data (Burnham and Anderson 2002). Converted parameter estimates indicated that the two species diverged ~11.36 million years ago (Ma) in the Miocene, but that intraspecific lineages within *P. strobiformis* diverged at T_2 at ~2.29 Ma in the early Pleistocene (**Fig. 3**). Also, *P. strobiformis* periphery had the largest population size estimate (N_p), while *P. flexilis* was inferred to have experienced a reduction in population size (N_F) through time.

Table 2. Model likelihoods and AIC model selection results for the single best replicate $\partial a \partial i$ run of each model, with the best-supported model (minimum AIC) shown in boldface.

Model	Ln composite likelihood	k	AIC	ΔAIC$_i$
M_1	−883.143112	6	1778.29	65.44
M_2	−886.227416	7	1786.45	73.60
M_3	−888.003307	7	1790.01	77.16
M_4	**−847.424540**	**9**	**1712.85**	**0.00**
M_5	−885.428135	9	1788.86	76.01
M_6	−883.949484	10	1787.90	75.05
M_7	−892.210862	9	1806.42	93.57
M_8	−869.824520	14	1757.65	44.80
M_9	−884.511096	11	1791.02	78.17
M_{10}	−902.279445	9	1828.56	115.71
M_{11}	−922.814525	11	1873.63	160.78

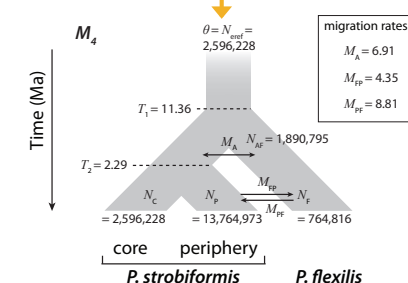**Fig. 3.** The best-supported $\partial a \partial i$ model plotted with optimized values of divergence time estimates (T_j) in units of millions of years ago (Ma), converted reference effective population size (θ; after conversion, N_{ref}), lineage population sizes (N_j), and migration rates (M_{ij}). Abbreviations: C, core; F, *P. flexilis*; P, periphery.

Conclusions

Our results support a pattern of *P. strobiformis*–*P. flexilis* speciation with gene flow, as well as low–moderate ongoing gene flow broadly consistent with predictions of the **bounded hybrid superiority model**. Incorporating genomic islands of differentiation through parameterizing heterogeneous migration also produced *much* worse models with essentially no weight of evidence compared with the best $\partial a \partial i$ model (**Table 2**). Thus, while genomic islands of differentiation are possible in a tension zone experiencing gene flow, they seem unlikely to have formed in this system through differential divergence or introgression of loci. This is consistent with numbers of migrants per generation (M_{ij}) estimated in $\partial a \partial i$, which are not strongly asymmetric between lineages at T_1 or T_2 (**Fig. 3**).

These findings are also consistent with biogeography studies of the desert southwest suggesting that montane 'sky-island' forest ecosystems expanded along lower elevations during glacial periods such as the Last Glacial Maximum (LGM), providing opportunities for gene flow between presently isolated montane lineages (e.g. Knowles 2000; Mastretta-Yanes *et al.* 2015, refs. therein). Boreal forest trees of the Mexican Highlands including our focal taxa may have been more likely to experience continuous gene flow, rather than post-glacial secondary contact, as lineages were repeatedly connected as cold and humid habitats expanded during Pleistocene glacial periods, as indicated by climate models and phylogeographic data (e.g. reviewed in Mastretta-Yanes *et al.* 2015).

Literature Cited

Barton, N. H., and G. M. Hewitt. 1985. *Annu. Rev. Ecol. Syst.* 16:113–148.
Benkman, C. W., R. P. Balda, and C. C. Smith. 1984. *Ecology* 65:632–642.
Burnham, K. P., and D. R. Anderson. 2002. *Model Selection and Multimodal Inference: A Practical Information Theoretic Approach*, 2nd Edn. Springer-Verlag, New York.
Christe, C., K. N. Stölting, M. Paris, C. Fraïsse, N. Bierne, and C. Lexer. 2017. *Mol. Ecol.* 26:59–76.
De La Torre, A. R., T. Wang, B. Jaquish, and S. N. Aitken. 2014. *New Phytol.* 201:687–699.
Gross, B. L., and L. H. Rieseberg. 2005. *J. Hered.* 96:241–252.
Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante. 2009. *PLoS Genetics* 5:e1000695.
Knowles, L. L. 2000. Evolution 54:1337–1348.
Lackey, A. C. R., and J. W. Boughman. 2017. *Evolution* 71: 357–372.
Lindtke, D., and C. A. Buerkle. 2015. *Evolution* 69:1987–2004.
Losos, J. B., S. J. Arnold, G. Bejerano, E. D. Brodie, D. Hibbett, H. E. Hoekstra, *et al.* 2013. *PLoS Biol.* 11.
Mastretta-Yanes, A., A. Moreno-Letelier, D. Piñero, T. H. Jorgensen, and B. C. Emerson. 2015. J. Biogeogr. 42:1586–1600.
Moore, W. S. 1977. *Q. Rev. Biol.* 52:263–277.
Moreno-Letelier, A., and T. G. Barraclough. 2015. *Evol. Ecol.* 29:733–748.
Moreno-Letelier, A., A. Ortiz-Medrano, and D. Piñero. 2013. *PLoS One* 8:e78228.
Puritz, J. B., C. M. Hollenbeck, and J. R. Gold. 2014. *PeerJ* 2:e431.
Schluter, D., and G. L. Conte. 2009. *Proc. Natl. Acad. Sci.* 106:9955–9962.
Tine, M., H. Kuhl, P.-A. Gagnaire, B. Louro, E. Desmarais, R. S. T. Martins, *et al.* 2014. *Nature Comm.* 5:5770.
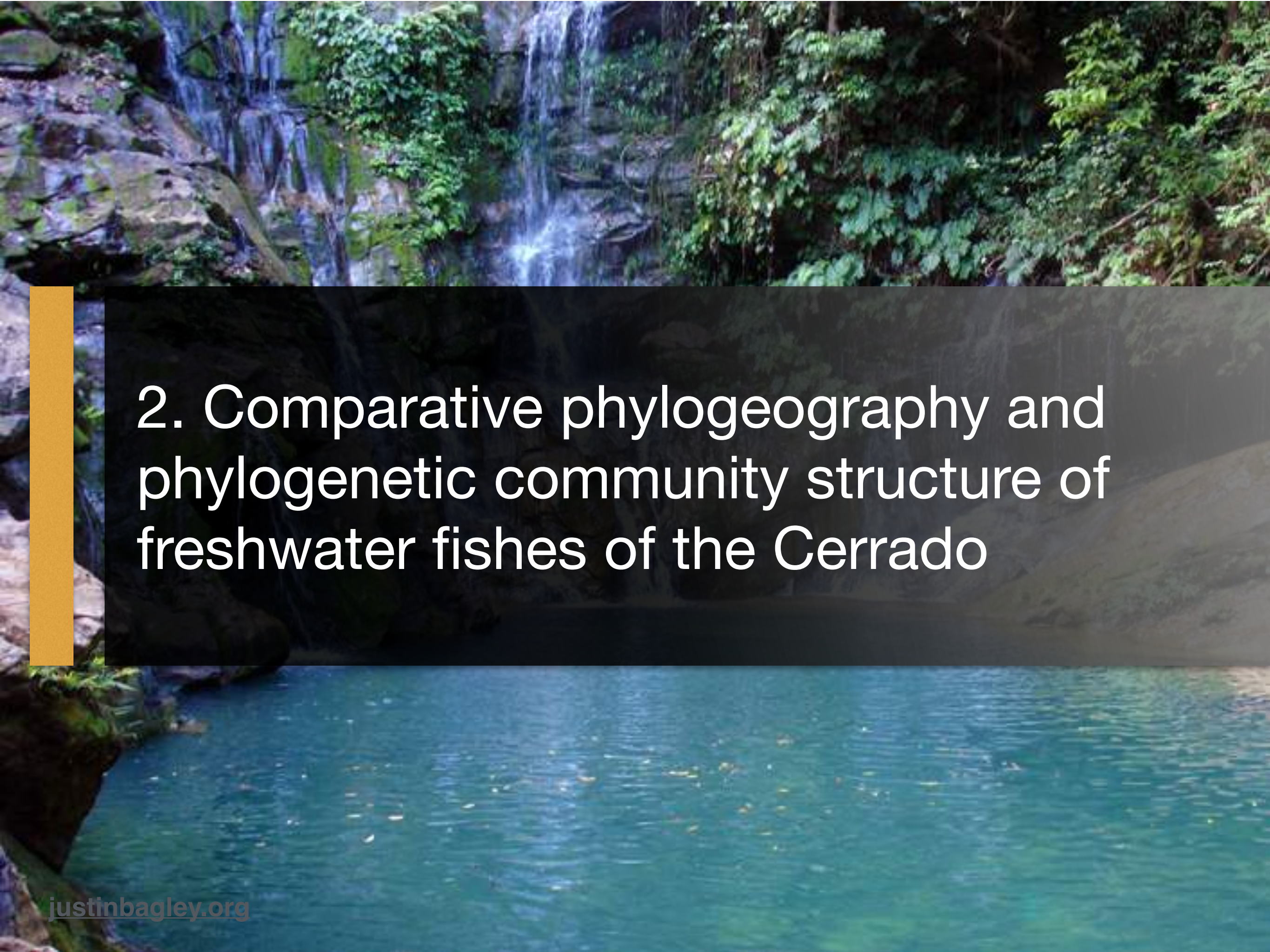
For further information

Please contact jcbagley@vcu.edu, follow @JBagz1 on Twitter, and visit www.justinbagley.org. The QR code at right links to an online, PDF version of this poster.

2. Comparative phylogeography and phylogenetic community structure of freshwater fishes of the Cerrado

Comparative Biogeography of Cerrado Stream Fish Communities

This project is funded by a Young Talent Fellowship grant from the Brazil's CNPq, as well as ancillary funding (e.g. Fundação CAPES) and represents a collaboration between researchers at the University of Brasília (UnB), State University of São Paulo (UNESP), and Federal University of Amazonas (UFAM), led by Justin Bagley.

The overarching aim of the project is to use comparative phylogeography and community phylogenetics to understand the influence of ecological and historical factors (such as river capture) on the genetic diversity and community composition of the headwater stream fish assemblage.

Sub-projects:

· **Single species and comparative ddRAD-seq phylogeography to test genetic predictions of river capture in single fish species**

· **Molecular-based community phylogenetics analyses of regional fish communities**

Study system: Paraná-Tocantins-São Francisco (PTSF) triple divide

Example focal species

Hypostomus sp. 2, Loricariidae

- Small range
- Endemic to PTSF study area
- Ongoing taxonomic description

Photo credit: P. Aquino (center, *top*); I. Pinheiro (*bottom left*).