MS. JOYCE RODRIGUES PRADO (Orcid ID: 0000-0002-2025-5479)

DR ALEXANDRE REIS PERCEQUILLO (Orcid ID: 0000-0002-7892-8912)

Article type: Research Paper

Title: Similar but different: revealing the relative roles of species-traits versus biome properties structuring genetic variation in South American marsh rats

Joyce R. Prado¹, Alexandre R. Percequillo¹,2, Andréa T. Thomaz³, L. Lacey Knowles³

¹Departamento de Ciências Biológicas, Escola Superior de Agricultura ‘Luiz de Queiroz’, Universidade de São Paulo, São Paulo, Brazil

²Department of Life Sciences, The Natural History Museum, London, UK

³Department of Ecology and Evolutionary Biology, The University of Michigan, Ann Arbor MI, USA

Corresponding Author: Joyce R. Prado
Orcid id: orcid.org/0000-0002-2025-5479
email: joyceprra@gmail.com

¹Present address: Departamento de Ciências Biológicas, Universidade Federal do Espírito Santo, Vitória, ES, Brazil.

²Present address: Biodiversity Research Centre and Department of Zoology, University of British Columbia, Vancouver, BC, Canada.

RH: species-traits versus biome properties structuring genetic variation

Word count of main text: 6409

Number of Figures: 6

Number of Tables: 0

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/JBI.13529

This article is protected by copyright. All rights reserved
ACKNOWLEDGEMENTS

This work was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; Proc: 2014/22444-0, 2012/24099-3, 2009/16009-1, 2016/20055-2). We are also grateful to the curators and researchers that generously granted us tissues samples. We also would like to thank Luciana Resende, Anna Papadopoulou, and Melisa Olave for all the help with bioinformatics processing and giving helpful feedback on earlier drafts of the manuscript.

ABSTRACT

Aim: Wetland habitats, and the ecological restrictions imposed by them, structure patterns of genetic variation in constituent taxa. As such, genetic variation may reflect properties of the specific biomes species inhabit, or shared life history traits among species may result in similar genetic structure. We evaluated these hypotheses jointly by quantifying the similarity of genetic structure in three South American marsh rat species (Holochilus), and test how genetic variation in each species relates to biome-specific environmental space and historical stability.

Location: South America.

Taxon: Rodentia

Methods: Using complementary analyses (Mantel tests, dbRDA, Procrustes, covariance structure of allele frequencies, and ENMs) with 8,000–32,000 SNPs per species, we quantified the association between genomic variation and geographic and/or environmental differences.

Results: Significant association between genetic variation and geography was identified for all species. Similarity in the strength of the association suggests connectivity patterns dictated by shared species-traits predominate at the biome scale. However, substantial amounts of genetic variation are not explained by geography. Focusing on this portion of the variance, we demonstrate a significant quantitative association between genetic variation and the environmental space of a biome, and a qualitative association with varying regional stability. Specifically, historically stable areas estimated from ecological niche models are correlated with local levels of geographic structuring, suggesting that local biome-specific histories affect population isolation/connectivity.

Main conclusions: These tests show that although species exhibit similar patterns of genetic variation that are consistent with shared natural histories, irrespective of inhabiting different wetland biomes, local biome-specific properties (i.e., varying environmental conditions and
historical stability) contribute to departures from equilibrium patterns of genetic variation expected by isolation by geographic distance. The reflection of these biome-specific properties in the genetic structure of the marsh rats provide a window into the differences among South American wetlands with evolutionary consequences for their respective constituent assemblages.

KEYWORDS climatic changes, environment, phylogeography, rodent, next generation sequencing, wetland.

Introduction

In the tradition of using concordance across taxa to infer the common effects of abiotic factors in structuring genetic variation, comparative phylogeographic studies usually focus on a specific geographic region with expanded taxonomic sampling. The role of abiotic factors in structuring genetic variation can also be tested by controlling for the potentially confounding influence of biotic factors, for example, by studying a subset of species with similar biological traits, such as similar ecological or life history traits, within a given geographic region (Papadopoulou & Knowles, 2015). However, genetic variation may be shaped by species-specific traits in concert with the abiotic factors (Capurucho et al., 2013; Massatti & Knowles, 2014; Choueri et al., 2017), or regional differences in abiotic factors may mediate the effects of such factors on genetic variation (Papadopoulou & Knowles, 2016). For relatively understudied regions and taxa, accommodating these complexities becomes an important part of the study design because tests focused on a single component structuring genetic variation (i.e., abiotic as opposed to biotic factors) may be giving an especially incomplete picture given the lack of background information for determining appropriate hypotheses.

Here we test for the effects of both abiotic and biotic factors in structuring genetic variation of three species of marsh rats (genus *Holochilus*: Cricetidae) that are restricted to wetland habitats associated with different biomes throughout South America (Fig. 1). Marsh rats are semi-aquatic, large bodied rodents with a herbivorous diet (Hershkovitz, 1955; Gonçalves, Teta, & Bonvicino, 2015) that inhabit open and seasonally flooded grasslands. They share a number of characters associated with specialization for aquatic life and an
herbivorous diet. These include not only morphological traits (fusiform body shape, webbed and elongate hindfeet, loss of molars lophs; Hershkovitz, 1962; Stein, 1988; Domínguez-Bello & Robinson, 1991), but also behavioural traits linked to foraging and how the mammals seek shelter, as well as where they mate (Eisenberg, 1981). As a function of these shared traits, they may also exhibit similar patterns of genetic variation — that is, biotic factors related to being adapted to wetlands may play a role in the structuring of genetic variation across the species (Avise, 2000). However, the wetlands they inhabit differ. For example, two of the focal species (Holochilus sciureus and Holochilus vulpinus) inhabit patchy wetlands across forested biomes, whereas the other taxa (Holochilus chacarius) lives in an open contiguous wetland (Fig. 1). The forested wetlands include wetlands distributed along rivers throughout the Amazon biome, and wetland patches throughout the Atlantic Forest, where H. sciureus and H. vulpinus are distributed, respectively. However, H. vulpinus can also be found in the wetlands of a non-forested biome, the Pampas (Fig. 1). Unlike the forest-associated taxa, H. chacarius is distributed throughout the contiguous wetlands of Chaco/Pantanal biomes (Fig.1). These South American wetlands oscillate between flooded and unflooded phases, although the length, depth, frequency and timing of the phases vary across biomes (Junk, 2013).

In addition, the history of the wetlands differs. For example, the Pantanal is a biome whose formation is linked to precipitation changes since the Last Glacial Maximum (LGM; McGlue et al., 2015). For the forest associated wetlands within the Amazon, while some riverine forest persisted during more arid periods of the Pleistocene (Erwin & Adis, 1982), terra fime vegetation expanded, replacing the flooded forests, and interrupting the distribution and availability of flooded environments (Choueri et al., 2017; Thom et al., 2018). Likewise, expansions and contractions of forested areas in association with climatic change and sea level shifts no doubt impacted the wetlands of the Atlantic Forest, although we lack information specific to the wetland habitats. With a history of climatic (or sea level) driven change, the stability of wetland habitats varied not only across biomes, but also within biomes (Bush, Silman & Listopad, 2007; Quattrocchio, Borromeia, Deschamps, Grill, & Zavala, 2008; Whitney et al., 2011), as well as across biomes, during the Pleistocene. As such, the genetic structure of the marsh rats may reflect biome-specific properties of the different wetland habitats — that is, abiotic factors related to the wetlands themselves that are extrinsic to the shared characteristics of the marsh rat taxa.

We approach the question of the effects of both abiotic and biotic factors in structuring genetic variation of three species of South American marsh rats by quantitatively...
assessing not only the similarity of genetic structure across species, but also the correspondence between patterns of genomic variation and biome-specific aspects of the wetland habitats. Specifically, using 8,000–32,000 SNPs per species, we quantify (i) the similarity of genetic structure in three South American marsh rat species (*Holochilus*), and (ii) test how genetic variation in each species relates to (a) environmental and (b) historical stability differences across the biomes (i.e., biome-specific impacts on genetic structure of the marsh rat species). Such tests involve an evaluation of whether the data conform to levels of gene flow expected under an isolation by distance (IBD) model, where a similar correspondence between geographic and genetic data would suggest a similar migration history among the taxa reflecting the shared traits among the marsh rats (e.g., Peterson & Denno, 1998). In addition, the degree to which the biome-specific environmental properties and historical stability (as characterized by ecological niche models for the past and present) contribute to genetic structure properties is evaluated by multiple procedures for testing for a correspondence between the biome properties and deviations from IBD, including both quantitative and qualitative measures of association (He, Edwards & Knowles, 2013; Knowles, Massatti, He, Olson & Lanier, 2016).

Given the recent origin of the marsh rat taxa (Machado, Leite, Christoff, & Giuliano, 2013), these biome-specific properties, not more ancient Miocene and Pliocene climatic changes (Hoorn et al., 2010), are a logical starting point for testing the potential effects of biome-specific properties on genetic variation in constituent taxa. Moreover, with few information on the physiogamy of regions and potential impact of Pleistocene climatic changes on wetland habitats specifically (see Aleixo, 2006; McGlue et al., 2015; Ledru et al., 2016; Leite et al. 2016; Thom et al., 2018), differences in the genetic structure of constituent taxa across biomes, and any correspondence with regional properties can be mutually informative (e.g., Ferreira et al., 2018; Thom et al., 2018). As one of the first (and to our knowledge) only studies that brings the resolution of genomic data to these vast, but relatively understudied, biomes, our findings provide insights about not only the history of the taxa, but also the features of the different wetlands with evolutionary consequences for their respective assemblages.

**Material and methods**

**DNA extraction, amplification, and sequencing**
Genomic data was generated for 26 individuals and 9 populations of *H. chacarius*, 24 individuals and 6 populations of *H. sciureus*, and 18 individuals and 6 populations of *H. vulpinus*, sampled across their entire range (see Fig. 2, Table S1.1, Appendix S1 in Supporting Information). Tissues were collected or requested from museums (see Appendix S2 in Supporting Information). DNA was extracted from liver, muscle or skin using the Qiagen DNeasy Blood and Tissue Kit.

Four reduced representation libraries were sequenced using ddRADseq method (Peterson Weber, Kay, Fisher, & Hoekstra, 2012; Appendix S1). Briefly, DNA was double digested with the restriction enzymes EcoR1 and MseI and 150 base pair, single-end reads sequenced on four lanes of Illumina HiSeq2000, which produced 270 million raw reads at the Centre for Applied Genomics, Canada.

**Processing of Illumina Data and generating summary statistics**

Raw sequence reads were processed separately for each species in Stacks v.1.35 (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). The reads were demultiplexed and filtered using *process radtags*. Only reads with Phred score >10, unambiguous barcodes, and individuals with more than 500,000 reads were retained (additional details are given in Appendix S1). A *de novo* assembly of filtered reads with a minimum coverage depth of 6 were used to identify putative loci with the USTACKS. A catalogue of consensus loci was constructed for each species with CSTACKS with a distance between individuals for a given locus ≤ 2. Alleles were called for loci using SSTACKS.

Characterizations of genetic variation and population summary statistics were generated with the POPULATIONS program (Catchen et al., 2013). At this step, all loci present in at least two populations were identified and exported in Variant Call Format (vcf).

This first dataset (hereafter referred as the full dataset) was used to calculate population genetic diversity statistics, such as nucleotide diversity (*π*), major allele frequency, observed heterozygosity (*H*O), and Wright’s inbreeding coefficient (*F*IS) at each locus. Population-level summaries of genetic diversity were also characterized (average *π*, *H*O and *F*IS), and a one-way ANOVA was used to test for significant differences in genetic diversity between species (conducted in R, Car package; R Core Team 2014). A second dataset with one randomly chosen SNP and a maximum of 20% missing data was generated with the toolset PLINK v.1.07 (Purcell et al., 2007) and used for all other analyses (hereafter referred to as the putative unlinked-SNP dataset).
The full dataset consists of 359,728 SNPs and 26 individuals of *H. chacarius*, 357,050 SNPs and 24 individuals of *H. sciureus* and 160,879 SNPs and 18 individuals of *H. vulpinus*. The unlinked-SNPs dataset consists of 32,210 SNPs in 25 individuals of *H. chacarius*, 17,513 SNPs in 20 individuals of *H. sciureus* and 8,035 SNPs in 15 individuals of *H. vulpinus* (Table S1.2 in Appendix S1).

### Structuring of genetic variation

A series of complementary approaches were used to test for similar structuring of genetic variation across taxa and for a correspondence between properties of the wetland habitats and genetic variation. These tests were selected because they differ in their respective assumptions, and hence potential to capture different aspects of population genetic structure.

**Procrustes analyses.** Tests of the similarity among taxa in the structuring of genetic variation, including the fit to, but also the departures from, expectations of isolation by distance were conducted using Procrustes analyses (Wang, Zöllner, & Rosenberg, 2012). Specifically, an association between genetic variation and geography while retaining latitudinal and longitudinal information for sampled individuals was quantified using the `procrustes` and `protest` functions in VEGAN package (Oksanen et al., 2017). The strength of the association was compared across taxa by the similarity statistic $t_0$, which ranges from 0 to 1, and provides a basis for investigating the history of migration.

The statistical technique, involves a transformation of a genetic map onto a geographic map to maximize their similarity (i.e., minimize the sum of squared Euclidean distances between the two maps), where a principle component analysis (PCA) of the genetic data was used to generate individual-level coordinates of the first two components (PC1 and PC2). The position of individuals in the genetic map can be visualized, highlighting whether individuals depart from expected levels of gene flow (i.e., they are genetically too similar, or less similar to other individuals sampled across the landscape than expected given their geographic position). To explore how aspects of the migration history (i.e., departures from expected patterns of gene flow under isolation by distance) reflect the wetlands, the correspondence between the genetic map and the (i) environmental space and (ii) stability of a biome were examined. To test for an association between biome-specific environmental space and genetic variation, the residuals of the Procrustes analysis between genes and geography for each species were paired with a matrix of extracted PCA data of.
environmental variables (described below) for the sequenced individuals (e.g., Knowles et al., 2016). For examining how the stability/instability of habitat within biomes might impact the history of migration, we compared the general correspondence between the position of individuals in genetic space and areas of projected stability; stability is inferred by comparing ecological niche models for different periods of the past with the present, and are described below. In particular, deviations from expected patterns of gene flow (i.e., departures from IBD) was examined in relation to the mapped areas of historical habitat instability to assess whether there was any qualitative correspondence of genetic variation with this biome-specific property (see also Knowles et al., 2016).

The significance of the similarity statistic $t_0$ was evaluated for each species based on 10,000 permutations, where geographical locations were randomly permuted across the different sample localities. In each species, the sensitivity of the similarity statistic $t_0$ to particular populations was also evaluated by repeating the analyses excluding one population at a time, with replacement (which is represented by the similarity statistic $t''$ following Wang et al., 2012). We also computed a similarity score ($t'$; following Wang et al., 2012) between PCA coordinates for the complete data set and those for data sets in which one population was excluded to assess whether any populations had a disproportionate effect on the relationship between genes and geography (Wang et al., 2012; Knowles et al., 2016).

Covariance structure of allele frequencies. Spatial patterns of genomic variation were also examined based on an allele frequency covariance matrix using the program SpaceMix (Bradburd, Ralph, & Coop, 2016); comparison with the results from the PCA provides a measure of whether the PCAs, and hence Procrustes analyses, may be biased by sampling effects (see Novembre & Stephens, 2008). Following the developer’s recommendations, 10 “fast” independent chains were run for $5 \times 10^6$ MCMC iterations, without conditioning populations on their locations and with no admixture. This was followed by a “long” run of $10^8$ iterations, with parameters sampled every $10^7$ iterations, in which population locations were initiated at the origin (i.e., inferred from the “fast” runs), and all other parameters ($\alpha_0$, $\alpha_1$, $\alpha_2$, $\eta$, and $w$) were drawn randomly from their priors at the start of each chain.

Mantel tests and dbRDA. As a complementary test to the analyses described above, the correlation between pairwise $F_{ST}$-values and Euclidean geographic distances among populations, as well as associations between genetic distance and environmental resistance, was examined using a Mantel, partial Mantel tests, and dbRDA (Legendre & Anderson,
For the Mantel tests, a sequential population drop out procedure, in which the test was repeated excluding one population at time, was also conducted to confirm that the results were robust. For the partial Mantel test an environmental resistance matrix among populations was generated based on raster maps obtained with the ENMs using CIRCUITSCAPE v4.0 (Shah & McRae, 2008). The capscale R function was used in the dbRDA analysis to test for the relationship between pairwise genetic distances and corresponding climatic variables (represented by the PC1 of the 19 climate layers used in the ENMs), and removing the effect of geographic distance separating populations (He et al., 2013). In addition, a second partial Mantel test and dbRDA analyses were performed with the environmental variables extracted from the PCA data for the locations of sequenced populations (He et al., 2013) to examine whether the environment might make a significant contribution to patterns of genetic variation, after controlling for the effects of geography.

Characterizing historical stability and environmental space across distributional areas

To characterize historical stability and compare the environmental space of the distributional areas of each Holochilus taxa, environmental niche models (ENMs) were generated from the 19 bioclimatic variables for the present (Bioclim), Holocene (6 kya; MPI-ESM-P; Bioclim) and the Last Glacial Maximum (LGM; 21 kya; MPI-ESM-P; Bioclim; Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) using Maxent 3.3.3k (Phillips, Anderson & Schapire, 2006). Georeferenced occurrence data representative of the ranges of each species was used; vetted data were obtained by direct examination of specimens (Appendix S2) or from taxon-specific bibliographic sources (Hershkovitz, 1955; Pardiñas & Teta, 2011; Pardiñas, Teta, Voglino, & Fernández, 2013; D’Elía, Hanson, Mauldin, Teta, & Pardiñas, 2015).

A Principal Component Analysis (PCA) was used to identify a subset of environmental variables with < 0.7 of correlation based on analyses from the prcomp function in R; among correlated variables, the variable with the highest score was retained. Occurrence data were rarefied using SDMToolBox at a resolution of 10 km to reduce spatial autocorrelation. To avoid overfitting, and considering the semi-aquatic habit of Holochilus, the geographical extent of the environmental layers used as a mask corresponded to a 100km buffer surrounding the respective river basins where the species are distributed. To achieve a balance between goodness-of-fit and model complexity, we used the ENMevaluate function from ENMeval package (Muscarella et al., 2014), and we tested models over combinations of
regularization parameters from 0.5 to 3 in intervals of 0.5 and combinations of features
parameters (Auto, Linear, Quadratic, Hinge, Linear + Quadratic and Linear + Quadratic +
Hinge, according to Maxent recommendations). Regularization and features parameters were
chosen using Akaike Information Criterion (AIC; Warren & Seifert, 2011) and Area Under
the receiver–operator Curve (AUC; Swets, 1988). Each model parameter class was replicated
10 times for cross-validation. For each model we extracted binary predictions, where suitable
habitat presence was inferred according to the significance threshold for each model from
Maxent.

Climatically stable areas were inferred from the intersection of the binary predictions
under current and past climate scenarios (see Table S3.1, Fig. S3.1, Appendix S3 in
Supporting Information for additional details). Variation in current environmental conditions
across each biome was quantified with a PCA of the 19 Bioclim variables using the
rasterPCA function from the DISMO package in R (Hijmans, Phillips, Leathwick, & Elith,
2016), with input variables rescaled from 0 to 1 (so that the PCs are not sensitive to
differences in the units). For a graphical presentation of this variation across the landscape,
the ggRGB function from RStoolbox package in R (Leutner & Horning, 2016) was used to
construct maps, with the red layer corresponding to PC1, green layer corresponding to PC2
and the blue layer corresponding to PC3. To visualize how the position of sampling localities
spans the environmental space, the environmental dispersion of sequenced individuals
relative to the total environmental space for each species was mapped using the
environmental data (PC1 and PC2 of the PCA performed with the bioclimatic variables) and
the occurrence points of sequenced individuals (see Lanier, Massatti, He, Olson, & Knowles,
2015).

Results
Population genetic summary statistics
Values of population genetic summary statistics were broadly overlapping among
taxa (Table S4.1, Fig. S4.1, Appendix S4 in Supporting Information), except for a statistically
significant difference in $H_{obs}$, with higher values in H. vulpinus compared with the other
species. Genetic diversity, as measured by $\pi$, was similar among taxa, varying from 0.1 to
0.2. Inbreeding coefficients, $F_{is}$, were also similar across taxa with the exception of two
populations; specifically, S_MAM in H. sciureus and YAC in H. chacarius had substantially
higher inbreeding coefficients, although neither exhibited reduced genetic diversity.
**Historical stability and environmental space across distributional areas**

The current environmental characteristics of the biomes inhabited by the three species clearly differ (Fig. 3). The environmental space of the Atlantic Forest and Pampas area inhabited by *H. vulpinus* differs the most from the other regions. There is some resemblance between the environmental space occupied by *H. vulpinus* and the north-western region of *H. sciureus*. However, these two areas are separated by a large region with different environmental characteristics (see the two bluish areas are separated by a large green area; Fig. 3). Comparison of the environmental conditions where populations were sampled to the rest of biome shows that the sites where genetic samples were collected are generally representative of the environmental range encompassed by the species’ distributions, spanning the entire dispersion of environmental values of the PCA (Fig. 4).

Estimates of habitat stability based on similarity in spatial distribution of suitable habitat from the LGM, the Holocene and the present, show differences in the stability of the habitat across biomes, as well as within each biome. Specifically, the habitat across most of the distributional range of *H. vulpinus* is projected to be stable, whereas only the southwestern Amazon is predicted to be stable for *H. sciureus*, and only small and patchy areas were inferred to be stable for *H. chacarius* (Fig. 2).

**Structuring of genetic variation**

Regarding the geographic distribution of genetic variation within each species, several patterns are reinforced across methods. For example, when the position of individuals along PC1 and PC2 from a genetic PCA is projected onto the geographic distribution of sampled individuals (Fig. 5), it highlights how similarity in the multidimensional genetic space is related to where an individual occurs geographically, confirming the Procrustes results (Figs. 2 and 6). Likewise, this similarity is also evident when comparing the results from the spatial covariance of alleles analyses (SpaceMix results) with those from a PCA of genetic variation (Fig. S4.2 in Appendix 4). This general correspondence suggests that the PCAs (and therefore the Procrustes analyses) are not subject to biases that can be introduced from the sampling distribution of individuals (Novembre & Stephens, 2008). However, the position of some individuals of *H. chacarius* in multivariate genetic space differs somewhat from the spatial covariance of alleles. Specifically, we note that the GOL population occupies a more distinctive position relative to other populations in the geogenetic map of Spacemix compared with its overlap with other populations in the PCA (Fig. S4.2 in Appendix 4). However, results from the sequential population drop-out procedure in the Procrustes
This article is protected by copyright. All rights reserved
Genetic deviations/displacements in geographic and environmental space

In *H. vulpinus*, when we consider the deviations of individuals in genetic space from expectations based on their geographic location (i.e., the length of the lines connecting a sampled population to a sequenced individual; Fig. 2), the deviations primarily vary along a latitudinal axis. Moreover, the individuals tend to occupy the central area of the species distribution and the position of these individuals in the PC genetic space show a strong correspondence with areas of stability (Fig. 2), with the exception of one population in the southeast (BAR; Fig. 2) whose individuals deviate in a southerly direction.

In contrast, the directions of departures in genetic space relative to where individuals of *H. sciureus* were sampled geographically tend to follow a longitudinal axis (Fig. 2). Nevertheless, this species also shows some correspondence between the position of individuals in genetic space and areas of projected stability like *H. vulpinus*. Likewise, the deviations of individuals sampled in what is projected to be stable areas historically (N_MAM and S_MAM) tend to show relatively small departures from geographic expectations.

Lastly, in *H. chacarius* the displacement of many individuals to the central part of the species range was observed even though the sampled populations correspond to areas of projected stability (Fig. 2). Only some of the northern populations (PNP and COR, but not POC) are somewhat distinct from this general cluster.

Tests of a correlation between the residuals from the Procrustes analysis of genes and geography with the environmental distance among populations suggests that environmental differences contribute to some of the genetic differences observed among individuals in each species (Table S4.3 in Appendix 4). As with the tests of isolation by distance that did not retain the relative position of sampled populations in geographic space, neither the partial Mantel nor dbRDA detected significant contribution of environmental differences to genetic distances (Tables S4.4 and S4.5 in Appendix 4).

Discussion

Similarity in the structuring of population genetic variation (Fig. 2) and population genetic diversities (see Table S4.1, Fig. S4.1 in Appendix 4) highlight the contribution of common natural history traits to the similar migration histories of the taxa. However, biome-specific effects on genetic structure explain a substantial amount of genetic variation not explained solely by gene flow associated with shared traits among the taxa. Specifically,
quantitative and qualitative analyses (Figs. 2 and 3) show that variation of the environment and historical stability within biomes contribute to patterns of genetic structure.

Below we discuss the insights these findings offer about the South American wetlands and their constituent taxa as one of the first, and to our knowledge, only study that brings the resolution of data from next-generation sequence technologies to these vast, but relatively understudied, landscapes. In addition, we highlight the niche our work fills in comparative phylogeography: the study of ecologically similar taxa from different regions, as opposed to a tradition of comparing taxa (often with differing ecologies) from a single region, to make inferences about the processes underlying observed patterns of genetic variation (Avise, 2000; Knowles, 2009; Hickerson et al., 2010).

Insights from genomic analyses about South American biomes

Despite differences in the distribution of wetland habitats across the biomes inhabited by the taxa (e.g., patches of wetlands versus large contiguous wetlands), the migration history of the species is similar. Specifically, the strength of the association between genetic variation and geography was similar across the biomes (when the outlier population GUI in H. sciureus was excluded). Moreover, even when considering all populations, the Procrustes analyses indicated the genetic isolation between geographically distant populations in the species inhabiting the two biomes of forest–based patches of wetlands was not more similar to each other relative to the species in the contiguous wetlands. Instead, the strength of the association between genes and geography was most similar in H. chacarius, which inhabits the large contiguous wetlands, and H. vulpinus, which inhabits patches of wetlands in the Atlantic Forest and in the open vegetation of the Pampas. Together, these results emphasize that gene flow decreases with geographic distance similarly across taxa, irrespective of differences across biomes. This similar migration history suggests that shared characters of the marsh rats, that is biotic factors, play a key role in structuring genetic variation when genetic variation is measured at the scale of the species entire distribution. It may be that the effect of the current configuration of habitats (contiguous versus patchy) may be discernible at different spatial scales, such as at a local landscape level. Such scale-dependencies of biotic traits in migration history is something we would like to investigate in the future; however, we cannot address this question with the available sampling.

Notwithstanding the similar migration histories suggested by the similar strength of the genes and geography association across taxa, the association between biome-specific properties with deviations from expectations under isolation by distance point to potential
hypotheses about how the biomes may impact genetic structure. In particular, the magnitude of deviations from a pattern of isolation by distance differs across the landscape in each taxon, suggesting localized differences in population connectivity and/or persistence. The most dramatic effect is across Amazonia, where the more northern population of *H. sciureus* tend to be genetically quite similar despite fairly large distances separating the sampled individuals (Fig. 2). This geographic area that individuals map based on their genetic makeup is consistent with an estimated area of habitat stability based on ENMs for the past and present (Fig. 2), which suggests a potential role of climatic shifts, and coincidentally is estimated to be less stable relative to the southern sampled populations (Mayle, Burbridge, & Killeen, 2000; Urrego, Silman, & Bush, 2005; Bush et al., 2007; Arruda et al., 2017). In contrast, in southern regions of the distribution of *H. sciureus*, there is a strong correspondence between the genetic and geographic position of individuals (Fig. 2), and their general genetic distinctiveness relative to the more northern populations suggests a history of relative regional isolation within the biome (e.g., Knowles et al. 2016). Differences in fit of northern and southern regions to expected patterns of gene flow under an isolation by distance model, in concert with corresponding differences in duration of stability inferred from the ENMs (Fig. 2), suggests variation in the historical stability within a biome may be contributing to genetic structure of constituent taxa. Specifically, our genetic analyses suggest a dynamic of recent expansion may characterize the wetlands from the northern part of the Amazonian biome, in contrast to relative population persistence within the southern region. This pattern agrees with some, but not all, aspects of inferences about past regional stability from other studies. For example, the area corresponding to our LOR, JUR, and MAD populations (Fig. 2) corresponds to an area that is hypothesized to have been associated with a dynamic geological history (Aleixo & Rossetti, 2007; Leite & Rogers, 2013), and stable past hydroclimate variation (see Cheng et al., 2013). Although the sampling in the east is limited, the genetic clustering of the GUI population with other western Amazonian sampled populations suggests however that the wetlands in the eastern area (or at least the north-east; Fig 2) is also relatively unstable, unlike the results from some landscape genetic and geologic studies (Aleixo & Rossetti, 2007; Leite & Rogers, 2013).

Although less dramatic, patterns of genetic structure in *H. chacarius* show a somewhat similar pattern with populations from the northern part of their range exhibiting the greatest deviations under isolation by distance (individuals from the north tend to be genetically more similar to individuals from the southern than expected), with most of the estimated regions of stability located in the south (Fig. 2). This complex history suggests the
expansion of wetland across this biome may be fairly recent (see Bezerra & Mozeto, 2008), with more sparse vegetation and intermittent torrential flows along the alluvial fans in the past (Whitney et al., 2011; Assine et al., 2015). In contrast, while H. vulpinus shows the larger continuous areas of stability among all species (see Arruda et al., 2017 for proposed expansion scenarios), it nonetheless shows a correspondence between genes and geography that is similar to the other two species (i.e. the $t_0$ of H. vulpinus is 0.69, while the $t_0$ of H. sciureus and H. chacarius is 0.52 and 0.64, respectively). Such similarity again emphasizes that differences in stability by itself, or continuity of wetlands (as opposed to being patchy), is not sufficient to override patterns of genetic variation within each species that are consistent with migration history where gene flow decreases with distance (Fig. 2).

In addition to the potential effect of historical stability, multiple lines of evidence point to other characteristics of the biomes that may affect patterns of genetic variation, albeit at a local level (as opposed to the regional scale of an entire biome, given similar degrees in the strength in the association between genes and geography). Environmental differences across the landscape show a significant effect on the position of individuals in genetic space, after controlling for the effect of geography. However, it is noteworthy that this effect, as with test of a general correspondence between genes and geography, is only detectable when retaining the relative distance of populations latitudinally and longitudinally (i.e., with Procrustes analyses). The lack of significance when reducing the geographic separation of populations to a one–dimensional axis (Mantel and dbRDA analyses; Table S4.4 and S4.5 in Appendix 4) highlights how connectivity patterns do not mirror expectations based on random diffusion across a landscape (Excoffier, Foll, & Petit, 2009). That is, a longitudinal distance versus a latitudinal distance is not equivalent in terms of the impact on population differences. As such, our work highlights that within each of the biomes, connectivity patterns vary locally in each species.

Patterns of environmental heterogeneity across the landscape (Fig. 3) and/or shifts in habitat stability over time (Fig. 2) can cause populations separated by similar geographic distances to differ in levels of connectivity (McRae, 2006; He et al., 2013). Although our tests are insightful in that they identify patterns suggestive of environmental factors influencing connectivity, as correlative analyses they do not provide tests of the process itself (Knowles et al., 2016). Determining which of the dynamics might produce genetic variation consistent with what is observed (i.e., applying a model selection framework to distinguish between the geographic distribution of wetland habitats versus shifting climates; see He et al., 2013; Knowles & Massatti, 2017), is beyond the scope of this study without additional
sampling for conducting spatially explicit model–based analyses. However, as with the
guiding theme of this work, tests to identify concordant processes across species will be
especially exciting to understand the extent to which genetic variation in the taxa might arise
from common processes, despite differences in the biomes themselves.

Similar taxa, but different biomes, in comparative phylogeography.

Although the study of multiple taxa across different regions predominates in historical
biogeography and macroecology (Leite et al., 2014; Arregoitia, Fisher, & Schweizer, 2017),
comparative phylogeographic analyses have focused primarily on co–distributed taxa from a
single area (Avise, 1992; Knowles, 2009; Hickerson et al., 2010). There are several potential
explanations for this tradition and one is related to sampling efforts for phylogeographic
inference: population sampling poses logistical constraints such that a more circumscribed
geographic region is simply more feasible. The focus on more circumscribed region also
reflects the traditional motivation behind the comparative phylogeography – making
inferences about the history of a region. It is through tests of concordance across taxa from a
particular region that specific biogeographic barriers (Avise, 1992) or areas of long-term
stability (Carnaval, Hickerson, Haddad, Rodrigues, & Moritz, 2009) might be inferred. In
fact, such tests of concordance across taxa with differing ecologies have been used to identify
how abiotic factors might supersede any ecological differences among taxa (Naka, Bechtoldt,
Henriques, & Brumfield, 2012) or reject the role of historical barriers in driving divergence
patterns (Smith et al., 2014).

As a complement to the study of co–distributed taxa, there is a precedent for studying
taxa from different regions in comparative phylogeography. For example, analysis of
evolutionarily independent regions can be considered as natural replicates and can be used for
making generalizations about geologic events that have broadly affected landscape histories
(Bermingham & Moritz, 1998), while controlling for discord that may arise from differences
in species–specific traits to mitigate poor predictive power when concordance across regions
is not supported (see Papadopoulou & Knowles, 2016). Here we adopted a strategy of similar
species, but different regions as an inferential framework. Moreover, with a complementary
set of tests and the genetic resolution provided by thousands of loci, we address how genetic
patterns vary (i.e., we move beyond concordant versus discordant binary). As such, we were
able to identify concordance in genetic variation (similar strength in the association between
genes and geography, despite differences in the wetlands from the different biomes) that
suggests a highly specialized semi–aquatic life may result in similar genetic isolation by

This article is protected by copyright. All rights reserved
distance. However, inspection of local deviations from isolation by distance identified additional biome-specific effects related to environmental differences and historical stability. That is, rather than biotic versus abiotic factors predominating on the genetic structure, our framework identifies how both components can influence genetic variation, as well as their respective geographic scales of influence. Regarding this latter point, the biotic factors predominate at the regional scale (at the level of genetic structure across the entire biome), whereas abiotic factors appear to contribute to local departures from isolation by distance (see Fig. 2). These scale-specific effects are especially interesting considering the qualitative correspondence between regions of instability and departures from isolation by distance, which may be suggestive of particular populations that might have been vulnerable to environmental change (Rocha et al. 2014; Choueri et al. 2017; Harvey, Aleixo, Ribas, & Brumfield, 2017).

With very few genetic studies on organisms that inhabit the flood-dominated landscapes in South America, the present work adds to our knowledge, but there is admittedly much that remains unknown. For example, the effect on terrestrial organisms (as opposed to aquatic or semi-aquatic species) may not show the same relative role of biotic and abiotic factors or geographic scale at which they might predominate, given that flooding and rivers may act as barriers in terrestrial organisms rather than as routes for connectivity (Lima, Lima-Ribeiro, Timoco, Terribile, & Collevatti, 2014; Rocha et al. 2014). This suggests the importance of a nuanced view of interpreting concordance (or the lack thereof) across taxa. Specifically, taxa may show opposing patterns of connectivity and yet this discord could arise from deterministic processes (i.e., not the idiosyncrasies of history; see Massatti & Knowles, 2014) because what constitutes routes of connectivity in a given wetland biome depends upon species-specific traits (e.g., for terrestrial versus aquatic taxa).

**FIGURE LEGENDS**

Figure 1. Examples of the open biomes of the central, southern and northern region of South America inhabited by *Holochilus* (see Fig. 2 for distributional map), such as (a) Poconé and (b) Miranda, in Pantanal biome, both habitats occupied by *H. chacarius*; (c) Caçapava do Sul in Pampa biome, within the range of *H. sciureus*, and (d) São José do Norte in Atlantic Forest biome, within the range of *H. vulpinus*; (e) lower Xingú River, and (f) Japurá River both in Amazon biome, inhabited by *H. sciureus*; (g) *H. chacarius*.
Figure 2. Map of the distributions with sampled populations marked for each of the three marsh rats species, with each species color coded (Holochilus chacarius, Holochilus sciureus, and Holochilus vulpinus are shown in black, green, and orange, respectively); sampling locations span each species’ range, which are largely non-overlapping with each taxa occupying different biomes. On the right plots of Procrustes–transformed PCA’s of genomic variation with each individual mapped in the genomic PC–space (marked by circles; note the genetic position of some individuals are largely overlapping in some cases) relative to the geographical location of sampled populations (marked by triangles) showing the deviation in the genomic PC–space from the expected pattern of genetic variation based on geography (i.e., the length of the line connecting individuals to their geographical location represents the magnitude of the deviation), as well as how the association between genes and geography differed among species, with $t_0 = 0.69$, 0.64, and 0.52 for H. vulpinus, H. chacarius, and H. sciureus, respectively. Additionally, the plots for each taxon are shown on the projected stability of each region, where stability is defined as areas that remained suitable overtime (see Fig S3.1 in Appendix 3 for distributional maps from ENMs for each geologic period). Stable areas since LGM, 21 kya, are marked in orange, yellow marks areas that have been stable since Holocene, 6 kya, relative to the unstable areas marked in green (i.e., projected suitable areas for the present, but not the past).

Figure 3. Map of the environmental variation across the region where the three species are distributed (orange dots represent H. vulpinus, black dots H. chacarius, and green dots H. sciureus), where differences in color depict geographic regions that differ the most from each other. Specifically, PC1, PC2 and PC3 of bioclimatic variables across the landscape were rescaled between 0 and 1, and the RGB color composite was calculated and plotted in the map with PC1 set as the red scale, PC2 as the green scale, and PC3 as the blue scale. Colored dotes correspond to the populations presented in Fig. 2. Note that the border used to characterize the relative difference in environment corresponded to the same area used to generate the ENMs in each species (see Fig. 2) and does not include the Andes.

Figure 4. Dispersion in environmental space of the sampled populations used in our genetic analyses (marked as colored dots) relative to the PC values for Holochilus sampling locations used in the ENMs. For H. sciureus PC1 is strongly positively correlated with Minimum Temperature of Coldest Month (Bio6) and explains most of the variation among populations (i.e., 55.49%), whereas PC2 explains relatively little variation among populations (19.66%).
For *H. vulpinus* PC1 is strongly positively correlated with the Annual Precipitation (Bio12) and explains most of the variation among populations (48.56%), whereas PC2 explains 24.26% of the variation among populations. For *H. chacarius* PC1 is strongly positively correlated with the Minimum Temperature of Coldest Month (Bio6) and explains most of the variation among populations (54.4%), whereas PC2 explains 21.8% of variation among populations.

Figure 5. Distribution of individuals along PC1 and PC2 of genomic variation color coded, and the percentage of variance explained by each PC are shown on a map (i.e., different colors correspond to individuals with the greatest genomic difference along PCs) for each species. Elevation differences are shown in grey scale on each map as well.

Figure 6. Comparison of the changes in the strength of the association between genes and geography with the exclusion of individual populations (i.e. $t''$) relative to when all populations are analyzed (i.e. $t_0$). Values for each species are standardized by $t_0$ (i.e. 0 on y-axis corresponds to $t_0$) such that positive values indicate a stronger association between genes and geography when a population is excluded, whereas negative values indicate a weaker association. Bar colors represent sampling populations following the colored names of populations in Fig. 2.
Fig 3
Fig 4
Fig 5

a) *H. sciureus*

b) *H. vulpinus*

c) *H. chacarius*


This article is protected by copyright. All rights reserved


unravel the demographic history of a Neotropical swamp palm through the Quaternary.


https://doi.org/10.1111/evo.12491


https://10.1126/science.290.5500.2291


https://doi.org/10.1111/j.0014-3820.2006.tb00500.x


https://doi.org/10.1111/2041-210X.12261


https://doi.org/10.1038/ng.139


This article is protected by copyright. All rights reserved


**BIOSKETCHES**

The authors share an interest in the study of processes that structure the genetic variation among taxa and across geography. LLK, JRP and ARP conceived the idea; JRP collected the...
genomic data; JRP and ATT processed the genomic data; JRP, LLK and ATT performed the analyses; all authors worked on the writing.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix 1 Summaries of geographical information and genomic sampling
Appendix 2 Summaries of samples collected and scientific collections
Appendix 3 Summaries of ENM settings and projections of current and LGM distributions
Appendix 4 PC maps of genetic variation and summaries of genetic variation
(a) *H. sciureus*

(b) *H. vulpinus*

(c) *H. chacarius*