Local adaptation in mainland anole lizards: Integrating population history and genome–environment associations

Ivan Prates1,2 | Anna Penna3 | Miguel Trefaut Rodrigues4 | Ana Carolina Carnaval2

1Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, District of Columbia
2Department of Biology, City College of New York and Graduate Center, City University of New York, New York, New York
3Department of Anthropology, University of Texas at San Antonio, San Antonio, Texas
4Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

Correspondence
Ivan Prates, Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC.
Email: ivanprates@gmail.com

Funding information
Division of Environmental Biology, Grant/Award Number: DEB 1120487, DEB 1343578 and DEB 1601271; Fundação de Amparo à Pesquisa do Estado de São Paulo. Grant/Award Number: 03/10335-8, 11/50146-6 and 2013/50297-0

Abstract
Environmental gradients constrain physiological performance and thus species’ ranges, suggesting that species occurrence in diverse environments may be associated with local adaptation. Genome–environment association analyses (GEAA) have become central for studies of local adaptation, yet they are sensitive to the spatial orientation of historical range expansions relative to landscape gradients. To test whether potentially adaptive genotypes occur in varied climates in wide-ranged species, we implemented GEAA on the basis of genomewide data from the anole lizards Anolis ortonii and Anolis punctatus, which expanded from Amazonia, presently dominated by warm and wet settings, into the cooler and less rainy Atlantic Forest. To examine whether local adaptation has been constrained by population structure and history, we estimated effective population sizes, divergence times, and gene flow under a coalescent framework. In both species, divergence between Amazonian and Atlantic Forest populations dates back to the mid-Pleistocene, with subsequent gene flow. We recovered eleven candidate genes involved with metabolism, immunity, development, and cell signaling in A. punctatus and found no loci whose frequency is associated with environmental gradients in A. ortonii. Distinct signatures of adaptation between these species are not associated with historical constraints or distinct climatic space occupancies. Similar patterns of spatial structure between selected and neutral SNPs along the climatic gradient, as supported by patterns of genetic clustering in A. punctatus, may have led to conservative GEAA performance. This study illustrates how tests of local adaptation can benefit from knowledge about species histories to support hypothesis formulation, sampling design, and landscape gradient characterization.

Key words
Amazonia, Anolis, Atlantic Forest, gene flow, phylogeography, population genomics

1 | INTRODUCTION

Environmental variation across geographic space limits species’ ranges, therefore affecting ecological processes and large-scale biogeographic patterns (Ghalambor, Huey, Martin, Tewksbury, & Wang, 2006). Because physiological tolerances to abiotic conditions constrain organismal fitness (Navas, 2002), it has been suggested that the occurrence of any given species across various environments may be associated with local adaptation (Bridle & Vines, 2007; Hohenlohe et al., 2010; Huey, Gilchrist, Carlson,
Souza, Lima-Ribeiro, & Solferini, 2015). Moreover, pronounced
encompass these regions span a climatic gradient from warm and
Atlantic Forest (Carnaval et al., 2014). Species whose ranges now
experience largely distinct climates (Ledo & Colli, 2017; Sobral-
estis. Several species colonized the Atlantic Forest from Amazonia,
while documenting how population structure and history
constrain differentiation, can shed light on the role of local adap-
tation in the establishment of species in diverse habitats.

To examine local adaptation in heterogeneous landscapes, anal-
yses of genome–environment associations (GEAA) have become
a central tool (Rellstab, Gugerli, Eckert, Hancock, & Holderegger,
2015). GEAA studies test for correlations between allele frequen-
cies at multiple loci and environmental predictors across a species’
range (Frichot, Schoville, Bouchard, & François, 2013) and are able
to account for neutral patterns of population genetic structure that
can mimic signatures of selection (Rellstab et al., 2015). However,
these analyses are sensitive to the spatial orientation of historical
range expansions relative to that of the environmental gradients: A
simulation study found that GEAA studies perform best (have lower
false discovery rates) when the axes of historical expansion and eco-
logical variation are aligned in space (Frichot, Schoville, Villemereuil,
Gaggiotti, & François, 2015). This finding has important implications
for empirical studies. First, it suggests that knowledge about the his-
tory of occupation of the landscape is crucial to analyses of adap-
tive genomic differentiation on the basis of GEAA. Second, it implies
that studies of local adaptation will benefit from targeting species
whose biogeographic trajectories have led to a spatial and ecological
arrangement that maximizes GEAA performance.

A system that fits these GEAA requirements well is that of organ-
isms that have undergone range expansions in South American
forests. Several species colonized the Atlantic Forest from Amazonia,
subsequently expanding southward toward subtropical regions
(Batalha-Filho, Fjeldså, Fabre, & Miyaki, 2013; Costa, 2003; Gehara
et al., 2014; Prates, Rivera, Rodrigues, & Carnaval, 2016; Prates, Xue,
et al., 2016). It has been shown that Amazonia and the Atlantic Forest
experience largely distinct climates (Ledo & Colli, 2017; Sobral-
Souza, Lima-Ribeiro, & Solferini, 2015). Moreover, pronounced
climatic differences distinguish the northern from the southern
Atlantic Forest (Carnaval et al., 2014). Species whose ranges now
encompass these regions span a climatic gradient from warm and
wet conditions in Amazonia to cooler and less rainy settings in the
Atlantic Forest. Therefore, these organisms represent an appropri-
ate system to examine signatures of local adaptation associated with
species occurrence along environmental gradients on the basis of
GEAA. However, the role of genotypes that confer adaptation in
shaping the distribution of tropical species remains underexplored
(Damasceno, Strangas, Carnaval, Rodrigues, & Moritz, 2014; Moritz,
Patton, Schneider, & Smith, 2000; Teixeira, 2016).

This investigation asks whether genetic differentiation identified
as potentially adaptive is associated with climatic differentiation
along the ranges of the lizards Anolis ortonii and Anolis punctatus.
While these two species belong to anole clades that diverged around
50 mya (Prates et al., 2017; Prates, Rodrigues, Melo-Sampaio, &
Carnaval, 2015), it has been shown that they underwent similar his-
tories of colonization of the Atlantic Forest from Amazonia (Prates,
Rivera, et al., 2016; Prates, Xue, et al., 2016). These lizards are well
suited for studies of adaptation. Natural history data indicate that
South American forest anoles exhibit low body temperatures when
active, and perform little to no behavioral thermoregulation (Vitt,
Avila-Pires, Espósito, Sartorius, & Zani, 2003; Vitt, Avila-Pires,
Zani, Sartorius, & Espósito, 2003; Vitt, Cristina, Avila-Pires, Zani,
& Espósito, 2002; Vitt, Sartorius, Avila-Pires, & Espósito, 2001)—
which points to a role of genetically determined physiological traits
in shaping ecological tolerances (e.g., Navas, 2002).

To explore GEAA in this system, we first infer population struc-
ture and history within each anole species, estimating levels of
population gene flow—a force that could oppose local adaptation.
Then, we test whether species establishment in distinct climates is
associated with potentially adaptive genetic differentiation between
populations. To characterize the climatic gradients presently occu-
pied by A. ortonii and A. punctatus, we also estimate climatic space
occupancy over their ranges.

2 | MATERIAL AND METHODS

2.1 | Molecular sampling

We sampled 46 individuals of Anolis punctatus and 23 of Anolis or-
tonii (specimen and locality information in Supporting Information
Table S1; Supporting Information deposited in the Dryad Digital
Repository: https://doi.org/10.5061/dryad.1bj51s9), encompassing
a substantial portion of their distributions (Ribeiro-Júnior, 2015).
Reduced representation sequence data were generated as described by Prates, Xue, et al. (2016). Briefly, genomic DNA
was extracted from liver fragments preserved in 100% ethanol
through a high-salt extraction protocol following proteinase
and RNAase treatment. After examining DNA fragment size using
agarose gels, DNA concentration was measured in a Qubit 2 fluo-
rometer (Invitrogen, Waltham) and diluted to ensure a final con-
centration of 30–100 ng DNA/µl in a total volume of 30 µl (in TE
buffer). A restriction site-associated DNA library was generated
through a genotype-by-sequencing protocol (GBS; Elshire et al.,
2011) at the Institute of Biotechnology at Cornell University.
Genomic DNA was digested with the EcoT22I restriction enzyme,
and the resulting fragments were tagged with individual barcodes,
PCR-amplified, multiplexed, and sequenced on a single lane on an
Illumina HiSeq 2500 platform (rapid run mode). The number of sin-
gle-end reads per individual ranged from around 500,000 to six
million, with a read length of 100 base pairs. De-multiplexed raw
sequence data were deposited in the Sequence Read Archive (accession PRJNA492310).

We used iPyrad v. 0.7.23 (Eaton, 2014; available at https://ipyrad.readthedocs.io) to de-multiplex and assign reads to individuals based on sequence barcodes (allowing no mismatches from individual barcodes), perform de novo read assembly (clustering similarity threshold > 0.9), align reads into loci, and call single-nucleotide polymorphisms (SNPs). To filter out poor-quality reads and reduce base-calling error, we implemented a minimum Phred quality score (≥33), minimum sequence coverage (≥10×), minimum read length (≥70 bp), maximum proportion of heterozygous sites per locus (≤0.25), and maximum number of heterozygous individuals per locus (≤5) while ensuring that variable sites had no more than two alleles. Following the de-multiplexing step in iPyrad, read quality and length were assessed for each sample using FastQC (available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). For phylogenetic and genetic clustering analyses, a single SNP was randomly extracted per locus to reduce linkage disequilibrium across sites and to maximize sampling of independent SNP histories. Previous to GEAA and genetic clustering analyses, we used VCFtools v. 0.1.13 (Danecek et al., 2011) to filter out SNPs whose minor allele frequency (MAF) was lower than 0.1 to prevent the inclusion of SNPs that may correspond to sequencing errors (Ahrens et al., 2018).

2.2 | Inferring population genetic structure

To approximate the number of populations in A. ortonii and A. punctatus, we estimated the number of genetic clusters (k) within each species separately using sNMF v. 1.2 (Sparse Nonnegative Matrix Factorization; Frichot, Mathieu, Trouillon, Bouchard, & François, 2014). We tested k = 1–10, with 100 replicates for each k. The run with the lowest entropy value, estimated by masking 5% of the samples, was considered to identify the best k (Frichot et al., 2014). To examine the robustness of sNMF to the regularization parameter (α), we ran preliminary analyses with α = 1, 10, 100, and 1,000; the best entropy scores were obtained with α = 100. Up to 10% missing data were allowed (i.e., each DNA site was present in at least 90% of sampled individuals). The resulting datasets were composed of 2,580 unlinked SNPs for A. ortonii and 1,807 for A. punctatus.

To further characterize population genetic clustering, we performed a principal component analysis (PCA) based on the SNP data using the pca function of the R package LEA v. 2 (Landscape and Ecological Association; Frichot & François, 2015).

2.3 | Inferring population trees

To obtain a population tree for the estimation of gene flow under a historical demographic framework (see below), we inferred phylogenetic relationships between the genetic clusters (thereafter populations) identified by sNMF. Phylogenetic analyses were performed for A. punctatus only, for which three populations were recovered; in the case of A. ortonii, where only two populations were found based on the genetic data (see Section), they were treated as sister to each other in downstream analyses.

To inform historical demographic parameter estimation (see below), a population tree was inferred for A. punctatus using SNAPP (SNP and AFLP Package for Phylogenetic Analysis; Bryant, Bouckaert, Felsenstein, Rosenberg, & RoyChoudhury, 2012) as implemented in BEAST v. 2.4.8 (Bayesian Evolutionary Analysis by Sampling Trees; Bouckaert et al., 2014). A gamma-distributed prior (given by parameters shape, a, and scale, ρ) was applied for the θ parameter (= 4Neμ, where Ne corresponds to the effective population size and μ to the per-generation mutation rate), using a = 1 and ρ = 0.05 (with mean = 0.05). A gamma prior with a = 2 and ρ = 100 (with mean = 200) was set to the rate of the Yule diversification process (λ) given an expected tree height (τ = Tμ, where T corresponds to the total tree time in number of generations) of ~0.004 substitutions per site based on recent phylogeographic study that included A. punctatus (Prates, Rivera, et al., 2016). To help define λ prior ranges, we used the python script Pyule (available at https://github.com/joaks1/pyule). Default settings were implemented for the mutation parameters. Tree rooting followed Prates, Xue, et al. (2016). To help define prior distributions implemented in SNAPP and G-PhoCS (see below), we used the qgamma, dgamma, and plot functions in R v. 3.3.3 (R Core Team, 2018).

We performed two independent SNAPP runs of five million generations each, sampling every 1,000 steps. After ensuring convergence between runs and stationarity of model parameters in Tracer v. 1.6 (Drummond, Suchard, Xie, & Rambaut, 2012) and discarding 10% of generations as burn-in, we visualized posterior tree densities in DensiTree v. 2.4.8 (Bouckaert et al., 2014). Due to computational constraints, population trees were inferred based on five individuals from each of the populations inferred by sNMF, while allowing no missing data. Selected individuals were sampled in sites with no evidence of admixture as inferred by sNMF. The resulting dataset was composed of 2,927 unlinked SNPs across 15 samples of A. punctatus.

2.4 | Estimating historical demographic parameters

To estimate historical demographic parameters such as gene flow, we used G-PhoCS v. 1.3 (Generalized Phylogenetic Coalescent Sampler; Gronau, Hubisz, Gulko, Danko, & Siepel, 2011), which uses the entire nucleotide sequences of loci (as opposed to SNPs only). Populations of A. ortonii and A. punctatus were defined as the populations inferred through sNMF; historical relationships between populations were defined based on SNAPP results for A. punctatus, while the two populations of A. ortonii were treated as sister to one another. To assess historical population gene flow, we included bidirectional migration bands for every population pair (i.e., one migration band in each direction).

In G-PhoCS analyses, we applied gamma distributions (given by parameters shape, a, and rate, ρ) for the θ and τ priors, setting a = 1 and ρ = 20 (with mean = 0.05). The gamma prior for each migration band, m (=M/μ, where M corresponds to the proportion of individuals
in one population that arrived from another population in each generation, and \( \mu \) is the per-generation mutation rate), was set with \( \alpha = 1 \) and \( \beta = 2 \times 10^{-7} \) (with mean \( = 5,000,000 \), which corresponds to \( 1/1,000 \) of the individuals in one population having arrived from another population in each generation). To convert mutation rate-scaled parameter estimates to absolute effective population sizes and divergence times, we assumed an average substitution rate of \( 2.4 \times 10^{-9} \) substitutions per site per year, as estimated by a study that included South American anole species (Prates, Rivera, et al., 2016). To convert estimated coalescent times from the number of generations to years, we assumed a generation time of one year in anoles (Muñoz et al., 2013; Tollis, Ausubel, Ghimire, & Boissinot, 2012).

We performed G-PhoCS analyses for each species separately implementing a Markov Chain Monte Carlo of 400,000 generations and sampling every 100 generations. The stationarity of each run and the posterior distribution of model parameters were assessed as described for SNAPPP analyses. Due to computational constraints, G-PhoCS analyses used 2,000 randomly selected loci while allowing for no more than 10% missing data.

2.5 | Choice of environmental variables

As environmental predictors in GEAA, we used bioclimatic variables from the WorldClim database (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) that describe spatial patterns of temperature and precipitation variation. We limited our analysis to environmental variables that are easily interpretable and that showed clear geographic variation across the distribution of our study anoles: annual mean temperature, maximum temperature of the warmest month, mean temperature of the warmest quarter, mean temperature of the coldest quarter, annual precipitation, precipitation of the wettest month, and precipitation of the wettest quarter. Values were extracted from the collection site of each lizard sample using QGIS v. 2.18.15 (available at https://github.com/agis/QGIS). Because temperature and precipitation variables were often correlated (pairwise Pearson correlation coefficients >0.7), we followed best practices for GEAA and implemented a PCA on all variables, using the first principal component as a predictor in GEAA (Frichot et al., 2013). PCA was applied using the pcomp function in R.

To visualize and compare environmental space occupancy across the range of A. ortonii and A. punctatus, we used plots of the climatic PC1 against latitude based on sites sampled for genetic data. To better characterize the climatic spaces occupied by these two species, we also compiled a broader dataset that includes georeferenced museum specimens, with a total of 89 and 144 unique records for A. ortonii and A. punctatus, respectively (Supporting Information Table S2).

2.6 | Performing genome-environment association analyses

To test for associations between environmental predictors and allele frequencies, we used LFMM (Latent Factor Mixed Linear Models; Frichot et al., 2013). LFMM jointly models the effects of environmental variables and neutral genetic structure on allele frequencies, incorporating population structure in the form of unobserved (latent) factors. LFMM has low rates of false positives and is robust to the effects of isolation-by-distance and a variety of sampling designs and underlying demographic histories (Rellstab et al., 2015). To define the number of latent factors used in the LFMM analyses, we explored values around the best number of genetic clusters (k) inferred by sNMF, as recommended by Frichot et al. (2013). Specifically, we performed preliminary runs incorporating one to six latent factors. Two and five latent factors provided better control of the effects of neutral genetic structure in A. ortonii and A. punctatus, respectively (see below), and were used in final analyses.

To ensure that false discoveries (i.e., spurious associations between allele frequencies and environmental predictors) were adequately controlled for, we corrected association p-values based on empirical genomic inflation factors (\( \lambda \)). These factors are used to modify the baseline null hypothesis in GEAA with the goal of limiting inflation related to population structure and other confounding factors (François, Martins, Caye, & Schoville, 2016). To select \( \lambda \), we inspected histograms of corrected p-values (Supporting Information Figure S1) to ensure that they exhibited a uniform distribution over the interval [0,1] (as expected under the null hypothesis of selective neutrality for most loci) with a peak close to 0 (which corresponds to potentially selected loci; see François et al., 2016). Specifically, we implemented \( \lambda = 1.5 \) and 0.45 for A. ortonii and A. punctatus, respectively. A list of candidate SNPs was generated using the Benjamini-Hochberg algorithm (Benjamini & Hochberg, 1995) by assuming a maximum false discovery rate of \( 10^{-5} \).

LFMM was implemented using the lfmm function of the LEA R package, with 10 independent runs of 50,000 iterations each and 25,000 iterations as burn-in. A maximum of 10% missing data was allowed. To reduce the potential effects of missing sites in GEAA, we performed haplotype imputation based on the ancestry coefficients estimated by SNMF, using the impute function of the LEA package (as recommended by LFMM developers). Because LFMM makes no assumptions about linkage among SNPs (Dr. Olivier François, personal communication), all SNPs were included in GEAA analyses. After the filtering steps, the genetic dataset was composed of 3,412 SNPs in 2,580 loci for A. ortonii, and 3,249 SNPs in 1,920 loci for A. punctatus.

R and Unix shell scripts used in all analyses are available online through GitHub (https://github.com/ivanprates/2018_Anolis_ EcolEvol).

2.7 | Candidate gene identity and function

LFMM analyses recovered SNPs whose frequencies are linked to environmental gradients (see Section 3.4). To examine whether the candidate loci harboring these SNPs correspond to genes, we used the BLASTN algorithm (McGinnis & Madden, 2004) as implemented in Ensembl v. 91 (available at https://useast.ensembl.org/Multi/Tools/Blast) to map each sequence against the reference
genome of *Anolis carolinensis* (assembly AnoCar2.0; Alföldi et al., 2011). For blasting, we used a maximum e-value of $10^{-5}$ while masking low complexity (e.g., repetitive) regions. In the case of mapped loci that correspond to protein-coding regions, we then examined gene identity and function as summarized by gene ontology terms based on the UniProt database (available at https://www.uniprot.org/uniprot/).

3 | RESULTS

3.1 | Population genetic structure

Clustering analyses using SNMF inferred higher levels of population structure in *Anolis punctatus* than in *Anolis ortonii*, with three and two populations recovered in each species, respectively (Figure 1a,b). However, the two species show similarities in the patterns of population structure across space. For instance, Atlantic Forest samples belong to a distinct population relative to Amazonian samples in both species, which suggests genetic differentiation across forest blocks. On the other hand, only one population of *A. ortonii* was recovered in Amazonia, whereas two populations of *A. punctatus* were inferred in that region. In *A. punctatus*, one population is widely distributed in central and eastern Amazonia, also including samples from western Amazonia near the Brazil-Peru border, whereas another one extends from southwestern Amazonia (around the Madeira River) northwards (Figure 1b).

Genetic clustering analyses recovered admixture (as indicated by individual ancestry coefficients) between Atlantic Forest and Amazonian populations in both *A. ortonii* and *A. punctatus* (Figure 1a,b), with higher levels of admixture in the latter. In the case of *A. punctatus*, admixed individuals were sampled in the northern extreme of the Atlantic Forest, close to Amazonia’s eastern edge, and in western Amazonia (Figure 1b).

![Figure 1](image)

**Figure 1** Genetic clustering based on all SNPs from *Anolis ortonii* (a) as well as all SNPs (b) and candidate SNPs only (c) from *Anolis punctatus*. Proportions in pie charts on maps correspond to ancestry coefficients estimated by genetic clustering analyses. Gray areas on map indicate South American rainforests. Red arrows indicate *A. punctatus* sample MTR 20798 from Pacaraima on the Brazil-Venezuela border in the Guiana Shield region, a locality that is climatically similar to Atlantic Forest sites (see Figure 3); this sample is genetically more similar to eastern Amazonian samples based in the entire SNP dataset, yet more similar to Atlantic Forest samples based on the candidate SNPs only.

3.2 | Population history

A scenario of early separation between Amazonian and Atlantic Forest populations in *A. ortonii* differs from the pattern seen in *A. punctatus*. In *A. punctatus*, phylogenetic analyses using SNAPP found that the Atlantic Forest population is nested among Amazonian samples (Figure 2). This Atlantic population is the sister of the population that occurs in eastern Amazonia; the clade formed by them is sister to that population that extends from southwestern Amazonia (around the Madeira River) northwards.

G-PhoCS analyses inferred large effective population sizes in all regions for both *A. ortonii* and *A. punctatus*, estimated as hundreds of thousands to millions of individuals (Figure 2). Coalescent
times between the Atlantic Forest population and the most closely related Amazonian lineage date back to the mid-Pleistocene in both anole species. We found a signal of historical gene flow between populations that occur in Amazonia and the Atlantic Forest in both species, with levels of population gene flow being overall higher in *A. punctatus* than in *A. ortonii* (Figure 2).

### 3.3 | Climatic space occupancy

The first component of the PCA based on temperature and precipitation variables (PC1), used as a predictor variable in GEAA, explains >70% of the total climatic variation across sampled sites in both species. Positive scores in climate PC1 correspond to lower temperature and precipitation values (Supporting Information Table S3). Plots of climatic space occupancy along a latitudinal gradient based on PC1 suggest that both species experience similar climates over their range in Amazonia (Figure 3). However, while *A. punctatus* occupies cooler and less rainy environments in the southern end of its distribution in the Atlantic Forest, *A. ortonii* is restricted to warmer and more humid sites in the northern Atlantic Forest, which is climatically more similar to Amazonia. This pattern holds when considering both sites sampled for genetic data and a broader dataset that also includes hundreds of occurrence records from museum specimens (Figure 3).

### 3.4 | Genome–environment association analyses

GEAA controlling for neutral genetic structure using LFMM found allele frequencies of 86 SNPs in 39 loci to be significantly associated with climatic gradients in *A. punctatus*. Among these loci, 30 blasted against regions of the annotated reference genome of *A. carolinensis*. Eleven loci uniquely mapped to known protein-coding genes (Table 1); two candidate loci mapped non-specifically.

---

**FIGURE 2** Population history (from SNAPP) and historical demographic parameters (from G-PhoCS) inferred for *Anolis ortonii* (a) and *A. punctatus* (b). Parameters are the time of coalescence between populations (in millions of years, Mya), effective population sizes (in millions of individuals, M), and migration rates (in migrants per generation, m/g). Colors of terminals correspond to genetic clusters in Figure 1.

**FIGURE 3** Environmental space occupancy along latitude based on climatic PC1 for *Anolis ortonii* and *A. punctatus*. Samples used in genetic analyses are indicated with a black dot. Higher PC scores correspond to drier and colder sites. Dashed line indicates the approximate region of pronounced north–south climatic turnover in the Atlantic Forest as identified by Carnaval et al. (2014). Red arrow indicates *A. punctatus* sample MTR 20798 from Pacaraima, a mid-elevation site (820 m above sea level) in the Guiana Shield region that overlaps climatically with Atlantic Forest sites (horizontal axis).
## Table 1
Candidate loci identified by genome–environment association analyses using LFMM for *Anolis punctatus*

<table>
<thead>
<tr>
<th>GBS locus</th>
<th><em>Anolis carolinensis</em> chromosome (or scaffold)</th>
<th>Position (bp)</th>
<th>Overlapping gene</th>
<th>Match length (bp)</th>
<th>E-value</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>48380</td>
<td>LGa</td>
<td>3,044,340–3,044,411</td>
<td>CXADR-like membrane protein gene (CLMP)</td>
<td>72</td>
<td>1E−20</td>
<td>93.06</td>
</tr>
<tr>
<td>52588</td>
<td>GL343225</td>
<td>1,347,775–1,347,853</td>
<td>PHD finger protein 2 (PHF2)</td>
<td>79</td>
<td>1E−22</td>
<td>92.41</td>
</tr>
<tr>
<td>57548</td>
<td>4</td>
<td>91,951,599–91,951,684</td>
<td>Dihydropyrimidine dehydrogenase gene (DPYD)</td>
<td>86</td>
<td>1E−19</td>
<td>89.53</td>
</tr>
<tr>
<td>58069</td>
<td>5</td>
<td>140,581,178–140,581,220</td>
<td>N-deacetylase and N-sulfotransferase 3 gene (NDST3)</td>
<td>43</td>
<td>9E−06</td>
<td>90.70</td>
</tr>
<tr>
<td>65078</td>
<td>1</td>
<td>111,909,958–111,910,047</td>
<td>Teneurin transmembrane protein 2 (TENM2)</td>
<td>90</td>
<td>1E−38</td>
<td>97.78</td>
</tr>
<tr>
<td>68654</td>
<td>3</td>
<td>58,901,943–58,902,027</td>
<td>Inositol polyphosphate–5-phosphatase A (INPP5A)</td>
<td>85</td>
<td>2E−16</td>
<td>88.24</td>
</tr>
<tr>
<td>72522</td>
<td>GL343384</td>
<td>248,338–248,397</td>
<td>Lymphoid enhancer binding factor 1 (LEF1)</td>
<td>55</td>
<td>7E−06</td>
<td>87.27</td>
</tr>
<tr>
<td>74703</td>
<td>GL343290</td>
<td>1,535,974–1,721,555</td>
<td>Collagen type V alpha 2 chain (COL5A2)</td>
<td>80</td>
<td>1E−13</td>
<td>87.50</td>
</tr>
<tr>
<td>74803</td>
<td>1</td>
<td>197,640,485–197,640,538</td>
<td>Malic enzyme 1 gene (ME1)</td>
<td>54</td>
<td>9E−15</td>
<td>94.44</td>
</tr>
<tr>
<td>75555</td>
<td>2</td>
<td>120,299,259–120,299,348</td>
<td>Deducator of cytokinesis 2 gene (DOCK2)</td>
<td>90</td>
<td>1E−19</td>
<td>88.89</td>
</tr>
<tr>
<td>77416</td>
<td>1</td>
<td>101,850,853–101,850,977</td>
<td>Rho GTPase activating protein 15 (ARHGAP15)</td>
<td>114</td>
<td>2E−50</td>
<td>97.37</td>
</tr>
</tbody>
</table>

*Note.* Only genes that successfully blasted against known protein-coding regions of the genome of *A. carolinensis* are shown.
to more than four genes; and the remaining loci blasted against non-coding regions, which may correspond to regions that regulate gene expression (e.g., enhancers or transcription factor recognition sites) or that are physically linked to genes that underwent selection.

In the case of A. ortonii, GEAA found no evidence of genomic differentiation associated with environmental gradients. In this species, no SNPs were linked with temperature and precipitation variation across space, even when relaxing the false discovery rate from 10^{-5} to 10^{-3}.

Inspection of read quality using FastQC indicated similar high-quality reads for individuals of both species (not shown), suggesting that differences in DNA quality do not account for distinct GEAA results between A. punctatus and A. ortonii.

### 3.5 Patterns of candidate allele sharing among populations

In A. punctatus, genetic clustering analyses based solely on the candidate SNPs suggest no allele sharing between populations in distinct regions (Figure 1c), partially mirroring sNMF results based on the entire SNP dataset (Figure 1b). Interestingly, one sample from the Guiana Shield region (MTR 20798 from Pacaraima, state of Roraima in Brazil) clusters with Atlantic Forest samples in a clustering analysis based on the candidate SNPs only (Figure 1c). By contrast, this sample clusters with eastern Amazonian samples in an analysis based on the entire SNP dataset (Figure 1b). A genetic PCA based on these two SNP datasets shows the same pattern (Supporting Information Figure S2).

### 4 DISCUSSION

Based on genomewide data of A. ortonii and A. punctatus, two widespread South American anole lizard species, we infer distinct genetic populations in Amazonia and the Atlantic Forest, suggesting that separation of these forests following a period of contact has favored genetic divergence (Figure 1). In both species, historical demographic analyses indicate large effective population sizes, mid-Pleistocene colonizations, and post-divergence population gene flow (Figure 2). Lastly, we find evidence of associations between environmental gradients and allele frequency patterns in A. punctatus, but not in A. ortonii (Table 1), which is consistent with local adaptation in A. punctatus only. Analyses of environmental space occupancy suggest that the two species experience largely similar climates in Amazonia and the northern Atlantic Forest, yet A. punctatus occupies cooler and less humid localities that are not occupied by A. ortonii in the southern Atlantic Forest (Figure 3).

### 4.2 Candidate genes point to ecologically relevant physiological processes

GEAA results provide insights about physiological processes that may have experienced selection in response to climatic regimes in A. punctatus. The candidate genes identified play important roles in energy metabolism, immunity, development, and cell signaling. For instance, the malic enzyme 1 gene (ME1) encodes an enzyme that links the glycolytic and citric acid pathways (Jiang, Du, Mancuso, Wellen, & Yang, 2013) while generating NADPH necessary for fatty acid biosynthesis (Bukato, Kochan, & Świerczyński, 1995). The collagen type...
V alpha 2 chain gene (COL5A2) encodes a structural component of the extracellular matrix that is involved with skeletal, skin, and eye development through collagen fibril organization (Andrikopoulos, Liu, Keene, Jaenisch, & Ramirez, 1995). The product of the dedicator of cytokinesis 2 gene (DOCK2) is crucial in immune response by mediating the differentiation of T cells and lymphocyte migration to sites of infection (Fukui et al., 2001; Kunisaki et al., 2006). Finally, the coxsackie- and adenovirus receptor-like membrane protein gene (CLMP) encodes a transmembrane component of cell-to-cell junctional complexes that is essential for intestine development (Raschperger, Engstrom, Pettersson, & Fuxe, 2003). By pointing to ecologically relevant physiological processes, candidate genes recovered by GEAA can guide eco-physiological investigations of South American lizards. Future experimental studies that compare physiological tolerances between individuals from distinct habitats have the potential to reveal links between patterns of genetic variation, organismal function, and species distributions (Navas, 2002).

Similar to this study, other investigations found differences in the frequency of alleles that underlie ecologically relevant physiological processes between populations that inhabit contrasting habitats (e.g., Yoder et al., 2014; Benestan et al., 2016; Dennenmoser, Vamosi, Nolte, & Rogers, 2017; Guo, Lu, Liao, & Merilä, 2016). A few of these studies involved anole lizards. For instance, in the North American Anolis carolinensis, colonization of colder temperate regions from a tropical Caribbean source has been linked to adaptive shifts in several genes, including loci that are involved with development, energy metabolism, and immune response (Campbell-Staton, Edwards, & Losos, 2016). Additionally, evidence of thermal adaptation along an altitudinal gradient has been found in Anolis cybotes from Hispaniola, involving genes that play a role in bone formation and development, lipid metabolism, and regulation of blood pressure (Rodríguez et al., 2017). Similar to the case of A. punctatus, these examples support that adaptation to cooler climates has played an essential role in range expansions across anole taxa, including mainland and Caribbean forms that span altitudinal and latitudinal gradients. It is currently unclear, however, whether potentially beneficial polymorphisms preceded the colonization of new geographic regions by these species.

4.3 | Parallels between Guiana Shield and Atlantic Forest anoles

An individual of A. punctatus (MTR 20798) sampled in a mid-elevation site (820 m above sea level) in the Guiana Shield, on the Brazil-Venezuela border, grouped with Atlantic Forest samples in a clustering analysis based solely on candidate SNPs (Figure 1c, Supporting Information Figure S2b). By contrast, this sample clustered with eastern Amazonian samples (which are geographically closer) in an analysis based on the entire SNP dataset (Figure 1b, Supporting Information Figure S2a). Interestingly, the site where this individual was collected is climatically more similar to Atlantic Forest than to lowland Amazonian sites (Figure 3). These results may suggest that A. punctatus in the Guiana highlands and in the Atlantic Forest have independently acquired alleles that are beneficial under cooler or less rainy conditions, because this study (Figure 1b, Supporting Information Figure S2a) and previous analyses of historical relationships based on individual samples of A. punctatus (Prates, Rivera, et al., 2016; Prates, Xue, et al., 2016) found that Guiana Shield samples are not closely related to Atlantic Forest ones. Alternatively, this pattern may reflect differences in how neutral and beneficial alleles spread in geographic space, as demonstrated in other animal species that occur in contrasting habitats (Hoekstra, Drumm, & Nachman, 2004). Future analyses of local adaptation in populations that occur in the Guiana highlands will rely on improved sampling for this key region.

4.4 | Distinct GEAA patterns among species

Population structure and history can constrain local adaptation (Bridle & Vines, 2007; Lenormand, 2002), yet our analyses suggest that these factors do not account for different signatures of adaptation between the two anole species. G-PhoCS analyses recovered higher gene flow in A. punctatus than in A. ortonii, as well as older colonization of the Atlantic Forest in A. ortonii than in A. punctatus (Figure 2). Lower levels of gene flow and earlier colonization might have favored local adaptation in A. ortonii; however, GEAA found allele frequency patterns associated with climatic gradients in A. punctatus only. As a result, constraints related to population structure and history do not seem sufficient to explain discrepant GEAA results between the two species.

Alternatively, distinct GEAA patterns in A. ortonii and A. punctatus might stem from differences in the climatic regimes experienced by each species in parts of their range. In the coastal Atlantic Forest, A. ortonii is restricted to northern regions that are climatically more similar to Amazonia (Ledo & Colli, 2017; Sobral-Souza et al., 2015; Figure 3). By contrast, A. punctatus extends into colder and less rainy areas in the southern Atlantic Forest (Figure 3), as far as the state of São Paulo in Brazil’s southeast. As a result, our finding of several candidate SNPs in A. punctatus might be associated with local adaptation to cooler and less rainy conditions in the southern edge of this species’ range, as opposed to a pattern of physiological conservatism in A. ortonii, which shows a narrower range in the Atlantic Forest. Different species capacities to adapt to regional climates may hold the key to a pattern of pronounced taxonomic composition turnover around 19–20° of latitude in coastal Brazil (Carnaval et al., 2014). However, genetic clustering (Figure 1c) and PCA analyses (Supporting Information Figure S2b) based on the candidate SNPs flagged by GEAA do not support a pattern of differentiation between northern and southern Atlantic Forest samples of A. punctatus.

Lastly, it is possible that contrasting GEAA results in A. ortonii and A. punctatus have been influenced by sampling differences. Fewer individuals of A. ortonii were available for study, reflecting difficulties to sample this less abundant and less conspicuous species in the field. Limited sample sizes can strongly affect the power of GEAA to detect loci whose frequencies are associated with environmental conditions.
factors. This limitation is particularly important in the case of genetic variants that have weak environmental associations, which can be detected only with large sample sizes (Murcray, Lewinger, Conti, Thomas, & Gauderman, 2011). It is therefore possible that GEAA recovered only those candidate SNPs that are more strongly associated with temperature and precipitation gradients across the range of A. punctatus yet did not detect several other potentially beneficial SNPs in this species and also in A. ortonii. This situation illustrates the challenges of implementing GEAA on the basis of samples from natural populations, which can be hard to obtain in large numbers for certain organisms. Nevertheless, the suggestion of loci that show strong genome–environment associations in A. punctatus encourages further tests on the basis of larger sample sizes.

4.5 | GEAA in species that have expanded along environmental gradients

Our analyses of local adaptation may have been affected by conservative GEAA performance in species that underwent a history of range expansions along environmental gradients. When the axis of range expansion is parallel to the orientation of ecological gradients, the selected sites are expected to exhibit patterns of spatial structure that are similar to those of the neutral sites. As a result, the test’s null hypothesis (which relies on neutral background genetic variation) yields a stricter control of false positives (Frérot et al., 2015), and GEAA is expected to detect only the outlier loci that exhibit strong associations with the environmental predictor. A similar pattern of spatial structure between presumably selected and neutral sites agrees with our observation of concordant population clustering based on both candidate and neutral SNPs in A. punctatus (Figure 1b,c; Supporting Information Figure S2). While minimizing spurious genome–environment associations, this situation may also increase the rate of undetected adaptive genotypes. Conservative tests of adaptation in species that have expanded along environments gradients may be intensified in instances of limited spatial and individual sampling, as it may have been the case of A. ortonii.

This investigation illustrates how studies of adaptation on the basis of GEAA can benefit from knowledge about the history of landscape occupation by the species under investigation. Data on population structure and history can provide insight about how gene flow and natural selection interact and shape population genetic differentiation. Moreover, information about the direction and routes of colonization of new habitats can support spatial sampling design, help to characterize landscape gradients, and support the formulation of hypotheses about how organisms have responded to environmental variation in space.

ACKNOWLEDGMENTS

We thank Francisco Dal Vecchio, João Tonini, José Cassimiro, Marco Sena, Mauro Teixeira-Jr., Paulo Melo-Sampaio, Pedro V. L. Peloso, Renato Recoder, Roberta Damasceno, and Sergio Marques-Souza for field support. Ana Prudente and Annelise D’Angiolella provided samples deposited at the Museu Parelense Emilio Goeldi. Frederick Sheldon provided samples deposited at the Louisiana State University Museum of Natural Science. Hussan Zaher provided samples deposited at the Museu de Zoologia da Universidade de São Paulo. Olivier François, Bruno de Medeiros, Kevin P. Mulder, and Richard H. Adams provided advice on analysis implementation. Suggestions by Andrea Paz, Danielle Rivera, James B. Walsh, Jason Munshi-South, Kevin de Queiroz, Kevin P. Mulder, Kyle A. O’Connell, Laura Bertola, Maria L. Strangas, Mariana Vasconcellos, Matthew McElroy, Michael J. Hickerson, Rayna C. Bell, and Ryan K. Schott greatly improved this manuscript. Instituto Chico Mendes de Conservação da Biodiversidade issued collection permits (SISBIO 36753-1, 36753-4 and 27290-3). This work was co-funded by FAPESP (BIOTA, 2013/50297-0), NSF (DEB 1343578), and NASA through the Dimensions of Biodiversity Program. AP also acknowledges funding from a UTSA Graduate Presidential Distinguished Research Fellowship. MTR also acknowledges funding from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), grants 03/10335-8 and 11/50146-6. AC also acknowledges funding from NSF DEB 1120487. IP also acknowledges funding from an NSF Doctoral Dissertation Improvement Grant (DEB 1601271) and a Smithsonian Peter Buck Postdoctoral Fellowship.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS


DATA ACCESSIBILITY

All sequence data were deposited in the Sequence Read Archive (accession PRJNA492310). Supplementary tables and figures were deposited in the Dryad Digital Repository (https://doi.org/10.5061/dryad.1bj51s9). R and Unix shell scripts used in all analyses are available online through GitHub (https://github.com/ivanprates/2018_Anelis_EcoEvol).

ORCID

Ivan Prates http://orcid.org/0000-0001-6314-8852

REFERENCES

hypothesis revisited. Integrative and Comparative Biology, 46(1), 5-17. https://doi.org/10.1093/icb/icj003
species. PLoS ONE, 7(6), e38474. https://doi.org/10.1371/journal.pone.0038474


SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of the article.