Received: 6 May 2021 | Revised: 12 October 2021 | Accepted: 14 October 2021

DOI: 10.1111/jbi.14285

Journal of Biogeography WILEY

RESEARCH ARTICLE

Identifying traits that enable lizard adaptation to different habitats

Flavia M. Lanna[1] | **Guarino R. Colli**[2] | **Frank T. Burbrink**[3] | **Bryan C. Carstens**[1]

[1]Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, Ohio, USA

[2]Departamento de Zoologia, Universidade de Brasília, Brasília, Brazil

[3]Department of Herpetology, American Museum of Natural History, New York, New York, USA

Correspondence
Flavia M. Lanna, Department of Evolution, Ecology and Organismal Biology, The Ohio State University, 318 W. 12th Ave, Columbus, OH 43210, USA.
Email: flaviamollanna@gmail.com

Funding information
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; Conselho Nacional de Desenvolvimento Científico e Tecnológico; Fundação de Apoio à Pesquisa do Distrito Federal; USAID's PEER program; National Science Foundation

Handling Editor: David Chapple

Abstract

Aim: Species adapt differently to contrasting environments, such as open habitats with sparse vegetation and forested habitats with dense forest cover. We investigated colonization patterns in the open and forested environments in the diagonal of open formations and surrounding rain forests (i.e. Amazonia and Atlantic Forest) in Brazil, tested whether the diversification rates were affected by the environmental conditions and identified traits that enabled species to persist in those environments.

Location: South America, Brazil.

Taxon: Squamata, Lizards.

Methods: We used phylogenetic information and the current distribution of species in open and forested habitats to estimate ancestral ranges and identify range shifts relative to the current habitats. To evaluate whether these environments influenced species diversification, we tested 12 models using a Hidden Geographic State Speciation and Extinction analysis. Finally, we combined phylogenetic relatedness and species traits in a machine learning framework to identify the traits permitting adaptation in those contrasting environments.

Results: We identified 41 total transitions between open and forested habitats, of which 80% were from the forested habitats to the open habitats. Widely distributed species had higher speciation, turnover, extinction, and extinction fraction rates than species in forested or open habitats, but had also the lower net diversification rate. Mean body temperature, microhabitat, female snout–vent length and diet were identified as putative traits that enabled adaptation to different environments, and phylogenetic relatedness was an important predictor of species occurrence.

Main conclusions: Transitions from forested to open habitats are most common, highlighting the importance of habitat shift in current patterns of biodiversity. The combination of phylogenetic reconstruction of ancestral distributions and the machine learning framework enables us to integrate organismal trait data, environmental data and evolutionary history in a manner that could be applied on a global scale.

KEYWORDS
Diagonal of Open Formations, machine learning, Neotropics, random forest algorithm, range transitions, Squamata

© 2021 John Wiley & Sons Ltd

1 | INTRODUCTION

Understanding the processes that have shaped species' geographical distributions has intrigued scientists for centuries (Sclater, 1858; Wallace, 1854). Two main processes—adaptation and dispersal—can lead to a species occupying a novel habitat. Communities are formed where taxa evolve in situ and adapt to changes in their environment or where ancestors disperse into these areas from other regions. Regardless of which process account for the presence in a community, successful occupancy of a particular habitat requires tolerance to prevailing abiotic conditions, the acquisition of resources and the avoidance of predation. Since organismal traits play a crucial role in these actions, explicit consideration of organismal traits is likely to yield clues that tell us how species colonize and persist in novel habitats.

Organismal traits mediate the potential for adaptation and dispersal, leading to colonization, which influences how far species can disperse, cross geographical barriers and persist in the new areas. Furthermore, the evolution of critical traits may enable a species to diversify in new environments (Stroud & Losos, 2016). Diversification may also be facilitated by the absence of predators and competitors (Moore & Donoghue, 2007). Alternatively, species may adapt to a changing environment and persist through time in the same region, depending on the speed and intensity of environmental change and the standing genetic variation (Frank & Slatkin, 1992). Intense environmental pressures can influence the development of similar traits among distantly related species. For instance, different plant families inhabiting similarly dry environments in different parts of the Earth have leaves adapted to water storage (Eggli & Nyffeler, 2009). Differentiating between in situ adaptation and colonization is difficult, particularly without abundant fossil evidence or other records showing the past geographical distribution of a lineage.

Identifying traits that enable colonization is an essential step towards understanding species diversification. Numerous studies have investigated the patterns of colonization, their direction and species recolonization in areas with different environmental characteristics (Antonelli et al., 2018; Dutech et al., 2003; Gagnon et al., 2019; Hughes et al., 2013). For many plant species, transitions between environmentally different areas were less common than expected by chance. This finding may be explained by the niche conservatism hypothesis, where such shifts are rare due to the tendency of species to retain their ancestral ecological niches (Crisp et al., 2009). Other evidence supports broad shifts in habitat over time and the persistence of species. For example, studies in the Neotropics demonstrate transitions from wet and forested ancestral habitat to drier and open new habitat for multiple species groups (Antonelli et al., 2018; Zizka et al., 2020). These studies suggest that investigation into other groups is essential to generalize these patterns.

Due to differences in precipitation, temperature gradients and vegetation composition, the region encompassing the South American Diagonal of Open Formations (DOF) and its surrounding forests (Amazonia and Atlantic Forest) offers an excellent opportunity to investigate colonization patterns, the effect of habitat shifts on diversification rates, and to identify the traits that could have facilitated colonization and survival of species. The DOF extends diagonally across South America, connecting parts of Brazil, Paraguay, Bolivia and Argentina. The Brazilian DOF (Figure 1) encompasses environments seasonally stressed by droughts and composed chiefly of habitats with sparse vegetation (open habitats; Werneck, 2011). Importantly, these environments are heterogeneous and encompass patches of humid forested habitats (Mesquita et al., 2017). Although forested biomes (Figure 1) have higher precipitation and are composed chiefly of forested habitats (Franchito et al., 2009), they also include areas of open habitats (Moraes et al., 2020). Hence, it is crucial to account for habitat heterogeneity when assessing the drivers of adaptation and colonization within broad regions.

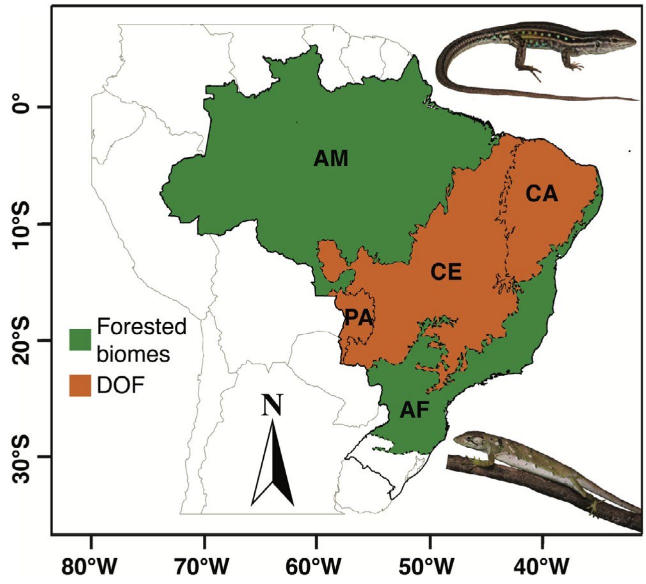

FIGURE 1 Georeferenced location of the Brazilian Diagonal of Open Formations (DOF; orange) and forested biomes (green). A dotted line separates biomes. AM: Amazonia, AF: Atlantic Forest, CA: Caatinga, CE: Cerrado and PA: Pantanal. The Chaco is not shown on the map due to its reduced area in the Brazilian DOF. Photographs from Ricardo Marques

Both open and forested habitats have changed through time as a response to climate and landscape modifications. For example, Andean uplift (during Eocene and Miocene) likely influenced the climate in South America by increasing aridity in many regions (Armijo et al., 2015) and modifying the Amazon drainage patterns (Hoorn et al., 2010). In the Middle Miocene, the Cerrado biome went through changes in vegetation composition by expanding C4 plants, associated with the establishment of savanna-like vegetation worldwide (Edwards et al., 2010; Graham, 2011). The marine incursions in Middle–Late Miocene inundated the lowland parts of the Chaco region (Hulka et al., 2006) and are hypothesized to have caused the local extinction of species in these lowlands (Garda & Cannatella, 2007). Pleistocene climatic fluctuations, characterized by periods of high variation in temperature and precipitation (Hewitt, 1996), significantly affected the extent of forested and

13652699, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14285 by Ohio State University Ohio Sta, Wiley Online Library on [07/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

open vegetation in South America (Costa et al., 2018). These climatic and landscape modifications likely offered an opportunity for species to disperse and colonize new areas. However, the colonization history is still understudied for many taxonomic groups in South America.

Lizards are useful for understanding how traits vary among environments given their ectothermic physiology (Deutsch et al., 2008; Sinervo et al., 2010). Lizards depend on external sources of heat to thermoregulate and achieve homeostasis (Deutsch et al., 2008; Huey et al., 2009, 2012). Brazil has 276 known lizard species from 15 taxonomic families (Costa & Bérnils, 2018), where most of the targeted DOF and the surrounding forested biomes are located. Presumably, extant species in this region are the product of adaptations that allowed their ancestors to survive in these drastically different environments.

Previous investigations have shown the impact of the environment on trait variation across the globe, such as clutch size (number of eggs laid in each reproduction event—smaller clutches in Amazonian populations than in non-Amazonian populations; larger clutches in the Caatinga than in the Cerrado populations; Garda et al., 2012; Rand, 1982; Vitt & Colli, 1994), body size (in some species it decreases with latitude and increases with temperature, and increases latitudinally in others; Ashton & Feldman, 2003; Azócar et al., 2015; Oufiero et al., 2011) and coloration (light phenotypes in dunes and dark phenotypes in dark soil habitats; Rosenblum et al., 2010). Here, we investigate: (1) the direction of the colonization events among open and forested habitats in the Brazilian DOF and their forest counterparts (i.e. Amazonia and Atlantic Forest); (2) the differences in diversification rates between open and forested habitats; and (3) how traits could have facilitated species colonization and persistence in open and forested regions. Given the putative early origin of the forested habitats, we expect more forest dispersions into younger open habitats. We also expect that traits associated with microhabitat and temperature will significantly predict colonization of forested and open habitats due to differences in vegetation cover and solar radiation. Exploring the evolution of traits in a predictive framework will lead to an improved understanding of the degree to which traits facilitate the colonization of novel habitats.

2 | MATERIALS AND METHODS

2.1 | Data collection

We used a comprehensive list of Brazilian lizard species (Costa & Bérnils, 2018) to identify candidate taxa to examine the influence of traits on the colonization of open and forested habitats. For each species we derived habitat information from our field experience with these taxa and the International Union for Conservation of Nature and Natural Resources website (IUCN—https://www.iucnredlist.org/). Since our objective was to identify traits that enabled species to colonize and persist in the open and forested habitats, we selected species associated with one or both habitats in the Brazilian DOF, Amazonia and Atlantic Forest.

We compiled lizard traits from Meiri (2018). Traits consisted of life history, morphology and physiology characteristics, such as diet, body size and body temperature respectively (Table 1). To fill in missing traits, we mainly used the Reptile Database website (Uetz et al., 2020), the Museu Virtual do Cerrado website (http://www.mvc.unb.br/pesquisa/especies/conheca-as-especies/jag) and the study by Mesquita et al. (2017). We used traits available for more than 50 species and were considered non-redundant (e.g. clutch size was excluded, while the smallest clutch size and largest clutch size were used). The traits not used were as follows: clutch size, smallest and largest mean clutch size, breeding age (months), youngest age at first breeding (months), oldest age at first breeding (months) and mean body temperature of active animals in the wild.

2.2 | Ancestral range estimation

To identify colonization patterns and estimate the diversification time of species in forested and open habitats, we combined species habitats and genetic data to estimate the probability of ancestral ranges. We used habitat information to create an occurrence matrix according to the presence and absence of each species in the forested habitats and the open habitats. We trimmed the consensus tree of the squamate phylogeny (Tonini et al., 2016) to contain only the species selected for this analysis using the function *keep.tip* from the R package 'ape' v5.4 (Paradis et al., 2004). To estimate species ancestral range, we used the R package 'BioGeoBEARS' v1.1.2 (Matzke, 2013). We tested six different models: DEC, a dispersal-extinction-cladogenesis model run using a Maximum Likelihood approach; DIVALIKE, a likelihood interpretation of the parsimony DIVA that generates dispersal-vicariance models; and BAYAREALIKE, a simplified likelihood interpretation of the Bayesian BAYAREA model; each with and without the parameter to control for the probability of founder-event speciation (+J parameter; DEC+J, DIVALIKE+J, BAYAREALIKE+J). We chose the model with the lowest Akaike information criterion (AIC) and AIC corrected for small sampling size (AICc) scores as the best model.

To determine ancestral range, we calculated the probability of each estimated range in each ancestral node of the phylogenetic tree. To quantify the number of colonization events from the open to the forested habitats and vice versa, we used the ancestral range probability map to count the range transitions. We considered the range with the highest probability in each node as the most probable range in that node. Transitions from forested to open habitats were counted every time nodes changed from forested habitats either to open habitats only or to both habitats. Likewise, transitions from open to forested habitats were counted every time nodes changed from open habitats either to forested habitats only or to both habitats.

13652699, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14285 by Ohio State University Ohio Sta, Wiley Online Library on [07/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Names, units of measurement, and descriptions of species traits are shown. Modified from Meiri (2018)

Traits	Unit	Trait description
Maximum SVL	mm	Maximum snout–vent length for the species (size measurement)
Female SVL	mm	Mean snout–vent length for females
Hatchling neonate SVL	mm	Midpoint of the snout–vent length for hatchling/neonate
Leg development	cg	Number of functioning legs for each species; species can be 'four-legged' (four functioning legs), 'leg reduced' (one or two pairs of reduced legs), 'two limbs' (have only forelimbs or hindlimbs) or 'no limbs' (have no limbs)
Activity time	cg	Time in the day in which the species is active. Species can be diurnal, nocturnal or cathemeral
Microhabitat	cg	Substrate in which the species lives and conducts most of its activities. Species can be terrestrial, arboreal, fossorial, saxicolous, semi-arboreal or semi-aquatic
Diet	cg	Diet preference. Species can be carnivorous (eats mostly or only animal matter), herbivorous (eats mostly or only plants) or omnivorous (eats both animal matter and plants)
Foraging mode	cg	Behaviour the species uses to acquire food. Species can have 'sit and wait' behaviours (ambush predator), 'active' behaviour (chases the prey) or 'mixed' behaviour (a combination of both 'sit and wait' and 'active' behaviours)
Reproductive mode	cg	Mode of reproduction. Species can be 'oviparous' or 'viviparous'
Smallest clutch size	#	Smallest number of eggs laid or offspring produced in each clutch or litter
Largest clutch size	#	Largest number of eggs laid or offspring produced in each clutch or litter
Minimum mean body temperature	°C	Lowest mean body temperature recorded in the field while the lizard was active
Maximum mean body temperature	°C	Highest mean body temperature recorded in the field while the lizard was active

Abbreviations: #, number; °C, degree Celsius; cg, categorical trait; mm, millimetres; SVL, snout–vent length.

2.3 | Identifying shifts in the diversification rate associated with species distributions

We also investigated if the environmental pressure of contrasting habitats influenced species diversification rates. We estimated state-dependent speciation and extinction rates from the phylogeny using the *GeoHiSSE* (Hidden Geographic State Speciation and Extinction) function (Caetano et al., 2018) of the 'hisse' v1.9.19 package (Beaulieu & O'Meara, 2016). GeoHiSSE analysis uses habitats as traits, allows shared trait states to account for generalist species and includes hidden traits that allow shifts in the diversification rates to be related to unmeasured traits, not forcing the correlation between diversification rate shifts to the traits under study (Caetano et al., 2018).

The hidden trait is a subdivision of the species that were divided into habitats. In other words, within each habitat, there is another category separating the species in two or more states. For example, species within a geographical area could be separated into categories of migrants and no migrants. While the species trait associated with the hidden trait or the biological relevance of the hidden trait is unknown, the presence of a hidden trait implies that some factor associated with this subdivision is necessary to explain the different rates in the model and allows these rates to vary within geographical areas.

We used all species that occur only in the open habitats, only in the forested habitats and in both habitats from our dataset, and the same phylogeny (Tonini et al., 2016) used in the 'BioGeoBEARS' analysis (using the consensus tree). We fitted 12 models (Caetano et al., 2018) that varied from null models (considering all rates the same for all states), full models (all parameters are free), containing area-independent or area-dependent variation of the diversification rates (we considered the areas as habitats), and separating extinction and extirpation (range reduction) rates for endemic lineages (+extirpation) (Table 2). We then evaluated the models using the AIC and AICc. We also used the marginal reconstruction algorithm to calculate model averages and diversification rates for each species with the *MarginReconGeoSSE* and *GetModelAveRates* functions within the 'hisse' package (Beaulieu & O'Meara, 2016).

2.4 | Identifying important traits to predict species occurrence in open and forested habitats

We aimed to determine if the occurrence of species in forested or open habitats could be predicted by species traits and phylogenetic relatedness using the random forest classification method. The advantage of using the random forest algorithm is that it considers different distributions of characters (continuous and categorical—including binary and multiple category characters), and thus is helpful in situations where it is unclear which combination of these characters will provide the most information relevant to the response variable (the occurrence of species in forested or open habitats). The random forest approach builds multiple decision trees to evaluate whether a particular trait can predict the response (Biau, 2012; Liaw & Wiener, 2002). It returns the mean decrease accuracy (MDA) estimate, which shows the decrease in the accuracy of the prediction function after removing a specific trait from the analysis, characterizing the most important traits for the predictive function.

13652699, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14285 by Ohio State University Ohio Sta, Wiley Online Library on [07/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Traits identified as relevant for the classification model could indicate their association with the colonization and persistence of species to their current habitats. Two categories were selected as the response variables for the model: presence in the open vegetation habitats and presence in the forested habitats. We used a subset of our species dataset (235 species) that contained species currently occurring only in one of those two habitats and absent in the other habitats. We excluded species that occur in both open and forested habitats (nine species) because we only have trait information on the species level and not for individuals or populations. Therefore, we could not separate the traits of individuals in forested habitats from those in open habitats.

To possibly account for traits inherited by species that might be absent from our random forest model, we incorporated phylogenetic relatedness into the models as a proxy for shared but unscored traits. We trimmed the consensus tree of the squamate phylogeny (Tonini et al., 2016) to contain only the species present in our trait dataset using the function *keep.tip* from the 'ape' package (Paradis et al., 2004). Then we used the Phylogenetic Eigenvector Regression method (PVR; Diniz-Filho et al., 1998) to transform the phylogenetic tree into a pairwise distance matrix and convert the distances into eigenvalues using a principal coordinate analysis (PCoA) with the functions *cophenetic* and *pcoa*, respectively, from the 'ape' package (Paradis et al., 2004). We selected the first five axes with more cumulative eigenvalues as variables in the random forest analysis.

We then used the random forest classification method through the 'randomForest' v4.6-14 package (Liaw & Wiener, 2002) to create and evaluate the models. We used the MDA estimate to measure variable importance and the out-of-bag (OOB) error rate and the classification error to evaluate the model's overall quality. In addition to the MDA estimate, we also identified the relevant variables using a wrapper algorithm implemented in the function *Boruta* in the package 'Boruta' v7.0.0 (Kursa & Rudnicki, 2010). This method uses a similar approach to the MDA in the package 'randomForest' with the addition of a set of new variables with randomized values (with no correlation with the response variable). We ran the random forest classifier using the *Boruta* function, increasing the 'maximal number of importance source runs' (maxRuns) parameter to correct unresolved variables' importance when needed. The function compares the importance of the model variables with shadow attributes, considering as relevant those with importance greater than the most important shadow attribute. All variables less important than the most important shadow attribute are considered non-relevant to the model (Kursa & Rudnicki, 2010).

We tested two approaches because the random forest algorithm does not allow missing data, and our trait dataset has missing information for some species. First, we tested models varying in a combination of traits and species; we created a model with all the variables and without the species with missing values for any variable. Then we evaluated, using the *Boruta* function, the importance of model variables and built a model containing only the important variables (those that are more important than the shadow attributes). This improved the model accuracy and allowed the inclusion of species previously excluded due to missing values for one or some of the non-important variables. The last step of identifying non-relevant variables was repeated until all the variables in the model were relevant. Eigenvalues used as phylogenetic relatedness among species vary according to the species used to generate them, and different random forest models had different species compositions due to missing trait data. Therefore, for each set of species used in each random forest model, we transformed the

TABLE 2 GeoHiSSE model descriptions. Models and descriptions are obtained from the study by Caetano et al. (2018). Models are organized according to importance (AIC, AICc and AIC weights). For detailed information on each model, see Caetano et al. (2018)

Models	Description	Free parameters	AIC	AICc	AIC weights
10	GeoHiSSE+extirpation, two hidden rate classes, full model	19	2184.47731	2187.87016	1
11	CID—GeoHiSSE+extirpation, five hidden rate classes, null model	15	2387.95135	2390.05661	6.55E-45
4	GeoHiSSE, two rate classes, full model	15	2406.76936	2408.87462	5.37E-49
5	CID—GeoHiSSE, five hidden rate classes, null model	13	2414.72929	2416.3119	1.00E-50
9	CID—GeoHiSSE+extirpation, three hidden rate classes, null model	11	2415.3374	2416.47534	7.40E-51
12	CID—GeoHiSSE+extirpation, two hidden rate classes	9	2424.85812	2425.62735	6.34E-53
3	CID—GeoHiSSE, three hidden rate classes, null model	9	2437.0968	2437.86603	1.39E-55
2	Original GeoSSE, full model	7	2442.62301	2443.09759	8.80E-57
8	GeoSSE+extirpation, full model	9	2446.14106	2446.91029	1.52E-57
6	CID—GeoHiSSE, two hidden rate classes	7	2448.61978	2449.09435	4.39E-58
1	CID—original GeoSSE	4	2506.41609	2506.58346	1.24E-70
7	CID—GeoSSE+extirpation	6	2531.73559	2532.09002	3.92E-76

Abbreviation: CID, character-independent diversification.

13652699, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14285 by Ohio State University Ohio Sta, Wiley Online Library on [07/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

pairwise distances into eigenvalues only for the species used in each model and applied only the eigenvalues of the first five axes (cumulative eigenvalue >80%) as predictors in the model.

Second, we used the function *missForest* in the package 'missForest' v1.4 (Stekhoven & Bühlmann, 2012) to impute trait missing values to create a complete model using all the variables and all the species. To run the imputation function, we excluded the response variable of the random forest model (occurrence in open or forested habitats). This avoids circularity caused by using the response variable to predict missing values of predictor variables and then using the imputed predictors to model the response. One of the biological data features is that they are not independent due to their evolutionary history; therefore, we included the eigenvalues axes to account for phylogenetic relatedness. The imputation function uses a random forest model to train the dataset based on the available values and uses the model to predict the missing values. Both categorical and numerical variables are accepted. It returns an OOB error rate estimation for the imputation. After imputing the missing values, we created two random forest models, one with all variables and another with only the relevant variables according to the *Boruta* function.

We built each random forest classifier model with 2000 trees and we ran each model 100 times to account for the variation in the multiple decision trees. We calculated the mean values of the OOB error rate, the classification error of each response variable and the mean and standard deviation for the MDA scores at the end of the 100 runs. To understand how the essential traits vary between the habitats, we used only the species used in the best model (lowest OOB error rate) and the traits classified as important for the model and plotted their distributions according to each habitat.

3 | RESULTS

3.1 | Data summary

Our dataset was composed of 244 species and represented approximately 88% of the species that occur in Brazil (Costa & Bérnils, 2018). These species represented 15 of 38 taxonomic families recognized worldwide for lizards (Uetz et al., 2020) and all the families occurring in Brazil (Costa & Bérnils, 2018). Of these 244 species, only nine use both open and forested habitats, while 129 use only forested habitats and 106 use only open habitats. From all species in this study, 74.6% had genetic information available when the phylogeny was constructed, while 25.4% had their genetic data imputed (Tonini et al., 2016).

Our trait dataset was composed of the 235 species that occurred only in the open or forested habitats and 13 life-history, morphological and physiological traits (Table 1). Only two of the traits had information for all the species—maximum snout–vent length (SVL) and leg development. All the other traits had different levels of missing data, ranging from 4.2% for microhabitat to 62.9% for both minimum and maximum mean body temperature.

3.2 | Ancestral range and transitions estimation

The estimation of species ancestral ranges was conducted using a tree with all 244 species, in which 235 species occur only in one habitat (129 in forested and 106 in open habitats) and nine species occurring in both open and forested habitats. The dispersal-extinction-cladogenesis with the founder-event speciation (DEC+J) model was selected as the best model for estimating ancestral ranges.

The forested habitats were estimated as the ancestral range for all the species (i.e. most ancestral node) with a probability of ~53%, followed by ~38% of being both ranges and 9% of being the open habitats (Figure S1.1). We recorded a total of 41 range transitions between the open and the forested habitats, in which most of them were from the forested to the open habitats (33 of 41 transitions; 80%; Figure 2). According to the consensus tree, the first transition was from the forested to the open habitats approximately 82 million years ago (Ma). Since there was no genetic information for all the species on the tree and some were imputed, this age may vary in the different posterior trees. Therefore, this represents an estimation of the age using the consensus tree. Range transitions varied in time, and range transitions involving congeneric species occurred more recently (Figure S1.1).

3.3 | Geography-dependent speciation

We found a significant dependence of the species' occurrence on the diversification rates. The best-fitting model of the GeoHiSSE analysis was model 10, which included area-dependent diversification, two hidden rate classes, and extirpation and extinction rates separated for endemic lineages (Table 2). The best model included area-dependent diversification, which emphasizes the importance of the habitats on the diversification rates (Table 2). A model of area independence would have performed better in the case of no or minor importance of the habitats in these rates.

The habitats—open, forested and both—had species separated into two hidden traits (A and B). The hidden traits had a different turnover, extinction fraction and dispersal rates within each habitat. Species rates in open, forested and both habitats varied within the groups and hidden traits (A and B). The species from the hidden trait A (in the three habitats) presented higher turnover rates and slightly higher extinction fraction rates than those in the hidden trait B (Table S1.1).

To compare the diversification rates among the habitats while integrating the hidden traits, we averaged the models according to their AIC weights. The speciation, turnover, extinction and extinction fraction rates for the species in both habitats were higher than those in forested and open habitats. On the other hand, the net diversification rate was higher for the species that occur in forested habitats, followed by the species that occur in open habitats. We found a lower net diversification rate for species occurring in both habitats, indicating similar rates of speciation and extinction (Figure 3).

13652699, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14285 by Ohio State University Ohio Sta, Wiley Online Library on [07/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

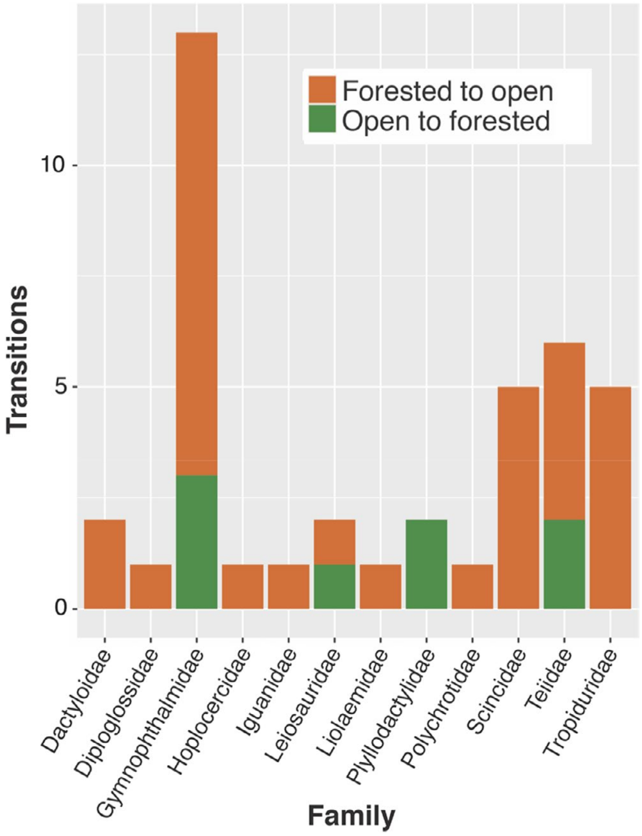

FIGURE 2 Range transitions to a single family and within families. Forty transitions from the BioGeoBEARS analysis are shown. Three of the 15 families (Alopoglossidae, Gekkonidae, and Sphaerodactylidae) had no inferred range transitions. The transition that occurred in the common ancestor of the Gekkonidae and the Phyllodactylidae is not shown. See Figure S1.1 for more information on ancestral range reconstruction

3.4 | Random forest classification

We generated five random forest models to predict the occurrence of the species in open and forested habitats and identify the critical variables for this prediction. Three models had species with missing data and non-relevant variables removed, and two were based on the imputed data and without non-relevant variables. Eigenvalues were calculated for the species dataset used in each model, and the first five PCoA axes were used as characters in the models (>80% cumulative eigenvalue) representing phylogenetic relatedness among species.

The first model tested comprised 18 variables, including five eigenvalue axes and 66 species (27 in open habitats and 39 in forested habitats). The OOB error of this model was 12.33%, with a classification error of 22% for the species in the open habitats and 5% for the species in the forested habitats. The classification errors were asymmetrical, which likely resulted from the smaller number of open habitats species used in the model. Some of the variables, such as leg development, smallest clutch size and reproductive mode, were relatively unimportant to the model (MDA close to zero; Figure S1.2). The other variables varied according to the importance level. The Boruta analysis recovered six variables as non-relevant and 11 as relevant, including all the five eigenvalue axes of phylogenetic relatedness (Figure 4a).

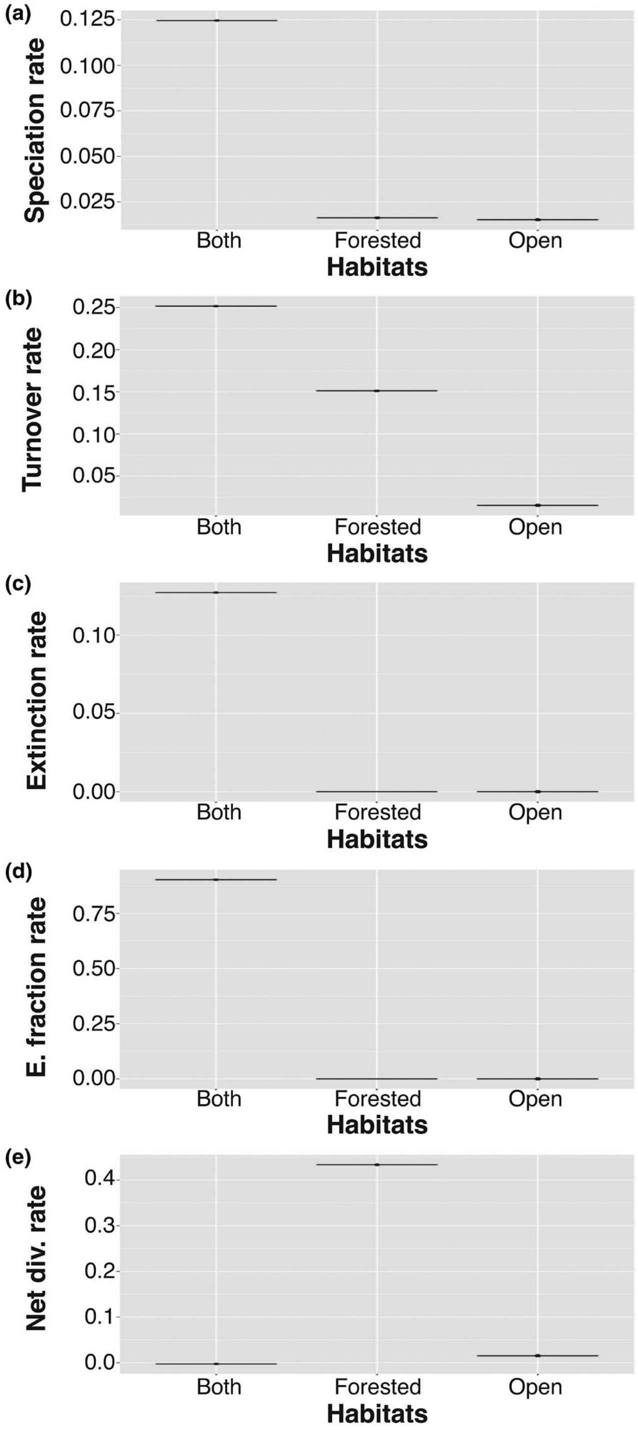

FIGURE 3 The variation of speciation (a), turnover (b), extinction (c), extinction fraction (d) and net diversification (e) rates after model averaging for species from both habitats, forested habitats and open habitats

After removing the non-relevant predictor variables from model 1 (Figure 4a), the second model included 11 variables and 82 species (39 in the open habitats and 43 in the forested habitats). The

13652699, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14285 by Ohio State University Ohio Sta, Wiley Online Library on [07/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

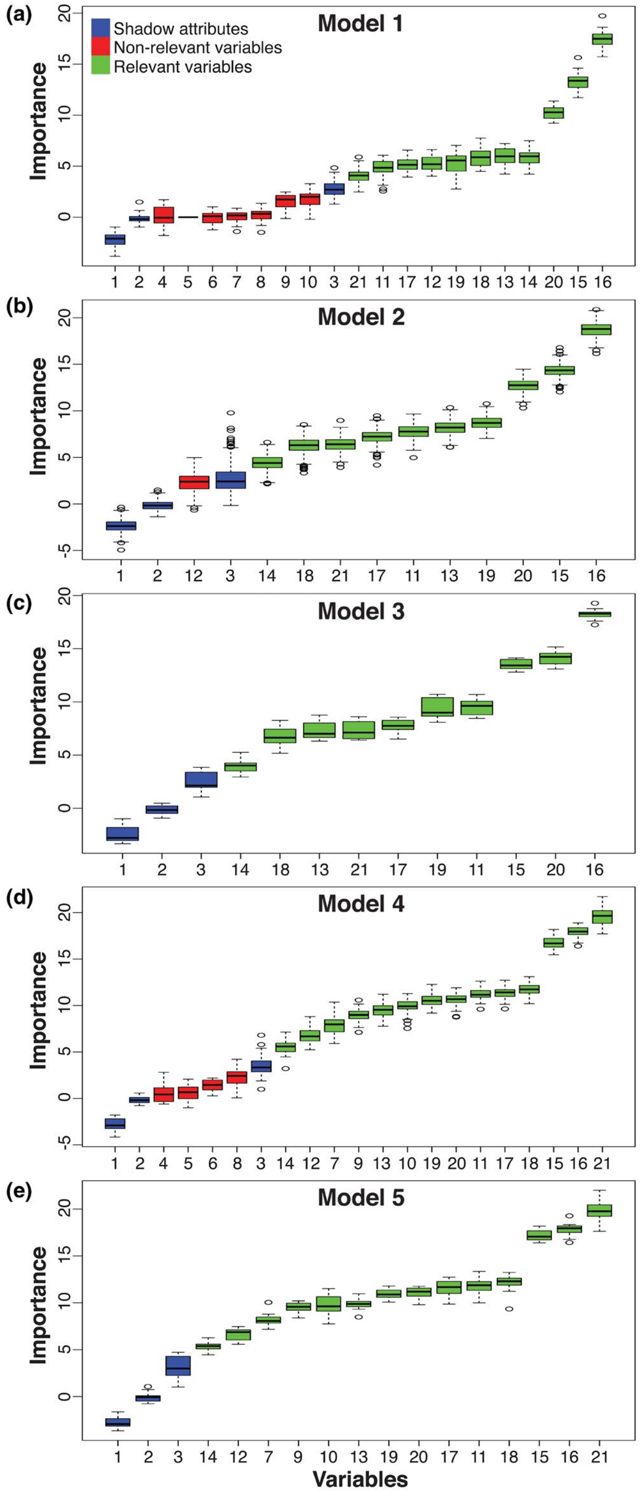

FIGURE 4 Variable importance of random forest models according to the Boruta analysis. The *y*-axis corresponds to the importance of each variable. Relevant traits are displayed in green; non-relevant traits are in red; and the minimum, average and maximum shadow attributes are in blue. The *x*-axis corresponds to the variables. Numbers are associated with the variables. a–e corresponds to variable relevance and importance for models 1–5 respectively. 1–3: minimum, average and maximum shadow attributes respectively; 4: reproductive mode; 5: leg development; 6: foraging mode; 7: maximum SVL; 8: smallest clutch; 9: hatchling/neonate SVL; 10: activity time; 11: female SVL; 12: largest clutch; 13: microhabitat; 14: diet; 15: minimum mean body temperature; 16: maximum mean body temperature; 17–21: phylogenetic relatedness variables (eigenvalue axes 1–5 respectively)

removal of non-relevant variables decreased the OOB (9.58%) and the classification errors (open habitats: 15% and forested habitats: 4%).

Due to the increase in the number of species, we ran the Boruta function one more time to check if all the variables were relevant to the model, and one of the variables was classified as non-relevant (largest clutch size, Figure 4b). We then removed this variable and ran a third model, which included 10 variables (all relevant variables; Figure 4c) and 85 species (39 in the open habitats and 46 in the forested habitats). The OOB error decreased from 9.58% to 8.22%. The classification error for the species from forested habitats decreased (2%), while the classification error for the species from open habitats remained the same (15%).

The imputation function returns two measurements of error, one for the categorical and one for the continuous variables. The continuous variables had an error (NRMSE—normalized root mean squared error) of 7%, while the classification error for the categorical variables (PFC—proportion of falsely classified entries) was 11%. Using the values imputed, totalling 235 species (106 in the open habitats and 129 in the forested habitats), we ran two random forest models (models 4 and 5). The first (model 4) included all the variables. The OOB error was higher than the models without imputed data (OOB: 15.63%). The classification error was also higher than the other models (open habitats: 23% and forested habitats: 9%). The Boruta analysis classified four variables as non-relevant (Figure 4d). After removing those variables, we ran the fifth model (second with imputed trait data), and both the OOB and the classification errors were similar (OOB: 15.76%, open habitats: 22% and forested habitats: 10%). All the variables were relevant in this model (Figure 4e). Although the data are different, overall, the models without imputed data performed better than those with the imputed data.

In these models, minimum and maximum body temperature, microhabitat, female SVL and diet were the essential variables for all models according to the MDA estimations (the second figure in Appendix 1 as Figure S1.2). The phylogenetic relatedness variables were also critical classifiers for the models, indicating the importance of other lineage-specific traits not included in our dataset.

Considering the relevant traits used in the best model (with the lowest OOB error rate—model 3), species from open habitats have higher minimum and maximum mean body temperatures (Figure 5a,b respectively) than species from forested habitats, tend to be more saxicolous and terrestrial and had a more omnivorous and herbivorous diet. In contrast, species from the forested habitats are more semi-arboreal, semi-aquatic, arboreal and fossorial (Figure 5e), and have a more carnivorous diet (Figure 5d). The female SVLs were slightly different between the habitats, in which species

13652699, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14285 by Ohio State University Ohio Sta, Wiley Online Library on [07/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

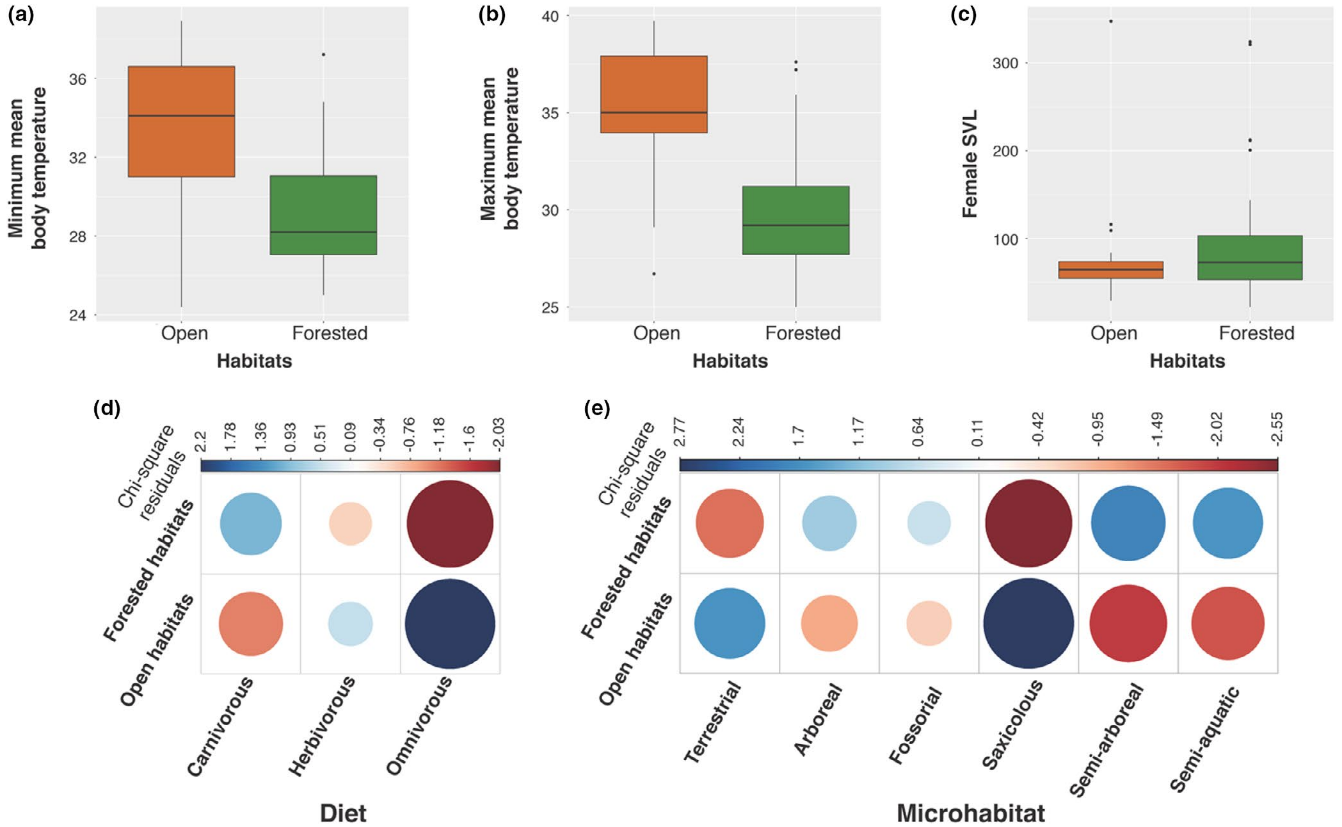

FIGURE 5 Comparison of species from open (39 species) and forested habitats (46 species) for the most important predictors according to best random forest model (third model): minimum mean body temperature (a), maximum mean body temperature (b), female SVL (c), diet (d) and microhabitat (e). (a–c): boxplot showing mean body temperature (minimum [a] and maximum [b]) and female SVL (c) for the species in the open habitats (orange) and the forested habitats (green). (d, e) represents the chi-square residuals; colours represent the residuals according to positive (blue) to negative (red) association between habitats (open and forested habitats) and diet (d) or and microhabitat (e); the size of the circles corresponds to the amount of cell contribution

in the forested habitats varied more in female size, and species in the open habitats had a narrower size variation (Figure 5c).

4 | DISCUSSION

We examined lizard colonization patterns among habitats, determined if particular habitats influence species diversification, and identified traits that facilitated these colonization events and species survival in their current ranges. Our results suggest that the forested habitats were the ancestral range for the lizard species investigated in this study, and we also found evidence for a more typical range transition from the forested habitats to the open habitats. Widely distributed species had higher speciation, turnover, extinction, and extinction fraction rates than species in forested or open habitats, but had also the lower net diversification rate. Our results identified five important species traits (i.e. minimum and maximum mean body temperature, microhabitat, female SVL and diet) that could have facilitated colonization and persistence of species to their current range and highlighted the importance of phylogenetic relatedness to explain the occurrence of the species.

4.1 | Historical biogeography, range transitions and influence of the habitats on the diversification rate

The open habitats in South America are younger than the forested habitats (Azevedo et al., 2020), consistent with our results showing that the forested habitats are the ancestral ranges for most lizards investigated. Also consistent is the finding that transitions from forested to open habitats were more common than the reverse. Due to their younger age, the open habitats could have afforded more empty niches and allowed species' colonization from the forested habitats. This pattern of more transitions from forested to open habitats has been recovered in other studies. For example, Antonelli et al. (2018) estimated dispersal events for six taxonomic groups of species among Neotropical biomes and found that Amazonia was the source of species for all the groups. Unlike our investigation, these authors separated the habitats into biomes and recovered dispersal events from Amazonia to most other areas, open and forested. Similarly, Zizka et al. (2020) identified more transition events of Bombacoideae plants from evergreen forests (forested habitats) to seasonally dry biomes (open habitats), with multiple colonization events into the open habitats, as observed here for lizards. Zizka et al. followed Antonelli et al. in analysing their data at the

13652699, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14285 by Ohio State University Ohio Sta, Wiley Online Library on [07/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

biome level. This approach might confound the actual differences in habitat within these biomes. For example, many species occur in open habitats within Amazonia , and simply including taxa with occurrences in forested and open habitats in this named biome would likely complicate efforts to identify traits associated with habitat preference. To understand how the environment shapes the processes of dispersion and colonization of species, we argue that future studies should include species habitat and not the general biomes in the analyses.

4.2 | Important traits for colonization and persistence of species in different environments

Identifying traits that explain species distributions or highlighting how they fit different environments has previously been investigated for small species groups (e.g. Colli et al., 2003; Garda et al., 2012). Large-scale studies, particularly those involving species from multiple genera, usually focus on identifying patterns and processes of species distributions, range transitions between areas and habitats, or testing a single trait chosen a priori to identify its variance according to a process or pattern (e.g. Amado et al., 2018; Burbrink & Myers, 2014; Igea & Tanentzap, 2020; Kennedy et al., 2017; Slavenko et al., 2019; Velasco et al., 2020; Zizka et al., 2020). Here we identified traits that putatively influenced species colonization and persistence between two contrasting Neotropical environments. Our results suggest that five species traits are essential for the occupation of open or forested habitats. Minimum and maximum mean body temperature, female SVL, microhabitat and diet likely influence the survival of these species in each region. We also identified that phylogenetic relatedness is also crucial to understand species occurrence.

Lizards from the open habitats have a higher mean body temperature (minimum and maximum) than the lizards in the forested habitats. This finding agrees with the general nature of forested habitats, which have a closed canopy resulting in less solar radiation penetration there. There are also more saxicolous and terrestrial lizards in the open habitats, suggesting that saxicolous microhabitats help shelter from the sun to avoid water loss and overheating. In contrast, as expected, the denser vegetation and canopy cover in the forested habitats enable more arboreal and semi-arboreal lizards in these regions. Forested habitats have more standing water than open habitats, and there are more semi-aquatic lizards in these habitats. The greater carnivorous diet in the forested habitats versus the more omnivorous diet in the open habitats could be explained by the difference in prey availability in the two habitats, which could also be correlated with climate (mostly precipitation) and vegetation cover. Even though female size was a relevant variable, there are no apparent differences among the species from the two habitats, only that species from the open habitats had a narrower variation in female SVL, in contrast to the broader variation in size from species in the forested habitats (Figure 4). This broader variation in size could be due to a possibly higher number of niches available in the forested habitats when compared to the open habitats.

Our findings demonstrate that phylogenetic relatedness is vital for predicting the occurrence of the species in open or forested habitats (Figure 4). This result highlights the role of niche conservatism in Neotropical diversification, indicating that species retain their niche preferences even after speciation and tend to occur in similar environments (Wiens & Graham, 2005). Indeed, most species within some genera or even families are restricted to a single region.

While considering transitions between habitats using only extant species and their current distributions, it is essential to consider other factors. Species may have colonized the open or forested habitats from other regions or even other continents, not transitioning between these two habitats. For example, some of the species from the family Gekkonidae colonized South America directly from the African continent (Gamble et al., 2011).

5 | CONCLUSIONS

As reported for other groups, our results demonstrate a predominant pattern of transitions from forested to open habitats. Mean body temperature, female SVL, microhabitat and diet were important predictors of species occurrence in open or forested environments. We emphasize that our approach should be able to identify phenotypic transitions across different regions across the globe. The machine learning approach presented here corresponds to the first step of the comparative phylogenomic approaches proposed by Smith et al. (2020). Therefore, identifying phenotypic transitions among clades in a macroevolutionary perspective represents a framework to study the genetic basis for phenotypes, which is especially important for non-model organisms where traits responsible for adaptation are rarely known. The traits identified in this study can be used in the future comparative genomic investigations to understand the genomic basis of adaptation.

ACKNOWLEDGEMENTS

The authors thank the National Science Foundation (NSF) for financial support via a grant to BCC (DEB-1831319) and FTB (DED-1831241). GRC thanks CAPES, CNPq, Fundação de Apoio à Pesquisa do Distrito Federal (FAPDF) and USAID's PEER program under cooperative agreement AID-OAA-A-11-00012 for financial support. They thank members of the Carstens Lab for comments and suggestions on the manuscript and colleagues on the Dimensions of Biodiversity grant. No permits were necessary for this work.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The trait dataset, files and scripts used in the analyses are available via the Dryad archive (https://doi.org/10.5061/dryad.t1g1jwt3c).

13652699, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14285 by Ohio State University Ohio Sta, Wiley Online Library on [07/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ORCID

Flavia M. Lanna https://orcid.org/0000-0002-7263-872X
Guarino R. Colli https://orcid.org/0000-0002-2628-5652
Frank T. Burbrink https://orcid.org/0000-0001-6687-8332
Bryan C. Carstens https://orcid.org/0000-0002-1552-227X

REFERENCES

Amado, T. F., Bidau, C. J., & Olalla-Tárraga, M. A. (2018). Geographic variation of body size in New World anurans: Energy and water in a balance. *Ecography*, *41*, 1–11. https://doi.org/10.1111/ecog.03889

Antonelli, A., Zizka, A., Carvalho, F. A., Scharn, R., Bacon, C. D., Silvestro, D., & Condamine, F. L. (2018). Amazonia is the primary source of Neotropical biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(23), 6034–6039. https://doi.org/10.1073/pnas.1713819115

Armijo, R., Lacassin, R., Coudurier-Curveur, A., & Carrizo, D. (2015). Coupled tectonic evolution of Andean orogeny and global climate. *Earth-Science Reviews*, *143*, 1–35. https://doi.org/10.1016/j.earscirev.2015.01.005

Ashton, K. G., & Feldman, C. R. (2003). Bergmann's rule in nonavian reptiles: Turtles follow it, lizards and snakes reverse it. *Evolution*, *57*(5), 1151–1163. https://doi.org/10.1111/j.0014-3820.2003.tb00324.x

Azevedo, J. A. R., Collevatti, R. G., Jaramillo, C. A., Strömberg, C. A. E., Guedes, T. B., Matos-Maraví, P., Bacon, C. D., Carillo, J. D., Faurby, S., & Antonelli, A. (2020). On the young savannas in the land of ancient forests. In V. Rull & A. C. Carnaval (Eds.), *Neotropical diversification: Patterns and processes* (p. 820). Springer.

Azócar, D. L. M., Perotti, M. G., Boninho, M. F., Schulte, J. A. II, Abdala, C. S., & Cruz, F. B. (2015). Variation in body size and degree of melanism within a lizards clade: Is it driven by latitudinal and climatic gradients? *Journal of Zoology*, *295*, 243–253. https://doi.org/10.1111/jzo.12193

Beaulieu, J. M., & O'Meara, B. C. (2016). Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology*, *65*(4), 583–601. https://doi.org/10.1093/sysbio/syw022

Biau, G. (2012). Analysis of a random forest model. *Journal of Machine Learning Research*, *13*, 1063–1095. https://doi.org/10.5603/AIT.a2017.0074

Burbrink, F. T., & Myers, E. A. (2014). Body size distributions at community, regional or taxonomic scales do not predict the direction of trait-driven diversification in snakes in the United States. *Global Ecology and Biogeography*, *23*(4), 490–503. https://doi.org/10.1111/geb.12139

Caetano, D. S., O'Meara, B. C., & Beaulieu, J. M. (2018). Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution*, *72*(11), 2308–2324. https://doi.org/10.1111/evo.13602

Colli, G. R., Mesquita, D. O., Rodrigues, P. V. V., & Kitayama, K. (2003). Ecology of the gecko *Gymnodactylus geckoides amarali* in a Neotropical savanna. *Journal of Herpetology*, *37*(4), 694–706. https://doi.org/10.1670/180-02A

Costa, G. C., Hampe, A., Ledru, M.-P., Martinez, P. A., Mazzochini, G. G., Shepard, D. B., Werneck, F. P., Moritz, C., Carnaval, A. C., & Fortin, M.-J. (2018). Biome stability in South America over the last 30 kyr: Inferences from long-term vegetation dynamics and habitat modelling. *Global Ecology and Biogeography*, *27*(3), 285–297. https://doi.org/10.1111/geb.12694

Costa, H. C., & Bérnils, R. S. (2018). Répteis do brasil e suas unidades Federativas: lista de espécies. *Revista Herpetologia Brasleira*, *7*(1), 11–57.

Crisp, M. D., Arroyo, M. T. K., Cook, L. G., Gandolfo, M. A., Jordan, G. J., McGlone, M. S., Weston, P. H., Westoby, M., Wilf, P., & Linder, H. P. (2009). Phylogenetic biome conservatism on a global scale. *Nature*, *458*(7239), 754–756. https://doi.org/10.1038/nature07764

Deutsch, C. A., Tewksbury, J. J., Huey, R. B., Sheldon, K. S., Ghalambor, C. K., Haak, D. C., & Martin, P. R. (2008). Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(18), 6668–6672. https://doi.org/10.1073/pnas.0709472105

Diniz-Filho, J. A. F., Sant'Ana, C. E. R., & Bini, L. M. (1998). An eigenvector method for estimating phylogenetic inertia. *Evolution*, *52*(5), 1247–1262. https://doi.org/10.1111/j.1558-5646.1998.tb02006.x

Dutech, C., Maggia, L., Tardy, C., Joly, H. I., & Jarne, P. (2003). Tracking a genetic signal of extinction-recolonization events in a Neotropical tree species: *Vouacapoua americana* Aublet in French Guiana. *Evolution*, *57*(12), 2753–2764. https://doi.org/10.1111/j.0014-3820.2003.tb01517.x

Edwards, E. J., Osborne, C. P., Strömberg, C. A. E., Smith, S. A., Bond, W. J., Christin, P.-A., Cousins, A. B., Duvall, M. R., Fox, D. L., Freckleton, R. P., Ghannoum, O., Hartwell, J., Huang, Y., Janis, C. M., Keeley, J. E., Kellogg, E. A., Knapp, A. K., Leakey, A. D. B., Nelson, D. M., … Tipple, B. (2010). The origins of C4 grasslands: Integrating evolutionary and ecosystem science. *Science*, *328*, 587–591.

Eggli, U., & Nyffeler, R. (2009). Living under temporarily arid conditions - succulence as an adaptive strategy. *Bradleya*, *27*(27), 13–36. https://doi.org/10.25223/brad.n27.2009.a10

Franchito, S. H., Rao, V. B., Vasques, A. C., Santo, C. M. E., & Conforte, J. C. (2009). Validation of TRMM precipitation radar monthly rainfall estimates over Brazil. *Journal of Geophysical Research Atmospheres*, *114*, 1–9. https://doi.org/10.1029/2007JD009580

Frank, S. A., & Slatkin, M. (1992). Fisher's fundamental theorem of natural selection. *Trends in Ecology & Evolution*, *7*(3), 92–95. https://doi.org/10.1016/0169-5347(92)90248-A

Gagnon, E., Ringelberg, J. J., Bruneau, A., Lewis, G. P., & Hughes, C. E. (2019). Global succulent biome phylogenetic conservatism across the pantropical *Caesalpinia* group (Leguminosae). *New Phytologist*, *222*(4), 1994–2008. https://doi.org/10.1111/nph.15633

Gamble, T., Bauer, A. M., Colli, G. R., Greenbaum, E., Jackman, T. R., Vitt, L. J., & Simons, A. M. (2011). Coming to America: Multiple origins of New World geckos. *Journal of Evolutionary Biology*, *24*(2), 231–244. http://onlinelibrary.wiley.com/store/ https://doi.org/10.1111/j.1420-9101.2010.02184.x/asset/j.1420-9101.2010.02184.x.pdf?v=1&t=ibax4cg8&s=932d0d39faf58122d6b5dd336e9ed817e2b3fb8d

Garda, A. A., & Cannatella, D. C. (2007). Phylogeny and biogeography of paradoxical frogs (Anura, Hylidae, Pseudae) inferred from 12S and 16S mitochondrial DNA. *Molecular Phylogenetics and Evolution*, *44*(1), 104–114. https://doi.org/10.1016/j.ympev.2006.11.028

Garda, A. A., Costa, G. C., França, F. G. R., Giugliano, L. G., Leite, G. S., Mesquita, D. O., Nogueira, C., Tavares-Bastos, L., Vasconcellos, M. M., Vieira, G. H. C., Vitt, L. J., Werneck, F. P., Wiederhecker, H. C., & Colli, G. R. (2012). Reproduction, body size, and diet of *Polychrus acutirostris* (Squamata: Polychrotidae) in two contrasting environments in Brazil. *Journal of Herpetology*, *46*(1), 2–8. https://doi.org/10.1670/10-288

Graham, A. (2011). The age and diversification of terrestrial New World ecosystems through Cretaceous and Cenozoic time. *American Journal of Botany*, *98*(3), 336–351. https://doi.org/10.3732/ajb.1000353

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, *58*, 247–276. https://doi.org/10.1006/bijl.1996.0035

Hoorn, C., Wesselingh, F. P., ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J., Sanmartín, I., Sanchez-Meseguer, A., Anderson, C. L., Figueiredo, J. P., Jaramillo, C., Riff, D., Negri, F. R., Hooghiemstra, H., Lundberg, J., Stadler, T., Särkinen, T., & Antonelli, A. (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, *330*, 927–931. https://doi.org/10.1126/science.1194585

13652699, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14285 by Ohio State University Ohio Sta, Wiley Online Library on [07/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Huey, R. B., Deutsch, C. A., Tewksbury, J. J., Vitt, L. J., Hertz, P. E., Pérez, H. J. Á., & Garland, T. (2009). Why tropical forest lizards are vulnerable to climate warming. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 1939–1948. https://doi.org/10.1098/rspb.2008.1957

Huey, R. B., Kearney, M. R., Krockenberger, A., Holtum, J. A. M., Jess, M., & Williams, S. E. (2012). Predicting organismal vulnerability to climate warming: Roles of behaviour, physiology and adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 1665–1679. https://doi.org/10.1098/rstb.2012.0005

Hughes, C. E., Pennington, R. T., & Antonelli, A. (2013). Neotropical plant evolution: Assembling the big picture. *Botanical Journal of the Linnean Society*, *171*(1), 1–18. https://doi.org/10.1111/boj.12006

Hulka, C., Gräfe, K. U., Sames, B., Uba, C. E., & Heubeck, C. (2006). Depositional setting of the middle to Late Miocene Yecua Formation of the Chaco foreland basin, southern Bolivia. *Journal of South American Earth Sciences*, *21*(1–2), 135–150. https://doi.org/10.1016/j.jsames.2005.08.003

Igea, J., & Tanentzap, A. J. (2020). Angiosperm speciation cools down in the tropics. *Ecology Letters*, *23*, 692–700. https://doi.org/10.1111/ele.13476

Kennedy, J. D., Borregaard, M. K., Jønsson, K. A., Holt, B., Fjeldså, J., & Rahbek, C. (2017). Does the colonization of new biogeographic regions influence the diversification and accumulation of clade richness among the Corvides (Aves: Passeriformes)? *Evolution*, *71*(1), 38–50. https://doi.org/10.1111/evo.13080

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, *36*(11), 1–13. https://doi.org/10.18637/jss.v036.i11

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*, 18–22.

Matzke, N. J. (2013). Probabilistic historical biogeography: New models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography*, *5*(4), 242–248. https://doi.org/10.21425/f5fbg19694

Meiri, S. (2018). Traits of lizards of the world: Variation around a successful evolutionary design. *Global Ecology and Biogeography*, *27*(10), 1168–1172. https://doi.org/10.1111/geb.12773

Mesquita, D. O., Costa, G. C., Garda, A. A., & Delfim, F. R. (2017). Species composition, biogeography, and conservation of the Caatinga lizards. In J. M. C. Silva, I. R. Leal, & M. Tabarelli (Eds.), *Caatinga* (pp. 151–180). Springer. https://doi.org/10.1007/978-3-319-68339-3_6

Moore, B. R., & Donoghue, M. J. (2007). Correlates of diversification in the plant clade Dipsacales: Geographic movement and evolutionary innovations. *The American Naturalist*, *170*, S28–S55. https://doi.org/10.1086/519460

Moraes, L. J. C. L., Ribas, C. C., Pavan, D., & Werneck, F. P. (2020). Biotic and landscape evolution in an Amazonian contact zone: Insights from the herpetofauna of the Tapajós River Basin, Brazil. In V. Rull & A. C. Carnaval (Eds.), *Neotropical diversification: Patterns and processes* (p. 820). Springer.

Oufiero, C. E., Gartner, G. E. A., Adolph, S. C., & Garland, T. Jr (2011). Latitudinal and climatic variation in body size and dorsal scales counts in *Sceloporus* lizards: A phylogenetic perspective. *Evolution*, *65*(12), 35903607. https://doi.org/10.1111/j.1558-5646.2011.01405.x

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412

Rand, A. S. (1982). Clutch and egg size in Brazilian iguanid lizards. *Herpetologica*, *38*, 171–178.

Rosenblum, E. B., Römpler, H., Schöneberg, T., & Hoekstra, H. E. (2010). Molecular and functional basis of phenotypic convergence in white lizards at White Sands. *Proceedings of the National Academy of Sciences*, *107*(5), 2113–2117. https://doi.org/10.1073/pnas.0911042107

Sclater, P. L. (1858). On the general geographical distribution of the members of the class Aves. *Zoological Journal of the Linnean Society*, *2*(7), 130–136. https://doi.org/10.1111/j.1096-3642.1858.tb02549.x

Sinervo, B., Méndez-de-la-Cruz, F., Miles, D. B., Heulin, B., Bastiaans, E., Villagrán-Santa Cruz, M., Lara-Resendiz, R., Martínez-Méndez, N., Calderón-Espinosa, M. L., Meza-Lázaro, R. N., Gadsden, H., Avila, L. J., Morando, M., De la Riva, I. J., Sepulveda, P. V., Rocha, C. F. D., Ibargüengoytía, N., Puntriano, C. A., Massot, M., … Sites, J. W. (2010). Erosion of lizard diversity by climate change and altered thermal niches. *Science*, *328*, 894–899. https://doi.org/10.1126/science.1184695

Slavenko, A., Feldman, A., Allison, A., Bauer, A. M., Böhm, M., Chirio, L., Colli, G. R., Das, I., Doan, T. M., LeBreton, M., Martins, M., Meirte, D., Nagy, Z. T., Nogueira, C. D. C., Pauwels, O. S. G., Pincheira-Donoso, D., Roll, U., Wagner, P., Wang, Y., & Meiri, S. (2019). Global patterns of body size evolution in squamate reptiles are not driven by climate. *Global Ecology and Biogeography*, *28*(4), 471–483. https://doi.org/10.1111/geb.12868

Smith, S. D., Pennell, M. W., Dunn, C. W., & Edwards, S. V. (2020). Phylogenetics is the new genetics (for most of biodiversity). *Trends in Ecology and Evolution*, *35*(5), 415–425. https://doi.org/10.1016/j.tree.2020.01.005

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118. https://doi.org/10.1093/bioinformatics/btr597

Stroud, J. T., & Losos, J. B. (2016). Ecological opportunity and adaptive radiation. *Annual Review of Ecology, Evolution, and Systematics*, *47*(1), 507–532. https://doi.org/10.1146/annurev-ecolsys-121415-032254

Tonini, J. F. R., Beard, K. H., Ferreira, R. B., Jetz, W., & Pyron, R. A. (2016). Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. *Biological Conservation*, *204*, 23–31. https://doi.org/10.1016/j.biocon.2016.03.039

Uetz, P., Freed, P., & Hošek, J. (2020). The reptile database. Retrieved September 2020, from http://www.reptile-database.org

Velasco, J. A., Villalobos, F., Diniz-Filho, J. A. F., Poe, S., & Flores-Villela, O. (2020). Macroecology and macroevolution of body size in *Anolis* lizards. *Ecography*, *43*(6), 812–822. https://doi.org/10.1111/ecog.04583

Vitt, L. J., & Colli, G. R. (1994). Geographical ecology of a Neotropical lizard: *Ameiva ameiva* (Teiidae) in Brazil. *Canadian Journal of Zoology*, *72*, 1986–2008.

Wallace, A. R. (1854). On the monkeys of the Amazon. *Journal of Natural History Series 2*, *14*(84), 451–454. https://doi.org/10.1080/037454809494374

Werneck, F. P. (2011). The diversification of eastern South American open vegetation biomes: Historical biogeography and perspectives. *Quaternary Science Reviews*, *30*(13–14), 1630–1648. https://doi.org/10.1016/j.quascirev.2011.03.009

Wiens, J. J., & Graham, C. H. (2005). Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution, and Systematics*, *36*, 519–539. https://doi.org/10.1146/annurev.ecolsys.36.102803.095431

Zizka, A., Carvalho-Sobrinho, J. G., Pennington, R. T., Queiroz, L. P., Alcantara, S., Baum, D. A., Bacon, C. D., & Antonelli, A. (2020). Transitions between biomes are common and directional in Bombacoideae (Malvaceae). *Journal of Biogeography*, *47*, 1310–1321. https://doi.org/10.1111/jbi.13815

BIOSKETCH

Flávia M. Lanna is a PhD student interested in macroecology, evolutionary biology, community phylogenetics and biogeography, with particular attention to Neotropical lizards. The authors share a common interest in phylogeography, biogeography and ecology of disparate groups of organisms such as reptiles,

13652699, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14285 by Ohio State University Ohio Sta, Wiley Online Library on [07/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

amphibians, mammals and plants, using genetic, morphological and environmental data combined with cutting-edge analytical techniques (e.g. machine learning). In particular, this study is part of a project that aims to understand the adaptation and diversification of organisms across dry environments in Brazil.

Author contributions: B.C.C. and F.M.L. conceived the ideas; G.R.C. and F.M.L. collected the data; all authors designed the methodology; F.M.L. analysed the data and led the writing with the assistance of all the authors.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Lanna, F. M., Colli, G. R., Burbrink, F. T., & Carstens, B. C. (2022). Identifying traits that enable lizard adaptation to different habitats. *Journal of Biogeography*, 49, 104–116. https://doi.org/10.1111/jbi.14285

13652699, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jbi.14285 by Ohio State University Ohio Sta, Wiley Online Library on [07/09/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License