ECOLOGY LETTERS

Ecology Letters, (2013) **16**: 175–182 doi: 10.1111/ele.12025

LETTER

Quantifying the roles of ecology and geography in spatial genetic divergence

Ian J. Wang,[1]* Richard E. Glor[2] and Jonathan B. Losos[1]

[1]*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, 02138, USA*

[2]*Department of Biology, University of Rochester, Rochester, NY, 14627, USA*

**Correspondence: E-mail: ianwang@fas.harvard.edu*

Abstract

Investigating the properties of ecological landscapes that influence gene flow among populations can provide key insights into the earliest stages of biological divergence. Both ecological and geographical factors can reduce gene flow, which can lead to population divergence, but we know little of the relative strengths of these phenomena in nature. Here, we use a novel application of structural equation modelling to quantify the contributions of ecological and geographical isolation to spatial genetic divergence in 17 species of *Anolis* lizards. Our comparative analysis shows that although both processes contributed significantly, geographical isolation explained substantially more genetic divergence than ecological isolation (36.3 vs. 17.9% of variance respectively), suggesting that despite the proposed ubiquity of ecological divergence, non-ecological factors play the dominant role in the evolution of spatial genetic divergence.

Keywords

Biogeography, ecological isolation, gene flow, genetic divergence, isolation by distance, structural equation modelling.

Ecology Letters (2013) **16**: 175–182

INTRODUCTION

Understanding the factors that contribute to population genetic divergence is a long-standing goal in ecology and evolution (Wright 1921; Mayr 1963; Coyne & Orr 2004). Patterns of genetic variation often reflect spatial variation in gene flow, which can be influenced by ecological landscapes in two important ways. Spatially separated populations may experience isolation-by-distance (IBD; Wright 1943), in which landscape barriers and geographical distances cause restricted gene flow, and isolation-by-environment (IBE; Wang & Summers 2010), in which gene flow among populations inhabiting different ecological environments is limited either by selection against dispersers moving between them or by individual preference to remain in a particular environment due to local adaptation (Dobzhansky 1937). IBD predicts a correlation between genetic divergence and geographical factors such as landscape barriers and geographical distance, and IBE predicts a correlation between genetic divergence and environmental dissimilarity, because greater environmental differences between populations are expected to be associated with stronger divergent selection and reduction in the success of dispersers (Crispo *et al.* 2006; Lee & Mitchell-Olds 2011). Of course, geographical and environmental isolation are not mutually exclusive, and spatial genetic divergence among populations can result from reduced gene flow associated with both geographical and ecological factors (e.g. Coyne & Orr 2004; Crispo *et al.* 2006; Thorpe *et al.* 2008).

Despite the large body of research on patterns of genetic divergence, few studies have quantified the contributions of IBD and IBE to spatial genetic divergence, especially at comparative scales necessary to investigate the strengths and prevalence of these relationships in nature (Thorpe *et al.* 2008; Schluter 2009; Sobel *et al.* 2010). Historically, studies of population divergence have focused on IBD (Wright 1943), while recent studies have argued that IBE plays a dominant role in the evolution of genetic divergence (Thorpe *et al.* 2008; Nosil 2012). The relative contributions of these two factors is still debated, and understanding their effects on the reduction of gene flow among populations can inform our understanding of how landscapes and environments shape patterns of genetic variation in nature (Cushman *et al.* 2006; Wang & Summers 2010; Lee & Mitchell-Olds 2011).

Because populations that are geographically distant also tend to occupy different environments, disentangling the effects of IBD and IBE is inherently difficult and remains a major challenge. However, the rise of modern spatial statistical methods and the increasing availability of high-resolution geographical and environmental data layers now makes it possible to accurately describe geographical and ecological landscapes and to simultaneously estimate the effects of IBD and IBE on spatial genetic divergence. The genetic consequences of IBD and IBE are often best studied in widespread organisms with spatially structured population divergence (Thorpe *et al.* 2004; Sobel *et al.* 2010). Thus, comparative analyses of such species can provide valuable insights into the evolution of genetic divergence among populations.

In this study, we use a novel implementation of SEM to quantify the contributions of IBD and IBE to spatial genetic divergence in 17 widespread species in the diverse adaptive radiations of *Anolis* lizards on the Greater Antilles. We used this approach to test predictions stemming from three alternative hypotheses: (1) geographical distances and barriers limit gene flow between populations such that IBD contributes significantly to spatial genetic divergence, (2) divergent selection in different environments reduces gene flow between populations such that IBE contributes significantly to spatial genetic divergence and (3) geographical distances and divergent selection each influence gene flow such that IBD and IBE both contribute significantly to spatial genetic divergence. In all cases, we quantified the relative strengths of IBD and IBE for each species. Previous work suggests that anole species often inhabit heterogeneous environments (Losos & Ricklefs 2009), have substantial levels of genetic variation (e.g. Glor *et al.* 2003; Kolbe *et al.* 2004) and adapt to local ecological conditions (e.g. Thorpe *et al.* 2004; Calsbeek *et al.* 2006), making them excellent study systems for examining the evolution of spatial genetic divergence (Ogden &

Thorpe 2002; Thorpe *et al.* 2008). By investigating numerous widespread species across four large islands, we were able to conduct a comparative analysis of the factors affecting spatial genetic variation across ecologically diverse species and landscapes.

MATERIALS AND METHODS

Genetic data

We assembled neutral DNA sequence data sets for 17 widespread species of *Anolis* lizards, each inhabiting one of the four Greater Antillean islands. For 12 species, we acquired sequence data from previously published studies: *Anolis allisoni* (ND2 and tRNAs; Glor *et al.* 2004), *Anolis cooki* (ND2 and cytB; Jezkova *et al.* 2009), *Anolis chlorocyanus*, *Anolis crisatellus*, *Anolis cybotes* (ND2 and tRNAs; Kolbe *et al.* 2007), *Anolis distichus dominicensis*, *Anolis distichus ignigularis* (ND2 and tRNAs; Glor & Laport 2012), *Anolis krugi* (ND2 and tRNAs; Rodriguez-Robles *et al.* 2010), *Anolis poncensis* (ND2 and cytB; Jezkova *et al.* 2009), *Anolis porcatus* (ND2 and tRNAs; Glor *et al.* 2004), *Anolis sagrei* (ND2 and tRNAs; Kolbe *et al.* 2004) and *Anolis whitemani* (ND2, tRNAs, ND1, COI and control region; Glor *et al.* 2003). For five other species, *Anolis equestris*, *Anolis garmani*, *Anolis grahami*, *Anolis lineatopus* and *Anolis stratulus*, we collected DNA sequences from the ND2 gene and adjacent $tRNA^{Trp}$ and $tRNA^{Ala}$ following previously published methods (Glor & Laport 2012). To be included in our study, a species had to have genetic data collected from multiple localities with precise GPS coordinates providing < 100 m accuracy (at least four significant decimal-degree digits) or with locality information detailed enough to acquire coordinates with this accuracy (Fig. 1). In practice, most localities were georeferenced on-site using hand-held GPS units. For this study, we consider the subspecies of *A. distichus* to be distinct species, based on a recent report recommending that they be elevated to species status (Glor & Laport 2012).

For our mtDNA data set, we calculated between-locality genetic divergence using a maximum composite likelihood model of nucleotide evolution that allowed for gamma-distributed rate variation among sites in MEGA (Tamura *et al.* 2007). We then used these data to construct a matrix of genetic distances between localities in each species. For three species for which nDNA data were available (*A. allisoni*, *A. chlorocyanus* and *A. porcatus*), we also constructed nDNA genetic distance matrices and concatenated mtDNA and nDNA genetic distance matrices. We considered point localities to be the relevant sampling units in our study because our method does not rely on using a metapopulation model of reproductively isolated populations and, like many other approaches to examining spatial genetic patterns, can be applied to species with individuals distributed continuously on a landscape (Rousset 2000; Cushman *et al.* 2006). In practice, *Anolis* lizards are typically highly territorial and have small home ranges (Schoener & Schoener 1982; Johnson *et al.* 2009), suggesting that these species will have compact genetic neighbourhoods (Wright 1943) that should generally be smaller than the distances between our sampling localities.

Environmental data

We acquired geographic information systems (GIS) data layers for a total of 24 environmental and geographical variables, including 19 bioclimatic variables with 1-km resolution from the WorldClim database (http://www.bioclim.org), four vegetation variables from the MODIS land cover database with 250-m resolution (leaf area index, normalised difference vegetation index, tree density and herbaceous density; http://modis.gsfc.nasa.gov/) and a digital elevation model with 30-m resolution from the USGS EROS database (http://eros.usgs.gov). We extracted values for each variable at every locality for each species using ArcGIS (ESRI) and calculated differences between localities to create a dissimilarity matrix for each variable.

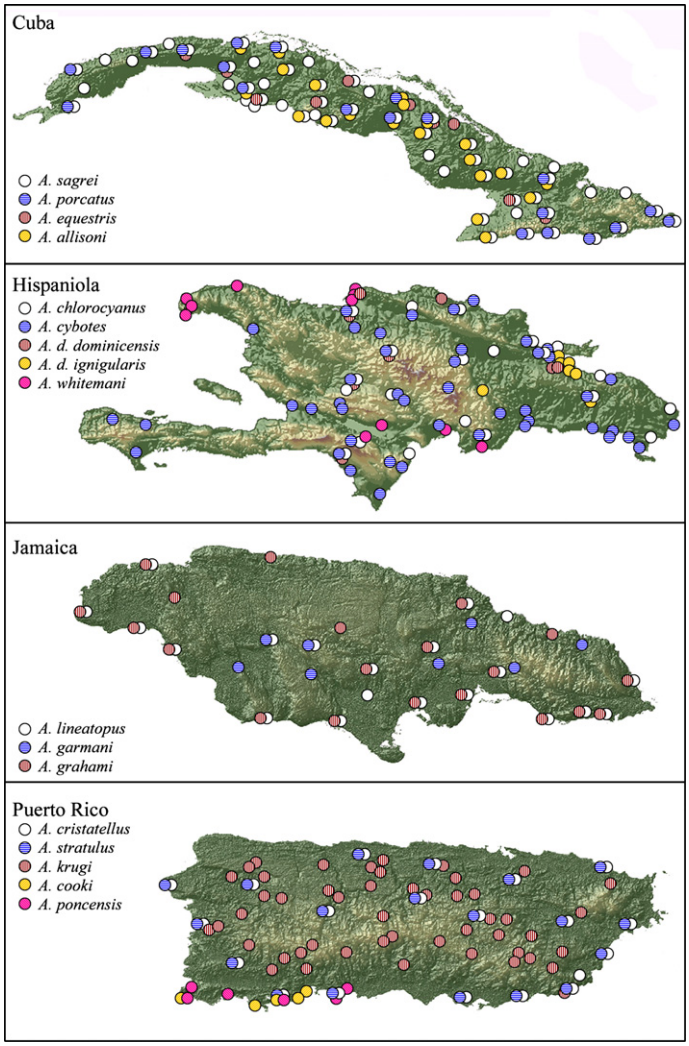

Figure 1 Sampling localities for the 17 species examined in this study from the four Greater Antillean islands (Cuba, Hispaniola, Jamaica and Puerto Rico). Localities are represented as coloured and patterned circles on a topographic relief map. Overlapping circles indicate multiple species collected from the same site.

Geographical distances

Resistance-based (also known as cost-weighted) distances have been shown to more closely reflect gene flow between populations than direct distances for a variety of organisms (Cushman *et al.* 2006; McRae & Beier 2007). These distances account for variation in the ease of movement across a heterogeneous landscape by assigning relative resistances (or costs) to different landscape features (McRae & Beier 2007; Storfer *et al.* 2010). To assign resistances to each cell in a landscape data layer for each species, we followed a method

for translating niche model suitability scores into resistances (Wang *et al.* 2008). We first created a niche model for each species using MAXENT (Phillips *et al.* 2006), which uses occurrence data and environmental layers to predict the environmental suitability of each cell in the study area. As input data, we used the 24 environmental and geographical data layers and an average of 110 occurrence records for each species (Table 1). Occurrence records were collected from our data, previously published studies and the HerpNet/VertNet species database of museum records from 55 institutions (www.vertnet.org). To evaluate how accurately the resulting GIS layer reflects the suitability of the landscape to the species being modelled, we performed a receiver operating characteristic (ROC) analysis (Phillips *et al.* 2006) and found that it indicated a good fit (AUC > 0.858 in all cases) to the species distribution model for each species (Phillips *et al.* 2006). MAXENT estimates environmental suitability values ranging from 0–1, and higher scores indicate more suitable habitat. We used the reverse of these values, because lower suitability should have a higher resistance, to assign resistances to each cell in the study area of each species.

We then used the resulting resistance layers to obtain geographical distance matrices using two alternative measures of geographical distance: (1) least-cost path distances calculated using ArcGIS and (2) circuit distances calculated using CircuitScape (McRae & Beier 2007). Least-cost path distance is calculated by searching for the path that minimises the total cumulative cost (or resistance) between two points (Wang *et al.* 2009), and circuit distance is calculated by summarising the costs of all possible paths between two points (McRae & Beier 2007). For least-cost path distances, our estimates were adjusted for the additional distance between points imposed by topographical relief based on the digital elevation model; this adjustment is not currently available in CircuitScape. These analyses resulted in two matrices of the geographical distances between localities for each species.

Although the geographical distance and environmental dissimilarity estimates both use the same set of 24 environmental variables, those variables are actually utilised very differently. In one case (environmental dissimilarity), only the raw values at specific points are considered, whereas in the other (geographical distance), the full data layers are processed to infer another variable, habitat suitability, which is then used in measuring the weighted distances along paths (or sets of paths) between localities. Thus, geographical distance and environmental dissimilarity are not necessarily correlated. For instance, if two populations inhabited localities with identical scores for each environmental variable, they would have an environmental dissimilarity of zero but a geographical distance greater than zero as long as they have any spatial separation. Nevertheless, to make sure that any associations with geographical distance were not driven by the inclusion of environmental variables, we also calculated topographic distances between localities.

Structural equation modelling

To estimate the relative contributions of geographical distance and environmental dissimilarity to genetic divergence, we utilised a structural equation modelling (SEM) framework (Grace 2006). Originally developed by the geneticist Sewall Wright (1921), SEM is a statistical framework for evaluating complex relationships between multiple variables that uses a series of regression and model-fitting analyses to calculate correspondence among any number of variables whose relationships are hypothesised *a priori* (Grace 2006; Santos & Cannatella 2011). SEM is, therefore, an ideal framework for testing hypotheses about the contributions of a large number of different variables associated with geographical distance and environmental variation to genetic divergence. Specifically, we estimated the strength of IBD from the effects of geographical distance on genetic divergence while controlling for environmental dissimilarity,

Table 1 Data used in structural equation modelling (SEM) and their associated results for 17 *Anolis* species. Data include the number of sampling localities (Loc.), the number (Seq.) and length in base pairs (bp) of sequences used in genetic comparisons and the number of occurrence points (Occ.) used in niche modelling. For each species, SEM was used to quantify the proportion of genetic divergence explained by isolation-by-distance (IBD) and isolation-by-ecology (IBE), presented as maximum-likelihood estimates ± standard errors (values in italics are non-significant). Also listed are the sums of IBD and IBE (Total), the covariation between these variables (Covar.) and the primary contributors (Contrib. vars.) to the environmental dissimilarity latent variable (individual factor loading coefficients > 0.7)

Study system		Data				Results				
Species	Island	Loc.	Seq.	bp	Occ.	IBD	IBE	Total	Covar.	Contrib. vars.
Anolis allisoni	Cuba	20	94	1173	59	0.464 ± 0.074	0.237 ± 0.080	0.701	0.198	Precip.
Anolis equestris	Cuba	10	13	1219	31	0.438 ± 0.137	0.194 ± 0.071	0.632	0.335	Temp., Precip., Veg.
Anolis porcatus	Cuba	20	59	1173	110	0.796 ± 0.046	0.060 ± 0.010	0.856	0.101	Precip.
Anolis sagrei	Cuba	56	197	1475	108	0.475 ± 0.023	0.140 ± 0.020	0.615	0.240	Temp., Elev.
Anolis chlorocyanus	Hispaniola	21	59	1504	277	*0.030 ± 0.008*	0.189 ± 0.014	0.219	0.014	Precip.
Anolis cybotes	Hispaniola	47	114	1109	681	0.405 ± 0.029	0.044 ± 0.028	0.449	0.167	Temp., Elev.
Anolis distichus dominicensis	Hispaniola	8	18	1462	145	*0.047 ± 0.024*	0.200 ± 0.098	0.247	0.042	Temp., Elev.
Anolis distichus ignigularis	Hispaniola	7	25	1462	65	0.310 ± 0.173	0.477 ± 0.180	0.787	0.302	Temp.
Anolis whitemani	Hispaniola	12	17	1921	46	0.075 ± 0.061	0.219 ± 0.065	0.294	0.114	Temp., Precip., Elev.
Anolis garmani	Jamaica	14	21	1213	14	0.698 ± 0.075	*0.001 ± 0.002*	0.699	0.059	Temp., Elev., Veg.
Anolis grahami	Jamaica	20	48	1548	36	0.487 ± 0.063	*0.006 ± 0.004*	0.493	0.022	Temp., Elev.
Anolis lineatopus	Jamaica	20	71	1548	42	0.345 ± 0.028	*0.005 ± 0.005*	0.350	0.009	Temp., Elev.
Anolis cooki	Puerto Rico	6	52	2124	21	0.529 ± 0.216	0.115 ± 0.102	0.644	0.513	Temp., Precip.
Anolis cristatellus	Puerto Rico	20	89	1436	71	0.157 ± 0.065	0.394 ± 0.158	0.551	0.093	Precip.
Anolis krugi	Puerto Rico	52	208	2001	72	0.242 ± 0.027	*0.001 ± 0.002*	0.243	0.086	Temp., Elev.
Anolis poncensis	Puerto Rico	6	56	2129	20	0.416 ± 0.220	0.327 ± 0.225	0.743	0.248	Temp., Precip.
Anolis stratulus	Puerto Rico	19	92	1452	71	0.301 ± 0.076	0.213 ± 0.077	0.514	0.048	Temp.
mean		21.1	72.5	1526.5	109.9	0.363 ± 0.208	0.179 ± 0.146	0.543	0.152	

and we estimated the strength of IBE from the effects of environmental dissimilarity on genetic divergence while controlling for geographical distance.

One of the advantages of SEM is that it can use latent variables – variables that are not directly measured but instead inferred from a set of observed variables. In our model, we described geographical distance and environmental dissimilarity as latent variables because we wanted to infer the best-fitting values rather than assume that geography or environment could be summarised by some arbitrary method. Essentially, geographical distance and environmental dissimilarity are both complex variables that are difficult to measure, because determining the biologically relevant variables is complicated and we had no *a priori* expectation for how those variables should be weighted. Thus, modelling these as latent variables was a more sensible and conservative approach. We inferred the geographical distance latent variable from our two measurements of distance (least-cost path distance and circuit distance), and we inferred the environmental dissimilarity latent variable from our set of 24 environmental variables. Prior to analysis, all observed variables were standardised. We also defined regression pathways between geographical distance and genetic divergence, and between environmental dissimilarity and genetic divergence, and a covariance pathway between geographical distance and environmental dissimilarity (Fig. 2). The estimated coefficients of these pathways can be used to quantify the relative effects of IBD and IBE on spatial genetic divergence, while controlling for covariation between geographical distance and environmental dissimilarity ('Covar.' in Fig. 2).

We solved the model by simultaneously estimating all parameters using maximum-likelihood estimation (MLE) in the 'lavaan' package in R (http://cran.r-project.org/web/packages/lavaan/). By estimating the parameters simultaneously, we were able to optimise the fit of each parameter, including the regression pathways connecting latent variables and the indicator pathways used to manifest the latent variables, while controlling for covariation among variables. So, we estimated the best fit of geographical distance and environmental dissimilarity to genetic divergence while determining the best estimates of geographical distance and environmental dissimilarity from their respective indicator variables. Standard errors for each parameter were determined by calculating Huber–White robust standard errors (White 1980) based on the residuals of the fitted model.

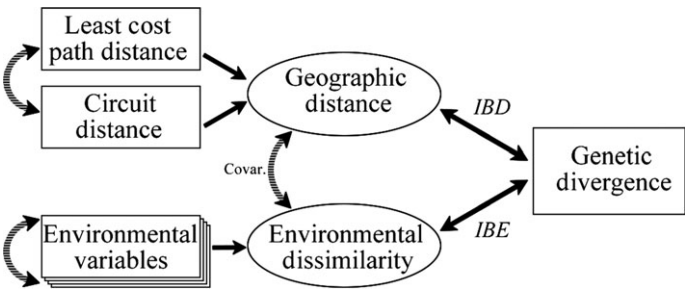

Figure 2 Graphical representation of the structural equation model used to quantify the relationships between geographical distance, environmental dissimilarity and genetic divergence. Observed variables are enclosed in boxes, whereas latent variables are enclosed in circles (geographical distance and environmental dissimilarity). Single-headed arrows show the observed variables used to infer each latent variable. Solid double-headed arrows indicate regression pathways, and dashed double-headed arrows indicate covariance pathways.

We incorporated a matrix permutation procedure for significance testing because of the distance matrix format of our data and assessed the overall goodness of fit for each model using the Akaike information criterion (AIC), which provides a metric for model selection by comparing the likelihoods of a set of models while penalising for their numbers of parameters. Lower AIC scores indicate closer fit to the true model, and models with AIC scores that exceed the lowest model by 10 or more are not supported (Burnham & Anderson 1998).

To test whether geographical distance and environmental dissimilarity contributed significantly to the model, we compared the AIC from the full model to the AICs from a model excluding the geographical variables and from one excluding the environmental variables. We performed the SEM analysis using models that tested three alternative hypotheses: (1) only geographical distance contributes significantly to genetic divergence, (2) only environmental dissimilarity contributes significantly to genetic divergence and (3) geographical distance and environmental dissimilarity each contribute significantly to genetic divergence (Fig. 2). We performed this analysis on mtDNA sequence divergence from all 17 species included in our study. We examined the results for each species individually (Table 1), and, to draw comparisons between islands, we also averaged the results of all species on each island.

Method validation

The development of methods for the study of spatial genetic variation is still very fluid, and universally suitable analytical frameworks are yet to emerge (Storfer *et al.* 2010). In general, evaluating a new method requires understanding its accuracy, its assumptions and its advantages and disadvantages. To confirm the consistency of our results with other similar methods, we performed two additional multiple regression-based analyses on our mtDNA data set. First, we performed generalised dissimilarity modelling (GDM; Ferrier *et al.* 2007), using the 'GDM' package in R (Ferrier *et al.* 2007). GDM allows for nonlinear relationships between predictor and response variables, but only currently implements direct distance (instead of resistance-based distances), because the geographical distance matrix is computed from raw GPS coordinates. Second, we performed a variation partitioning analysis (Legendre & Fortin 2010) using the 'vegan' package in R (Oksanen *et al.* 2007). Variation partitioning uses multiple regression to assign proportions of variation in a response variable to a set of explanatory variables, and canonical redundancy analysis to assess the significance of each proportion. Although similar in principal to our SEM method, there are notable differences. For instance, SEM uses latent variables, which allows for some variables to be inferred from others and for the relationships among all variables to be estimated simultaneously. Because GDM and variation partitioning do not use latent variables, for these analyses, we first reduced the number of predictor variables in our data set by performing a principal components analysis (PCA) on the 24 environmental variables using the 'prcomp' function in R. In addition, unlike GDM and variation partitioning, SEM estimates error around the maximum-likelihood estimate for each parameter and provides an information-theoretic output for assessing parameter importance. Thus, for our goals, SEM provides several distinct advantages over other similar methods. However, while SEM is powerful for isolating the effects of individual variables, it cannot be used to estimate the effects of variation shared among variables. For

our application, this means that we can use SEM to disentangle the effects of geographical and environmental distances on genetic divergence individually, but that we cannot quantify the effects of shared variation between geographical and environmental distances on genetic divergence, a drawback that also exists with GDM but not variation partitioning. Finally, an assumption of each methods is that inferences made using current data, both environmental and genetic, reflect the processes that produced genetic divergence. Thus, these analyses are most reliably applied to genetic markers that capture recent and ongoing processes and when the contemporary geography and environment of the study area has not changed substantially since patterns of genetic variation were established.

To test whether the form in which we used environmental variables affected the results, we repeated the SEM analysis for each species using environmental PC axes or topographic distances. To visualise the relationships between geographical distance, environmental dissimilarity and genetic divergence, we plotted genetic divergence vs. geographical distance, genetic divergence vs. environmental dissimilarity and geographical distance vs. environmental dissimilarity for each species (Fig. S1). For these plots, we averaged least-cost path and circuit distances to estimate geographical distance and calculated the distances between points for each population plotted against the two major environmental PC axes to estimate environmental dissimilarity.

Finally, to compare results from different molecular markers, we repeated the analysis on mtDNA and nDNA sequences together and on nDNA sequences alone from the three species (*A. allisoni*, *A. chlorocyanus* and *A. porcatus*) for which extensive nDNA sampling was available. Because nDNA data were not available from all localities for these species, analyses on nDNA alone and concatenated mtDNA and nDNA sequences were performed on a subset of the localities used in mtDNA analyses (Table S3). To allow comparisons between different molecular markers, we also reran the SEM analysis on mtDNA data alone from the subset of localities with nDNA data.

RESULTS

Genetic, geographical and environmental data

We assembled mtDNA sequence data sets for 17 widespread anole species and nDNA sequence data sets for three species. For our mtDNA data sets, we acquired an average sequence length of 1527 bp from 72.5 individuals from 21 localities per species. We found a wide range of variation (from < 0.5% between some localities to > 15% among others) with an average of 5.6% sequence divergence between localities within each species. The 24 environmental GIS layers used in our study indicated that each of the four Greater Antillean islands is environmentally heterogeneous, with each species typically being distributed across a broad range of environmental conditions (Fig. 1). The two cost-weighted distances we calculated (least-cost path and circuit distances) were highly correlated ($r^2 > 0.65$ in all cases) and roughly equally weighted in estimating a geographical distance latent variable (Table S1).

Structural equation modelling

The SEM analysis revealed that IBD explained a significant proportion of genetic divergence for 15 species, and IBE explained a significant proportion of genetic divergence in 13 species (Table 1; Fig. 3). In each case, the full model including both IBD and IBE was a significantly better fit to the data based on model-fit scores (AIC) than a model including only one or the other (Table S2). There was also, as expected, some covariation between geographical distance and environmental dissimilarity, but this was typically low (Mean = 15.2%; Table 1). However, for three species (*A. cooki*, *A. distichus ignigularis* and *A. equestris*), covariation between these variables was fairly high (> 0.3), and these results should be interpreted with caution. On average, IBD explained 36.3% of genetic divergence (range = 3.0–79.6%; Table 1; Fig. 3), whereas IBE explained 17.9% (range = 0.1–47.7%; Table 1; Fig. 3). We also identified the individual variables that contributed to making up the environmental dissimilarity latent variable in each species (Table 1; Table S1). Across islands, the average results were fairly consistent, except that IBE actually contributed more than IBD in the species from Hispaniola (0.173 ± 0.172 IBD vs. 0.226 ± 0.157 IBE), and IBE contributed very little in any of the species from Jamaica (0.510 ± 0.178 IBD vs. 0.004 ± 0.003 IBE). For species from Cuba (0.543 ± 0.169 IBD vs. 0.158 ± 0.076 IBE) and Puerto Rico (0.329 ± 0.146 IBD vs. 0.216 ± 0.154 IBE), IBD contributed 3.5 times and 1.5 times as much, respectively, as IBE to spatial genetic divergence.

Method validation

In general, results from GDM and variation partitioning, both of which have been recently described as powerful methods for studying genetic divergence (Ferrier *et al.* 2007; Thomassen *et al.* 2010; Legendre & Fortin 2010), were consistent with those inferred by SEM (Table S3). In addition, as expected, when we used simple topographic distances instead of resistance-based distances, the contribution from geography was lower, but the effects of environment were virtually unchanged (Table S3). When we performed a

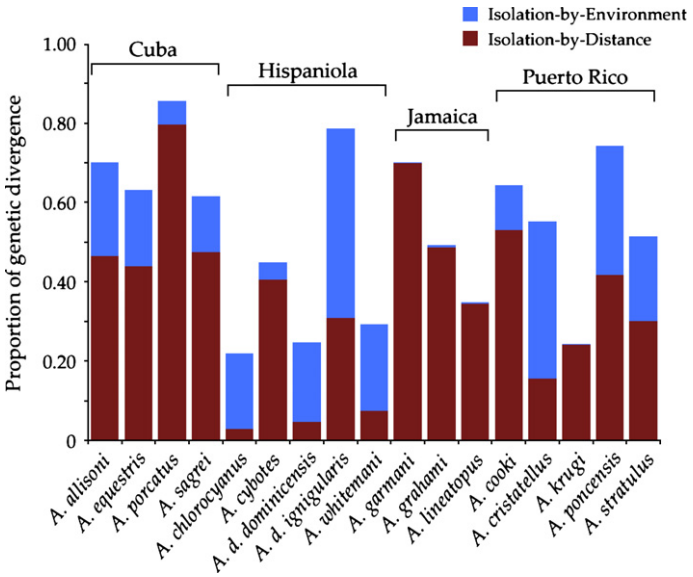

Figure 3 Proportions of spatial genetic divergence explained by isolation-by-distance (red) and isolation-by-environment (blue) based on SEM analysis. For each species, the height of the red bar is the proportion of genetic divergence explained by geographical distance (IBD), the height of the blue bar is the proportion explained by environmental dissimilarity (IBE) and the total height of the bar is the total proportion explained by both together.

PCA on the 24 environmental variables and entered the resulting PC axes as environmental dissimilarity variables – instead of simultaneously estimating the factor loadings for each variable when solving the model – the contribution from IBE was slightly lower, whereas the effects of IBD were basically the same (Table S3). Finally, SEM analysis of mtDNA and nDNA data produced concordant results in each case. For *A. porcatus*, analysis on nDNA data independently resulted in estimates that were very similar to those estimated from mtDNA alone for the strengths of IBD (67.3 vs. 77.6%) and IBE (8.8 vs. 5.7%) (Table 2). For *A. allisoni* and *A. chlorocyanus*, we did not find sufficient variation in nDNA sequences to permit convergence of the MLE during SEM analysis of nDNA alone. However, the addition of nDNA to the mtDNA data set resulted in only minor differences compared with mtDNA alone (Table 2).

DISCUSSION

Structural equation modelling

Disentangling the effects of IBD and IBE has been a major challenge in identifying the processes that drive spatial genetic divergence (Crispo *et al.* 2006; Thorpe *et al.* 2008; Wang & Summers 2010). We used a structural equation modelling approach to isolate and quantify the relative strengths of these effects, highlighting the importance of simultaneously considering the roles of both ecological and geographical factors in driving biological diversification. For most species, we found support for the hypothesis that both geography and ecology contribute to spatial genetic divergence, but, on average, IBD explained about twice as much genetic divergence as IBE (mean 36.3 vs. 17.9%; Table 1; Fig. 3). In most cases, these two factors together explained a large proportion of genetic variation (Fig. 3), suggesting that our SEM analysis captured the key underlying processes. Hence, although several recent studies have suggested a dominant role for IBE (e.g. Thorpe *et al.* 2008; Surget-Groba *et al.* 2012), our results clearly show that IBD still plays a key role in the evolution of spatial divergence.

There are, however, exceptions to this trend. IBE actually contributed more than IBD for four Hispaniolan taxa (*A. chlorocyanus*, *A. distichus domincensis*, *A. distichus ignigularis* and *A. whitemani*) and one Puerto Rican species (*A. cristatellus*). Meanwhile, for three Jamaican species (*A. garmani*, *A. grahami* and *A. lineatopus*), IBE played a very minor role, explaining an average of 0.4% of genetic divergence. The reasons for these differences are unclear. On one hand, overall environmental heterogeneity (variation averaged across all environmental variables after standardisation) was greatest on Hispaniola, and Hispaniola had the highest average contribution from environmental dissimilarity (22.6%), but on the other, environmental heterogeneity was not significantly different between the remaining three islands. Variation in environmental variables among sampling localities for each species was also not significantly different between Cuba, Jamaica and Puerto Rico, suggesting that the low contributions from IBE were not simply due to low levels of environmental variation.

In general, the results for each species within islands, except for Hispaniola, were fairly consistent, suggesting that closely related species have similar evolutionary responses to a particular landscape. We also found, though, that results were broadly consistent across islands as well, suggesting that how organisms interact with their environments and the underlying features of the landscape both play important roles. On average, and for 11 of 12 species on three of the four islands (Cuba, Jamaica, and Puerto Rico), IBD played a greater role than IBE. This result joins several recent studies in suggesting that the ubiquity of IBE may be overemphasised and that the important drivers of genetic divergence may vary substantially across different organisms (Crispo *et al.* 2006; Hendry 2009). In fact, in many cases, both factors may contribute to some extent to spatial genetic divergence (Lee & Mitchell-Olds 2011).

For those species for which IBE played a significant role, simultaneously estimating the contribution of each environmental GIS layer in the SEM allowed us to assess the degree to which individual environmental variables contributed to these results (Table 1; Table S1). Spatial variation in adaptations to thermal environment has been shown in some *Anolis* lizards (Hertz & Huey 1981), but morphological and genetic variation within some species seem to be more closely tied with spatial variation in precipitation or vegetation (Thorpe 2002; Calsbeek *et al.* 2006; Thorpe *et al.* 2008). For most species in our study, temperature variables played the biggest part in describing the environmental dissimilarity latent variable, but precipitation variables contribute significantly to environmental dissimilarity in eight species. Understanding how individual environmental variables can produce adaptive differences that lead to IBE remains a major challenge (Coyne & Orr 2004; Sobel *et al.* 2010), and our results demonstrate that doing so can contribute substantially to understanding the evolution of spatial genetic divergence.

Table 2 Results of structural equation modelling (SEM) analyses that included nuclear DNA sequence data (nDNA). Data used include the number of localities (Loc.) and the number (Seq.) and length in base pairs (bp) of DNA sequences. For each species, SEM was used to quantify the proportion of genetic divergence explained by isolation-by-distance (IBD) and isolation-by-environment (IBE), presented as maximum-likelihood estimates ± standard errors. Also listed are the primary contributors (Contrib. vars.) to the environmental dissimilarity latent variable (individual factor loading coefficients > 0.7). Results of analyses on mtDNA data alone are included for comparison

Study system		Data			Results			
Species	Marker	Loc.	Seq.	bp	IBD	IBE	Total	Contrib. vars.
Anolis allisoni	mtDNA	14	62	1173	0.448 ± 0.086	0.229 ± 0.091	0.677	Precip.
	mtDNA + nDNA	14	21	1674	0.437 ± 0.093	0.254 ± 0.116	0.691	Precip.
Anolis chlorocyanus	mtDNA	5	14	1504	0.028 ± 0.048	0.172 ± 0.110	0.200	Precip.
	mtDNA + nDNA	5	6	3266	0.056 ± 0.041	0.203 ± 0.125	0.259	Precip.
Anolis porcatus	mtDNA	15	41	1173	0.776 ± 0.067	0.057 ± 0.036	0.833	Precip.
	mtDNA + nDNA	15	26	2847	0.765 ± 0.078	0.064 ± 0.031	0.839	Precip.
	nDNA	15	26	1674	0.673 ± 0.180	0.088 ± 0.065	0.841	Precip.

Method validation

All three approaches we used to validate our novel SEM methodology produced results consistent with those from our method, confirming the robustness of our approach to different assumptions and sources of data. In each case, the slight differences we observed are in line with the differences expected based on previous studies of genetic divergence (McRae & Beier 2007; Wang *et al.* 2009). Our validation results also showed consistency between the results drawn from mtDNA data sets and those including nDNA data. So, we have some evidence that the inferences drawn from our mtDNA data sets should be good representations of the results from overall genetic variation. In any case, although there are potential drawbacks for using mtDNA-only data sets (or any single gene data set) for some analytical goals, these typically pertain to methods like phylogenetic inference, in which any single gene tree may present an 'incorrect' view of population history due to the random nature of lineage sorting. For this study, we used mtDNA sequences as indicators of population genetic structure and genetic divergence, rather than for inferring phylogenetic history, and these data should be robust for these purposes because mtDNA are known to be sensitive indicators of population structure (Zink & Barrowclough 2008; Karl *et al.* 2012). In addition, simulations have shown that unlinked neutral loci can consistently detect IBE when divergent selection between environments is strong and dispersal is intermediate (Thibert-Plante & Hendry 2010). We do not measure either selection or dispersal directly in this study, but previous work has shown that divergent selection on *Anolis* lizards can be strong (Calsbeek & Irschick 2007), and F_{ST} estimates from some species suggest moderate levels of dispersal (Ng & Glor 2011; Surget-Groba *et al.* 2012). Thus, because they typically provide more variable sites than nDNA sequences, mtDNA sequence data continue to be valuable for estimating population structure and relative levels of genetic divergence among populations (Zink & Barrowclough 2008), and for the conditions typical in *Anolis* lizards, they should also be reliable for detecting IBE. In any case, the SEM method we describe could easily be applied to many additional forms of genetic data, including microsatellite, SNP or genomic datasets, and future studies using these data should illuminate the factors influencing spatial genetic divergence even further.

While for this study, we examined patterns of gene flow through neutral genetic divergence, our approach could also easily be extended to examining variation in genes potentially under selection. Studies examining many genetic loci could look for outliers that have a significantly different signal from the background pattern of loci from across the genome. Those linked to genes under selection might be expected to exhibit greater degrees of ecological isolation because of divergent selection in different environments. Likewise, populations in similar environments should have less divergence in genes under selection than in neutral genes. Close examinations of the environmental variables contributing to these effects could prove especially valuable for linking the genes under selection to specific environmental forces of selection, and should provide insights into the geographical and ecological factors that contribute to adaptive genetic divergence.

CONCLUSIONS

Our results provide strong evidence that both IBD and IBE contributed to the evolution of spatial genetic divergence in the adaptive radiations of *Anolis* lizards on the Greater Antilles. IBD, in which physical distance and barriers reduce gene flow between populations, played the strongest role. IBE, in which adaptive differences or natural selection against dispersers limit gene flow between divergent habitats, played a secondary role. Experimental evidence has shown that anoles avoid microhabitats in which they have reduced performance (Irschick & Losos 1999; Calsbeek & Irschick 2007), and our results suggest that the same phenomenon may also occur among environmentally distinct macrohabitats. Overall, these results show that understanding the evolution of spatial genetic variation requires examining both ways in which ecological landscapes influence patterns of gene flow among populations.

ACKNOWLEDGEMENTS

We thank K. Boronow, S. Campbell-Staton, G. Gartner, A. Harrison, T. Ingram, A. Kamath, T. Moore, M. Munoz, E. Sherratt and Y. Stuart for their advice on this project. This research was supported by an NSF Postdoctoral Fellowship in Bioinformatics (to IJW) and an NSF collaborative grant (to REG and JBL).

AUTHOR CONTRIBUTIONS

IJW, REG and JBL conceived and designed the study. REG contributed new genetic and species distribution data. IJW contributed the new spatial statistical methods and analysed the data. The authors wrote the manuscript together.

REFERENCES

Burnham, K.P. & Anderson, D.R. (1998). *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York.

Calsbeek, R. & Irschick, D.J. (2007). The quick and the dead: correlational selection on morphology, performance, and habitat use in island lizards. *Evolution*, 61, 2493–2503.

Calsbeek, R., Knouft, J. & Smith, T. (2006). Variation in scale numbers is consistent with ecologically based natural selection acting within and between lizard species. *Evol. Ecol.*, 20, 377–394.

Coyne, J.A. & Orr, H.A. (2004). *Speciation*. Sinauer Associates, Sunderland, MA.

Crispo, E., Bentzen, P., Reznick, D.N., Kinnison, M.T. & Hendry, A.P. (2006). The relative influence of natural selection and geography on gene flow in guppies. *Mol. Ecol.*, 15, 49–62.

Cushman, S.A., McKelvey, K.S., Hayden, J. & Schwartz, M.K. (2006). Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *Am. Nat.*, 168, 486–499.

Dobzhansky, T.G. (1937). *Genetics and the Origin of Species*. Columbia University Press, New York.

Ferrier, S., Manion, G., Elith, J. & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Divers. Distrib.*, 13, 252–264.

Glor, R.E. & Laport, R.G. (2012). Are subspecies of *Anolis* lizards that differ in dewlap color and pattern also genetically distinct? A mitochondrial analysis. *Mol. Phylogenet. Evol.*, 64, 255–260.

Glor, R.E., Kolbe, J.J., Powell, R., Larson, A. & Losos, J.B. (2003). Phylogenetic analysis of ecological and morphological diversifiation in Hispaniolan trunk-ground anoles (*Anolis cybotes* group). *Evolution*, 57, 2383–2397.

Glor, R.E., Gifford, M.E., Larson, A., Losos, J.B., Schettino, L.R., Lara, A.R.C. *et al.* (2004). Partial island submergence and speciation in an adaptive radiation: a multilocus analysis of the Cuban green anoles. *Proc. R. Soc. Lond. B*, 271, 2257–2265.

Grace, J.B. (2006). *Structural Equation Modeling and Natural Systems*. Cambridge University Press, Cambridge, UK.

Hendry, A. (2009). Ecological speciation! Or the lack thereof?. *Can. J. Fish. Aquat. Sci.*, 66, 1383–1398.

Hertz, P.E. & Huey, R.B. (1981). Compensation for altitudinal changes in the thermal environment by some *Anolis* lizards on Hispaniola. *Ecology*, 62, 515–521.

Irschick, D.J. & Losos, J.B. (1999). Do lizards avoid habitats in which performance is submaximal? The relationship between sprinting capabilities and structural habitat use in Caribbean anoles. *Am. Nat.*, 154, 293–305.

Jezkova, T., Leal, M. & Rodríguez-Robles, J.A. (2009). Living together but remaining apart: comparative phylogeography of *Anolis poncensis* and *A. cooki*, two lizards endemic to the aridlands of Puerto Rico. *Biol. J. Linn. Soc.*, 96, 617–634.

Johnson, M.A., Revell, L.J. & Losos, J.B. (2009). Behavioral convergence and adaptive radiation: effects of habitat use on territorial behavior in *Anolis* lizards. *Evolution*, 64, 1151–1159.

Karl, S.A., Toonen, R.J., Grant, W.S. & Bowen, B.W. (2012). Common misconceptions in molecular ecology: echoes of the modern synthesis. *Mol. Ecol.*, 21, 4171–4189.

Kolbe, J.J., Glor, R.E., Schettino, L.R., Lara, A.C., Larson, A. & Losos, J.B. (2004). Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, 431, 177–181.

Kolbe, J.J., Glor, R.E., Schettino, L.R., Lara, A.C., Larson, A. & Losos, J.B. (2007). Multiple sources, admixture, and genetic variation in introduced *Anolis* lizard populations. *Conserv. Biol.*, 21, 1612–1625.

Lee, C.-R. & Mitchell-Olds, T. (2011). Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Mol. Ecol.*, 20, 4631–464.

Legendre, P. & Fortin, M. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Mol. Ecol. Resour.*, 10, 831–844.

Losos, J.B. & Ricklefs, R.E. (2009). Adaptation and diversification on islands. *Nature*, 457, 830–836.

Mayr, E. (1963). *Animal Species and Evolution*. Harvard University Press, Cambridge, MA.

McRae, B.H. & Beier, P. (2007). Circuit theory predicts gene flow in plant and animal populations. *Proc. Natl. Acad. Sci. U.S.A.*, 104, 19885–19890.

Ng, J. & Glor, R.E. (2011). Genetic differentiation among populations of a Hispaniolan trunk anole that exhibit geographical variation in dewlap colour. *Mol. Ecol.*, 20, 4302–4317.

Nosil, P. (2012). *Ecological Speciation*. Oxford University Press, New York.

Ogden, R. & Thorpe, R.S. (2002). Molecular evidence for ecological speciation in tropical habitats. *Proc. Natl. Acad. Sci. U.S.A.*, 99, 13612–13615.

Oksanen, J., Kindt, R., Legendre, P., Ohara, B. & Stevens, M.H.H. (2007). *Vegan: Community Ecology Package. R package version 1.8.8*. Available: http://r-forge.r-project.org/projects/vegan

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model.*, 190, 231–259.

Rodriguez-Robles, J.A., Jezkova, T. & Leal, M. (2010). Climatic stability and genetic divergence in the tropical insular lizard Anolis krugi, the Puerto Rican "Lagartijo Jardinero de la Montaña". *Mol. Ecol.*, 19, 1860–1876.

Rousset, F. (2000). Genetic differentiation between individuals. *J. Evol. Biol.*, 13, 58–62.

Santos, J.C. & Cannatella, D.C. (2011). Phenotypic integration emerges from aposematism and scale in poison frogs. *Proc. Natl Acad. Sci. USA*, 108, 6175–6180.

Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science*, 323, 737–741.

Schoener, T.W. & Schoener, A. (1982). Intraspecific variation in home-range size in some Anolis lizards. *Ecology*, 63, 809–823.

Sobel, J.M., Chen, G.F., Watt, L.R. & Schemske, D.W. (2010). The biology of speciation. *Evolution*, 64, 295–315.

Storfer, A., Murphy, M.A., Spear, S.F., Holderegger, R. & Waits, L.P. (2010). Landscape genetics: where are we now?. *Mol. Ecol.*, 19, 3496–3514.

Surget-Groba, Y., Johansson, H. & Thorpe, R.S. (2012). Synergy between allopatry and ecology in population differentiation and speciation. *Int. J. Ecol.*, 2012, 1–10.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, 24, 1596–1599.

Thibert-Plante, X. & Hendry, A.P. (2010). When can ecological speciation be detected with neutral loci?. *Mol. Ecol.*, 19, 2301–2314.

Thomassen, H.A., Buermann, W., Milá, B., Graham, C.H., Cameron, S.E., Schneider, C.J. *et al.* (2010). Modeling environmentally associated morphological and genetic variation in a rainforest bird, and its application to conservation prioritization. *Evolutionary Applications*, 3, 1–16.

Thorpe, R.S. (2002). Analysis of color spectra in comparative evolutionary studies: molecular phylogeny and habitat adaptation in the St. Vincent anole (*Anolis trinitatis*). *Syst. Biol.*, 51, 554–569.

Thorpe, R.S., Malhotra, A., Stenson, A.G. & Reardon, J.T. (2004). Adaptation and speciation in Lesser Antillean anoles. In: *Adaptive Speciation* (eds Dieckmann, U., Doebeli, M., Metz, J.A.J. & Tautz, D.). Cambridge University Press, Cambridge, U.K, pp. 322–344.

Thorpe, R.S., Surget-Groba, Y. & Johansson, H. (2008). The relative importance of ecology and geographic isolation for speciation in anoles. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 363, 3071–3081.

Wang, I.J. & Summers, K. (2010). Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Mol. Ecol.*, 19, 447–458.

Wang, Y.-H., Yang, K.-C., Bridgman, C. & Lin, L.-K. (2008). Habitat suitability modelling to correlate gene flow with landscape connectivity. *Landscape Ecol.*, 23, 989–1000.

Wang, I.J., Savage, W.K. & Bradley Shaffer, H. (2009). Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (*Ambystoma californiense*). *Mol. Ecol.*, 18, 1365–1374.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817–838.

Wright, S. (1921). Correlation and causation. *J. Agric. Res.*, 1921, 10–24.

Wright, S. (1943). Isolation by distance. *Genetics*, 28, 114–138.

Zink, R.M. & Barrowclough, G.F. (2008). Mitochondrial DNA under siege in avian phylogeography. *Mol. Ecol.*, 17, 2107–2121.

SUPPORTING INFORMATION

Additional Supporting Information may be downloaded via the online version of this article at Wiley Online Library (www.ecologyletters.com).

Editor, Frederick Adler
Manuscript received 20 June 2012
First decision made 18 July 2012
Second decision made 16 September 2012
Manuscript accepted 4 October 2012