Recent Secondary Contacts, Linked Selection, and Variable Recombination Rates Shape Genomic Diversity in the Model Species *Anolis carolinensis*

Yann Bourgeois¹,*, Robert P. Ruggiero¹, Joseph D. Manthey¹,², and Stéphane Boissinot¹,*

¹New York University Abu Dhabi, United Arab Emirates
²Department of Biological Sciences, Texas Tech University

*Corresponding authors: E-mails: yann.x.c.bourgeois@gmail.com; stephane.boissinot@nyu.edu.

Accepted: May 23, 2019

Data deposition: This project has been deposited to the Sequencing Read Archive under the Bioproject designation PRJNA533001.

Abstract

Gaining a better understanding on how selection and neutral processes affect genomic diversity is essential to gain better insights into the mechanisms driving adaptation and speciation. However, the evolutionary processes affecting variation at a genomic scale have not been investigated in most vertebrate lineages. Here, we present the first population genomics survey using whole genome resequencing in the green anole (*Anolis carolinensis*). Anoles have been intensively studied to understand mechanisms underlying adaptation and speciation. The green anole in particular is an important model to study genome evolution. We quantified how demography, recombination, and selection have led to the current genetic diversity of the green anole by using whole-genome resequencing of five genetic clusters covering the entire species range. The differentiation of green anole’s populations is consistent with a northward expansion from South Florida followed by genetic isolation and subsequent gene flow among adjacent genetic clusters. Dispersal out-of-Florida was accompanied by a drastic population bottleneck followed by a rapid population expansion. This event was accompanied by male-biased dispersal and/or selective sweeps on the X chromosome. We show that the interaction between linked selection and recombination is the main contributor to the genomic landscape of differentiation in the anole genome.

Key words: *Anolis carolinensis*, recombination, divergence, selection.

Introduction

Nucleotide variation along a DNA sequence results from the interactions between multiple processes that either generate new alleles (e.g., recombination, mutation) or affect the fate of these alleles in populations (e.g., selection, demography, and speciation). The variable outcome of these interactions along the genome can result in heterogeneous patterns of diversity and divergence at both intra- and interspecific scales (Begun and Aquadro 1992; Nachman and Payne 2012; Cruickshank and Hahn 2014; Roux et al. 2014; Seehausen et al. 2014; Wolf and Ellegrén 2017). Given their importance in divergence and speciation, quantifying these processes has been at the core of evolutionary genomics for the last decade. With the advent of next-generation sequencing and the continuous development of novel analytical tools, it has become possible to properly quantify the impact of recombination (Booker et al. 2017; Kawakami et al. 2017), selection (Begun and Aquadro 1992; Barrett et al. 2008; Mullen and Hoekstra 2008; Cai et al. 2009), and demographic history (Gutenkunst et al. 2009; Excoffier et al. 2013; Roux et al. 2016) on diversity patterns in several vertebrates. Ultimately, such investigations have the power to answer outstanding biological questions such as the role of sex chromosomes, the nature of reproductive barriers, or the timing of gene flow and selection during differentiation (Wolf and Ellegrén 2017). Assessing the effects of these mechanisms in a range of organisms is crucial to inform current debates about their relative importance (Bierne et al. 2011; Kern and Hahn 2018; Pouyet et al. 2018; Jensen et al. 2019) with a broader set of empirical data.

© The Author(s) 2019. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Thorough analyses of the factors affecting genome diversity at the peri-specific level have been performed in a small number of vertebrate species (see Ellegren et al. 2012; Sousa et al. 2013; Poelstra et al. 2014; Booker et al. 2017; Han et al. 2017; Kawakami et al. 2017) but several major clades of vertebrates have not been investigated at all. Among those are the nonavian reptiles, a speciose group of vertebrates that harbor a wide diversity of morphology and adaptation. Anoles in particular have been abundantly studied to understand the mechanisms underlying adaptation. This neotropical group of squamates diversified during the Cenozoic, and constitute a model system for understanding speciation and adaptation in ectotherms (Losos et al. 2004; Glor et al. 2005; Wade 2012; Campbell-Staton et al. 2012; Tollis et al. 2012; Tollis and Boissinot 2014; Manthey et al. 2016). The analysis of anoles genomes has also provided considerable insights on genome evolution in vertebrates (Alfoldi et al. 2011; Fujita et al. 2011; Tollis and Boissinot 2011, 2013; Figuet et al. 2015; Costantini et al. 2016; Ruggiero et al. 2017). At last, they display a dramatic physiological, morphological, and behavioral diversity (Lailvaux et al. 2004; Glor et al. 2005; Wade 2012; Campbell-Staton et al. 2018; Lapiedra et al. 2018). Given this importance as a model species, we decided to perform a study on genome-wide variation in the green anole (Anolis carolinensis) to better understand its microevolutionary dynamics, expanding previous genetic, and genomic work (Tollis et al. 2012; Tollis and Boissinot 2014; Manthey et al. 2016).

The green anole is the first nonavian reptile for which the whole genome was sequenced (Alfoldi et al. 2011) and its population structure is relatively well known (Tollis et al. 2012; Tollis and Boissinot 2014; Campbell-Staton et al. 2016; Manthey et al. 2016; Ruggiero et al. 2017). Whole-genome resequencing of green anoles populations would be an opportunity to better understand the drivers and constraints that act on their radiation at a resolution that was not allowed by the previous genetic data sets.

The green anole colonized Florida from Cuba (Glor et al. 2005; Campbell-Staton et al. 2012; Tollis et al. 2012; Tollis and Boissinot 2014; Manthey et al. 2016) between 6 and 12 Ma and diversified into five genetic groups: South Florida (SF), Eastern Florida (EF), Western Florida (WF), Gulf Atlantic (GA), and Carolinas (CA). Populations in Florida likely diverged in allopatry on island refugia before coming back into contact due to sea-level oscillations during the Pleistocene. Colonization of the rest of North America seems to be more recent, with two clades having probably expanded in the last 500,000 years (Manthey et al. 2016). It is the only species in the Anolis genus to have colonized temperate climates without human intervention.

Here, we present results obtained from whole-genome resequencing of five genetic clusters of the green anole. We provide a detailed assessment of the multiple factors that are likely to impact the green anole’s genetic diversity at a genome-wide scale. We demonstrate that the combined effects of regional variation in recombination rate, linked selection, and migration are responsible for the heterogeneous genomic landscape of diversity and divergence in the green anole.

Materials and Methods

DNA Extraction and Whole Genome Sequencing

Whole genome sequencing libraries were generated from A. carolinensis liver tissue samples collected between 2009 and 2011 (Tollis et al. 2012), and Anolis porcatus and Anolis allisoni tissue samples generously provided by Breda Zimkus from the Museum of Comparative Zoology at Harvard University. For each of the 29 samples, DNA was isolated from ethanol preserved tissue using Ampure beads per the manufacturers’ protocol. Illumina TRU-Seq paired end libraries were generated using 200 ng of DNA per sample and sequenced at the NYUAD Center for Genomics And Systems Biology Sequencing Core (http://nyuad.nyu.edu/en/research/ infrastructure-and-support/core-technology-platforms.html; Last accessed on June 7, 2019) with an Illumina HiSeq 2500. Read quality was assessed with FastQCv0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc; Last accessed on June 7, 2019) and Trimmomatic (Bolger et al. 2014) was subsequently used to remove low quality bases, sequencing adapter contamination and systematic base calling errors. Specifically, the parameters “trimmomatic_adapter.fa: 2:30:10 TRAILING:3 LEADING:3 SLIDINGWINDOW:4:15 MINLEN:36” were used. Samples had an average of 1,519,339,234 read pairs, and after quality trimming 93.3% were retained as paired reads and 6.3% were retained as single reads. Sequencing data from this study have been submitted to the Sequencing Read Archive (https://www.ncbi.nlm.nih.gov/sra; last accessed: June 7, 2019) under the BioProject designation PRJNA533001.

Sequence Alignment and SNP Calling

Quality trimmed reads were aligned to the May 2010 assembly of the A. carolinensis reference genome (Broad AnoCar2.0/anoCar2; GCA_000090745.1; Alfoldi et al. 2011) and processed for SNP detection with the assistance of the NYUAD Bioinformatics Core, using NYUAD variant calling pipeline (last accessed in June 2017). Briefly, the quality-trimmed FastQ reads of each sample were aligned to the AnoCar2.0 genome using the BWA-mem short read alignment approach (Li and Durbin 2011) and resulting SAM files were converted into BAM format, sorted, and indexed using SAMtools (Li et al. 2009). Picard was then used to identify insertions, deletions, and duplications in the sorted BAM files (http://broadinstitute.github.io/picard/; Last accessed on June 7, 2019) and evaluated using SAMtools (stats and depth). Alignments contained an average of 204,459,544 reads that passed QC, 97.75% mapping and 91.93% properly
paired (supplementary table S1, Supplementary Material online). Each individual resequenced genome was then processed with GATK for indel realignment, SNP and indel discovery, and genotyping, following GATK Best Practices (Depristo et al. 2011; Van Der Auwera et al. 2014). GATK joint genotyping was conducted with HaplotypeCaller for increased sensitivity and confidence, and results were selectively compared with results generated from SAMtools mpileup (Li et al. 2009). Filtering was performed in VCFtools (Danecek et al. 2011), with the following criteria: a 6× minimum depth of coverage per individual, a 15× maximum average depth of coverage, no more than 40% missing data across all 29 samples, a minimum quality score of 20 per site, and a minimum genotype quality score of 20.

Population Structure
To assess genetic structure, we conducted a clustering analysis using discriminant analysis of principal components (DAPC) on a subset of ~6,500 SNPs with <20% missing data and randomly thinned every 10 kb to minimize linkage disequilibrium (LD) between markers while retaining enough variants for inference. DAPC (Jombart et al. 2010) first estimates principal components (PC) describing variance in SNP data sets, then performs a discriminant analysis on these PC axes to identify genetic groupings. We selected the clustering model with the highest support using the Bayesian Information Criterion (BIC). We retained two principal components that explained ~40% of the total variance, and two of the linear discriminants. The probability for each individual to be assigned to a specific cluster was summarized by a barplot with the function compoplot() provided with the DAPC R package. We also described relationships between individuals with the same data set using the network algorithm implemented in Splitstree v4 (Huson and Bryant 2006). Lastly, we filtered the entire SNP data set to include one million randomly sampled SNPs present in a minimum of 80% of the individuals for use as input in RAxML v8 (Stamatakis 2014). We used RAxML to create a maximum-likelihood phylogeny, using the GTRGAMMA model of sequence evolution, and 100 rapid bootstraps to assess support for the phylogeny with the highest likelihood.

We further examined patterns of diversity and the shape of the allele frequency spectrum in each cluster by computing two summary statistics, the average number of pairwise differences \(D_{st}\) (or nucleotide diversity) per bp, and Tajima’s \(D\), for nonoverlapping 5-kb windows using the software POGENOME (Pfeifer et al. 2014). We removed windows overlapping ambiguities in the green anole genome using BEDTOOLS v2.25.0 (Quinlan and Hall 2010).

Effective Sex-Ratio
Sex-biased contribution to the gene pool is a critical aspect of demographic dynamics and is often impacted by variation in social structure between populations. We used the algorithm implemented in KIMTREE (Gautier and Vitalis 2013; Clemente et al. 2018) to estimate branch lengths from our SNP data set and infer the effective sex-ratios (ESR) for each of the five genetic clusters. This method is robust to LD, small sample sizes, and demographic events such as bottlenecks and expansions. To increase the number of usable markers, and since the authors recommend working with recently diverged populations, we focused on the recent northward colonization, and included individuals from the East Florida, Gulf Atlantic, and Carolinas genetic clusters.

Briefly, the method builds a hierarchical Bayesian model to estimate the evolution of SNP frequencies along branches of a population tree provided by the user. Genetic drift along branches is estimated by a time-dependent diffusion approximation. In this framework, branch length \(\tau\) is proportional to the time since divergence in generations (\(t\)) scaled by the effective population size (\(N_e\)), such that \(\tau = \tau_0\). The method can jointly contrast allele frequencies between autosomal and sex-linked markers to estimate the relative contribution of males and females to each generation (ESR). The ESR can then be seen as a comparison of the effective population sizes estimates obtained from autosomes and the X chromosome.

We sexed individuals by taking advantage of the expected relationship between depths of coverage at autosomal and
sex-linked loci in males and females. Since females are XX and males XY, the latter are expected to display a two-times lower coverage at X-linked sites compared with autosomal loci (supplementary fig. 1, Supplementary Material online). We then adjusted allele frequencies for all X-linked scaffolds, including Linkage Group b (Alföldi et al. 2011) and several scaffolds (GL343282, GL343364, GL343550, GL343423, GL343913, GL343947, GL343338, GL343417) recently identified as belonging to the green anole’s sex chromosome (Rupp et al. 2017). We counted one haplotype per male and two per female. To obtain confidence intervals over ESR estimates, we generated 50 pseudoreplicated data sets by randomly sampling 5,000 autosomal and 5,000 sex-linked SNPs with no missing data. The algorithm was started with 25 pilot runs of 1,000 iterations each to adjust the parameters of the Monte Carlo Markov Chain (MCMC). The MCMC itself was run for 100,000 generations and sampled every 25 iterations after a burn-in of 50,000 iterations. Convergence for all parameters was assessed by visually inspecting posterior sampling in R (R Core team 2016). For each replicate i, we estimated the support for biased sex-ratio (Si) such as:

\[
S_i = 1 - 2 |p_i - 0.5 |
\]

with \( S_i < 0.05 \) being interpreted as a strong support for biased sex-ratio and where \( p_i \) is the proportion of posterior MCMC samples with an ESR > 0.5.

Model Comparison of Demographic Scenarios

None of the previous population genetics studies of green anoles has ever precisely quantified the strength nor the timing of gene flow between genetic clusters. We addressed this issue by comparing different demographic scenarios for two pairs of sister clades (EF and GA; EF and WF) that included the most individuals (at least 11). We used the diffusion approximation-based likelihood approach implemented in the \( \delta \alpha \delta \) software (Gutenkunst et al. 2009). We compared a set of scenarios of strict isolation (SI), isolation with migration (IM), ancient migration (AM) with one or two (PAM) periods of gene flow and secondary contact (SC) with one or two (PSC) periods of gene flow (see Christe et al. 2017 for a detailed summary). We added complexity to this set of basic scenarios by allowing for a combination of population expansion (prefix “ex”), heterogeneous asymmetric migration rates (suffix “2M2P”) and heterogeneous effective population size (suffix “2N”) among loci. These additions were made to incorporate the genome-wide effects of selection on linked neutral sites (so-called “linked selection”) and model genomic islands resisting gene flow (Cruickshank and Hahn 2014). We also tested scenarios with both asymmetric migration rates and heterogeneous population sizes but were unable to reach convergence. Overall, we compared 34 scenarios combining these features, using a set of scripts available on dryad (Christe et al. 2017) and a modified version of \( \delta \alpha \delta \) (v1.7.0) kindly provided by Christelle Fraisse (available at http://methodsopangen.com/wp-content/uploads/2017/12/dadi-1.7.0_modif.zip; Last accessed on June 7, 2019). We extracted for each pairwise comparison a set of ~12,000 SNPs with no missing data and thinned every 100,000 bp to meet the requirement of independence among loci that is needed to properly compare the composite likelihoods estimated by \( \delta \alpha \delta \). We extracted the unfolded joint sites frequency spectra (SFS) by polarizing alleles using A. porcatus and A. allisoni as references. We considered ancestral the allele found at a minimal frequency of 75% in those two individuals or found fixed in one of them if the other individual was missing. We note that the \( \delta \alpha \delta \) models include a parameter (O) estimating the proportion of correctly polarized sites. We evaluated each model 30 times and retained the replicate with the highest likelihood for model comparison. Models were compared using the Akaikes information criterion (AIC). For the best model, we calculated uncertainties over the estimated parameters using a nonparametric bootstrap procedure, creating 100 pseudo-observed data sets by resampling with replacement from the SFS. We used the procedure implemented in the dadi.Godambe.GIM_uncert() script to obtain a maximum-likelihood estimate of 95% confidence intervals (Coffman et al. 2016). \( \delta \alpha \delta \) parameters are scaled by the ancestral population size \( N_{\text{eq}} \). For the sake of comparison with SMC++, estimates, parameters were converted into demographic units by estimating the ancestral effective population size as the harmonic mean of the SMC++ estimates before splitting time for all pairs of populations.

Estimating Recombination Rates

We used the LDHat software (McVean et al. 2002) to estimate effective recombination rates \( \rho = 4N_{\text{r}}r \) with \( r \) the recombination rate per generation and \( N \) the effective population size along the green anole genome. This method has been successfully used to obtain recombination maps for data sets similar to ours in terms of sequencing depth and sample sizes (Auton et al. 2012). Unphased genotypes were converted into LDHat format using VCFTools (option –ldhat). Since LDHat assumes that samples are drawn from a panmictic population, we focused on the Eastern Florida clade for which sampling effort was the highest (n = 8 diploid individuals). We used precomputed likelihood lookup tables with an effective population mutation rate (\( \theta \)) of 0.001, which was the closest from the \( \theta \) value estimated from our data set (\( \theta \approx 0.004 \)) and used the lkgen module to generate a table fitting the number of observed samples (16 chromosomes). Recombination rates were estimated over 500-kb windows with 100-kb overlaps using the Bayesian reversible MCMC scheme implemented in the interval module. The chain was run for 1,000,000 iterations and sampled every 5,000 iterations with a large block
penalty of 20 to avoid overfitting and minimize random noise. The first 100,000 generations were discarded as burn-in. Convergence under these parameters was confirmed by visually inspecting MCMC traces for a subset of windows. We averaged $\rho$ estimates over nonoverlapping 100-kb windows, or over coding sequences for subsequent analyses.

Summary Statistics for Differentiation and LD
To assess whether the joint effects of selection and low recombination on diversity and differentiation, we computed two measures of divergence (\(F_{ST}\) and \(d_{XY}\)) over nonoverlapping 100-kb windows for the three divergent Floridian lineages. These lineages were chosen because of their relative demographic stability (see Results). We picked 100,000 bp to reduce spatial autocorrelation between statistics of adjacent windows, since no further substantial LD decay could be observed over this distance for the two populations with the largest sample sizes (supplementary fig. 2, Supplementary Material online), pairwise LD (measured as \(r^2\)) was computed using VCFTools (Danecek et al. 2011). Comparison between those two statistics for a given genomic region has been proposed as a way to disentangle the effects of gene flow and selection (Cruickshank and Hahn 2014). For the sake of simplicity, correlations between differentiation statistics and recombination were estimated using a Spearman’s correlation test in R, although we note that measurements cannot be fully considered independent. As a sanity check, we computed the ZZ statistics (Rozas et al. 2001) to assess whether LDHat estimates of $\rho$ were consistent with the genomic distribution of LD. This statistic contrasts LD between adjacent pairs of SNPs to LD calculated over all pairwise comparisons in a given window. High values are suggestive of increased intragenic recombination. We also computed the average frequency of polymorphic derived alleles (average DAF) in the EF cluster since it included the most individuals, using \(A.\ porcatus\) and \(A.\ allisoni\) to polarize alleles, and excluding sites with at least six individuals genotyped over eight. DAF has been recently used to estimate the effects of linked selection in humans, an excess of ancestral alleles being expected in regions under the influence of background selection (Pouyet et al. 2018).

Because $\rho$ is the effective recombination rate and depends on the effective population size, it is directly correlated to local reduction of diversity due to linked selection. Thus, a low value of $\rho$ may be observed in regions linked to selection even if $r$ itself is not significantly different from the rest of the genome. To reduce this correlation, we calculated $\rho/\theta_x$, with $\theta_x$ the nucleotide diversity computed in POPOGENOME. Since $\theta_x$ is an estimator of $4\mu x$, with $\mu$ the mutation rate, this statistic represents the ratio between $\rho$ and the mutation rate $\mu$ (see Wang et al. 2016 for an example).

We examined the average DAF across five quantiles of $\rho/\theta_x$ to assess whether lower rates of recombination were associated with changes in the frequency of derived alleles that may be due to linked selection. We also compared the average DAF between the genomic background and regions of high relative and low absolute divergence which may be candidates for stronger linked selection. These regions were defined as regions belonging both to the top 20% quantile for \(F_{ST}\) and the lowest 20% quantile for $d_{XY}$.

Results
Statistics for Whole Genome Resequencing
Twenty-seven green anoles (\(A.\ carolinensis\)) sampled across the species’ range and covering the five genetic clusters identified in previous analyses (Tollis and Boissinot 2014; Manthey et al. 2016) were chosen for whole-genome resequencing. We also included two samples from the closely related species \(A.\ porcatus\) and \(A.\ allisoni\) as outgroups. Sequencing depth was between 7.22× and 16.74×, with an average depth of 11.45× (supplementary table S1, Supplementary Material online). About 74,920,333 variants with <40% missing data were retained after the first round of filtering (Materials and Methods).

Population Structure and Nucleotide Variation Reveal a Reduced Diversity in Northern Populations
We first assessed geographic structure across the distribution of \(A.\ carolinensis\). To this end, we used >6,500 SNPs thinned every 10 kb and with <20% missing data and identified $k = 5$ as the most likely number of genetic clusters with DAPC (fig. 1A and supplementary fig. 3, Supplementary Material online). Three groups were identified within Florida, while individuals from the rest of North-America were assigned to two clusters. These groups were consistent with the clusters identified in previous genetic studies (Tollis et al. 2012; Tollis and Boissinot 2014; Manthey et al. 2016). Possible introgression from Carolinas was observed for two Gulf Atlantic individuals (fig. 1A). A maximum-likelihood phylogeny estimated in RAxML based on one million random SNPs and a network analysis of relatedness in Splitstree further supported this clustering (fig. 1B and C). Results closely matched previous findings, with South Florida (SF) being the sister clade of all other groups. The two northernmost clusters, Gulf Atlantic (GA) and Carolinas (CA), clustered together in the RAxML phylogeny. Eastern Florida (EF) constituted a paraphyletic group in the phylogeny in which GA and CA were nested. This is likely due to incomplete lineage sorting induced by the high and constant effective population sizes of populations from Florida (see below), or to ongoing or recent gene flow resulting in the inclusion of loci with different coalescence times. At last, the Western Florida (WF) cluster was basal to all other groups except South Florida (SF).

Nucleotide diversity was the lowest in GA and CA (table 1) despite the large geographic area covered by these two
genetic clusters. Average Tajima’s $D$ values ranged between $-0.8$ (WF) and $0.14$ (GA). Positive Tajima’s $D$ values suggest recent population contraction, while negative Tajima’s $D$ are expected in the case of recent population expansion (Tajima 1989), although both population substructure and linked selection may impact it. Northern clusters (CA and GA) displayed the highest average Tajima’s $D$, possibly due to reductions in population sizes and relaxation of linked selection (see below).

Recent Population Expansion and Male-Biased Sex-Ratios in Northern Populations

We then reconstructed the demographic history of each genetic cluster. To this end, we used the whole set of filtered SNPs with <40% missing data to infer past changes in effective population sizes ($N_e$) without any a priori demographic model with SMC++ (fig. 2A). All populations from Florida showed rather stable demographic trajectories, with some evidence for population expansion in EF and WF. Assuming a mutation rate of $2.1 \times 10^{-10}$bp per year (Tollis and Boissinot 2014), present population sizes were in the range of 500,000 to 5,000,000 individuals for each population, in accordance with previous analyses based on target capture markers (Manthey et al. 2016). Northern populations (CA and GA) showed a clear signature of expansion starting between 200,000 and 100,000 years ago, following a bottleneck that started between 500,000 and 1,000,000 years in the past. We also estimated the splitting times between the different groups but since this model assumes no gene flow after the split, the estimates are likely to be biased toward the present. The split between GA and CA occurred shortly before these populations expanded, in accordance with the previously proposed hypothesis of double colonization following the Gulf and Atlantic coasts (Tollis and Boissinot 2014). In Florida, divergence events took place between 3 and 2 Ma.
The relative order of splitting events was consistent with the topology obtained from our phylogeny and previous studies. Since anoles exhibit sex differences in dispersal (Johansson et al. 2008), we tested whether the recent colonization of new and possibly suboptimal habitats could lead to unequal contribution of males and females to the gene pool or selection at sex-linked loci (fig. 2B). We built a population tree and quantified genetic signatures of biased sex-ratio with the algorithm implemented in KIMTREE. We focused on the three populations that diverged most recently, GA, CA, and EF. Note that the length of branch $i$ ($t_i$) represents time in generations ($t$) scaled by the effective population size for this branch such as $t_i = t/2Ne_i$ (Clemente et al. 2018). Branch lengths were particularly high for the CA and GA lineages compared with EF, as expected in the case of stronger drift (fig. 2B). This is in line with their smaller effective population sizes and the bottleneck inferred by SMC+++. We found evidence for a strongly male-biased ESR in CA and GA, but not in EF which was slightly female-biased. Indeed, nucleotide diversity was substantially more reduced at sex-linked scaffolds in GA than in EF when compared with autosomal diversity (supplementary fig. 4, Supplementary Material online). Note that sex-ratios are the proportion of females effectively contributing to the gene pool along each branch of the tree and should not be interpreted directly in terms of census size. The GA cluster displayed the strongest bias, with an estimated ratio of less than one female for 100 males, suggesting strong sex-bias in the founding population or strong male-biased dispersal during population expansion. The CA cluster and the inner branch leading to CA and GA showed a ratio of approximately 10 females for 100 males. All 50 replicates displayed a high support for a male-biased sex-ratio in CA and GA, while only five replicates supported a female-biased sex-ratio in EF (i.e., the Markov chain almost systematically explored sex-ratios $>0.5$ in only five replicates).

**Gene Flow at SC Has Homogenized Green Anole Populations**

We tested whether gene flow and its interruption may have played a role in shaping the genomic landscape of differentiation in green anoles (fig. 3). We focused on two pairs of genetic clusters. The first comparison was between the EF and WF clusters, which are two populations with high and stable population sizes (according to SMC+++, that both live in subtropical Florida. This comparison should be suited to detect the long-term effects of interrupted gene flow since there is already evidence that these clades may have been isolated by sea rising during interglacial periods (Tollis and Boissinot 2014; Manthey et al. 2016). In addition, since populations from Florida have remained relatively stable over the last 2 Myr, our power to detect the expected correlations between recombination and diversity should be enhanced in the case of linked selection (Burri 2017; Torres et al.)
The second comparison was between the EF and GA clusters, the latter corresponding to a recent expansion northward and adaptation to temperate environments.

We ran $\Delta a/i$ models to assess the effects of differential gene flow and linked selection. We compared a set of 34 divergence scenarios, allowing gene flow and effective population sizes to vary with time and across loci. Briefly, heterogeneity in gene flow (suffix 2M2P) was implemented by dividing the site frequency spectrum into three sets of loci with proportions $1-P_{WF}$, $P_{1}$, and $P_{2}$. Assuming that the first population is WF and the second EF, the first set of loci ($1-P_{WF}-P_{EF}$) is modeled with all parameters from the base model. The two other sets are modeled with the same parameters but with effective population sizes reduced by a background selection factor ($bf$).

Strict-isolation models (SI) consistently displayed the lowest likelihood and highest AIC, clearly supporting a role for gene flow in homogenizing green anoles genomes. For the comparison between EF and WF, models including heterogeneous population sizes performed better than models with heterogeneous gene flow. Among scenarios with gene flow, SC with one and two periods of gene flow (SC and PSC) often received the highest support (fig. 3B and supplementary fig. 5, Supplementary Material online). Parameters estimated from the best models are shown in table 2. There was no substantial gain in likelihood when adding expansion to scenario of two SCs with background selection (PSC2N), and models with heterogeneous migration displayed lower likelihood. The PSC2N model supported a scenario where about $nr = 65\%$ of the genome was affected by selection at linked sites, suggesting a rather large effect of low recombination and purifying selection on diversity. These Eastern and Western Floridian genetic clusters experienced long periods of isolation lasting $\approx 24$ Myr, followed by periods of SC lasting $\approx 125,000$ years in total.

For the comparison between GA and EF, we confirmed the decrease in effective population size detected by SMC++ in GA compared with EF, with a present effective population size of $N_{WF} = 10,000$ and $N_{EF} = 5,000$. The decrease is not due to ancient gene flow, but rather to demographic processes, such as a smaller effective population size in GA compared with EF. However, the genetic differentiation between GA and EF is not due to low recombination, but rather to purifying selection acting on nonrecombinant sites.

**Fig. 3.**—(A) Graphic description of the six categories of $\Delta a/i$ models tested over pairs of green anole genetic clusters. Each model describes a scenario where two populations diverge from an ancestral one, with varying timing and strength of gene flow after their split. SI, strict isolation; AM, ancestral migration, where populations first exchange gene flow then stops $T_{iso}$ generations ago; PAM, ancestral migration with two periods of contact lasting $T_{iso}/2$ generations; SC, secondary contact where populations still exchange gene flow at present time; PSC, secondary contact with two periods of contact lasting $T_{sc}/2$ generations; IM, isolation with constant migration and no interruption of gene flow. Models were constrained so that $T_{iso}$ and $T_{sc}$ lasted at least $\approx 50,000$ years. Reproduced with the authorization of Christelle Fraisse. (B) Fitting of the best models for the EF ($N = 16$) versus GA ($N = 14$) and EF versus WF ($N = 8$) comparisons. Both models fit the observed data sets as indicated by the similar spectra between observation and simulation. The “2N” suffix means that background selection was added to the base model by modeling heterogeneous effective population sizes across loci. The “2M2P” suffix means that heterogeneity in gene flow was incorporated into the model. The “ex” suffix means that exponential population size change was introduced in the base model.
## Summary of Best-Supported Demographic Models

| Comparison | Model | Na12 | Na1 | Na2 | N1 | N2 | m2-1 | m1-2 | Tiso | Tsc | Tscg | O | logLikelihood | AIC |
|------------|-------|------|-----|-----|----|----|------|------|------|-----|-----|----|----------------|-----|
| GA vs. EF  | IM2M2Pex | 2,156,641 | 3,538,759 | 2,438,866 | 371,048 | 7,132,328 | 2.42E-07 | 2.57E-07 | NA   | NA  | NA  | NA | -1,045.99 | 2,113.97 |
| GA vs. EF  | IMex  | 2,156,641 | 5,971,380 | 2,137,291 | 364,329 | 6,755,470 | 1.93E-07 | 2.36E-07 | 612,167** | 1,962,652 | NA | NA | 0.97 | -1,048.14 | 2,114.29 |
| GA vs. EF  | SC2M2Pex | 2,156,641 | 4,214,885 | 2,296,060 | 367,662 | 7,049,892 | 2.12E-07 | 2.58E-07 | 139,258** | 229,611* | 1,723,615 | 0.91 | NA | 0.97 | -1,046.01 | 2,116.01 |
| GA vs. EF  | PSNex | 2,156,641 | 4,126,463 | 2,402,541 | 367,459 | 6,926,709 | 1.93E-07 | 2.35E-07 | 58,466** | 110,560 | 1,713,849 | NA | NA | 0.97 | -1,048.03 | 2,116.06 |
| GA vs. EF  | Sce  | 2,156,641 | 4,150,102 | 2,359,593 | 369,470 | 7,034,978 | 1.91E-07 | 2.36E-07 | 97,752** | 214,565* | 1,672,998 | 0.93 | NA | 0.97 | -1,046.12 | 2,116.24 |
| GA vs. EF  | IM2Nex  | 2,156,641 | 4,102,102 | 2,359,593 | 369,470 | 6,926,709 | 1.93E-07 | 2.36E-07 | 97,752** | 214,565* | 1,672,998 | 0.93 | NA | 0.97 | -1,046.12 | 2,116.24 |
| EF vs.WF  | PSC2N | 2,091,300 | NA | NA | 5,181,876 | 5,389,218 | 1.94E-06 | 6.94E-07 | NA | NA | NA | NA | 0.97 | 0.98 | 0.57 | 0.26 | 0.97 |
| EF vs.WF  | SC2N | 2,091,305 | NA | NA | 1.18E-06 | 4.25E-07 | 1,952,690 | 144,821* | NA | NA | NA | NA | 0.57 | 0.92 | 0.25 | 0.98 | 0.97 |

---

### Notes
- **PSC2N**, secondary contact with two periods in isolation and heterogeneous effective population sizes across the genome; **SCex**, secondary contact with an episode of population expansion following secondary contact; **SC2M2P** and **IM2M2P**, models of secondary contact and constant gene flow with heterogeneous migration rates along the genome. The ancestral size (\(N_2\)) before the split was calculated from the SMC

---

### Acknowledgments
This study was supported by the National Science Foundation (DEB-1256095) and the Pennsylvania State University. We thank the anonymous reviewers for their helpful comments on an earlier version of this manuscript.
correlation between recombination rate and the average frequency of derived alleles (DAF) in the EF cluster (Spearman’s rank correlation test, $\rho = 0.11$, $P$ value < 0.001), although we did observe an increase in the average DAF for very low recombination rates that may be due to linked positive selection (supplementary fig. 9, Supplementary Material online). We did not observe any significant shift toward an excess of derived variants in regions of high divergence and low diversity (supplementary fig. 10, Supplementary Material online).

Given the possible effects of linked selection on diversity, we performed a SMC++ analysis masking 100-kb regions in the top 33% or low 33% $q/hr$. We did not observe any significant shift toward an excess of derived variants in regions of high divergence and low diversity (supplementary fig. 10, Supplementary Material online).

Discussion

In this study, we investigated the processes responsible for genomic differentiation and variation in the green anole. We showed that a complex history of recent expansion and SCs associated with linked selection shaped the genomic landscape of differentiation and diversity. Our results provide an important assessment of the forces acting on the green anole genome.

A Dynamic Demographic History Has Shaped the Genomic Landscape of Differentiation

Green anole populations are strongly structured and it was hypothesized that successive splits and SC occurred in Florida during the Pleistocene (Tollis and Boissinot 2014; Manthey et al. 2016). Fluctuations in sea level may have generated temporary islands on which isolated populations could have diverged. At last, reconnection of Florida to the mainland would have provided the opportunity for expansion northward (Solis et al. 2006). Our results support this claim in three ways. First, splitting times estimated by SMC++ and $\delta a\delta i$ suggest a series of splits in Florida between 3 and 2 Ma, a time range during which successions of glacial and interglacial periods may have led to several vicariance events (Lane 1994; Petuch 2004). Second, the models receiving the highest support in $\delta a\delta i$ were the ones allowing for several events of isolation followed by SC in Florida. Third, we found clear signatures of population expansion in GA and CA at the beginning of the Late Pleistocene, a time when lowering sea levels would have facilitated colonization (Lane 1994; Petuch 2004). We acknowledge that the exact timing of these events depends on the mutation rate used, which was previously established based on rates of divergence for three intronic nuclear markers compared with a mitochondrial one (Tollis and Boissinot 2014). The mutation rate used here is in the lower range of what is expected for nuclear markers.
In Anolis roquet, male-biased dispersal is associated with competition, since males disperse more when density increases and competition for females is stronger (Johansson et al. 2008). In Anolis sagrei, smaller males tend to disperse more while females are more likely to stay in high quality territories, independently of female density (Calibek 2009). The green anole is a polygynous species, with sexual dimorphism and high levels of competition between males (Jenssen et al. 2000). It is therefore likely that competition within sexes may lead to unequal contribution of males and females to the gene pool.

Another nonexclusive possibility lies in the action of positive selection on the X chromosome in northern populations. The X chromosome is extremely small compared with autosomes in green anoles, probably not exceeding 20 Mb (Rupp et al. 2017). This means that even a few recent selective sweeps could have widespread effects on the entire chromosome, reducing diversity and the effective population size. Since the method implemented in KIMTREE compares estimates of effective population sizes between autosomes and X chromosome, this would result in an artificially biased sex-ratio. Sexual or natural selection may be responsible for this pattern, and our finding calls for further comparisons of sex-biased dispersal and behavior between populations of the green anole.

Selection and Recombination Shape Nucleotide Composition and Diversity at Linked Sites

We observed strong heterogeneity in recombination rates along the green anole genome. Our results show that this heterogenous recombination landscape plays an important role in shaping genetic diversity in anoles. Both purifying and positive selections are expected to reduce diversity and increase genetic differentiation at linked sites (Cruickshank and Hahn 2014). We found evidence for an effect of linked selection in shaping differentiation between genetic clusters in Florida. These clusters have had relatively stable effective population sizes over the last 2 Myr, which should limit the stochasticity induced by drift in explaining heterogeneous patterns of differentiation along the genome. In that case, signatures of linked selection such as high differentiation and low diversity should be easier to detect in regions of low recombination. Indeed, regions of high diversity that are characterized by high $d_{ST}$ displayed higher recombination rates and lower $F_{ST}$ in the green anole, while regions with high
were found in regions of low recombination and diversity. More specifically, a prominent role for background selection in shaping the differentiation landscape is supported by the correlations we observed between diversity, differentiation, and recombination. We acknowledge that positive selection may lead to similar signatures of high differentiation and low diversity in regions of low recombination due to the effects of selection at linked sites (Josephs and Wright 2016), and we did observe an increase in the average frequency of derived alleles in regions of very low recombination (supplementary fig. 9, Supplementary Material online). However, regions of both high divergence and low diversity did not seem to harbor any excess of derived alleles (supplementary fig. 10, Supplementary Material online) which suggests that positive selection is not the main driver of the differentiation landscape within Florida. Moreover, the agreement between landscapes of differentiation for all pairwise comparisons (fig. 4) suggests a reduction in diversity in the ancestral population rather than population-specific events of selection, consistent with background selection, or positive selection in the lineage ancestral to all green anoles.

Since green anoles constitute an important model species to understand the mechanisms of adaptation, adopting a cautious (i.e., nonadaptationist) interpretation of divergence landscapes is primordial (Burri 2017). Disentangling recent positive selection from the confounding effects of demography and background selection is especially challenging, even in species for which extensive genomic and functional studies have been performed, as in humans. Genome scans for positive selection have often failed to identify common outliers (Pavlidis et al. 2012), and require to carefully consider the demographic history of populations (Li et al. 2012; Pavlidis et al. 2012; Elyashiv et al. 2016; Schrider and Kern 2016; Sheehan and Song 2016; Bourgeois et al. 2018). The green anole is an important system to understand local adaptation in reptiles (Campbell-Staton et al. 2017) and the incorporation of our findings in future studies will be useful to properly test for signals of local adaptation. This will be done by taking into account the possible biases induced by demography and the impact of selection at linked sites. Further studies of positive selection will require more detailed analyses, building on the results we show in the present study.

Supplementary Material
Supplementary data are available at Genome Biology and Evolution online.

Acknowledgments
We thank three anonymous reviewers for their constructive comments that improved an earlier version of this article. We are grateful to Breda Zimkus from the Museum of Comparative Zoology Cryogenic Collection in Harvard and J. Rosado from the Herpetology Collection for providing the samples of A. porcus and A. allisoni. We also thank Christelle Fraisse for providing tutorials and the modified version of \textit{Fast} that was needed to compare demographic models. We thank Justin Wilcox for his comments on the article. We thank Marc Arnoux from the Genome Core Facility at NYUAD for assistance with genome sequencing. This research was carried out on the High-Performance Computing resources at New York University Abu Dhabi. This work was supported by New York University Abu Dhabi (NYUAD) research funds AD180 (to S.B.). The NYUAD Sequencing Core is supported by NYUAD Research Institute grant G1205-1205A to the NYUAD Center for Genomics and Systems Biology.

Literature Cited
Alfoldi J, et al. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature 477(7366):587–591.
Allo J, Donega S, Galtier N, Nabholtz B. 2017. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. Mol Biol Evol. 34(11):2762–2772.
Auton A, et al. 2012. A fine-scale chimpanzee genetic map from population sequencing. Science 336(6078):193–198.
Barrett RD, Rogers SM, Schluter D. 2008. Natural selection on a major armor gene in threespine stickleback. Science 322(5899):255–257.
Barton NH, Hewitt GM. 1985. Analysis of hybrid zones. Annu Rev Ecol Syst. 16(1):113–148.
Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in \textit{D. melanogaster}. Nature 356(6369):519–520.
Bienne N, Welch J, Loire E, Bonhomme F, David P. 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. Mol Ecol. 20(10):2044–2072.
Boitger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30(15):2144–2142.
Booher TR, Ness RW, Keightley PD. 2017. The recombination landscape in wild house mice inferred using population genomic data. Genetics 207(1):297–309.
Bourgeois Y, et al. 2018. Genome-wide scans of selection highlight the impact of biotic and abiotic constraints in natural populations of the model grass \textit{Brachypodium distachyon}. Plant J. 96(2):438–451.
Burri R. 2017. Interpreting differentiation landscapes in the light of long-term linked selection. Evol Lett. 1(3):118–131.
Burri R, et al. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of \textit{Ficedula} flycatchers. Genome Res. 25(11):1656–1665.
Cai JJ, Macpherson JM, Sella G, Petrov DA. 2009. Pervasive hitchhiking at coding and regulatory sites in humans. PLoS Genet 5(1):e1000336.
Calsbeek R. 2009. Sex-specific adult dispersal and its selective consequences in the brown anole, \textit{Anolis sagrei}. J Anim Ecol. 78(3):617–624.
Campbell-Staton SC, Bare A, Losos JB, Edwards SV, Chevron ZA. 2018. Physiological and regulatory underpinnings of geographic variation in reptilian cold tolerance across a latitudinal cline. Mol Ecol. 27(9):2243–2255.
Campbell-Staton SC, Edwards SV, Losos JB. 2016. Climate-mediated adaptation after mainland colonization of an ancestrally subtropical island lizard, \textit{Anolis carolinensis}. J Evol Biol. 29(11):2168–2180.
Campbell-Staton SC, et al. 2012. Out of Florida: mtDNA reveals patterns of migration and pleistocene range expansion of the green anole lizard (Anolis carolinensis). Ecol Evol. 2(9):2274–2284.

Campbell-Staton SC, et al. 2017. Winter storms drive rapid phenotypic, regulatory, and genomic shifts in the green anole lizard. Science 357(6350):495–498.

Christe C, et al. 2017. Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. Mol Ecol. 26(1):59–76.

Clemente F, Gautier M, Vitalis R. 2018. Inferring sex-specific demographic history from SNP data. PLoS Genet. 14:1–32.

Coffman AJ, Hsieh PH, Gravel S, Gutenkunst RN. 2016. Computationally efficient composite likelihood statistics for demographic inference. Mol Biol Evol. 33(2):591–593.

Costantini M, Greif G, Alvarez-Valin F, Bernardi G. 2016. The Anolis lizard genome: an amniote genome without isochores? Genome Biol Evol. 8(4):1048–1055.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol. 23(13):3133–3157.

Danecek P, et al. 2011. The variant call format and VCFtools. Bioinformatics 27(15):2156–2158.

Depristo MA, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 43(5):491–501.

Ellegren H, et al. 2012. The genomic landscape of species divergence in Ficedula flycatchers. Nature 491(7426):756–760.

Elyashiv E, et al. 2016. A genomic map of the effects of linked selection in Drosophila. PLoS Genet. 12:1–24.

Excoffier L, Dupanloup I, Hauerta-Sanchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9(10):e1003905.

Feder JL, Nosil P. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. Evolution 64(6):1729–1747.

Figuet E, Ballenghien M, Romiguier J, Galtier N. 2015. Biased gene conversion and GC-content evolution in the coding sequences of reptiles and vertebrates. Genome Biol Evol. 7(1):240–250.

Fujita MK, Edwards SV, Ponting CP. 2011. The version and GC-content evolution in the coding sequences of reptiles and birds. Mol Ecol. 64(6):1729–1747.

Glot RE, Losos JB, Larson A. 2005. Out of Cuba: overwater dispersal and genomic shifts in the green anole lizard. Science 307(5715):140–149.

Gunterkunst RN, Macey JR, et al. 1999. Molecular phylogenetics, tRNA evolution, and histone biogeography in Anguid lizards and related taxonomic families. Mol Phylogenet Evol. 12(3):250–272.

Hair F, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079.

Jombart T, et al. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11(1):94.

Josephs EB, Wright SI. 2016. On the trail of linked selection. PLoS Genet. 12:1–5.

Kawakami T, et al. 2017. Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. Mol Ecol. 26(16):4158–4172.

Kern AD, Hahn MW. 2018. The neutral theory in light of natural selection. Mol Biol Evol. 35(6):1366–1371.

Kolbe JJ, et al. 2017. An incipient invasion of brown anole lizards (Anolis sagrei) into their own native range in the Cayman Islands: a case of cryptic back-introduction. Biol Invasions. 19(7):1989–1998.

Laila Arabs SP, Herrel A, Vanhooydonck B, Meyers JJ, Inschick DJ. 2004. Performance capacity, fighting tactics and the evolution of life-stage male morphs in the green anole lizard (Anolis carolinensis). Proc Biol Sci. 271(1556):2501–2508.

Lane, E. 1994. Florida’s Geological History and Geological Resources. Special publication (Florida Geological Survey (1989)). Vol. 35. Published for the Florida Geological Survey; Tallahassee, FL.

Lapedra O, Schoener TW, Leal M, Losos JB, Kolbe JJ. 2018. Predator-driven natural selection on risk-taking behavior in anole lizards. Science 360(6392):1017–1020.

Larson EL, White TA, Ross CL, Harrison RG. 2014. Gene flow and the maintenance of species boundaries. Mol Ecol. 23(7):1668–1678.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. Nature 475(7357):493–496.

Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079.

Li J, et al. 2012. Joint analysis of demography and selection in population genetics: where do we stand and where could we go? Mol Ecol. 21(1):28–44.

Losos JB. 2009. Lizards in an evolutionary tree. Berkeley: University of California Press [Database].

Losos JB, Schoener TW, Spiller DA. 2004. Predator-induced behaviour shifts and natural selection in field-experimental lizard populations. Nature 432(7016):505–508.

Macyr JR, et al. 1999. Molecular phylogenetics, tRNA evolution, and historical biogeography in Anatid species and related taxonomic families. Mol Phylogenet Evol. 12(3):250–272.

Malinsky M, et al. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science 350(6267):1493–1498.

Manthey JD, Tollis M, Lemmon AR, Moriarty Leemon E, Boissinot S. 2016. Diversification in wild populations of the model organism Anolis carolinensis: a genome-wide phylogeographic investigation. Ecol Evol. 6(22):8115–8125.

McCoo MD, Neches RS, Seehausen OA. 2016. Evaluating genomic divergence and parallelism in replicate ecomorphs from young and old chilid adaptive radiations. Mol Ecol. 25(1):260–268.

McVean G, Awadalla P, Fearnhead P. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics 160(3):1057–1071.

Mullen LM, Hoeckstra HE. 2008. Natural selection along an environmental gradient: a classic cline in mouse pigmentation. Evolution 62(7):1555–1570.

Nachman MW, Payseur BA. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philos Trans R Soc Lond B Biol Sci. 367(1587):409–421.

Nadeau NJ, et al. 2012. Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. Philos Trans R Soc Lond B Biol Sci. 367(1587):343–353.
Noor MAF, Bennett SM. 2009. Islands of speciation or mirages in the desert. Examining the role of restricted recombination in maintaining species. Heredity 103(6):439–444.
Pavlidis P, Jensen JD, Stephan W, Stamatakis A. 2012. A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. Mol Biol Evol. 29(10):3237–3248.
Payseur BA, Krenz JG, Nachman MW. 2004. Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. Evolution 58(9):2064–2078.
Petuch, EJ. 2004. Cenozoic Seas: the View from Eastern North America.
Boca Raton: CRC Press.
Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol. 31(7):1929–1936.
Poelstra JW, et al. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344(6190):1410–1414.
Pouyet F, Aeschbacher S, Thiéry A, Excoffier L. 2018. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. Elife 7:1–21.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842.
Roux C, et al. 2014. Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a Mytilus hybrid zone. J Evol Biol. 27(8):1662–1675.
Roux C, et al. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. PLoS Biol 14(12):e2000234.
Rozas J, Gullaund M, Blandin G, Aguade M. 2001. DNA variation at the rp49 gene region of Drosophila simulans: evolutionary inferences from an unusual haplotype structure. Genetics 158(3):1147–1155.
Ruggiero RP, Bourgeois Y, Boissinot S. 2017. L1E insertion polymorphisms are abundant but at low frequencies across populations of Anolis carolinensis. Front Genet. 8:1–14.
Rupp SM, et al. 2017. Evolution of dosage compensation in Anolis carolinensis, a reptile with XXXY chromosomal sex determination. Genome Biol Evol. 9(1):231–240.
Schrider DR, Kern AD. 2016. SHHIC: robust identification of soft and hard sweeps using machine learning. PLoS Genet. 12:1–31.
Seehausen O, et al. 2014. Genomics and the origin of species. Nat Rev Genet. 15(3):176–192.
Sheehan S, Song YS. 2016. Deep learning for population genetic inference. PLoS Comput Biol. 12:1–28.
Solitis DE, Morris AB, McLachlan JS, Manos PS, Solitis PS. 2006. Comparative phylogeography of unglaciated eastern North America. Mol Ecol. 15(14):4261–4293.
Sousa VC, Carneiro M, Ferrand N, Hey J. 2013. Identifying loci under selection against gene flow in isolation-with-migration models. Genetics 194(1):211–233.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3):585–595.
Teeter KC, et al. 2008. Genome-wide patterns of gene flow across a house mouse hybrid zone. Genome Res. 18(1):67–76.
Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat Genet. 49(2):303–309.
Tine M, et al. 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat Commun. 5:5770.
Tollis M, Ausubel G, Ghimire D, Boissinot S. 2012. Multi-locus phylogeographic and population genetic analysis of Anolis carolinensis: historical demography of a genomic model species. PLoS One 7:1–14.
Tollis M, Boissinot S. 2011. The transposable element profile of the Anolis genome: how a lizard can provide insights into the evolution of vertebrate genome size and structure. Mob Genet Elements. 1(2):107–111.
Tollis M, Boissinot S. 2013. Lizards and LINEs: selection and demography affect the fate of L1 retrotransposons in the genome of the green anole (Anolis carolinensis). Genome Biol Evol. 5(9):1754–1768.
Tollis M, Boissinot S. 2014. Genetic variation in the green anole lizard (Anolis carolinensis) reveals island refugia and a fragmented Florida during the quaternary. Genetica 159:59–72.
Torres R, Stetter MG, Hernandez RD, Ross-Ibarra J. 2019. The temporal dynamics of background selection in non-equilibrium populations. bioRxiv. doi: https://doi.org/10.1101/618389
Van Der Auwera GA, et al. 2014. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinforma. 11:11.10.1–11.10.33.
Wade J. 2012. Sculpting reproductive circuits: relationships among hormones, morphology and behavior in anole lizards. Gen Comp Endocrinol. 176(3):456–460.
Wang J, Street NR, Scafﬁeld DG, Ingvarsson PK. 2016. Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American asps. Mol Biol Evol. 33(7):1754–1767.
Wolf JBW, Ellegren H. 2017. Making sense of genomic islands of differentiation in light of speciation. Nat Rev Genet. 18(2):87–100.

Associate editor: Takashi Gojobori