Genomic Basis of Convergent Island Phenotypes in Boa Constrictors

Daren C. Card [1,2,3], Richard H. Adams[1], Drew R. Schield[1], Blair W. Perry[1], Andrew B. Corbin[1], Giulia I.M. Pasquesi[1], Kristopher Row[1], Melissa J. Van Kleeck[4], Juan M. Daza[5], Warren Booth[6], Chad E. Montgomery[7], Scott M. Boback[8], and Todd A. Castoe[1,*]

[1]Department of Biology, University of Texas Arlington

[2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts

[3]Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts

[4]Department of Biology, Pima Community College, Tucson, Arizona

[5]Grupo Herpetológico de Antioquia, Instituto de Biología, Universidad de Antioquia, Medellín, Colombia

[6]Department of Biological Science, University of Tulsa, Oklahoma

[7]Biology Department, Truman State University, Kirksville, Missouri

[8]Department of Biology, Dickinson College, Carlisle, Pennsylvania

*Corresponding author: E-mail: todd.castoe@uta.edu.

Accepted: October 10, 2019

Data deposition: Raw transcriptomic sequencing data have been deposited at the NCBI SRA under the accession SRP148755. Raw genomic sequencing data have been deposited at the NCBI SRA under the accessions SRP103533 and SRP189211. Other data and relevant analysis scripts have been deposited at Figshare under the accession 10.6084/m9.figshare.9793013. Data files for the annotated genome of *Boa constrictor* have been deposited at Figshare under the accession 10.6084/m9.figshare.9793013. In an effort to aid other researchers interested in using the annotated *B. constrictor* reference genome, we have also created a more detailed data repository page linking the previously published genome assembly with the new repeat and gene annotation (see http://darencard.github.io/boaCon; last accessed October 22, 2019).

Abstract

Convergent evolution is often documented in organisms inhabiting isolated environments with distinct ecological conditions and similar selective regimes. Several Central America islands harbor dwarf *Boa* populations that are characterized by distinct differences in growth, mass, and craniofacial morphology, which are linked to the shared arboreal and feast-famine ecology of these island populations. Using high-density RADseq data, we inferred three dwarf island populations with independent origins and demonstrate that selection, along with genetic drift, has produced both divergent and convergent molecular evolution across island populations. Leveraging whole-genome resequencing data for 20 individuals and a newly annotated *Boa* genome, we identify four genes with evidence of phenotypically relevant protein-coding variation that differentiate island and mainland populations. The known roles of these genes involved in body growth (*PTPRS*, *DMGDH*, and *ARSB*), circulating fat and cholesterol levels (*MYLIP*), and craniofacial development (*DMGDH* and *ARSB*) in mammals link patterns of molecular evolution with the unique phenotypes of these island forms. Our results provide an important genome-wide example for quantifying expectations of selection and convergence in closely related populations. We also find evidence at several genomic loci that selection may be a prominent force of evolutionary change—even for small island populations for which drift is predicted to dominate. Overall, while phenotypically convergent island populations show relatively few loci under strong selection, infrequent patterns of molecular convergence are still apparent and implicate genes with strong connections to convergent phenotypes.

Key words: *Boa constrictor* genome annotation, convergent evolution, genetic drift, selection, insular populations, whole-genome resequencing, snakes.

© 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 Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Introduction

Although numerous examples of convergent phenotypic evolution have been documented in nature, examples of convergent molecular evolution have historically been far more rare (Losos 2011; Stern 2013; Rosenblum et al. 2014). Over the past few decades, however, convergent molecular evolution has been identified in genes related to coloration (Nachman et al. 2003; Hoekstra et al. 2006; Rosenblum et al. 2007, 2010; Linnen et al. 2009, 2013), toxin resistance (Zakon 2002, 2012; Jost et al. 2008; Feldman et al. 2009, 2012; Zhen et al. 2012; McGlothlin et al. 2014; Liebeskind et al. 2015; Reid et al. 2016), metabolism (Castoe et al. 2009), digestion (Stewart et al. 1987), oxygen transport (Hoffmann et al. 2010; Projecto-Garcia et al. 2013; Natarajan et al. 2016), and unique traits like beak morphology in birds (Grant et al. 2004), evolution of electric organs in fish (Gallant et al. 2014), and pelvic spine morphology in sticklebacks (Hohenlohe et al. 2010; Jones et al. 2012). These studies provide compelling evidence for the occurrence of convergent molecular evolution, yet few have attempted to provide a genome-wide perspective on the relative roles of migration, genetic drift, and adaptation in shaping patterns of convergence and divergence underlying phenotypic evolution. In this study, we address these questions by investigating genome-wide patterns of genetic variation in a system of island–mainland populations of a snake species that exhibits phenotypic convergence across multiple islands.

Island fauna often exhibit unique phenotypes compared with their mainland counterparts due to their isolation and the ecological uniqueness of island environments; this phenomenon is known as the "island syndrome" (Adler and Levins 1994; Lomolino et al. 2010). For example, variation in body size between mainland and island populations is well-documented and widespread across diverse island fauna (Lomolino et al. 2013). One theory for changes in body size on islands posits that evolution of body size arises from shifts in ecological niche that result from adaptation to alternative food resources (McNab 1994, 2002; Boback 2003; Köhler and Moyà-Solà 2009). For example, Keogh et al. (2005) documented multiple independent island body size changes that each correlate with the body sizes of prey on individual islands. However, body size may also correlate with other morphological traits important for occupying novel niches. For instance, a smaller body size and a more slender shape could be selectively advantageous in an arboreal setting for several reasons, such as increased stability and ability to utilize smaller branches (Jayne 1982; Cartmill 1985), and thus, other traits beneficial to arboreal organisms may covary with body size.

Snakes in the genus *Boa* are widespread throughout the New World and are renowned for their large size. *Boa imperator*, which is found throughout Central America (Hynková et al. 2009; Reynolds et al. 2014; Suárez-Atilano et al. 2014; Card et al. 2016), has colonized dozens of off-shore islands in the Caribbean (Henderson et al. 1995; Porras 1999), including several off the coasts of Belize and Honduras (fig. 1*A*–*C*). Island populations of *B. imperator* often differ substantially in ecology from adjacent mainland populations. The most striking difference between island and mainland populations is overall body size and mass, which are both smaller on islands (Boback 2005, 2006; Boback and Carpenter 2007; fig. 1*D*, Belize populations). Snakes from these islands also have attenuated snouts, narrower craniofacial morphology, and are more slender with longer tails compared with mainland populations (Boback 2006; fig. 1*E*, Belize populations). Common garden experiments have determined that these traits are heritable outside of natural island conditions (Boback and Carpenter 2007), suggesting a genetic basis for these convergent island phenotypes. The arboreal tendencies and associated morphological traits are thought to enable the distinctly arboreal feeding habits of island boas, which feed primarily on migratory passerine birds (Lillywhite and Henderson 2002; Boback 2005).

In this study, we investigate the genomic basis for repeated evolution of distinct eco-morphotypes of Central American insular boas (fig. 1*A*–*C*). We generate and analyze a new functional annotation of the *Boa* reference genome, high-density RADseq population genomic sampling of island and mainland populations, and 20 resequenced whole genomes to study patterns of genomic variation between island–mainland population pairs and test hypotheses about molecular evolution underlying convergent phenotypic evolution. Using these data, we addressed five main questions about the evolution of genomic and phenotypic variation across island populations: 1) How many independent origins explain the founding of dwarf island *Boa* populations in Belize and Honduras? 2) Is there evidence that natural selection, in addition to genetic drift, has contributed to genomic differentiation of island populations? 3) Is there evidence for convergent molecular evolution across multiple island populations? 4) Does convergent molecular evolution explain convergent island phenotypes? And 5) What is the relative contribution and scale of natural selection and convergent evolution in the context of island genetic and phenotypic divergence?

Materials and Methods

Assessing Patterns of Morphological Evolution in Central American Boas

Patterns of body size evolution have been previously assessed for several island populations in Belize, for the Cayos Cochinos population in Honduras, and for the mainland population samples collected throughout Belize (Boback 2006; Reed et al. 2007). We supplemented these existing data with snout-vent length (SVL) and mass measurements for 42

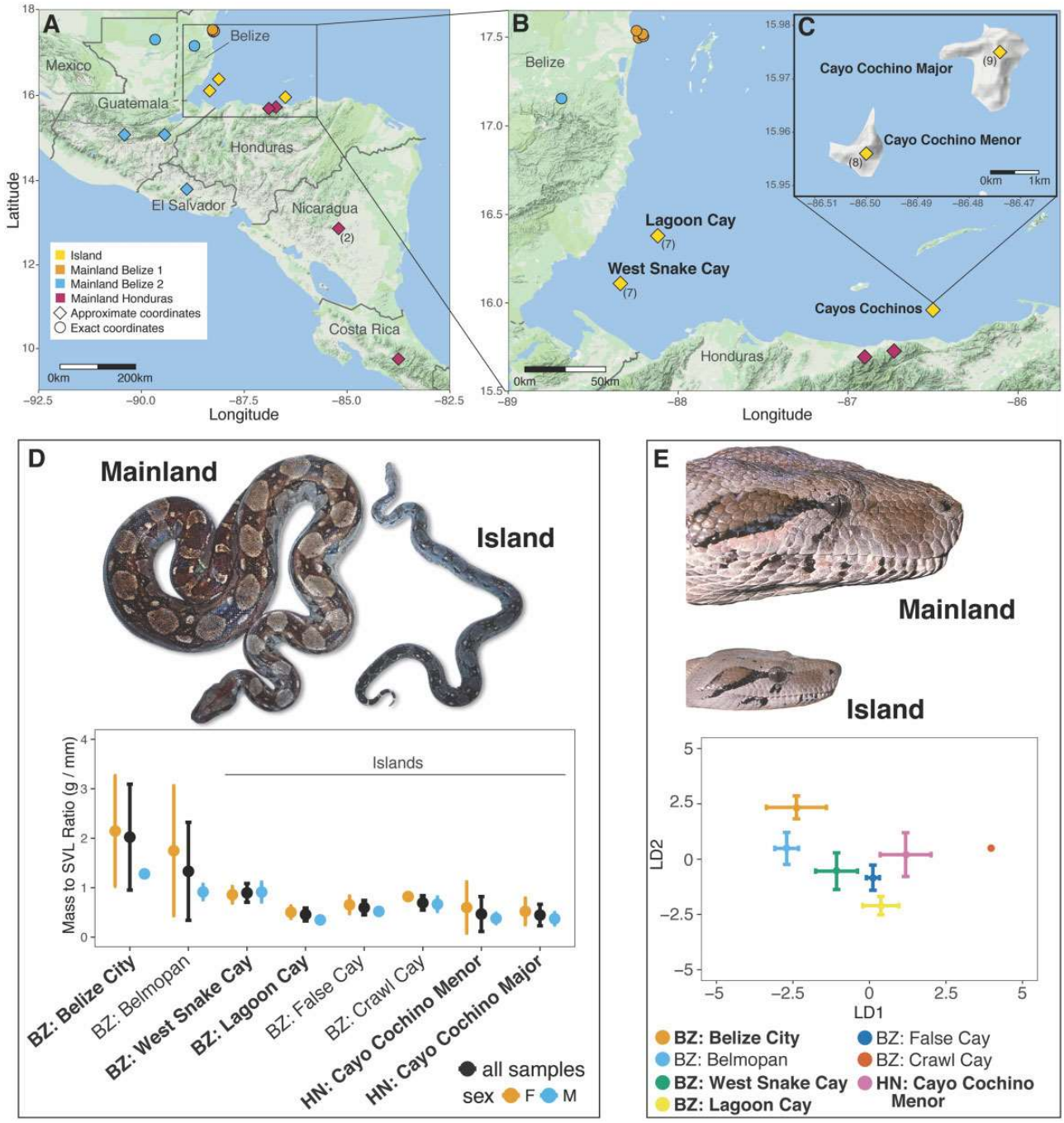

Fig. 1.—Summary of island boa study system. (*A*–*C*) Geographic representation of population sampling with sample sizes, including island populations located on Lagoon and West Snake Cays, Belize and Cayos Cochinos, Honduras. Points are colored by population and sample sizes are indicated in parentheses for a locality if the sample size is >1. (*D* and *E*) Overview of phenotypic differences between island and mainland boa populations, including body size (*D*) and craniofacial morphology (*E*). Phenotypic data were taken from previous studies (Boback 2005, 2006; Boback and Carpenter 2007; Reed et al. 2007) and collected from sampling of the Cayos Cochinos population. In panel (*D*), data from females and males are shown separately in orange and blue, respectively, and the distributions from the combined data are shown in black. Whiskers on each plot represent the standard deviation for phenotypic measurements. Bold text indicates populations samples analyzed in this study. BZ, Belize; HN, Honduras.

additional boas sampled from Cayo Cochino Menor ($N = 21$) and Cayo Cochino Major ($N = 21$). To better understand how body size varies across mainland populations, we grouped samples from mainland Belize geographically based on available locality/coordinate data. Two areas of the mainland had sufficient sampling to compare body size with island

populations: Belize City ($N = 7$) and Belmopan ($N = 15$), Belize. We used this expanded and repartitioned data set to study patterns of *Boa* variation across Belize and Honduras by estimating the mass to SVL ratio for a total of 184 individuals from all sampled island and mainland populations.

Craniofacial morphology has also been previously shown to vary between island and mainland populations and across island populations in Belize (Boback 2006), but analogous comparisons for Cayos Cochinos have not been conducted. To assess craniofacial morphology for Cayos Cochinos, we used previously unpublished measurements of craniofacial morphology for snakes ($N = 65$) captured on Cayo Cochino Menor. Measurements were based on the same eight morphological features as Boback (2006): head width, head length, labial, interocular, ocular, nares-ocular, rostral-ocular, and internares. Following Boback (2006), we standardized these measures by head length and we used a linear discriminant analysis in R v. 3.4.1 (R Core Team 2019) to visualize overall head shape. We provide all phenotypic data used in previous publications alongside newly assessed data in supplementary file S3, Supplementary Material online.

Annotation of a Boa Reference Genome

To facilitate the interpretation of our analyses of genomic variation, we generated new RNAseq data and annotated an existing genome assembly for *Boa constrictor*. We extracted RNA from nine tissue samples (supplementary table S1, Supplementary Material online) using the Trizol reagent (Invitrogen). RNAseq libraries were generated using an Illumina TruSeq RNAseq kit and we sequenced the multiplexed library using an Illumina HiSeq 2000 and 100 bp paired-end sequencing. We combined the resulting transcriptome data from these nine tissues with existing transcriptome data from male and female blood samples from NCBI (Vicoso et al. 2013) and used Trinity v. r20140717 (Haas et al. 2013) with default parameters and internal Trimmomatic quality trimming to assemble all Illumina reads into transcript contigs.

We annotated the highest quality genome assembly for *B. constrictor* generated during the Assemblethon2 project ("*snake assembly 7C*" produced by the SGA team; Bradnam et al. 2013). We identified repetitive content with RepeatMasker v. 4.0.6 (Smit et al. 2013) and used an iterative process within MAKER v. 2.31.8 (Holt and Yandell 2011) to annotate protein-coding genes using Augustus (Stanke and Waack 2003; Stanke et al. 2004) and our newly generated transcriptomic data (supplementary table S1, Supplementary Material online), together with gene models from other squamate genomes for gene prediction (supplementary table S2, Supplementary Material online). We identified the likely protein product for each gene model using BLAST (Altschul et al. 1990) queries against proteins in Unitprot/SwissProt (The UniProt Consortium 2017), the InterPro database (Jones et al. 2014; Mitchell et al. 2015), and from other vertebrate genomes. See the supplementary methods, Supplementary Material online, for a detailed description of genome annotation procedures.

Population Sampling, DNA Extraction, and Restriction-Site Associated DNA Library Preparation and Sequencing

We obtained tissue samples from 42 *Boa* individuals from mainland and island populations in Central America (see fig. 1 and supplementary table S3, Supplementary Material online, for population sampling information). DNA was extracted from blood, skin sheds, or tissues using either a Zymo Research Quick-gDNA Miniprep kit (Zymo Research, Irvine, CA) according to the manufacturer's protocol or a standard phenol–chloroform–isoamyl alcohol extraction. We conducted double digest Restriction-site Associated DNA sequencing (RADseq), following Peterson et al. (2012). *Pst*I and *Sau*3AI restriction enzymes were used to digest genomic DNA, and the resulting fragments were ligated to double-stranded adapters containing barcodes and unique molecular identifiers (UMIs; eight consecutive random nucleotides prior to the ligation site). Samples were pooled into groups of eight individuals for efficient size selection for fragments ranging from 570 to 690 bp using a Blue Pippin, a range expected to yield ~200,000 loci based on in silico digestion of the *B. constrictor* genome (Bradnam et al. 2013). We used a Bioanalyzer (Agilent, Santa Clara, CA) to quantify and pool libraries, which were sequenced using 100 bp paired-end reads on an Illumina HiSeq 2500.

RADseq Data Analysis and Variant Calling

We used the clone_filter module from the Stacks v. 1.42 pipeline (Catchen et al. 2011, 2013) to filter out PCR duplicates based on raw read UMIs. The process_radtags module from Stacks was used to parse reads by index, with default parameters, except with the "rescue" feature activated and the restriction digest site check disabled. Parsed reads were filtered for RADseq adapter and primer sequences and were quality trimmed using Trimmomatic v. 0.33 (Bolger et al. 2014) using the settings LEADING: 10 TRAILING: 10 SLIDINGWINDOW: 4:15 MINLEN: 36. We used NextGenMap (Sedlazeck et al. 2013) with default settings to map the quality-trimmed reads to the *B. constrictor* reference genome (Bradnam et al. 2013). Details of reads retained during RADseq read processing are provided in supplementary table S4, Supplementary Material online. We used the "GATK Best Practices" workflow (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013) to perform local indel realignment (with default settings) and joint genotyping of individual GVCFs using HaplotypeCaller to infer variants. We filtered the resulting variants using bcftools (Li et al. 2009; Li 2011) to produce a biallelic variant data set by excluding variants that fit any of the following conditions: SNPs within 3 bp of an INDEL, QUAL < 30, QD < 2, FS > 60.0,

MQ < 40.0, MQRankSum < −12.5, ReadPosRankSum < −8, FMT/DP < 5, DP > 500, DP < 100. We also used ThetaMater (Adams et al. 2018) to conduct posterior predictive simulation (PPS) to remove loci with evidence of unlikely (i.e., high) numbers of observed segregating sites (i.e., potential paralogs). The resulting data set contained a total of 187,221 variants across all 42 individuals (supplementary file S4, Supplementary Material online).

Evaluating Demographic Models to Assess Island Population Independence

To provide a demographic estimate and perspective to our analyses of genomic variation, we inferred a population-level tree using the program SNAPP (Bryant et al. 2012), our RADseq variant data, and the population assignments provided in supplementary table S3, Supplementary Material online. These population assignments were based on previous work described in Card et al. (2016), which inferred monophyletic groupings of individual island samples and three major clusters of mainland samples: 1) a monophyletic grouping of samples from Nicaragua and Costa Rica, which were further bolstered by the addition of two samples from North Honduras, as the mainland population associated with Cayos Cochinos (hereafter referred to as "mainland Honduras"); 2) a monophyletic grouping of samples from near Belize City, Belize that is sister to a clade of samples from Lagoon and Crawl cays, which we refer to as the "Belize 1" population; and 3) a nonmonophyletic set of samples that encompass samples from Central Belize and Guatemala that we refer to as the "Belize 2" population. The mainland Honduras clade includes pure-bred offspring of wild-caught samples from Nicaragua and Honduras, and while the inclusion of captive-bred samples is not ideal, we decided it is warranted to increase sample sizes. We quantified genome-wide heterozygosity of these pure-bred samples using our RADseq data (mean heterozygosity of 0.108) and find that it is similar to that observed in all wild-caught mainland samples (0.114) and in the two samples collected from North Honduras (0.106). We ran the SNAPP Markov Chain Monte Carlo (MCMC) algorithm for a total of 10^6 generations, sampling every 10^3 generation to obtain posterior estimates of the population tree, divergence times, and effective population size parameters. Posterior stationarity and convergence was assessed using Tracer (Drummond and Rambaut 2007). We discarded the first 25% of generations as burn-in and used the remaining MCMC samples to produce a maximum clade credibility consensus population tree with median node heights. In addition to SNAPP, which assumes no migration between populations, we conducted more detailed demographic model inference using 2D site-frequency spectra (2D SFS) analyses of our RADseq data in the program $\delta a \delta i$ (Gutenkunst et al. 2009) to evaluate eight competing demographic models (supplementary tables S5 and S6, Supplementary Material online). We conducted these model tests independently for two parallel analyses: one between the Lagoon and West Snake Cays off the coast of Belize and one between Cayo Cochino Menor and Mayor off the coast of Honduras. To minimize the impact of missing data while maximizing the number of shared variants, we downsampled to ten alleles (i.e., five individuals) per population in $\delta a \delta i$, which retained 3,565 variants for the Lagoon and West Snake Cay comparison, and 2,228 variants for the Cayo Cochino Menor and Mayor comparison. We used $\delta a \delta i$ to test the fit of each of the eight demographic scenarios to the empirical 2D AFS following the approach described in Portik et al. (2017) and Schield et al. (2017).

RADseq-Based Calculation of Population Genetic Statistics and Identification of Signatures of Selection

We calculated Weir and Cockerham's (1984) F_{ST} between populations pairs using our RADseq data and the pegas package (v. 0.10; Paradis 2010) in R v. 3.4.1 (R Core Team 2019). See supplementary file S5, Supplementary Material online, for empirical genome-wide F_{ST} measures for each population. To evaluate whether a strictly neutral model of divergence was sufficient to explain particularly high F_{ST} values that may be indicative of natural selection, we used the software package GppFst (Adams et al. 2017) to generate and compare a theoretical null distribution of F_{ST} and our empirical distributions. Using the Bayesian posterior probability distribution of population demographic parameters inferred via SNAPP using our RADseq data, we used GppFst to simulate a null distribution of F_{ST} under the neutral coalescent model, which can be used to identify F_{ST} intervals that appear poorly explained by drift alone (Schield et al. 2017). We generated neutral distributions of F_{ST} for three island systems (Lagoon Cay, West Snake Cay, and the combined Cayos Cochinos island populations) and their corresponding adjacent Belize or Honduras mainland populations, and, separately, between the Belize and Honduras mainland populations alone. For the purposes of this analysis, we combined all samples from Belize and Guatemala together to represent the mainland Belize population to improve statistical power. Previous analyses do indicate some population structure across this region (Card et al. 2016), but it is relatively weak and likely due to isolation by distance. For each of these analyses, we identified the 97.5% F_{ST} percentile of the empirical distribution and computed the probability of observing equal to or greater than the number of empirical variants given the simulated F_{ST} distributions. In other words, we used the null distributions of F_{ST} to compute P values for observing the empirical number of variants in our RADseq data with F_{ST} values that are greater or equal to the 97.5% percentile. When $P \leq 0.05$, this suggests an excess of variants in our empirical data set with high F_{ST} values that are poorly explained by drift alone.

Whole-Genome Resequencing Library Preparation, Sequencing, and Data Processing

We augmented our RADseq analyses by generating whole-genome resequencing (WGS) data for a subset of 20 individuals, including 2 individuals from each Belize island (Lagoon and West Snake Cays), 4 individuals from Cayos Cochinos, Honduras (2 each from Cayo Cochino Menor and Mayor), and 6 mainland samples each from the Belize and Honduras clades. Mainland samples were prioritized to provide as full an understanding as possible of segregating variation across the mainland populations. Shotgun genome libraries were produced using either a KAPA HyperPlus or an Illumina Nextera library preparation kit, multiplexed, and sequenced on an Illumina HiSeq X using 150 bp paired-end reads to an average depth of 12.7× (8.9× standard deviation). We followed essentially the same data analysis process outlined above for the RADseq data, but with duplicated reads resulting from PCR being filtered away following mapping to the *Boa* reference genome using the Picard MarkDuplicates tool. Details of reads retained during WGS read processing are provided in supplementary table S7, Supplementary Material online. Final variants were called using HaplotypeCaller based on the GVCF files of individual samples, and we filtered variants using identical settings to the RADseq data. The resulting variant data set contained 8,146,817 variants based on the *B. contrictor* genome assembly (supplementary file S6, Supplementary Material online).

Quantifying Parallel Island Allele Frequency Fluctuation from Genomic Data

Using both our RADseq and WGS data sets, we tested whether genomic regions that show high allelic differentiation between mainland and island population pairs were shared across different island populations (i.e., whether the same genomic regions were highly differentiated in different mainland–island pairs). For the RADseq data, we identified the variant positions with high F_{ST} values indicative of extreme differentiation between an island and its associated mainland population or between the two mainland populations (top 2.5% tail of empirical distributions). For the WGS data, we estimated the maximum allele frequency change between each island population and its adjacent mainland population for nonoverlapping 10 kb windows across the *Boa* genome (see supplementary file S7, Supplementary Material online, for these measures) and identified windows with extreme allele frequency changes of 0.90 or greater in each island population. For both data sets, we determined the number of instances where extreme F_{ST} values or allele frequency changes occurred in two or more island populations. To better understand whether more overlap was observed than is expected by random chance, we randomly selected loci (i.e., variants for RADseq data and windows for WGS data) from the empirical data set for each population at the same frequencies observed in the empirical data sets and measured the Jaccard index of overlap. By permutating this analysis 100 times, we established a distribution of expected Jaccard indices for each data set under a null model of no shared evolutionary patterns among mainland–island pairs, and we compared our empirical Jaccard index measurements to these null distributions.

Predicting the Effects of Coding Variation Estimated from WGS Data

Based on variants from our WGS data and gene models from our *B. constrictor* genome annotation, we used the Variant Effect Prediction (VEP v. 91.1; McLaren et al. 2016) program to identify the locations and infer the relative consequences and impacts of all identified variants according to established Sequence Ontologies (SOs; see supplementary file S8, Supplementary Material online, for results). Similarly, we also ran PROVEAN v. 1.1.5 (Choi 2012; Choi et al. 2012) to estimate the relative likelihood of a phenotypic impact of coding variants, based on evolutionary conservation inferred from the NCBI nonredundant protein database (downloaded January 29, 2018; see supplementary file S9, Supplementary Material online, for results). Following the recommendations of the creators of PROVEAN, we used a threshold of −2.5 for binary classification of deleterious (−2.5 or below) versus neutral (above −2.5) variation.

Identifying Phenotypically Relevant Genes in Regions of Shared Extreme Allele Frequency Fluctuation

To identify potentially phenotypically relevant genes (i.e., genes with coding variation) in regions of extreme allele frequency fluctuation of 0.90 or greater in each island population, we used our WGS data to search up to 100 kb in either direction of windows with parallel high allelic differentiation in all three independent island populations (fig. 2). Analyses of pairwise RADseq variant linkage disequilibrium (LD) across populations indicate that the best-fit LD decays to approximately half the maximum value at distances of 300 kb or greater in each population (supplementary fig. S1, Supplementary Material online), making a 100-kb window size conservative. We did not specifically investigate regulatory variation because protein-coding regions provide the most direct functional links between genotype and phenotype and because, unlike regulatory regions, objective criteria exist for predicting the penetrance of mutations in protein-coding regions. However, we note that our approach is capable of detecting signals of selection acting upon regulatory regions of the genome as well. For nearby protein-coding genes, we extracted functionally relevant variants that met two conditions: 1) showed a 0.75 or greater allele frequency fluctuation at least one island population and 2) were annotated as a high impact coding variant according to the VEP analysis or were annotated as a moderate coding impact variant (i.e., nonsynonymous variants) in the VEP analysis and

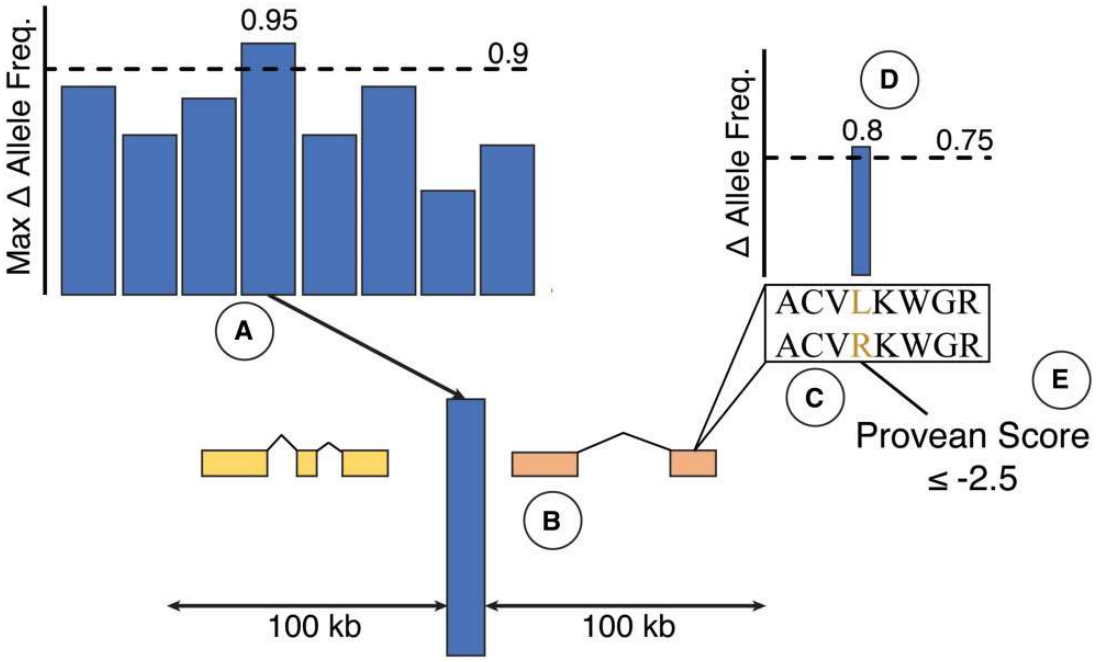

Fig. 2.—Workflow for identifying phenotypically relevant protein-coding variation shared between island populations. Graphical depictions of each step are indicated and correspond to the following steps: (*A*) about 10-kb windows with extreme allele frequency fluctuations (≥0.90) across all three island populations are identified; (*B*) genes within 100 kb of outlier windows are isolated; (*C*) only genes with nonsynonymous variation are retained; (*D*) nonsynonymous variants must fluctuation greatly (≥0.75) in at least one island population; and (*E*) widely fluctuating protein-coding variants must have a deleterious (less than or equal to −2.5) Provean score.

also had a deleterious PROVEAN score (fig. 2). Genes with penetrant, protein-coding alleles that are at much higher frequency in multiple island populations are likely targets of strong selection on islands and are likely linked to phenotypes that vary between island and mainland boa populations.

Assessing Function of Phenotypically Relevant Genes

Our posterior predictive simulations (PPS) indicated that neutral processes may explain a high degree of allelic differentiation for an appreciable number of loci (see Results and Discussion). Functional assessments of genes in regions of high allelic differentiation can help to identify genes with links to observed phenotypic shifts on islands, and thus help to distinguish loci that are likely to be truly under selection from loci with false signals of selection. To this end, we examined mouse phenotype data for phenotypically relevant genes using the Mouse Genome Informatics (Smith et al. 2018) batch query tool. To test for enrichment of particular mouse phenotypes in gene sets, we used Model organism Phenotype Enrichment Analysis (ModPhea; Weng and Liao 2017), with our search covering all phenotypic levels and using the full set of boa reference genome genes for the background. For all enrichment analyses, *P* values were corrected based on the method of Benjamini and Hochberg (1995) and we retained phenotypes with FDR *P* values <0.05 as significantly enriched.

Results and Discussion

Assessing Patterns of Morphological Evolution in Central American Boas

In this study, we extended previous analyses of the evolution of body size and craniofacial morphology across island and mainland boa populations in Belize and Honduras. Based on at least six samples per population, we found that body size is reduced across all island populations in comparison with mainland populations from Belize (fig. 1*D*). Although some mainland samples do exhibit small body sizes similar to that observed on islands, and overall greater variation in body size, island populations appear to have a restricted upper size limit compared with mainland populations (fig. 1*D*). Sexual dimorphism is evident in mainland Belize and in the Cayos Cochinos populations, as the variance in body size is higher in females than males, especially on the mainland (fig. 1*D*). Moreover, in many cases males on islands appear to have a more reduced body size than that on the mainland, though island female body sizes are more likely to overlap the range of female body sizes observed on the mainland (fig. 1*D*). Previous studies have documented this same pattern of body size evolution and sexual dimorphism (Boback 2006; Reed et al. 2007).

Previous analyses of craniofacial morphology in island and mainland populations from Belize indicated that craniofacial morphology varies between the island and mainland populations and also across island populations (Boback 2006). Our

analysis based on an expanded data set shows a similar pattern, with all island populations differing from the mainland Belize populations along the first linear discrimination (LD1) axis of variation and the False Cay and Lagoon Cay populations also varying noticeably from the mainland Belize populations along the LD1 axis (all Belize island populations have a similar position along the second linear discrimination [LD2] axis; fig. 1*E*). The LD1 axis corresponds with head length while the LD2 axis corresponds with head width (supplementary fig. S2, Supplementary Material online). Craniofacial morphology does vary between the two mainland populations by an amount similar to what is observed between individual island populations (fig. 1*E*). Importantly, craniofacial morphology in the populations from Cayo Cochino Menor, which has not been previously assessed, also varies from mainland Belize populations along LD1 (fig. 1*E*). Overall, our findings mirror previous studies, and add new insight into the distinct craniofacial morphology in the Cayo Cochino Menor population and a more nuanced view of craniofacial morphology across the mainland.

Annotation of a Boa Reference Genome

We annotated an existing genome assembly (contig and scaffold N50 values of 29.3 kb and 4.5 Mb, respectively) for *B. constrictor* (Bradnam et al. 2013). Our annotation inferred 31.61% of the genome as repetitive, with transposable elements and simple sequence repeats (microsatellites) composing 29.6% and 2.4% of the assembly, respectively. LINE elements (12.8%), DNA transposons (5.2%), LTR elements (2.3%), non-LTR elements (1.1%), and Penelope-like elements (1.0%) comprised significant portions of the genome (supplementary fig. S3 and table S8, Supplementary Material online). We identified 19,178 gene models and were able to confidently assign functional information (i.e., gene IDs) for 96.7% of annotated genes based on homology searches, including 93.18% of genes that were matched with human gene orthologs. Details on the results of several annotation-derived summaries are described in the supplementary results, Supplementary Material online. We have made this genome annotation publicly available (see http://darencard.github.io/boaCon; last accessed October 22, 2019 and Figshare doi: 10.6084/m9.figshare.9793013) and use it as the basis for our inferences of functionally relevant genomic differences among island and mainland populations inferred from our sampling of these populations.

Independent Origins of Island Dwarf Boa Populations Support Convergent Phenotypic Evolution

Boas have colonized at least 43 islands across Central America, but the exact number of independent island colonization events is unknown. In a previous study, we demonstrated that island populations from Belize (Lagoon and West Snake Cays) and Honduras (Cayos Cochinos) cluster into distinct Central American clades with significant divergence (4–5 Myr; Card et al. 2016), yet it has remained unclear whether different islands in either Belize or Honduras represent distinct populations and thus independent origins of dwarfism. Our demographic analyses suggest independent colonization and subsequent evolution of dwarfism in the two Belize island populations, as well as the Honduran Cayos Cochinos population, where we found evidence of ongoing gene flow between the Cayo Cochino Menor and Cayo Cochino Mayor (also known as Grande; fig. 3). Our SNAPP analysis of dense RADseq sampling yielded a consensus population-level tree for which all nodes were resolved with 100% posterior support, suggesting that the two Belizean island populations (Lagoon and West Snake Cays) represent two independent colonizations from their respective mainland population (fig. 3*A*). Demographic model tests of Lagoon Cay and West Snake Cay using RADseq variants inferred a best-fit model consisting of population divergence without gene flow, further indicating that these two populations have evolved independent of one another (fig. 3*B*). Importantly, our demographic analysis only tests for the presence of gene flow following divergence between Lagoon Cay and West Snake Cay, assuming that dwarfism evolved independently in each island population. Our analysis thus ignores two possibilities: 1) that dwarfism evolved once in an island population and a subset of that island population recolonized the mainland, forming a paraphyletic pattern of island dwarfism and 2) that dwarfism evolved once after isolation from the mainland but before the two islands became isolated from one another. The first possible confounding scenario is highly unlikely, since any buildup on genetic divergence on an early island would almost certainly be overwhelmed by gene flow with larger, existing mainland populations upon recolonization and would not persist as a clearly defined population cluster. Moreover, because selection for reduced body size is only apparent on small islands with limited, canopy-dwelling prey species, it is unlikely that dwarfism evolved once in an ancestral population inhabiting an early insular landmass. The distance between Lagoon Cay and West Snake Cay is relatively large (~62 km), and this distance dictates that the landmass would have been quite large and likely capable of supporting prey communities like those found on the mainland, making the second possible confounding demographic scenario improbable. Additionally, such a demographic scenario implies a monophyletic relationship between Lagoon Cay and West Snake Cay, which is not supported by our phylogenetic inference in SNAPP. In contrast, the two Honduran islands (Cayo Cochino Menor and Mayor) were found to be sister to one another in the SNAPP tree (fig. 3*A*), and the best-fit model identified by $\delta a \delta i$ consisted of secondary contact with asymmetric gene flow between these two populations (higher migration rates from Mayor to Menor; fig. 3*C*), suggesting that these two islands represent a single dwarf lineage. In this circumstance, it is harder to

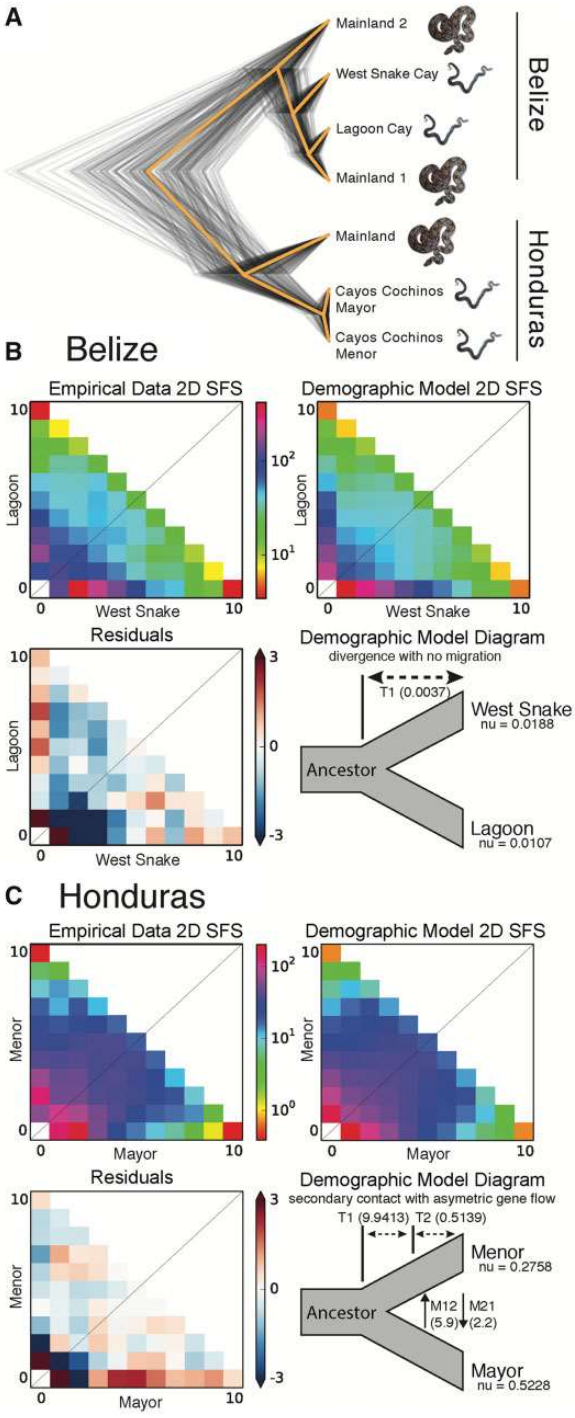

Fig. 3.—Demographic analysis of island populations establishes three independent instances of the evolution of dwarfism on islands. (*A*) DensiTree showing posterior topologies estimated from our RADseq data using SNAPP, with the consensus population phylogeny highlighted in orange. (*B*) Results of δaδi 2D SFS analysis of the RADseq data comparing plausible demographic relationships between Lagoon and West Snake Cays in Belize, which supports a model of divergence with no subsequent migration. (*C*) Results of δaδi 2D SFS analysis comparing plausible demographic relationships between the two Cayos Cochinos populations, which results in a best-supported model of ongoing gene flow between islands. In panels (*B*) and (*C*), site frequency spectra heatmaps are shown for the empirical data and simulated data based on the best-fit demographic model (top left and right, respectively). The residual heatmaps depict where the allele frequency spectra of the empirical and simulated differ (bottom left). A diagram depicting the best-supported, parameterized demographic scenario is sown at the bottom right. The inferred parameters *T* and *nu* reflect the timing of the demographic event in coalescent units (2*N* generations) and the effective population size in coalescent units (2*N* individuals), respectively.

rule out the possibility that dwarfism evolved before or after the two islands became geographically isolated from one another, as Cayo Cochino Menor and Major are quite close together (~2.5 km) and any previous landmass uniting the two likely would have been relatively small and ecologically similar to the modern islands.

These findings bring the confirmed number of independent dwarf boa populations to three: Lagoon Cay, West Snake Cay, and Cayos Cochinos. Considering the existence of additional island dwarf populations not sampled here in Belize (Boback 2005, 2006; Boback and Carpenter 2007) and elsewhere across Central America (Henderson et al. 1995; Porras 1999), three independent origins of island dwarf populations is likely the lower bound of the number of independently evolved island dwarf populations. Evidence for multiple independently evolved boa island populations with similar dwarfed phenotypes makes this system a particularly rich model to investigate the genetic basis of complex traits (e.g., body size and craniofacial morphology)—our analyses of genomic variation among island and mainland populations leverage these features to investigate links between molecular and phenotypic evolution that may explain the repeated evolution of similar island phenotypes.

Roles of Drift and Selection in Shaping the Evolution of Island Dwarf Populations

On islands, drift may have a particularly strong influence on population genetic variation due to the smaller population sizes typical of these populations, which is reflected in our estimated effective population sizes inferred using SNAPP (fig. 3). We estimated that the average allelic differentiation based on our RADseq data between each island and mainland population pair was variable across islands, with Cayos Cochinos being the most differentiated (median F_{ST} = 0.19), followed by Lagoon Cay (median F_{ST} = 0.03) and West Snake Cay (median F_{ST} = 0; fig. 4). Allelic differentiation between the mainland Belize (including both Belize populations) and Honduras populations was similar to that between island–mainland population pairs (median F_{ST} = 0.04), yet relatively small considering the geographic distance and divergence time between mainland populations compared with

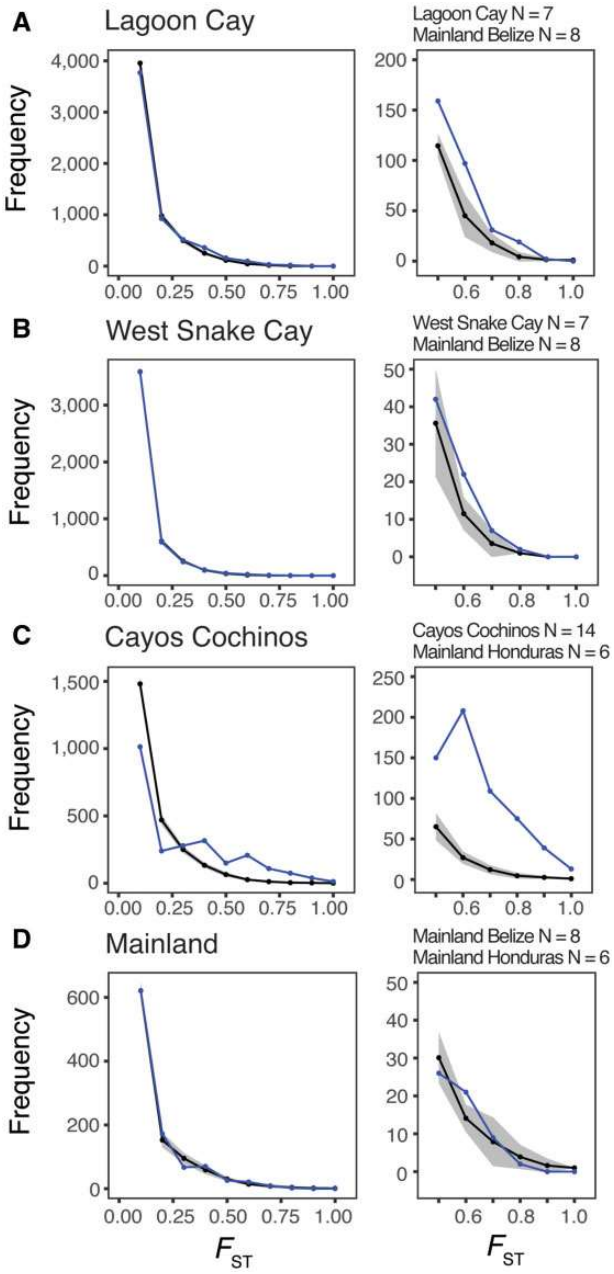

Fig. 4.—Evidence for genomic diversity stemming from natural selection versus neutral genetic drift in island populations. Panels present the distributions of F_{ST} values inferred from RADseq data from pairwise comparisons between island and mainland population pairs (*A*–*C*) and between the two mainland populations (mainland Belize samples combined; *D*). Sample sizes for each population in the comparison are indicated above the plots. Left-most panel provides the full F_{ST} distributions while the right panel focuses on F_{ST} values >0.5, which represent the most differentiated regions of the genome in each pairwise comparison. The black line and points represent the mean F_{ST} and the grey ribbons represent the 95% confidence interval that resulted from ten GppFst PPS runs. The blue line and points represent the empirical frequency of F_{ST} across bins. Statistically significant excess frequencies were observed in the bins with high F_{ST} values in comparisons between island and mainland island–mainland population pairs. Elevated measures of allelic differentiation observed between most island–mainland population pairs are consistent with the expected heightened effects of drift on these small populations compared with the mainland.

Considering evidence for the independent evolution of multiple island populations with convergent phenotypes, and the expectation that drift may be strong in island populations, we used our RADseq data to test for evidence that allelic differentiation of island *Boa* populations is due to natural selection, in addition to genetic drift. We conducted a simulation based on demographic information inferred from our high-resolution RADseq data set to understand neutral expectations of allelic differentiation that could be compared with our empirical results. Posterior predictive simulations (PPS), based on the neutral coalescent model using GppFst, suggested that genetic drift alone was capable of producing measures of F_{ST} as high as 0.75–1.0, depending on the specific island–mainland comparison. However, these extreme values were quite rare (i.e., <5% of PPS loci had F_{ST} >0.5; fig. 4). The 97.5% quantile threshold for empirical F_{ST} values varied from 0.35 to 0.75 among island–mainland comparisons (fig. 4 and supplementary table S9, Supplementary Material online). In each of these comparisons, the top 2.5% tail of empirical F_{ST} values in our RADseq data contained significantly more loci than expected given the simulated distributions of F_{ST} ($P < 0.05$ in all island–mainland comparisons). Using the simulated distributions of neutral F_{ST} values, we estimated that 16–52% of highly differentiated variants (F_{ST} > 0.50) in our empirical data could be explained by drift alone, suggesting the poor explanatory power of a strictly neutral model of divergence in generating high F_{ST} variants between island and mainland population pairs (fig. 4*A*–*C*). In contrast, we found no significant excess of variants in the top 2.5% tail of the F_{ST} distribution in the comparison of the two mainland populations (F_{ST} measured in Belize vs. Honduras mainland populations; $P = 0.42$), with the number of expected values due to drift almost exactly matching the number of observed loci (fig. 4*D*). Taken together, these results provide strong support for drift being a substantial driver of genetic divergence in island *Boa* populations, and yet drift alone does not appear to explain an appreciable fraction of highly differentiated regions of the genome between island and mainland populations. Instead, these results argue for combined roles of genetic drift and other processes (i.e., natural selection) in shaping patterns of genomic divergence

Fig. 4.—Continued
population pairs (*A*–*C*), while the same threshold did not yield excess frequencies in the comparison between mainland populations (*D*). These findings indicate that natural selection, on top of drift, has impacted allelic differentiation between island and mainland populations, but not between the two mainland populations.

in all three island populations of dwarf boas. These findings also suggest that even in relatively small island populations, where selection is expected to be less effective due to the dominating impact of drift, selection is also an appreciable force driving allelic differentiation—presumably due to the strength of selection on particular loci that may function in increasing the fitness of island boas to their similar island environments.

Considering that 16–52% of highly differentiated variants ($F_{ST} > 0.50$) in our empirical data could be explained by drift alone, many highly differentiated variants are likely false positives for evolving under the influence of natural selection. Previous studies have demonstrated that multivariate measures show increased power for detecting signals of selection (Lotterhos et al. 2017). In an effort to further distinguish true and false positive signals of selection, we used the software package MINOTAUR (Verity et al. 2017) to estimate a multivariate Mahalanobis distance measure based on measures of F_{ST} and three other univariate measures of nucleotide evolution between island and mainland populations: the absolute differences of the change in nucleotide diversity (π), Tajima's *D*, and observed heterozygosity, taking covariation in these nonindependent measures into account. We found that outlier measures of Mahalanobis distance (top 2.5% tail of each island–mainland comparison distribution) largely overlap with F_{ST} outliers, though unique loci are also identified among Mahalanobis outliers (supplementary fig. S4, Supplementary Material online). These results indicate that our F_{ST}-based selection scans are detecting true signals of selection in many cases, as least based on secondary evidence from multivariate measures. Considering these results and that F_{ST} is a direct measure of allelic differentiation between populations, we focused primarily on the results of our F_{ST} in downstream analyses. Moreover, we focused on shared patterns of high allelic differentiation across all three replicated island lineages, where false signals of selection are unlikely to persist.

Identifying Shared Patterns of Molecular Evolution across Island Populations

To begin to identify regions of the genome potentially linked to convergent island phenotypes, we searched for genomic regions with extreme allelic divergence between island and mainland populations and examined whether these regions overlapped in multiple island populations. Our RAD variant data showed some evidence for shared patterns of high allelic differentiation among islands, which only occurred in the comparison between Lagoon and West Snake Cays (supplementary fig. S5*A*, Supplementary Material online). In our empirical comparison, 11 highly differentiated loci (3.3% of highly differentiated island loci) were identified in both Lagoon and West Snake Cay, which greatly exceeds what is expected by random chance (supplementary fig. S5*B*, Supplementary Material online). These results suggest that drift alone is unlikely to adequately explain such a high degree of overlap in loci with extreme allele fluctuation between multiple island populations in Belize. These 11 highly differentiated loci were spread across 9 genomic scaffolds. Scaffold 1273 contained two highly differentiated loci separated by ~146 kb while scaffold 3122 contained two highly differentiated loci separated by ~2.8 Mb. About 31 genes were located within 100 kb of these 11 highly differentiated loci, of which 29 were confidently annotated with human gene IDs (supplementary file S10, Supplementary Material online). About 53 mouse phenotypes showed enrichment (FDR-corrected *P* value <0.05) based on these 29 genes, with several enriched phenotypes linked to craniofacial morphology (supplementary file S11, Supplementary Material online).

To further explore potential convergence across populations, we used our WGS data set derived from 20 individuals sampled from island and mainland populations. Similar to our approach with the RADseq data analysis, we used a genome-wide windowed approach to identify 10 kb regions of the genome with extreme allele frequency fluctuations (maximum is >0.90) between island and mainland populations. With the higher resolution of our WGS data, we found 4,278, 3,848, and 6,887 10-kb genomic regions with extreme allelic fluctuations (≥0.90) in the Lagoon Cay, West Snake Cay, and Cayos Cochinos populations, respectively, and 6,678 such regions between the two mainland populations. We found 238 shared regions between Lagoon and West Snake Cays, 285 between Lagoon Cay and Cayos Cochinos, and 259 between West Snake Cay and Cayos Cochinos (fig. 5*A*). For all interisland comparisons, the degree of overlap in genomic windows was significantly higher than expected based on randomly permutated data sets (fig. 5*B*), indicating that these genomic windows occurred at the same places in the genome more frequently than expected by chance among island populations.

Candidate Genes and Protein-Coding Variants Related to Convergent Island Phenotypes

Our analysis of highly differentiated genomic regions from our WGS data revealed evidence for extreme island–mainland allele frequency differentiation in 20 genomic regions (i.e., 10 kb windows) shared between all three island populations. These 20 genomic regions were located across 14 genomic scaffolds. In two cases, multiple regions were located in close proximity: five regions were located in an 80-kb area on scaffold 509 and two regions were located in a 50-kb area on scaffold 739. In one instance (scaffold 2231), two genomic regions were located distantly on the same scaffold (1.65 Mb separating the regions). About 47 genes were located within 100 kb of these genomic regions, including 42 that were confidently annotated with human gene IDs (supplementary file S12, Supplementary Material online). About 11 mouse phenotypes showed enrichment (FDR-corrected *P* value <0.05) based on

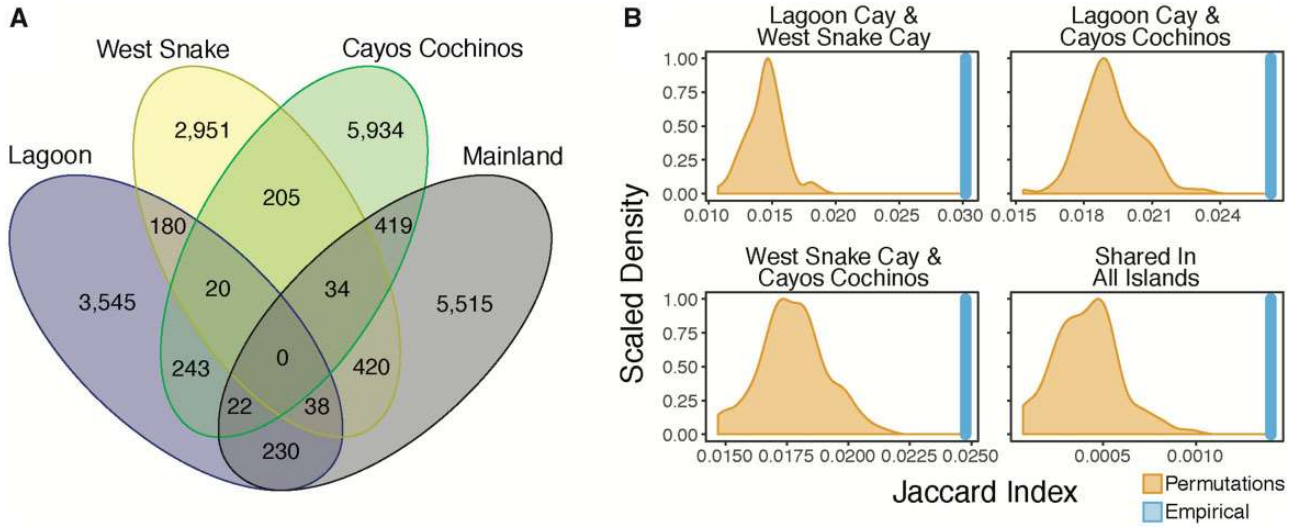

Fig. 5.—About 10-kb WGS windows with extreme fluctuations in allele frequencies in island populations are shared between islands. (*A*) Venn diagram summarizing the overlap of 10 kb genomic windows with extreme (≥0.90) allele frequency fluctuation between an island and its associated mainland population (labeled by island name) or between the two mainland populations (labeled "Mainland") based on WGS data. (*B*) Permutation analyses indicate that the empirical number of windows with extreme allelic fluctuation in pairwise or all three island populations is higher than expected by chance. In each panel, the permutation density distribution of the Jaccard index is shown in orange and the empirical Jaccard index is depicted with a blue vertical line.

these 47 genes (supplementary file S13, Supplementary Material online), though only one (MP: 0030384: short facial bone) showed an obvious link to one of the phenotypes examined in this study (craniofacial morphology). The paucity of genes enriched for mouse phenotypes that are linked to observed island phenotypes could be due to the small number of genes available for the enrichment analysis.

We surveyed our WGS data for potentially phenotypically relevant protein-coding variation in island–mainland comparisons within and around these 20 genomic regions as a means of identifying candidate genes with evidence of penetrant coding variation with relatively high likelihoods of being phenotypically relevant in an interpretable way. We identified four annotated genes within these regions with nonsynonymous allelic variants in island populations. Three of these genes contained coding variants with extreme (0.75 or greater) allele frequency fluctuations in at least one population and deleterious functional impacts (as assessed by VEP and PROVEAN): protein tyrosine phosphatase, receptor type S (*PTPRS*), myosin regulatory light chain interacting protein (*MYLIP*), and dimethylglycine dehydrogenase (*DMGDH*). *PTPRS* and *DMGDH* each contain a variant with high allele frequency fluctuation in West Snake Cay, and we found a variant in *MYLIP* with high allele frequency fluctuation in West Snake Cay and Cayos Cochinos (with modest variation in the mainland). A fourth gene, arylsulfatase B (*ARSB*) is the single instance of an allele exhibiting high-frequency fluctuations across all island populations that are absent in the mainland–mainland comparison (i.e., nonreference alleles are fluctuating at high frequency in island versus mainland populations). In some cases, this fluctuation was less extreme, but was still >0.5 allele frequency change between island and mainland populations. These four genes are excellent candidates for explaining key phenotypic traits that differentiate island boa populations, including their unique dwarf phenotypes, craniofacial morphology, and slender body form. Below, we describe the characteristics of each of these genes and their links to key island phenotypes, while integrating these findings with existing knowledge of genes and pathways impacting body size and craniofacial morphology.

Links between Regulation of the IGF-1/GH Pathway to Dwarfism in Island Boa Populations

Among the four identified candidate genes, both *PTPRS* and *DMGDH* play a role in regulating the Insulin-like Growth Factor/Growth Hormone (IGF-1/GH) pathway, an important regulator of vertebrate growth (Baker et al. 1993). The knockout of *PTPRS* in mice causes a significant reduction in circulating levels of insulin-like growth factor 1 (IGF-1) and growth hormone (GH; Elchebly et al. 1999; Batt et al. 2002). Accordingly, mice *PTPRS* null mutants exhibit reduced body size and weight, general retardation of growth, and decreased litter size (Elchebly et al. 1999). In the West Snake Cay population, *PTPRS* contains an indel segregating at high frequency that results in a frameshift mutation at protein residue 222, which was not observed in any other island or mainland population (figs. 6*A* and (7*A*); this frameshift variant was classified as high-impact by VEP. A second moderate-impact variant was observed at protein residue 434 and results in an Alanine to Valine substitution, which is at high frequency in the Lagoon Cay population and segregates at a

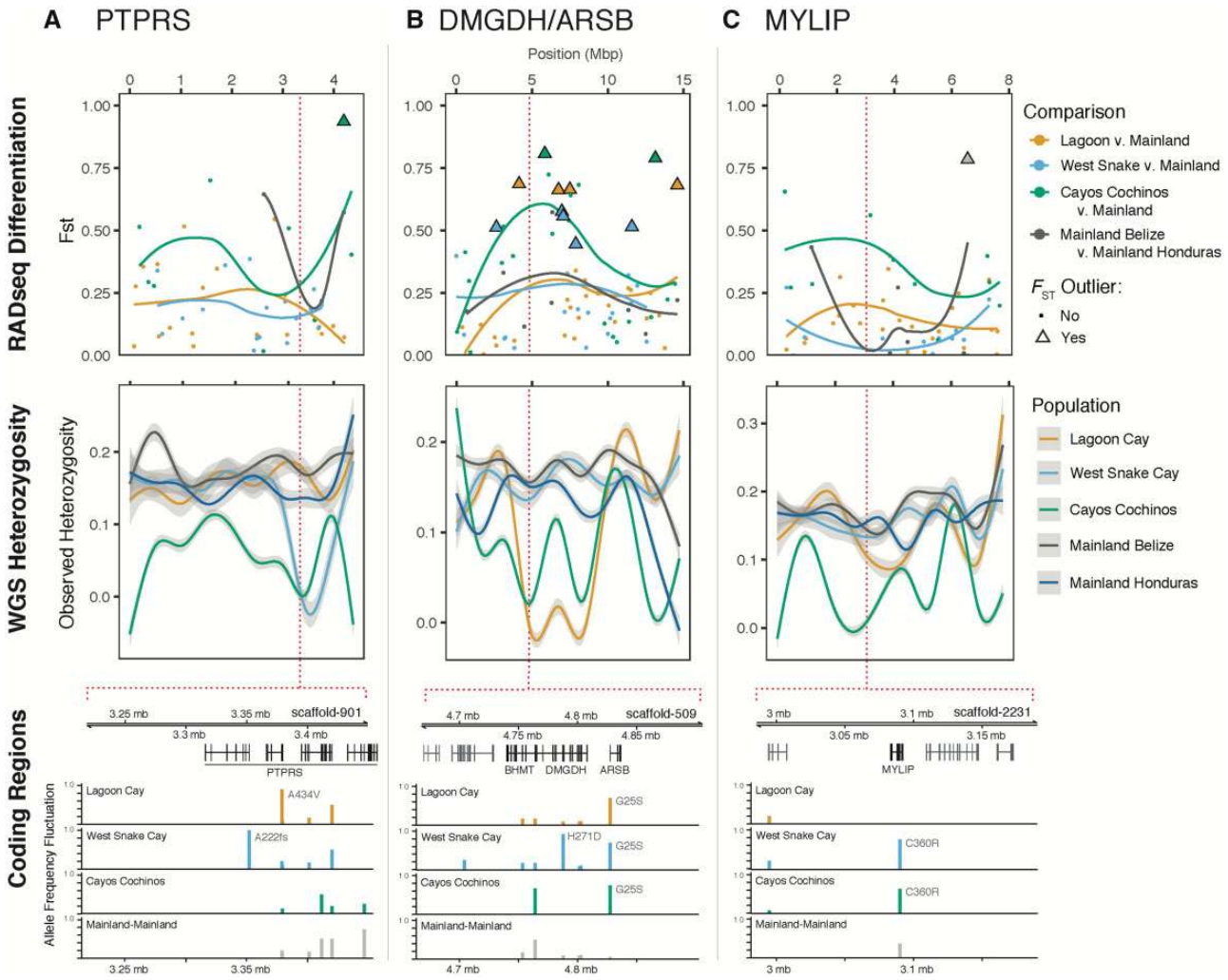

Fig. 6.—Genomic variation surrounding genes putatively underlying island traits. Each column represents the genomic region surrounding (*A*) *PTPRS*, (*B*) *DMGDH* and *ARSB*, and (*C*) *MYLIP*. The first row for each region depicts allelic differentiation in island–mainland population pairs and between mainland populations in Belize and Honduras based on RADseq data, with F_{ST} measurements indicative of selection (i.e., above the 97.5% quantile) indicated as triangles. Colored lines depict loess-smoothed trendlines (span = 1) showing comparison-specific trends across the entire genomic scaffold. The second row shows scaffold-wide observed heterozygosity in individual populations from WGS variants at loci across scaffolds with functionally relevant genes. The trendlines are based on a generalized additive model with the formula $y \sim s(x, \text{bs}=\text{"cs"})$, and the grey areas represent the 95% confidence interval for each trendline. The third row depicts the allele frequency fluctuation of coding sequence variation between island–mainland population pairs. Tracks at the top of each coding region panel summarize the genomic scale of the local regions and the gene models for the four focal genes. Relevant protein substitutions discussed in the main text are indicated using standard notation.

frequency of 0.1 in the mainland Belize population (figs. 6*A* and 7*A*; supplementary table S10, Supplementary Material online). Although classified as moderate-impact based on VEP, this second variant had a nondeleterious PROVEAN score of −0.012. The region of the genome containing *PTPRS* also contains a RAD locus located ~840 kb downstream from *PTPRS* with an exceptionally high F_{ST} that is in the top 2.5% of F_{ST} values identified as statistically significant in the Cayos Cochinos population based on our PPS analyses (figs. 6*A* and 7*A*). The genomic region around *PTPRS* also shows particularly low relative heterozygosity in both the West Snake and Cayos Cochinos populations—2 Mb regions surrounding *PTPRS* average 54% and 63% of the genome-wide average heterozygosity, respectively (figs. 6*A* and 7*A*). Together, our results suggest that distinct island-specific alleles in or adjacent to *PTPRS* may contribute to altering the function of this gene in island populations and driving the observed island dwarf phenotypes.

In addition to *PTPRS*, alleles of *DMGDH* demonstrate patterns of variation that may have functional, and potentially synergistic, impacts on development and growth of island boas. *DMGDH* functions in the catabolism of choline, and a

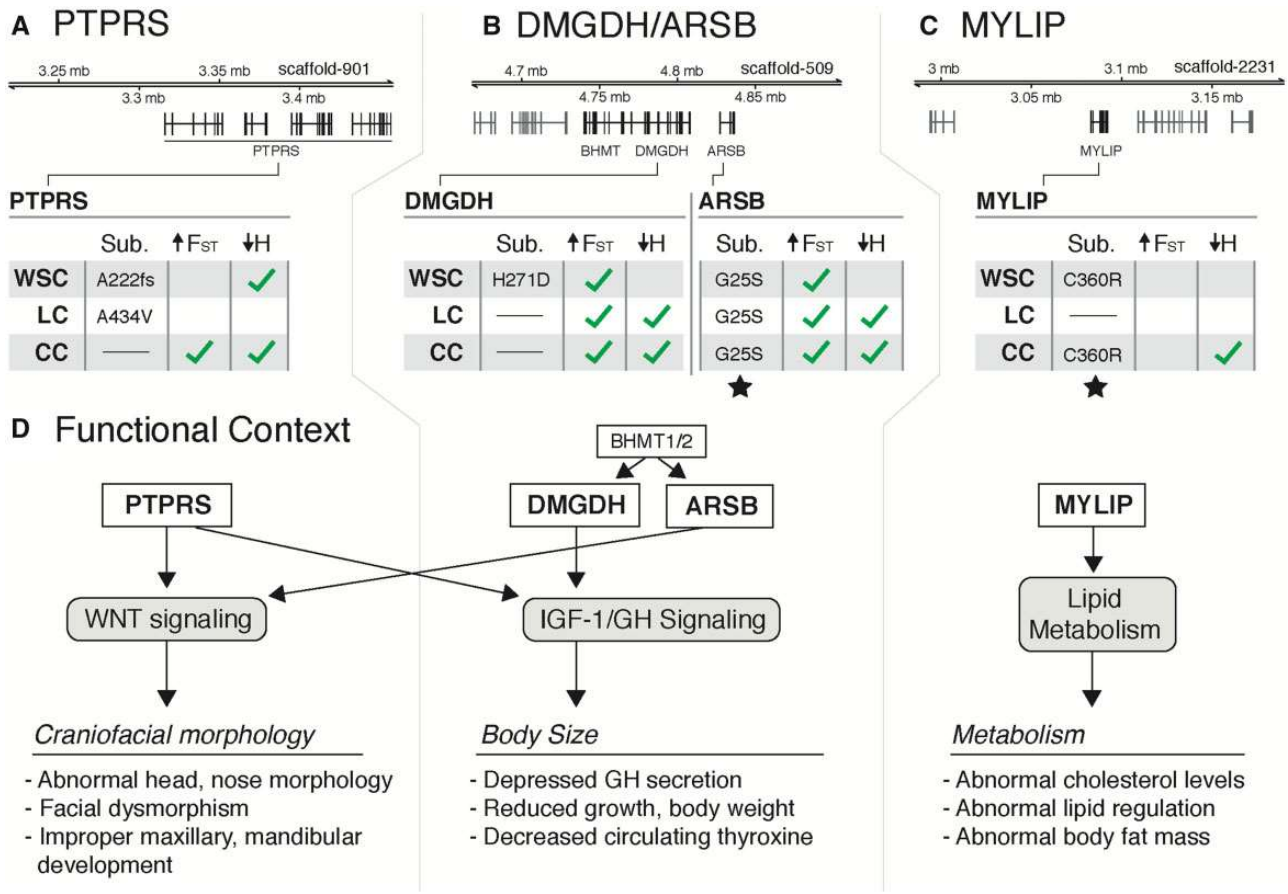

Fig. 7.—Summary of evidence for signatures of selection in phenotypically relevant gene regions and the broader functional context tying these genes to island phenotypes. As with figure 6, each column represents the genomic region surrounding (*A*) *PTPRS*, (*B*) *DMGDH* and *ARSB*, and (*C*) *MYLIP*. The top tracks indicate the genomic scale of the local regions and gene models for the four focal genes (see also fig. 6). For each gene, tables summarize amino acid or frameshift (fs) substitutions coding by nonsynonymous variants on islands (WSC, West Snake Cay; LC, Lagoon Cay; CC, Cayos Cochinos). Additional columns and green checkmarks in each table indicate whether these protein-coding regions had high allelic differentiation (i.e., RADseq F_{ST} values above the 97.5% quantile) or low heterozygosity based on the WGS data set. Stars beneath substitutions indicate that the substitution was found and varied greatly in allele frequency in two or more island populations. (*D*) Schematics of the functional context and interactions of phenotypically relevant genes in developmental signaling pathways with demonstrated importance in craniofacial morphology, body size, and fat/triglyceride regulation.

loss of function mutation in this gene in mice leads to decreased circulating thyroxine (Smith et al. 2018), resulting in depressed GH secretion, suppressed growth, and reduced body weight (Root et al. 1986; Amit et al. 1991; Choi et al. 2018). We found three nonsynonymous coding variants in *DMGDH* (protein residues 271, 585, and 667), although only one (protein residue 271) has a particularly deleterious PROVEAN score of −4.628 (figs. 6*B* and 7*B*). This variant results in a Histidine to Aspartic Acid substitution, which shows very high allelic differentiation in the West Snake Cay population (the nonreference allele is at high frequency on West Snake Cay, yet segregates at 0.083 in mainland Belize), but very low allelic differentiation in the other island–mainland comparisons (figs. 6*B* and 7*B*; supplementary table S10, Supplementary Material online). Although only the West Snake Cay population shows high allelic frequency shifts for the nonsynonymous *DMGDH* allele, the genomic region surrounding *DMGDH* contains a high density of putatively selected RAD-based variants (i.e., variants with F_{ST} values in the upper 2.5% quantiles) in all three island populations, and this region is characterized by particularly low heterozygosity in both Lagoon Cay and Cayos Cochinos populations (21% and 71% of genome-wide average heterozygosity, respectively; figs. 6*B* and 7*B*). Similar to *PTPRS*, these results suggest that different island populations have experienced selection for different *DMGDH* alleles, some of which are highly penetrant coding variants while others are not.

Selection on *IGF-1* alleles is known to impact body size in dogs (Sutter et al. 2007) and humans (Becker et al. 2013), and modulation of the function of this pathway appears to represents a recurrent target for selection in vertebrates. Our finding that multiple distinct alleles for *PTPRS* and *DMGDH* are

associated with dwarf island populations, together with evidence that selection may play a role at these loci in multiple island populations, provides an example of independent allelic solutions that may result in convergent signaling outcomes (i.e., modulation of IGF-1/GH pathway) leading to convergent dwarfism phenotypes. Neither *PTPRS* nor *DMGDH* are currently known to be associated with sex-biased phenotypes and thus it remains an open question whether either of these contributes to the patterns of sexual dimorphism evident in island populations. Collectively, our results supplement existing work on the genetics underlying body size that indicates many genes of large effect have some regulatory role in the IGF-1/GH pathway (Sutter et al. 2007; Becker et al. 2013).

The Potential Role of Wnt Signaling in Craniofacial Morphology of Island Boa Populations

Island boas possess unique snout attenuation, head width, and eye size compared with mainland populations (fig. 1*E*). These phenotypic traits are likely linked to the unique arboreal habits and hunting behavior of these island populations (Shine 1983; Lillywhite and Henderson 2002). Wnt signaling has been implicated in craniofacial development in many systems (Schmidt and Patel 2005; Brugmann et al. 2007, 2010; Kurosaka et al. 2014), and two genes identified in our analysis, *PTPRS* and *ARSB*, are known to have impacts on the Wnt pathway. In addition to the roles *PTPRS* can have on growth (discussed above), loss of *PTPRS* function in mice also causes alterations to BMP and Wnt signaling pathways, resulting in improper maxillary and mandibular development and changes to craniofacial morphology (Stewart et al. 2013). Accordingly, nonsynonymous variants observed in *PTPRS* in the two Belize island populations (Lagoon and West Snake), and evidence for selection in the Cayos Cochinos and West Snake Cay populations, may also be linked to phenotypic effects on craniofacial morphology through genomic variation in *PTPRS* via its interaction with Wnt signaling (figs. 6*A* and 7*A*).

A second gene, *ARSB*, is also involved in Wnt signaling related to craniofacial phenotypes, as well as in cell signaling that effects body size and mass. *ARSB* is associated with abnormal caudal vertebrae morphology, head and nose morphology, fat/triglyceride levels, and decreased birth and adult body size in mice (Smith et al. 2018), and is genetically linked to another of our four candidate genes, *DMGDH*, in most vertebrates; these two genes are located in close proximity to one another in the boa genome (~20 kb; figs. 6*B* and 7*B*). Reduced expression of *ARSB* has been linked to downstream increases in Wnt/β-catenin signaling (Bhattacharyya et al. 2017) through a proposed interaction with LDL-receptor-related protein 5/6 (Kawano et al. 2006; Veeck and Dahl 2012; Ueno et al. 2013). *ARSB* is the causative gene for the human disorder mucopolysaccharidosis type VI (Maroteaux Lamy Syndrome), which is associated with short stature and with facial dysmorphism (Azevedo et al. 2004). Similar phenotypes caused by mutations to an orthologous gene have also been noted in dogs (Wang et al. 2018).

Our results suggest that convergent molecular evolution at the amino acid level in and around *ARSB* may underlie some aspects of phenotypic convergent evolution across all three island populations. All island population-specific *ARSB* alleles contain a nonsynonymous Glycine to Serine substitution classified as moderate-impact by VEP, but with a nondeleterious PROVEAN score of 0.467. The Serine residue appears nearly at high frequency in all three island populations but segregates at 0.25–0.30 in both mainland populations (figs. 6*B* and 7*B*; supplementary table S10, Supplementary Material online). Given the close genomic proximity of *ARSB* and *DMGDH*, the region encompassing these two genes shares characteristics of genetic variation (discussed above) and evidence of selection acting on this region in multiple island populations, including a high density of putatively selected variants in all three island populations and low heterozygosity in both the Lagoon Cay and Cayos Cochinos populations (17% and 71% of genome-wide average heterozygosity in each population, respectively; figs. 6*B* and 7*B*).

Taken together, evidence for molecular convergence at the amino acid level and regional genomic patterns consistent with selection in all three island lineages implicate *ARSB* as a likely driver of convergent island phenotypes. Island-specific genomic variation patterns associated with both *ARSB* and *DMGDH* also suggest that Wnt signaling may represent an important nexus for molecular and pathway-level convergence mediating adaptation and phenotypic evolution of island populations. This conclusion is also consistent with previous studies that have found that Wnt signaling underlies adaptive craniofacial variation in the rapid evolution of cichlid fish in Lake Malawi (Parsons et al. 2014). Craniofacial variation in the African lake cichlid adaptive radiation is hypothesized to be driven by trophic adaptation, representing a key example of how evolutionary variation in Wnt signaling may underlie trophic adaptation. Similarly, craniofacial shifts in island boas appear to be driven by the unique, arboreal feeding ecology of snakes in these populations (Lillywhite and Henderson 2002; Boback 2005), and thus further highlight the broad potential for evolutionary variation in Wnt signaling to drive rapid trophic adaptation in vertebrates.

Links between Lipid Metabolism and Reduced Body Mass in Island Boas

The rarity and seasonality of prey, together with the less massive, slender phenotypes of island boas suggest that substantial differences in metabolism and fat storage may be a shared feature of island populations. The fourth candidate gene identified by our WGS analysis, *MYLIP*, plays a role in regulating lipid metabolism and body mass. Human GWAS studies have identified *MYLIP* in screens for low-density lipoprotein

cholesterol and total cholesterol (Weissglas-Volkov et al. 2011; Global Lipids Genetics Consortium et al. 2013; Surakka et al. 2015), and mice with null mutations in *MYLIP* show a number of phenotypes, including those linked to cholesterol levels, lipid regulation, and body fat mass (Smith et al. 2018).

Our comparisons of island and mainland boa populations identified a nonsynonymous coding variant (protein residue 360) in *MYLIP* with relatively high shifts in allele frequency on West Snake Cay (0.75 allele frequency shift from mainland Belize) and Cayos Cochinos (0.625 allele frequency fluctuation versus mainland Honduras; figs. 6*C* and 7*C*; supplementary table S10, Supplementary Material online). This nonsynonymous variant is classified as moderate-impact by VEP and has a deleterious PROVEAN score of −3.269. The scaffold containing *MYLIP* did not contain any putatively selected variants in our RAD data set in any island populations, although heterozygosity in the Cayos Cochinos population is moderately reduced in this region (2 Mb region surrounding MYLIP has an average heterozygosity that is 42% of the genome-wide average; figs. 6*C* and 7*C*). Our finding that a nonsynonymous *MYLIP* allelic variant with inferred deleterious impacts has convergent, high allele frequency fluctuation in both the West Snake Cay and Cayos Cochinos populations suggests that *MYLIP* may also be relevant in mediating island-specific phenotypes, such as body mass, fat storage, or fat metabolism. However, the lack of strong evidence for selection acting on this region in island populations raises the question of whether drift may have driven the elevated frequency of this variant in island populations, or if we instead failed to detect selection due to a lack of power (due to limited sampling or the age or strength of selection).

Genetically and Functionally Linked Gene Sets May Evolutionary Tune Island Phenotypes

The genomic and functional characteristics of our candidate genes, and associated variants, highlight the potential role of genetic linkage and overlapping functional interactions in driving rapid phenotypic convergence through modulation of relatively few genes. Two candidate genes for island phenotypes, *DMGDH* and *ARSB*, are found in close proximity in vertebrate genomes, including that of boas (fig. 6*B*), and this genomic region also contains two betaine—homocysteine S-methyltransferase genes adjacent to *DMGDH*: *BHMT* and *BHMT2* (figs. 6*B* and 7*B*). The *DMGDH*, *BHMT*, and *BHMT2* complex is associated with modulation of plasma betaine levels in humans (Hartiala et al. 2016). Betaine and choline also regulate insulin sensitivity, fat deposition, and energy metabolism (Millard et al. 2018). This group of three genes therefore plays an important role in physiological processes that impact body growth and fat deposition, two key traits that differentiate island and mainland boas. The fourth gene in this region, *ARSB*, plays no apparent role in betaine metabolism, but does impact body size and craniofacial morphology through alternative mechanisms. However, variation in the nearby *BHMT* gene is associated with differential methylation of *ARSB*, which can modulate *ARSB* expression (Lupu et al. 2017). The functional interplay between genes in this region and the collective ability of these genes to potentially alter a broad spectrum of distinctive traits that characterize island boas—body size, craniofacial morphology, and fat metabolism—suggests that this linked gene cluster may be an important conserved target of rapid adaptation to modulate island-specific traits and phenotypes broadly. Indeed, the strong signatures of selection in this region based on both RADseq and WGS further support the conclusion that variation in this genomic region plays a role in convergent island phenotypes in boas. Such genomic regions that contain functionally interrelated genes—referred to as "supergenes" in certain contexts (Thompson and Jiggins 2014)—have been shown to be important for other adaptive traits including self-incompatibility in plants (Takayama and Isogai 2005), assortative mating in white-throated sparrows (Thomas et al. 2008; Tuttle et al. 2016), and mimicry in butterflies (Joron et al. 2011; Kunte et al. 2014). Our results suggest that the *DMGDH–BHMT–ARSB* region could also be included in this emerging list of functionally dense targets for rapid phenotypic adaptation in vertebrates. Further work is needed to more definitely link genetic variation in this region to phenotypic differences between island and mainland populations, and follow-up functional assays would be valuable to determine if this region may be capable of modulating a broad spectrum of phenotypes with only a small number of mutations.

Experimental Considerations for Future Research on Convergent Island Phenotypes in Boas

Our results provide an exciting, yet preliminary, perspective on the potential connections between genotypic and phenotypic evolution across distinct island boa populations, and highlight the value of this island boa system for studying the genetic basis of complex traits and the propensity for molecular convergence. To more fully leverage this system, future studies would benefit from expanded sampling to increase power to detect and understand genetic shifts occurring across one or more island populations. In our case, the use of RADseq data facilitated the economy to sample more individuals to better quantify the relative contributions of drift and selection to the evolution of island populations but lacked power to identify causal genetic variation. Moreover, WGS provided some power to dissect underlying causal genetic variation, but at a cost that limited our ability to sample many individuals. Thus, expanding WGS sampling to include greater numbers of individuals from an increased number of populations would provide substantially increased power to uncover across-island and island-specific evolutionary patterns that contribute to both convergent and divergent island phenotypes. In

particular, additional sampling of other dwarf island populations not included in this study would further leverage the power of the natural replication of this system (Henderson et al. 1995; Boback 2005).

In this study, we primarily focused on the role of protein-coding variation in shaping patterns of convergence and divergence among mainland–island boa populations. Thus, a clear limitation of our study is a lack of insight into noncoding, regulatory variation that may play a fundamental role in shaping phenotypes. Previous studies indicate that changes in *cis*-regulatory elements (i.e., enhancers) are important for producing new gene expression patterns that can impact phenotypes (reviewed in Carroll 2008 and Wray 2007). However, identifying enhancers and associating them with gene expression is difficult due to the fact that these regions are relatively small and can be located up to 1 Mb away from the transcription start sites of genes that they regulate. Moreover, there is currently remarkably little knowledge about regulatory elements in reptiles, which led us to forego more focused analyses of regulatory evolution that very likely serves as a target for selection in island boa populations. Similarly, previous studies have documented instances of TE proliferation seeding regulatory switches important for the evolution of new, complex traits (Wagner and Lynch 2010; Sundaram et al. 2014; Chuong et al. 2017). Patterns of TE abundance and evolution may therefore be a nontrivial mechanism for rapid phenotypic evolution in small island populations, where the efficacy of purifying selection is reduced and could result in proliferation of active TE families. Indeed, analysis of copy-number changes of TE families across island populations in relation to relevant mainland populations in putatively selected versus neutrally evolving regions found significantly higher numbers of Maverick DNA transposons in the Lagoon Cay population and TcMar-Tigger DNA transposons in the West Snake Cay population (Bonferroni-corrected $P < 0.05$; supplementary table S11, Supplementary Material online). However, while this general pattern supports the possibility that TEs could play a role in island evolution, determining the penetrance (i.e., impact on phenotype) of TE insertions, and mutations in general, linked to regulatory regions is far less straight-forward than it is in protein-coding regions. Expanding our understanding of the presence and evolutionary dynamics of regulatory regions in nonavian reptiles, and especially boas, would be an exciting step forward that would enable more thorough interrogation of the role of regulatory elements, and genomic change overall, has played in the evolution of convergent and divergent phenotypic changes across islands.

Conclusion

Although examples of convergent evolution provide some of the most compelling demonstrations of natural selection in action, evidence of phenotypic convergence is relatively uncommon and molecular convergence is even more rare (Losos 2011; Stern 2013; Rosenblum et al. 2014). Questions therefore remain about the relative contribution and magnitude of natural selection and convergent molecular evolution in the context of phenotypic convergence. Given the general paradigm that there is a preponderance of genetic "solutions" for a phenotype, molecular convergence is not expected, but our results contribute to a growing body of literature challenging this assertion (Castoe et al. 2009; Hohenlohe et al. 2010; Jones et al. 2012). Our results suggest that genetic variation in dwarf island populations of boas is predominately shaped by drift, while signals of selection and convergence are less common, but nonetheless detectable in particular regions of the boa genome. We found that, in many cases, a simple model of drift alone was sufficient to explain high allelic differentiation without needing to invoke selection. Similarly, shared patterns of high allelic differentiation among islands were observed frequently at a genome-wide scale, highlighting the difficulties of teasing apart selection versus drift and further underscoring the importance of proper null model specification when conducting genome scans for selection. Although most of the genome of island boas may be diverging neutrally (or nearly so) from their mainland counterparts, our analysis of replicate sampling of island boa populations did detect signals of selection and convergence in three genomic regions with functional links to the unique biology of dwarf island *Boa* populations. Our results also provide an example of how variation in a small number of functionally interconnected genomic regions may have the potential to drive major shifts in phenotypes and facilitate rapid adaptation. More specifically, some genetically and functionally linked gene complexes (e.g., the *BHMT–DMGDH–ARSB* locus), in which we find strong evidence for selection and convergence across island populations, may be important for rapid adaptation and broad phenotypic evolution.

Phenotypic variation in adaptive traits replicated across island taxa presents a powerful opportunity to better understand how both convergent and divergent genetic changes propagate through molecular pathways to alter complex phenotypes. The results of this study highlight the utility of island systems in general, and island boas in particular, as "natural experiments" for addressing fundamental questions about adaptation and links between molecular and phenotypic evolution. Indeed, the replication inherent in island systems across a range of phylogenetic scales provides significant power to examine associations between genotype and phenotype. However, despite the ubiquity of insular body size variation across many diverse taxa, little work has addressed the underlying genetic basis of this phenotypic shift. Our results suggest that, despite similar phenotypes across island populations, convergent phenotypic evolution is largely driven by unique and island-specific evolutionary trajectories rather than dominated by convergent molecular evolution. Indeed,

we only detected convergent patterns in alleles with frequency fluctuations across all three island populations for one protein-coding variant. However, convergent signals of selection in other regions of the genome may be targeting regulatory variation. Determining the precise location, impact on gene expression, and penetrance (i.e., impact on phenotype) of putative *cis*-regulatory changes requires significant functional studies that are currently largely impractical in non-model reptilian species and thus beyond the scope of this study. Despite not interrogating signals of adaptation in regulatory regions, our results do provide an important narrative on convergence at the level of genes, genomic regions, and functional pathways. Specifically, we find evidence of selection in the same genomic regions across multiple island populations, suggesting that evolutionary changes to the function of the same sets of genes—but through different alleles and substitutions across island populations—underlie convergent island phenotypes. This study also provides candidate genes that may be important across vertebrates in modulating major phenotypic shifts observed in island populations. Our findings providing a valuable starting place to search for evidence across other vertebrate lineages for these implicated genes and pathways being relevant for explaining how organisms respond to the novel selective regimes of islands.

Supplementary Material

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

Acknowledgments

This work was supported by an NSF Award to T.A.C. (NSF IOS-1655735), an NSF Doctoral Dissertation Improvement Grant award (NSF DEB-1501747) to D.C.C. and T.A.C., a Beta Phi chapter of the Phi Sigma society to D.C.C., and a Lewis and Clark Award from the American Philosophical Society to D.C.C. We thank the numerous researchers at the University of Texas Arlington who collected and deposited *Boa* tissue samples at the Amphibian and Reptile Diversity Research Center, which formed the basis of this work. We thank A. Carnes for assistance in the field and the staff of the Honduran Coral Reef Fund for facilitating field work in Cayos Cochinos.

Author Contributions

D.C.C. and T.A.C. conceived the project; D.C.C., B.W.P., A.B.C., M.J.V.K., J.M.D., W.B., C.E.M., S.M.B., and T.A.C. collected the data; D.C.C., D.R.S., B.W.P., R.H.A., G.I.M.P., and K.R. analyzed the data; and D.C.C. and T.A.C. wrote the article. All authors reviewed the article and approved its submission.

Literature Cited

Adams RH, Schield DR, Card DC, Blackmon H, Castoe TA. 2017. GppFst: genomic posterior predictive simulations of F_{ST} and d_{XY} for identifying outlier loci from population genomic data. Bioinformatics 33(9):1414–1415.

Adams RH, Schield DR, Card DC, Corbin A, Castoe TA. 2018. ThetaMater: Bayesian estimation of population size parameter θ from genomic data. Bioinformatics 34(6):1072–1073.

Adler GH, Levins R. 1994. The Island syndrome in rodent populations. Q Rev Biol. 69(4):473–490.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Amit T, et al. 1991. Effects of hypo or hyper-thyroidism on growth hormone-binding protein. Clin Endocrinol. 35(2):159–162.

Azevedo A, et al. 2004. Clinical and biochemical study of 28 patients with mucopolysaccharidosis type VI. Clin Genet. 66(3):208–213.

Baker J, Liu J-P, Robertson EJ, Efstratiadis A. 1993. Role of insulin-like growth factors in embryonic and postnatal growth. Cell 75(1):73–82.

Batt J, Asa S, Fladd C, Rotin D. 2002. Pituitary, pancreatic and gut neuroendocrine defects in protein tyrosine phosphatase-sigma-deficient mice. Mol Endocrinol. 16(1):155–169.

Becker NS, et al. 2013. The role of *GHR* and *IGF1* genes in the genetic determination of African pygmies' short stature. Eur J Hum Genet. 21(6):653–658.

Benjamini Y, Hochberg, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach for Multiple Testing. J R Stat Soc B. 57(1):289–300.

Bhattacharyya S, et al. 2017. Chondroitin sulfatases differentially regulate Wnt signaling in prostate stem cells through effects on *SHP2*, phospho-*ERK1/2*, and Dickkopf Wnt signaling pathway inhibitor (*DKK3*). Oncotarget 8(59):100242–100260.

Boback SM. 2003. Body size evolution in snakes: evidence from island populations Montgomery. Copeia 2003(1):81–94.

Boback SM. 2005. Natural history and conservation of island boas (*Boa constrictor*) in Belize. Copeia 2005(4):879–884.

Boback SM. 2006. A morphometric comparison of island and mainland boas (*Boa constrictor*) in Belize. Copeia 2006(2):261–267.

Boback SM, Carpenter DM. 2007. Body size and head shape of island *Boa constrictor* in Belize: environmental versus genetic contributions. In: Henderson RW, Powell R, editors. Biology of the boas and pythons. Eagle Mountain (UT): Eagle Mountain Publishing. p. 102–117.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

Bradnam KR, et al. 2013. Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species. GigaScience 2(1):1–31.

Brugmann SA, et al. 2007. Wnt signaling mediates regional specification in the vertebrate face. Development 134(18):3283–3295.

Brugmann SA, et al. 2010. Comparative gene expression analysis of avian embryonic facial structures reveals new candidates for human craniofacial disorders. Hum Mol Genet. 19(5):920–930.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol Biol Evol. 29(8):1917–1932.

Card DC, et al. 2016. Phylogeographic and population genetic analyses reveal multiple species of *Boa* and independent origins of insular dwarfism. Mol Phylogenet Evol. 102:104–116.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134(1):25–36.

Cartmill M. 1985. Climbing. In: Hildebrand M, Bramble DM, Liem KF, Wake DB, editors. Functional vertebrate morphology. Cambridge (MA): Harvard University Press.

Castoe TA, et al. 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. Proc Natl Acad Sci U S A. 106(22):8986–8991.
Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and genotyping loci *de novo* from short-read sequences. G3 (Bethesda) 1:171–182.
Catchen JM, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. Mol Ecol. 22(11):3124–3140.
Choi H, Ryu K-Y, Roh J, Bae J. 2018. Effect of radioactive iodine-induced hypothyroidism on longitudinal bone growth during puberty in immature female rats. Exp Anim. 18–0013.
Choi Y. 2012. A fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein. In: Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine BCB '12. New York (NY): ACM. p. 414–417.
Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012. Predicting the functional effect of amino acid substitutions and indels. PLoS One 7(10):e46688.
Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. 18(2):71–86.
DePristo MA, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 43(5):491.
Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 7(1):214.
Elchebly M, et al. 1999. Neuroendocrine dysplasia in mice lacking protein tyrosine phosphatase σ. Nat Genet. 21(3):330–333.
Feldman CR, Brodie ED, Brodie ED, Pfrender ME. 2009. The evolutionary origins of beneficial alleles during the repeated adaptation of garter snakes to deadly prey. Proc Natl Acad Sci U S A. 106(32):13415–13420.
Feldman CR, Brodie ED, Brodie ED, Pfrender ME. 2012. Constraint shapes convergence in tetrodotoxin-resistant sodium channels of snakes. Proc Natl Acad Sci U S A. 109(12):4556–4561.
Gallant JR, et al. 2014. Genomic basis for the convergent evolution of electric organs. Science 344(6191):1522–1525.
Global Lipids Genetics Consortium, et al. 2013. Discovery and refinement of loci associated with lipid levels. Nat Genet. 45:1274–1283.
Grant PR, Grant BR, Markert JA, Keller LF, Petren K. 2004. Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. Evolution 58(7):1588–1599.
Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5(10):e1000695.
Haas BJ, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 8(8):1494.
Hartiala JA, et al. 2016. Genome-wide association study and targeted metabolomics identifies sex-specific association of *CPS1* with coronary artery disease. Nat Commun. 7(1):10558.
Henderson RW, Waller T, Micucci P, Puorto G, Bourgeois RW. 1995. Ecological correlates and patterns in the distribution of neotropical boines (Serpentes: Boidae): a preliminary assessment. Herpetol Nat Hist. 3:1.
Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313(5783):101–104.
Hoffmann FG, Opazo JC, Storz JF. 2010. Gene cooption and convergent evolution of oxygen transport hemoglobins in jawed and jawless vertebrates. Proc Natl Acad Sci U S A. 107(32):14274–14279.
Hohenlohe PA, et al. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6(2):e1000862.
Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12(1):491.
Hynková I, Starostová Z, Frynta D. 2009. Mitochondrial DNA variation reveals recent evolutionary history of main *Boa constrictor* clades. Zool Sci. 26(9):623–631.
Jayne BC. 1982. Comparative morphology of the semispinalis-spinalis muscle of snakes and correlations with locomotion and constriction. J Morphol. 172(1):83–96.
Jones FC, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484(7392):55–61.
Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240.
Joron M, et al. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature 477(7363):203–206.
Jost MC, et al. 2008. Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. Mol Biol Evol. 25(6):1016–1024.
Kawano Y, et al. 2006. Regulation of prostate cell growth and morphogenesis by Dickkopf-3. Oncogene 25(49):6528–6537.
Keogh JS, Scott IAW, Hayes C. 2005. Rapid and repeated origin of insular gigantism and dwarfism in Australian tiger snakes. Evolution 59(1):226–233.
Köhler M, Moyà-Solà S. 2009. Physiological and life history strategies of a fossil large mammal in a resource-limited environment. Proc Natl Acad Sci U S A. 106:20354–20358.
Kunte K, et al. 2014. *doublesex* is a mimicry supergene. Nature 507(7491):229–232.
Kurosaka H, Iulianella A, Williams T, Trainor PA. 2014. Disrupting hedgehog and WNT signaling interactions promotes cleft lip pathogenesis. J Clin Invest. 124(4):1660–1671.
Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27(21):2987–2993.
Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079.
Liebeskind BJ, Hillis DM, Zakon HH. 2015. Convergence of ion channel genome content in early animal evolution. Proc Natl Acad Sci U S A. 112(8):E846–E851.
Lillywhite HB, Henderson RW. 2002. Behavioral and functional ecology of arboreal snakes. In: Collins, JT, Seigel RA, editors. Snakes: ecology and behavior. Caldwell (NJ): The Blackburn Press. p. 1–48.
Linnen CR, et al. 2013. Adaptive evolution of multiple traits through multiple mutations at a single gene. Science 339(6125):1312–1316.
Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE. 2009. On the origin and spread of an adaptive allele in deer mice. Science 325(5944):1095–1098.
Lomolino MV, et al. 2013. Of mice and mammoths: generality and antiquity of the island rule. J Biogeogr. 40(8):1427–1439.
Lomolino MV, Riddle BR, Whittaker RJ, Brown JH. 2010. Biogeography. 4th ed. Sunderland (MA): Sinauer Associates.
Losos JB. 2011. Convergence, adaptation, and constraint. Evolution 65(7):1827–1840.
Lotterhos KE, et al. 2017. Composite measures of selection can improve the signal-to-noise ratio in genome scans. Methods Ecol Evol. 8(6):717–727.
Lupu DS, et al. 2017. Altered methylation of specific DNA loci in the liver of *Bhmt*-null mice results in repression of *Iqgap2* and *F2rl2* and is associated with development of preneoplastic foci. FASEB J. 31(5):2090–2103.

McGlothlin JW, et al. 2014. Parallel evolution of tetrodotoxin resistance in three voltage-gated sodium channel genes in the garter snake *Thamnophis sirtalis*. Mol Biol Evol. 31(11):2836–2846.
McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9):1297–1303.
McLaren W, et al. 2016. The Ensembl variant effect predictor. Genome Biol. 17(1):122.
McNab BK. 1994. Resource use and the survival of land and freshwater vertebrates on oceanic islands. Am Nat. 144(4):643–660.
McNab BK. 2002. Minimizing energy expenditure facilitates vertebrate persistence on oceanic islands. Ecol Lett. 5(5):693–704.
Millard HR, et al. 2018. Dietary choline and betaine; associations with subclinical markers of cardiovascular disease risk and incidence of CVD, coronary heart disease and stroke: the Jackson Heart Study. Eur J Nutr. 57(1):51–60.
Mitchell A, et al. 2015. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. 43(D1):D213–D221.
Nachman MW, Hoekstra HE, D'Agostino SL. 2003. The genetic basis of adaptive melanism in pocket mice. Proc Natl Acad Sci U S A. 100(9):5268–5273.
Natarajan C, et al. 2016. Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. Science 354(6310):336–339.
Paradis E. 2010. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26(3):419–420.
Parsons KJ, Taylor AT, Powder KE, Albertson RC. 2014. Wnt signalling underlies the evolution of new phenotypes and craniofacial variability in Lake Malawi cichlids. Nat Commun. 5(1):3629.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One 7(5):e37135.
Porras L. 1999. Island boa constrictors (*Boa constrictor*). Reptiles 7:48–61.
Portik DM, et al. 2017. Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. Mol Ecol. 26(19):5245–5263.
Projecto-Garcia J, et al. 2013. Repeated elevational transitions in hemoglobin function during the evolution of Andean hummingbirds. Proc Natl Acad Sci U S A. 110(51):20669–20674.
R Core Team. 2019. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: http://www.R-project.org/; last accessed October 22, 2019.
Reed RN, et al. 2007. Ecology and conservation of an exploited insular population of *Boa constrictor* (Squamata: Boidae) on the Cayos Cochinos, Honduras. In: Henderson RW, Powell R, editors. Biology of the boas and pythons. Eagle Mountain (UT): Eagle Mountain Publishing. p. 289–403.
Reid NM, et al. 2016. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science 354(6317):1305–1308.
Reynolds RG, Niemiller ML, Revell LJ. 2014. Toward a Tree-of-Life for the boas and pythons: multilocus species-level phylogeny with unprecedented taxon sampling. Mol Phylogenet Evol. 71:201–213.
Root AW, Shulman D, Root J, Diamond F. 1986. The interrelationships of thyroid and growth hormones: effect of growth hormone releasing hormone in hypo- and hyperthyroid male rats. Acta Endocrinol. 113(Suppl 4):S367–S375.
Rosenblum EB, Hoekstra HE, Nachman MW. 2007. Adaptive reptile color variation and the evolution of the *MC1R* gene. Evolution 58:1794–1808.
Rosenblum EB, Parent CE, Brandt EE. 2014. The molecular basis of phenotypic convergence. Annu Rev Ecol Evol Syst. 45(1):203–226.
Rosenblum EB, Römpler H, Schöneberg T, Hoekstra HE. 2010. Molecular and functional basis of phenotypic convergence in white lizards at White Sands. Proc Natl Acad Sci U S A. 107(5):2113–2117.
Schield DR, et al. 2017. Insight into the roles of selection in speciation from genomic patterns of divergence and introgression in secondary contact in venomous rattlesnakes. Ecol Evol. 7(11):3951–3966.
Schmidt C, Patel K. 2005. *Wnts* and the neural crest. Anat Embryol. 209(5):349–355.
Sedlazeck FJ, Rescheneder P, von Haeseler A. 2013. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. Bioinformatics 29(21):2790–2791.
Shine R. 1983. Arboreality in snakes: ecology of the Australian elapid genus *Hoplocephalus*. Copeia 1983(1):198–205.
Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0. Available from: http://repeatmasker.org/, last accessed October 22, 2019.
Smith CL, Blake JA, Kadin JA, Richardson JE, Bult CJ. 2018. Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. Nucleic Acids Res. 46(D1):D836–D842.
Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32(Web Server):W309–W312.
Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19(Suppl 2):ii215–ii225.
Stern DL. 2013. The genetic causes of convergent evolution. Nat Rev Genet. 14(11):751–764.
Stewart C-B, Schilling JW, Wilson AC. 1987. Adaptive evolution in the stomach lysozymes of foregut fermenters. Nature 330(6146):401–404.
Stewart K, Uetani N, Hendriks W, Tremblay ML, Bouchard M. 2013. Inactivation of LAR family phosphatase genes *Ptprs* and *Ptprf* causes craniofacial malformations resembling Pierre-Robin sequence. Development 140(16):3413–3422.
Suárez-Atilano M, Burbrink F, Vázquez-Domínguez E. 2014. Phylogeographical structure within *Boa constrictor imperator* across the lowlands and mountains of Central America and Mexico. J Biogeogr. 41:2371–2384.
Sundaram V, et al. 2014. Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Res. 24(12):1963–1976.
Surakka I, et al. 2015. The impact of low-frequency and rare variants on lipid levels. Nat Genet. 47(6):589–597.
Sutter NB, et al. 2007. A single *IGF1* allele is a major determinant of small size in dogs. Science 316(5821):112–115.
Takayama S, Isogai A. 2005. Self-incompatibility in plants. Annu Rev Plant Biol. 56(1):467–489.
The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45:D158–D169.
Thomas JW, et al. 2008. The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (*Zonotrichia albicollis*) is a complex rearrangement and suppressor of recombination. Genetics 179(3):1455–1468.
Thompson MJ, Jiggins CD. 2014. Supergenes and their role in evolution. Heredity 113(1):1–8.
Tuttle EM, et al. 2016. Divergence and functional degradation of a sex chromosome-like supergene. Curr Biol. 26(3):344–350.
Ueno K, et al. 2013. microRNA-183 is an oncogene targeting *Dkk-3* and *SMAD4* in prostate cancer. Br J Cancer. 108(8):1659–1667.
Van der Auwera GA, et al. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Hoboken (NJ): John Wiley & Sons, Inc.

Veeck J, Dahl E. 2012. Targeting the Wnt pathway in cancer: the emerging role of *Dickkopf*-3. Biochim Biophys Acta. 1825(1):18–28.
Verity R, et al. 2017. minotaur: a platform for the analysis and visualization of multivariate results from genome scans with R Shiny. Mol Ecol Resour. 17(1):33–43.
Vicoso B, Emerson JJ, Zektser Y, Mahajan S, Bachtrog D. 2013. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. PLoS Biol. 11(8):e1001643.
Wagner GP, Lynch VJ. 2010. Evolutionary novelties. Curr Biol. 20(2):R48–R52.
Wang P, et al. 2018. Mucopolysaccharidosis type VI in a great dane caused by a nonsense mutation in the *ARSB* gene. Vet Pathol. 55(2):286–293.
Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.
Weissglas-Volkov D, et al. 2011. The N342S *MYLIP* polymorphism is associated with high total cholesterol and increased LDL receptor degradation in humans. J Clin Invest. 121(8):3062–3071.
Weng, MP, Liao BY. 2017. modPhEA: model organism Phenotype Enrichment Analysis of eukaryotic gene sets. Bioinformatics 33(21):3505–3507.
Wray GA. 2007. The evolutionary significance of *cis*-regulatory mutations. Nat Rev Genet. 8(3):206–216.
Zakon HH. 2002. Convergent evolution on the molecular level. Brain Behav Evol. 59:250–261.
Zakon HH. 2012. Adaptive evolution of voltage-gated sodium channels: the first 800 million years. Proc Natl Acad Sci U S A. 109(Suppl 1):10619–10625.
Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. 2012. Parallel molecular evolution in an herbivore community. Science 337(6102):1634–1637.

Associate editor: Federico Hoffmann