Supplemental Material (not copyedited or formatted) for: Timothy J. Thurman, Todd M. Palmer, Jason J. Kolbe, Arash M. Askary, Kiyoko M. Gotanda, Oriol Lapiedra, Tyler R. Kartzinel, Naomi Man in't Veld, Liam J. Revell, Johanna E. Wegener, Thomas W. Schoener, David A. Spiller, Jonathan B. Losos, Robert M. Pringle, Rowan D. H. Barrett. 2023. "The Difficulty of Predicting Evolutionary Change in Response to Novel Ecological Interactions: A Field Experiment with Anolis Lizards." The American Naturalist 201(4). DOI: https://doi.org/10.1086/723209.

Online Supplement: The difficulty of predicting evolutionary change in response to novel ecological interactions: a field experiment with *Anolis* lizards

The American Naturalist

Timothy J. Thurman[1,2,3,*], Todd M. Palmer[4],
Jason J. Kolbe[5], Arash M. Askary[1],
Kiyoko M. Gotanda[1,6,7], Oriol Lapiedra[8],
Tyler R. Kartzinel[9,10], Naomi Man in't Veld[11],
Liam J. Revell[12,13], Johanna E. Wegener[5],
Thomas W. Schoener[14], David A. Spiller[14],
Jonathan B. Losos[15], Robert M. Pringle[16,*],
Rowan D. H. Barrett[1,*]

1. Redpath Museum and Department of Biology, McGill University. Montréal, QC, Canada;

2. Smithsonian Tropical Research Institute. Panamá, República de Panamá;

3. Current address: University of Arkansas for Medical Sciences, Little Rock, AR, USA;

4. Department of Biology, University of Florida, Gainesville, FL, USA;

5. Department of Biological Sciences, University of Rhode Island, Kingston, RI, USA;

6. Department of Zoology, University of Cambridge. Cambridge, United Kingdom;

7. Current address: Department of Biological Sciences, Brock University, St. Catharines, ON, Canada;

8. Centre for Ecological Research and Applied Forestries, Cerdanyola del Vallès, Catalonia 08193, Spain;

9. Department of Ecology, Evolution, and Organismal Biology, Brown University, Providence, RI, USA;

10. Institute at Brown for Environment and Society, Brown University, Providence, RI, USA;

11. Unaffiliated;

12. Department of Biology, University of Massachusetts at Boston, Boston, MA, USA;

13. Facultad de Ciencias, Universidad Católica de la Santísima Concepción, Concepción, Chile;

14. Department of Evolution and Ecology, University of California, Davis, CA, USA;

15. Department of Biology and Living Earth Collaborative, Washington University, St. Louis, MO, USA;

16. Department of Ecology & Evolutionary Biology, Princeton University, Princeton, NJ, USA;

∗ Corresponding authors: Timothy J. Thurman (timothy.thurman@mail.mcgill.ca), Robert M. Pringle (rpringle@princeton.edu), Rowan D. H. Barrett (rowan.barrett@mcgill.ca).

Ethics

This work was conducted under permits from the BEST Commission and the Bahamas Department of Agriculture. Princeton University's Institutional Animal Care and Use Committee provided guidance on and approved the study protocols (permits 1922-13 and 1922-F16).

Study organisms

Anolis sagrei- The brown anole, *A. sagrei*, is a small lizard native to the Bahamas. *Anolis* species are often classified into different "ecomorphs," groupings of lizards with common morphological, ecological, and behavioral traits which appear to be well-suited for various structural habitats (Williams, 1972; Beuttell and Losos, 1999). Brown anoles are "trunk-ground" ecomorphs: they are generally found on or close to the ground, and have skeletal traits suited to moving on broad perches such as rocks and tree trunks. Brown anoles are sexually dimorphic with males being larger than females (Butler and Losos, 2002). On small islands in the Bahamas, year-to-year survival rates are often extremely low such that brown anoles are effectively an annual species with 1 generation per year (Calsbeek and Smith, 2007; Cox and Calsbeek, 2010).

Anolis smaragdinus- The green anole, *A. smaragdinus*, is a "trunk-crown" ecomorph. Thus, it has a more arboreal lifestyle than brown anoles, using narrower perches in taller vegetation, and has morphological traits to match. In both the field and in the literature, *A. smaragdinus* is often confused with its close relative *A. carolinensis*, which is native to the southeast United States (Les and Powell, 2014).

Leiocephalus carinatus- The curly-tailed lizard, *L. carinatus*, is a large-bodied, ground dwelling lizard. It is native to the Bahamas, however much less is known about its ecology compared to *Anolis* species. Curly-tailed lizards are known predators of *Anolis* lizards, both generally (Schoener et al., 1982) and in our experiment specifically (Pringle et al., 2019). They also consume a wide variety of other foods and prey items (Schoener et al., 1982).

Experiment details

Our experimental islands are located in the Exuma chain of the Bahamas near Staniel Cay, an area that has been used for studies of anoles since the 1970s (e.g., Schoener and Schoener, 1983). We selected 16 islands for our study (Table S1, Figure S1) that were chosen because they: (1) were small enough to be effectively surveyed, (2) had existing brown anole populations (5 islands had been experimentally colonized in the 1970s, the remainder were naturally-occurring), (3) had no other lizard or predator populations, and (4) had vegetation greater than 2m tall (considered necessary to support *A. smaragdinus* populations, Losos and Spiller, 1999). After an initial population census in 2011, we assigned islands to treatments randomly. We first used Google Earth Pro to calculate the vegetated area of each island (excluding the bare rock around the island edge). We stratified our random assignment to ensure that each treatment contained 2 replicates among the 8 largest islands and two replicates among the 8 smallest.

We captured green anoles for transplantation from Staniel Cay (which has no curly-tailed lizards) and captured curly-tailed lizards from nearby Thomas Cay (which has all 3 species present). Thus, the curly-tailed lizards were accustomed to the anole species and the green anoles were accustomed to brown anoles. However, brown anoles were naive to the introduced competitor, and both anoles were ecologically naive to the predator. Transplanted lizards were randomly assigned to islands. We released green anoles in groups of 10 or 11 (equal sex ratio, with the 11th individual a juvenile). We released unsexed curly-tailed lizards in groups of 5-7. These numbers were chosen to mimic colonization by small founder populations.

Census and population estimation

To estimate population sizes, we used a mark-resight procedure developed for Caribbean *Anolis* (Heckel and Roughgarden, 1979). A team of 3-6 researchers comprehensively searched islands for all lizard species over three consecutive days. Using a squirt gun, we marked lizards with a day-specific color of non-toxic, water-soluble paint (blue, red, and yellow for days 1, 2, and 3, respectively). Following Pringle et al. (2015), we estimated population size as the mean of the three possible Chapman estimates:

$$n_{Chap} = \frac{(M+1)(C+1)}{R+1} - 1 \tag{S1}$$

where n_{Chap} is the Chapman estimate of population size, M and C are the numbers of individuals marked on the first and second visits, and R is the number of marked individuals re-sighted on the second visit.

During each census, we recorded data on habitat use for each lizard spotted. Specifically, we noted: (1) the sex of the lizard, (2) whether the lizard was on the ground, (3) perch height (estimated to nearest 5 cm, 0 for individuals found on the ground), (4) perch diameter (estimated to the nearest 0.5 cm, not recorded for individuals on the ground), and (5) maximum vegetation height within a 1-m radius (recorded only 2014-2016).

Lizard sampling

On all sampling trips, we captured lizards by hand or with small nooses made of dental floss or thread. In 2016, we recorded habitat use information (e.g., perch height and diameter) for captured lizards. Each lizard was given a unique ID (which contained no treatment information). We kept lizards individually in disposable plastic containers and transported them to Staniel Cay for phenotyping.

We anesthetized lizards with isoflurane, then used a portable X-ray to capture an X-ray image of the full lizard for measuring skeletal traits. We scanned the underside of each lizard for measurements of toepad traits. Finally, we took a small piece of tail tip (0.5-1.25cm) from each lizard and preserved it either in DMSO or froze it at -20°C in 100% ethanol for genetic analysis. After leaving the Bahamas tissue samples were stored at -80°C for long-term storage.

Phenotypic measurement details

We measured phenotypic traits which either (1) are associated with ecomorph type in *Anolis*, (2) affect organismal performance, or (3) have been found to be under natural selection. We measured snout-vent length (Beuttell and Losos, 1999; Losos et al., 2004; Calsbeek, 2009), head width and length (Beuttell and Losos, 1999; Herrel et al., 2006, 2007), pectoral and pelvic width (Beuttell and Losos, 1999), forelimb length (Beuttell and Losos, 1999; Donihue et al., 2018), hindlimb length (Beuttell and Losos, 1999; Losos et al., 2004, 2006), and the number of scales on the toepad of the third toe of the fore foot and the fourth toe of the hind foot (Beuttell and Losos, 1999; Glossip and Losos, 1997; Stuart et al., 2014).

TJT performed all skeletal measures (Figure S3). The number of toepad lamellae were counted twice: once by TJT and once by ACAV. We attempted to measure all traits for all lizards, but we did not take measurements in cases where (1) X-rays or scans were too unclear or contained anomalies such that we could not locate landmarks, (2) individuals had bones that appeared to have healed after a fracture, and (3) some part of the trait of interest was missing or damaged such that it could not be measured. For lizards that were measured multiple times (including by multiple measurers), we averaged across the multiple measurements for each trait. In the case of bilateral traits, we first averaged the multiple measurements for each side, and then took the average of the left and right sides. We excluded juvenile lizards from phenotypic analysis following Kolbe et al. (2012), excluding males with SVL < 38.5 mm and females with SVL < 35 mm.

Repeatability of phenotypic measures

To ensure that our phenotypic measurements were repeatable, we measured a subset of X-rays and scans multiple times and estimated repeatability of all trait measurements following Quinn and Keough (2002) and Whitlock and Schluter (2009). Repeatability is equal to:

$$\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_E^2} \tag{S2}$$

where σ_α^2 is the variance among groups (here the groups are the individual lizards) and σ_E^2 is the variance within groups (Whitlock and Schluter, 2009). This value, also called the intraclass correlation, ranges from 0 (no repeatability) to 1 (complete repeatability), and is the correlation between measurements taken within the same group (i.e., lizard). Within-group variance is estimated from the mean-squared error (or residual)

term of a single-factor, random-effects ANOVA in which the trait measurement is the response variable and lizard ID is the factor (Whitlock and Schluter, 2009). Because X-rays and scans were measured different numbers of times, we estimated σ_α^2 as:

$$\frac{\mathrm{MS}_{\mathrm{groups}} - \mathrm{MS}_{\mathrm{residual}}}{(\sum n_i - \sum n_i^2 / \sum n_i)/(p-1)} \tag{S3}$$

where n_i is the number of measurements of individual i and p is the total number of individuals that were measured multiple times (Quinn and Keough, 2002). X-ray and scan quality varied across years, so we estimated repeatability within each year for each trait. For our counts of toepad lamellae, we calculated repeatability separately for each measurer. For estimating heritability, we measured 74 x-rays and 96 scans multiple times. All measurements were highly repeatable (Table S2).

Linear mixed models

We fit three three types of linear mixed models, all using the `R` package `brms`. For all model types and for all traits, we fit models separately by sex.

We fit two types of models to examine the effect of predators, competitors, and their interaction on univariate phenotypes. The first model type was used for SVL and number of lamellae, which do not require correction for body size. It included year (as a factor), presence/absence of predator, presence/absence of competitor, and all possible interactions as main effects. We included island as a random effect. In the syntax of `brms`, the model equation is:

```
trait.value ~ pred*comp*year + (1 | island)
```

When analyzing hindlimb length and all other skeletal traits, we corrected for body size by adding SVL as a covariate. We allowed the effect of body size to vary across islands using partial pooling, and we standardized SVL to have a mean of 0 and a standard deviation of 1 to improve model fitting and interpretation. In the syntax of `brms`, the equation for this model is:

```
hindlimb.length ~ SVL + pred*comp*year + (1 + SVL | island)
```

For our multivariate analyses, we removed the effect of body size from size-dependent traits by taking residuals from a third linear model. This model included SVL and allowed the effect of SVL to vary across islands using partial pooling, but it did not include the experimental terms of predator, competitor, and

year. As before, we standardized SVL prior to model fitting. In the syntax of `brms`, the equation for this model is:

```
trait.value ~ SVL + (1 + SVL | island)
```

For all models, we used a normal prior of $N(0, 100)$ for the intercept. For all fixed effect coefficients, we used a normal prior with $\mu = 0$ and σ scaled to be one-half the standard deviation of the trait being analyzed. These weakly informative priors encapsulate our prior belief that the fixed effects are unlikely to shift trait means by more than 1 standard deviation, and help guard against over-fitting. For all other parameters, we used the `brms` default priors. We used the Student's t-distribution as the likelihood. This is similar to using a Gaussian likelihood, but more robust to possible outliers in the data because the Student's t-distribution has heavier tails (Gelman and Hill, 2007). We tested whether our conclusions were robust to prior choice by fitting models with two other priors for our fixed effects: uniform flat priors, and less-informative normal priors with $\mu = 0$ and σ scaled to be equal to the standard deviation of the trait being analyzed. These alternative priors did not significantly change the conclusions of our study (results not shown).

For all models, we fit four independent chains with 1000 iterations of warmup and 1000 iterations of sampling each. We assessed convergence using the Gelman-Rubin $\hat{R}$ statistic (Gelman and Rubin, 1992) and ensured that the ratio of effective samples to total samples for each chain was > 0.1. For hypothesis testing, we assessed statistical significance for the parameters of interest by determining whether the 95% highest posterior density interval (HPDI) for that parameter contained 0.

Size independence of hindlimb lamella number

The number of scales on the hindlimb toepad is thought to be fixed at birth and thus independent of body size (Hecht, 1952). Nevertheless, some studies of *Anolis* correct for body size when analyzing toepad scale number (e.g., Stuart et al., 2014; Winchell et al., 2016). In our analyses in the main text, we did not correct for body size. To ensure that this did not affect our results, we ran models correcting for body size, using the same models as outlined above for hindlimb length. Body size had no statistically significant effect on hindlimb lamella number (Table S5). Including body size as a covariate does alter the significance of the effect of predators: it is marginally non-significant when we correct for body size (Table S5).

Additional univariate analyses

We analyzed the effects of predators and competitors on all traits, including those for which we did not make *a priori* predictions. For most effects for most traits, we find no significant effects (Table S6). There are two significant effects that are similar to results we report in the main text. There is a significant positive effect of predators on male femurs, consistent with the positive (though nonsignificant) effect of predators on hindlimbs that we report in the main text (effects of predators on the tibia and 4th toe of the hindlimb are positive but non-significant). Biomechanical studies of performance in *Anolis* (e.g., Losos and Sinervo, 1989; Losos, 1990*a*,*b*; Losos and Irschick, 1996) have focused on the whole hindlimb: the performance effects of changes in the femur alone are unclear. We also find that predators have a significant negative effect in males on the number of forelimb lamellae: this recapitulates the effect we report on number of hindlimb lamellae.

We find three novel significant effects (Table S6). Competitors had a negative effect on male pectoral width and male head width. Pectoral width is generally narrower in arboreal ecomorphs (Beuttell and Losos, 1999), though its functional importance for performance is less well-studied than for other *Anolis* traits, making the possible selective reason for this effect unclear. Narrower heads are also associated with more arboreal ecomorphs, and head size and shape are related to bite force, with narrower heads having weaker bite forces (Herrel et al., 2006). Bite force is related to feeding performance through the size and hardness of prey (Herrel et al., 2006), and competitors did cause brown anoles to change their diet composition (albeit in a more terrestrial, not arboreal, manner). Finally, we find a significant negative interaction of predators and competitors on female femur length. Shorter femurs would be the more "arboreal" trait, so this could perhaps be taken as evidence for increased selection for arboreal traits when female brown anoles must simultaneously avoid predators and compete with green anoles. But this interpretation is speculative, given that femur length alone has not been strongly linked to performance.

Power analysis

We performed a post-hoc power analysis of our LMMs to determine if the null results we found may be due to insufficient statistical power. Power analysis is traditionally associated with frequentist statistics, not the Bayesian framework used in our LMMs, and as such there are few tools for Bayesian power analysis. Instead, we performed our power analysis in a frequentist framework using the R package `simr` (Green and MacLeod, 2016). `simr` fits frequentist LMMs using the `lme4` R package and uses repeated Monte

Carlo simulations to evaluate statistical power (Bates et al., 2015; Green and MacLeod, 2016). We used the `powerSim` function to evaluate the power of frequentist LMMs of equivalent form and with equal sample sizes as our Bayesian ones. For each trait, sex, and statistical term of interest, we calculated power curves across a range of effect sizes (Figure S4).

Next, we considered our statistical power in the context of our *a priori* predictions of evolution in response to predators (the effect for which we have the most prior information). The breeder's equation can be used to predict the change in phenotype for a trait under natural selection, given information about the strength of selection and the heritability of the trait. Estimates of heritability for SVL are generally high but vary widely: h^2 = 0.291-1 for males, h^2 = 0.107-0.74 for females (Calsbeek and Smith, 2007; Calsbeek and Bonneaud, 2008; McGlothlin et al., 2019). The heritability of relative hindlimb length is less well studied, but has also been found to be high: h^2 = 0.78 for males, h^2 = 0.41 for females (Calsbeek and Bonneaud, 2008). We are unaware of estimates of heritability for number of hindlimb lamellae. Using low (0.107), moderate (0.55), and high (0.74) estimates of h^2 across these traits, we used the breeder's equation to estimate the variance-standardized selection gradient that would be needed to elicit an evolutionary response (over 5 generations) which we could detect with 80% power (eqn. 4, Hereford et al., 2004):

$$\frac{R}{\sigma_z} = h^2 \beta_\sigma$$

where R is the evolutionary response (here, the effect size we could detect with 80% power, divided by 5 generations), σ_z is the phenotypic standard deviation of the trait, calculated by taking the average of per-island standard deviations, h^2 is the narrow sense heritability, and β_σ is the standardized selection gradient.

We find that, if heritability is low (0.107), we have power only to detect strong selection, though power to detect weaker selection increases substantially with higher heritability (Table S7). Previous studies have found found strong selection on SVL and relative hindlimb length after the introduction of predators, with the absolute value of per-population estimates of selection gradients ranging from ~0.1 to ~0.5 for SVL and from ~0.2 to ~1.5 for relative hindlimb length (Losos et al., 2004, 2006). We thus conclude that, provided heritability is not unusually low on our islands, our study had reasonable power to detect the phenotypic changes we were predicting, given the strong selection imposed by the introduction of predators in previous experiments.

Additional multivariate phenotypic analyses

We performed some analyses of sex-specific multivariate phenotypic change in addition to those reported in the main text. In brief, we ran the same linear models and t-tests testing for the effects of predators and competitors on ΔD, θ, and Ω as we did at the population level, but separately for the PCVs for each sex. These within-sex tests revealed that, as at the population level, competitors and predators did not generally promote parallel multivariate phenotypic changes. For Ω, predators and competitors led to divergence relative to controls. Likewise for θ, predators and competitors led to either less parallelism in direction or had non-significant effects. In females, the effects of predators and competitors were non-significant for ΔD. In males, predators and competitors had a significant effect of smaller ΔD. As at the population level, then, predators and competitors do not tend to promote predictable sex-specific phenotypic change.

Library preparation and sequencing

We extracted DNA from individual lizard tail tips using a standard phenol-chloroform based method. We used individual sterilized razor blades to chop a portion of the tail tip into 0.5mm pieces. Those pieces were left overnight at 55°C in tissue digestion buffer and proteinase K. From the digested tissue, we used 25:24:1 phenol:chloroform:isoamyl alcohol washes and centrifugation to isolate DNA, then ethanol washes to precipitate DNA. Extracted DNA was resuspended in molecular-grade water or TE and cleaned using NucleoSpin® gDNA cleanup kits (Takara Bio) according to the manufacturer's instructions.

When selecting samples for library preparation, we distributed individuals from islands and timepoints as evenly as was feasible across libraries and sequencing lanes to avoid batch effects. Overall, we made 28 libraries. For each library, we digested at least 300ng of genomic DNA per sample for three hours using the restriction enzymes NlaIII and MluCI. We performed a bead cleanup using NaCl-PEG diluted SpeedBeads to remove small fragments (Rohland and Reich, 2012), then quantified DNA using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Thermo Fisher Scientific). We normalized individual samples to 90ng using a Biomek FX^{P} liquid handling robot (Beckman Coulter), then ligated adaptors to the digested DNA. We used the 48 individually-barcoded P1 adaptors from Peterson et al. (2012). Our P2 adaptors were custom designed to include degenerate base regions (DBRs), short stretches of random bases which allow PCR duplicates to be bioinformatically removed (which is not otherwise possible with ddRAD sequencing). Our adaptors were modified slightly from the design of Vendrami et al. (2017). The sequences of the two strands are:

`flex_DBR_P2.1`

`GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNMMGGACG`

`flex_DBR_P2.2_bio`

`/5Phos/AATTCGTCCIINNNNNNAGATCGGAAGAGCGAGAACAA/3Bio/`

where N represents a random base (A,C,T, or G), M is a random base of either A or C, /5Phos/ is a phosphate group on the 5′ end of the primer, I is inosine, a synthetic base which binds with all 4 natural bases, and /3Bio/ is a biotin tag on the 3′ end of the primer. After adaptor ligation we pooled individual samples into single libraries and performed a second bead cleanup to remove small fragments. For each library, we size-selected fragments with inserts of 400-450bp using a PippinPrep (Sage Science). We performed a bead cleanup with Dynabeads (Thermo Fisher Scientific), which select only fragments with P1-P2 adaptor combinations. We PCR amplified libraries using the AccuPrime™ *Taq* DNA Polymerase System (Invitrogen). We divided each library into 4 separate 50 μl reactions and used a PCR program with 12 amplification cycles to minimize the chance of PCR duplicates. We used the PCR primers and 12 library barcodes from Peterson et al. (2012) as well as 4 custom library barcodes for a total of 16 library barcodes. Combined with the 48 individual barcodes, this allows for multiplexing of up to 768 samples per lane. After PCR amplification, we did one final round of bead cleanup, after which each library was ready for sequencing.

Bioinformatic pipeline

We demultiplexed the raw reads from each library using the `process_radtags` tool in `Stacks` v1.46 (Catchen et al., 2011, 2013), using the `--inline_index`, `--disable_rad_check`, and `-P` options. Next, we used the `clone_filter` tool in Stacks v1.46 to remove potential PCR duplicates. Again, we used the `--inline_index` option and set the length of our paired-end oligo DBR sequence as 8 bp with the `--oligo_len_2` option. We trimmed extra bases from the reverse reads using the `fastx_trimmer` from the FASTX Toolkit v0.0.14 found at http://hannonlab.cshl.edu/fastx_toolkit/index.html. Finally, we used `cutadapt` v2.1 (Martin, 2011) to remove possible adaptor contamination, requiring 10 bp of adaptor sequence before trimming (`--overlap 10`).

We aligned these processed reads to the preliminary *Anolis sagrei* genome using the default settings of the `BWA-MEM` algorithm in `bwa` v0.7.15 (Li and Durbin, 2009). We used `samtools` v1.5 to convert the `sam` output

from `bwa` into sorted `bam` files, keeping only alignments with a mapping quality of at least 20 (Li et al., 2009).

To analyze genome-wide population structure, we first used `ANGSD` to estimate genotype likelihoods for all individuals using the following options: `-GL` 1: use the `samtools` genotype likelihood model; `-doGLF` 2: output in `BEAGLE` GLF format; `-doCounts` 1: count alleles; `-dosaf` 1: perform multisample GL estimation of the site frequency spectrum; `-doMajorMinor` 1: infer major and minor alleles from the genotype likelihoods; `-doMAF` 2: calculate allele frequencies assuming a fixed major and unknown minor allele; `-SNP_pval` 1e-6; only keep SNPs that are likely to be variable according to a likelihood ratio test with p value $< 10^{-6}$; `-uniqueOnly` 1: use only reads that map uniquely; `-minIndDepth` 5: minimum per-individual read depth of 5; `-minMapQ` 30: minimum read mapping quality of 30; `-minQ` 20: minimum base quality score of 20; `-minInd` set to 700 (~70% of the total number of individuals). We then used `PCAngsd` v0.98 to perform the genome-wide PCA from these genotype likelihoods, using the `-minMaf` option to set the minimum minor allele frequency to 2% and using default settings otherwise. This analysis used 183599 SNPs.

For our analysis of allele frequency change, we used `ANGSD` to estimate allele frequencies directly from these `bam` files (Korneliussen et al., 2014). For each island/year combination, we ran `ANGSD` with the following options: `-P` 1: use a single processor; `-GL` 2: use `ANGSD`'s implementation of the early `GATK` method for calculating genotype likelihoods; `-doMaf` 1: estimate allele frequencies using the method from Kim et al. (2011) with a fixed major allele and known minor allele; `-doMajorMinor` 4: set the reference allele as the major allele; `-uniqueOnly` 1: use only reads that map uniquely; `-minIndDepth` 5: minimum per-individual read depth of 5; `-minMapQ` 30: minimum read mapping quality of 30; `-minQ` 20: minimum base quality score of 20; `-minMaf` 0.05; minimum minor allele frequency of 5%, `-SNP_pval` 1e-6; only keep SNPs that are likely to be variable according to a likelihood ratio test with p value $< 10^{-6}$; and `-minInd` set to ~70% (with rounding) of the total number of individuals per population and timepoint, such that SNPs must be in ~70% of individuals to be retained.

For the CMH test, which requires called genotypes, we first created read pileups from the `bam` files using the `samtools mpileup` program (Li, 2011). We used the `-C` 50 flag to adjust mapping quality as suggested in the manual, the `-E` flag to recalculate base quality, and the `-t` flag to calculate and add allelic depth (AD), strand bias (SP), and depth (DP) to the output. We only included reads with a minimum mapping quality of 20 (`-q 20`) and bases with a minimum quality of 30 (`-Q 30`). We then called variants and calculated genotype posterior probabilities (`-f GP`) using the multiallelic SNP caller (`-m`) in `bcftools` v1.5 (Li, 2011;

Danecek et al., 2014).

We used `vcftools` v0.1.14 (Danecek et al., 2011) to filter these raw called variants. We retained only (1) SNPs (`--remove-indels`); (2) with two alleles (`--min-alleles` and `--max-alleles` of 2); (3) a minor allele frequency greater than 5% (`--maf` 0.05); (4) a minimum genotype quality score of 25 (`--minQ` 25); (5) a maximum of 50% missing genotypes across the experiment (`--max-missing` 0.5); (6) a minimum mean depth per sample of 5 (`--min-meanDP` 5); and (7) a maximum total depth of 42, which is 2 standard deviations higher than the mean coverage (`--maxDP` 42). After filtering, we retained 85888 SNPs for analysis.

We used `PLINK` v1.9 to perform the CMH test on the four islands in each treatment (Chang et al., 2015; Purcell and Chang, 2017). We used `PLINK` to convert our filtered VCF file into `PLINK` `bed/bim/fam` files, then ran the CMH test in `PLINK` using the `--mh` module. We used the `--allow-extra-chr` and `--allow-no-sex` flags to override `PLINK`'s default checks for human chromosome names.

Estimating variance effective population size

We estimated the average variance effective population size, N_e, for each island over the course of our experiment following Jorde and Ryman (2007). Specifically, we estimated N_e as:

$$N_{\mathrm{e}} = \frac{t}{2F}$$

where t is the number of generations (5) and F is an estimate of the variance in the change in allele frequency (Δp) across timepoints. We estimated F as:

$$F = \frac{\sum_{i=1}^{a} (x_i - y_i)^2}{\sum_{i=1}^{a} z_i (1 - z_i)}$$

where x and y are the frequencies for a given allele at the two timepoints, z is the average frequency of the allele across timepoints, and $i_1, i_2, ..., i_\alpha$ indexes SNPs.

Association mapping of focal traits

The extent to which parallel phenotypic changes will lead to parallel change in allele frequency depends on the genetic architecture of the trait. We examined the genetic architecture of our main focal traits to see if there were associated SNPs for which we could track allele frequency changes. We performed genotype-

phenotype association mapping using the Bayesian Sparse Linear Mixed Model (BSLMM) approach of the program `GEMMA` v0.97. (Zhou and Stephens, 2012; Zhou et al., 2013). Importantly, `GEMMA` estimates a genetic relatedness matrix to account for population structure during association mapping (Zhou et al., 2013). We first converted our VCF file of filtered SNPs into a BIMBAM mean genotype file using a custom Perl script (V. Soria-Carrasco, pers. communication). This mean genotype file uses genotype likelihoods to calculate the posterior mean genotype for each individual. The posterior mean genotype is a value between 0 (no copies of the alternate allele) and 2 (2 copies of the reference allele). This approach provides some of the benefits of the more complete likelihood models as implemented in a program like `angsd`, that is, it accounts for some genotyping uncertainty. For samples from which we had genotypic data but could not measure a phenotype, we encoded the phenotype value as missing data but kept all genetic data for use in the calculation of the genetic relatedness matrix. Then, we ran GEMMA using the following options: `-bslmm` 1: use the BSLMM model, `-w` 5000000: 5 million warm-up iterations, `-s` 25000000: 25 million sampling iterations, `-rpace` 1000: thinning interval of 1000 to target 25k samples from the posterior total, `-wpace` 10000: writing interval of 10000, and `-seed`, the seed for random number generation. We ran `GEMMA` three times to create three independent chains, each with a unique seed. GEMMA returns the full posterior distribution of its hyperparameters, and we used these distributions to check for chain convergence using the Gelman-Rubin $\hat{R}$ statistic (Gelman and Rubin, 1992). For individual SNPs, `GEMMA` calculates a posterior inclusion probability (PIP), the posterior probability that the SNP makes a significant individual contribution to the phenotype, and β, the individual SNP's effect on the phenotype. For each SNP, we averaged these estimates across the three chains. We considered a SNP to be significantly associated with a trait if it had an average PIP > 0.95. However, we found no SNPs that met this criterion (all posterior inclusion probabilities < 0.95), and thus could not test for parallel allele frequency changes at individual SNPs associated with our traits.

Additional genetic analyses

We performed two sets of additional genetic analyses not reported in the main text. Given that some phenotypic non-parallelism is due to differences across sex, we examined sex-specific allele frequency changes. Using the same ANGSD pipeline described above, we estimated allele frequencies separately for each sex for each timepoint on each island. We then calculated all pairwise correlations in the change in allele frequency for each sex (i.e., $r^{male}_{\Delta p}$ and $r^{female}_{\Delta p}$) across populations. However, as with the per-population correlations in allele frequency change, we did not find an increased correlation in allele frequency change

within treatments, and linear models found that predators and competitors had no significant effect on sex-specific correlations in allele frequency (results not shown).

We also examined genetic parallelism only for SNPs found within genes: these genes may be more likely to be affected by selection, and thus may be more likely to display parallel changes in allele frequency. Though the draft *A. sagrei* genome is not annotated, we used `LiftOff` (Shumate and Salzberg, 2021) to transfer the *A. carolinensis* annotation to the *A. sagrei* draft genome. We then re-did our analysis of genetic parallelism, examining the correlations in allele frequency change only at SNPs within genes. For the CMH test, we used the harmonic mean method of combining dependent tests to test for a genome-wide signal of parallelism, but considering only SNPs within genes. In both analyses, examining only genic SNPs did not reveal any signals of parallelism. Per-population correlations in allele frequency change were similar to the pattern seen when considering all SNPs. Likewise, there was no signal of genome-wide parallelism for genic SNPs in the CMH test (all $p > 0.05$).

Supplemental tables and figures

Table S1: Names, locations, and characteristics of experimental islands.

alias	island	latitude	longitude	treatment	vegetated area (m2)	initial A. sagrei abundance
CON1	936	24.19	-76.47	CON	2772	327.4
CON2	ANDREW	24.17	-76.45	CON	1758	164.2
CON3	332	24.21	-76.50	CON	1450	41.8
CON4	5	24.15	-76.45	CON	1333	105.9
PRED1	930	24.19	-76.47	PRED	2582	311.0
PRED2	WBC	24.15	-76.45	PRED	1575	153.3
PRED3	314	24.24	-76.52	PRED	1400	132.5
PRED4	204	24.23	-76.49	PRED	487	57.6
COMP1	311	24.24	-76.49	COMP	2241	456.8
COMP2	6	24.15	-76.45	COMP	1851	311.0
COMP3	931	24.21	-76.49	COMP	1070	241.4
COMP4	305	24.24	-76.50	COMP	603	199.4
ALL1	926	24.16	-76.47	ALL	3320	333.0
ALL2	922	24.19	-76.47	ALL	1648	149.9
ALL3	1	24.15	-76.47	ALL	1429	248.6
ALL4	312	24.24	-76.52	ALL	640	125.9

Table S2: Estimates of repeatability for measured phenotypic traits across years

trait	measurer	2011	2016
femur	TJT	1.000	0.999
fore.limb	TJT	1.000	1.000
fore.scale	ACAV	0.877	0.895
fore.scale	TJT	0.946	0.910
head.length	TJT	1.000	0.999
head.width	TJT	1.000	0.999
hind.limb	TJT	1.000	0.999
hind.scale	ACAV	0.862	0.915
hind.scale	TJT	0.922	0.922
humerus	TJT	0.999	0.998
pectoral.width	TJT	1.000	0.999
pelvic.width	TJT	0.999	0.999
SVL	TJT	1.000	1.000
tibia	TJT	1.000	0.998
toe.III	TJT	0.998	0.995
toe.IV	TJT	1.000	0.999
ulna	TJT	1.000	0.998

Table S3: Number of individuals phenotyped (excluding juveniles) and genotyped for each island in each year

alias	island	year	phenotyped	genotyped
CON1	936	2011	29	27
CON1	936	2016	35	62
CON2	ANDREW	2011	27	30
CON2	ANDREW	2016	46	49
CON3	332	2011	12	12
CON3	332	2016	33	38
CON4	5	2011	29	31
CON4	5	2016	39	51
PRED1	930	2011	26	24
PRED1	930	2016	37	42
PRED2	WBC	2011	29	28
PRED2	WBC	2016	47	48
PRED3	314	2011	29	32
PRED3	314	2016	23	33
PRED4	204	2011	16	19
PRED4	204	2016	9	20
COMP1	311	2011	21	18
COMP1	311	2016	33	31
COMP2	6	2011	29	32
COMP2	6	2016	40	41
COMP3	931	2011	29	35
COMP3	931	2016	37	59
COMP4	305	2011	23	25
COMP4	305	2016	35	36
ALL1	926	2011	29	31
ALL1	926	2016	55	60
ALL2	922	2011	22	22
ALL2	922	2016	17	20
ALL3	1	2011	27	26
ALL3	1	2016	44	47
ALL4	312	2011	30	31
ALL4	312	2016	30	33

Table S4: Results for LMMs of the effects of predators, competitors, and their interaction on brown anole traits for which we made predictions.

trait	sex	term	posterior mean	posterior median	lower 95% CI	upper 95% CI	significant
SVL	F	pred:year	-0.686	-0.697	-1.477	0.152	false
SVL	F	comp:year	0.146	0.141	-0.710	0.878	false
SVL	F	pred:comp:year	0.193	0.189	-0.843	1.250	false
SVL	M	pred:year	-1.762	-1.757	-3.021	-0.637	true
SVL	M	comp:year	-0.395	-0.393	-1.598	0.677	false
SVL	M	pred:comp:year	0.158	0.176	-1.388	1.949	false
hind.limb	F	pred:year	0.004	0.005	-0.374	0.392	false
hind.limb	F	comp:year	0.048	0.050	-0.324	0.410	false
hind.limb	F	pred:comp:year	-0.309	-0.309	-0.819	0.186	false
hind.limb	M	pred:year	0.391	0.392	-0.129	0.819	false
hind.limb	M	comp:year	0.128	0.129	-0.303	0.557	false
hind.limb	M	pred:comp:year	-0.428	-0.427	-1.087	0.181	false
hind.scale	F	pred:year	-0.081	-0.078	-0.381	0.202	false
hind.scale	F	comp:year	0.004	0.003	-0.269	0.293	false
hind.scale	F	pred:comp:year	-0.174	-0.176	-0.536	0.190	false
hind.scale	M	pred:year	-0.375	-0.377	-0.636	-0.106	true
hind.scale	M	comp:year	-0.199	-0.196	-0.458	0.047	false
hind.scale	M	pred:comp:year	0.320	0.320	-0.016	0.688	false

Table S5: Results for LMMs of the effects of predators, competitors, and their interaction on hindlimb lamella number while correcting for body size (SVL).

trait	sex	term	posterior mean	posterior median	lower 95% CI	upper 95% CI	significant
hind.scale	F	SVL	0.035	0.037	-0.074	0.134	false
hind.scale	F	pred:year	-0.057	-0.054	-0.349	0.233	false
hind.scale	F	comp:year	0.014	0.012	-0.276	0.278	false
hind.scale	F	pred:comp:year	-0.191	-0.188	-0.542	0.184	false
hind.scale	M	SVL	0.066	0.068	-0.027	0.164	false
hind.scale	M	pred:year	-0.262	-0.264	-0.543	0.014	false
hind.scale	M	comp:year	-0.229	-0.227	-0.486	0.041	false
hind.scale	M	pred:comp:year	0.288	0.287	-0.072	0.655	false

Table S6: Results for LMMs of the effects of predators, competitors, and their interaction on brown anole traits for which we did not make predictions.

trait	sex	term	posterior mean	posterior median	lower 95% CI	upper 95% CI
femur	F	pred:year	0.047	0.047	-0.066	0.155
femur	F	comp:year	-0.016	-0.017	-0.121	0.087
femur	F	pred:comp:year	-0.163	-0.163	-0.319	-0.023
femur	M	pred:year	0.169	0.169	0.016	0.315
femur	M	comp:year	0.019	0.020	-0.118	0.160
femur	M	pred:comp:year	-0.125	-0.126	-0.334	0.081
fore.limb	F	pred:year	-0.026	-0.026	-0.267	0.185
fore.limb	F	comp:year	-0.122	-0.121	-0.337	0.078
fore.limb	F	pred:comp:year	-0.010	-0.013	-0.307	0.263
fore.limb	M	pred:year	0.092	0.091	-0.176	0.365
fore.limb	M	comp:year	-0.062	-0.062	-0.315	0.182
fore.limb	M	pred:comp:year	-0.058	-0.056	-0.451	0.317
fore.scale	F	pred:year	0.171	0.168	-0.083	0.445
fore.scale	F	comp:year	-0.046	-0.044	-0.298	0.216
fore.scale	F	pred:comp:year	-0.102	-0.101	-0.434	0.232
fore.scale	M	pred:year	0.005	0.003	-0.284	0.261
fore.scale	M	comp:year	-0.281	-0.282	-0.528	-0.028
fore.scale	M	pred:comp:year	0.218	0.219	-0.113	0.560
head.length	F	pred:year	-0.001	-0.001	-0.099	0.106
head.length	F	comp:year	-0.094	-0.093	-0.198	0.006
head.length	F	pred:comp:year	0.099	0.099	-0.045	0.231
head.length	M	pred:year	0.071	0.070	-0.044	0.218
head.length	M	comp:year	-0.039	-0.038	-0.162	0.085
head.length	M	pred:comp:year	0.184	0.186	-0.008	0.361
head.width	F	pred:year	-0.080	-0.081	-0.165	0.003

Table S6: Results for LMMs of the effects of predators, competitors, and their interaction on brown anole traits for which we did not make predictions. *(continued)*

trait	sex	term	posterior mean	posterior median	lower 95% CI	upper 95% CI
head.width	F	comp:year	-0.043	-0.042	-0.126	0.038
head.width	F	pred:comp:year	0.026	0.026	-0.090	0.137
head.width	M	pred:year	-0.034	-0.033	-0.137	0.058
head.width	M	comp:year	-0.169	-0.169	-0.260	-0.077
head.width	M	pred:comp:year	0.091	0.090	-0.057	0.222
humerus	F	pred:year	-0.075	-0.076	-0.166	0.022
humerus	F	comp:year	-0.050	-0.050	-0.137	0.045
humerus	F	pred:comp:year	0.061	0.060	-0.064	0.180
humerus	M	pred:year	0.059	0.058	-0.069	0.181
humerus	M	comp:year	0.051	0.051	-0.054	0.163
humerus	M	pred:comp:year	-0.076	-0.077	-0.242	0.096
pectoral.width	F	pred:year	-0.011	-0.012	-0.107	0.078
pectoral.width	F	comp:year	-0.047	-0.048	-0.140	0.044
pectoral.width	F	pred:comp:year	0.026	0.027	-0.098	0.152
pectoral.width	M	pred:year	0.034	0.035	-0.075	0.145
pectoral.width	M	comp:year	-0.126	-0.127	-0.241	-0.021
pectoral.width	M	pred:comp:year	0.029	0.029	-0.130	0.188
pelvic.width	F	pred:year	0.044	0.044	-0.029	0.114
pelvic.width	F	comp:year	-0.027	-0.027	-0.096	0.044
pelvic.width	F	pred:comp:year	-0.004	-0.003	-0.102	0.082
pelvic.width	M	pred:year	-0.019	-0.018	-0.084	0.051
pelvic.width	M	comp:year	-0.034	-0.034	-0.099	0.031
pelvic.width	M	pred:comp:year	-0.005	-0.007	-0.096	0.095
tibia	F	pred:year	-0.002	-0.003	-0.114	0.099
tibia	F	comp:year	-0.025	-0.024	-0.126	0.079
tibia	F	pred:comp:year	-0.026	-0.024	-0.177	0.113

Table S6: Results for LMMs of the effects of predators, competitors, and their interaction on brown anole traits for which we did not make predictions. *(continued)*

trait	sex	term	posterior mean	posterior median	lower 95% CI	upper 95% CI
tibia	M	pred:year	0.084	0.083	-0.041	0.212
tibia	M	comp:year	0.054	0.054	-0.060	0.175
tibia	M	pred:comp:year	-0.112	-0.112	-0.278	0.072
toe.III	F	pred:year	0.003	0.003	-0.075	0.074
toe.III	F	comp:year	0.014	0.014	-0.055	0.089
toe.III	F	pred:comp:year	-0.001	0.001	-0.105	0.089
toe.III	M	pred:year	-0.011	-0.012	-0.103	0.084
toe.III	M	comp:year	-0.061	-0.061	-0.155	0.021
toe.III	M	pred:comp:year	0.078	0.078	-0.051	0.211
toe.IV	F	pred:year	-0.030	-0.030	-0.211	0.145
toe.IV	F	comp:year	0.027	0.028	-0.137	0.203
toe.IV	F	pred:comp:year	-0.112	-0.113	-0.326	0.133
toe.IV	M	pred:year	0.131	0.130	-0.075	0.322
toe.IV	M	comp:year	0.001	0.002	-0.184	0.195
toe.IV	M	pred:comp:year	-0.145	-0.145	-0.419	0.139
ulna	F	pred:year	0.001	0.001	-0.081	0.075
ulna	F	comp:year	-0.011	-0.011	-0.089	0.060
ulna	F	pred:comp:year	-0.023	-0.024	-0.123	0.079
ulna	M	pred:year	0.041	0.041	-0.060	0.135
ulna	M	comp:year	-0.024	-0.024	-0.116	0.061
ulna	M	pred:comp:year	0.010	0.010	-0.120	0.149

Table S7: The minimum standardized selection gradient on each trait due to the introduction of predators that we have 80% power to detect. Minimum strength of selection varies according to whether we assume a high (h^2 = 0.74), moderate (h^2 = 0.55), or low (h^2 = 0.107) level of heritability.

		min selection gradient		
trait	sex	high h2	moderate h2	low h2
SVL	M	0.222	0.299	1.535
SVL	F	0.222	0.299	1.535
relative hindlimb length	M	0.089	0.120	0.614
relative hindlimb length	F	0.163	0.220	1.130
hindlimb lamellae	M	0.226	0.304	1.565
hindlimb lamellae	F	0.262	0.352	1.811

Table S8: Results for linear model of the effects of predators, competitors, and their interaction on the magnitude of phenotypic change, D, for each population

term	estimate	std.error	statistic	p.value
(Intercept)	2.259	0.313	7.209	0.000
predPRESENT	-0.257	0.443	-0.580	0.572
compPRESENT	-0.428	0.443	-0.967	0.353
predPRESENT:compPRESENT	0.450	0.627	0.718	0.486

Table S9: Results for linear model of the effects of predators, competitors, and their interaction on differences in the magnitude of phenotypic change, ΔD, in pairwise comparisons within each treatmeant.

term	estimate	std.error	statistic	p.value
(Intercept)	0.969	0.241	4.026	0.001
predPRESENT	-0.325	0.340	-0.957	0.350
compPRESENT	-0.243	0.340	-0.714	0.484
predPRESENT:compPRESENT	-0.075	0.481	-0.156	0.878

Table S10: Results for linear model of the effects of predators, competitors, and their interaction on differences in the direction of phenotypic change, θ, in pairwise comparisons within each treatmeant.

term	estimate	std.error	statistic	p.value
(Intercept)	40.506	5.109	7.929	0.000
predPRESENT	17.723	7.225	2.453	0.023
compPRESENT	26.081	7.225	3.610	0.002
predPRESENT:compPRESENT	-46.188	10.217	-4.521	0.000

Table S11: Results for linear model of the effects of predators, competitors, and their interaction on the amount of phenotypic convergence, Ω, in pairwise comparisons within each treatmeant.

term	estimate	std.error	statistic	p.value
(Intercept)	-0.373	0.099	-3.783	0.001
predPRESENT	0.343	0.140	2.459	0.023
compPRESENT	0.337	0.140	2.412	0.026
predPRESENT:compPRESENT	-0.354	0.197	-1.791	0.088

Table S12: Results for linear model of the effects of predators, competitors, and their interaction on the amount of correlation in allele frequency change in pairwise comparisons within each treatmeant for all SNPs.

term	estimate	std.error	statistic	p.value
(Intercept)	0.078	0.036	2.143	0.045
predPRESENT	-0.111	0.052	-2.147	0.044
compPRESENT	-0.002	0.052	-0.031	0.976
predPRESENT:compPRESENT	0.030	0.073	0.413	0.684

Table S13: Estimates of variance effective population size, N_e, for each island.

alias	island	Ne
CON1	936	109
CON2	ANDREW	104
CON3	332	73
CON4	5	90
PRED1	930	98
PRED2	WBC	129
PRED3	314	79
PRED4	204	44
COMP1	311	48
COMP2	6	89
COMP3	931	97
COMP4	305	98
ALL1	926	145
ALL2	922	78
ALL3	1	88
ALL4	312	107

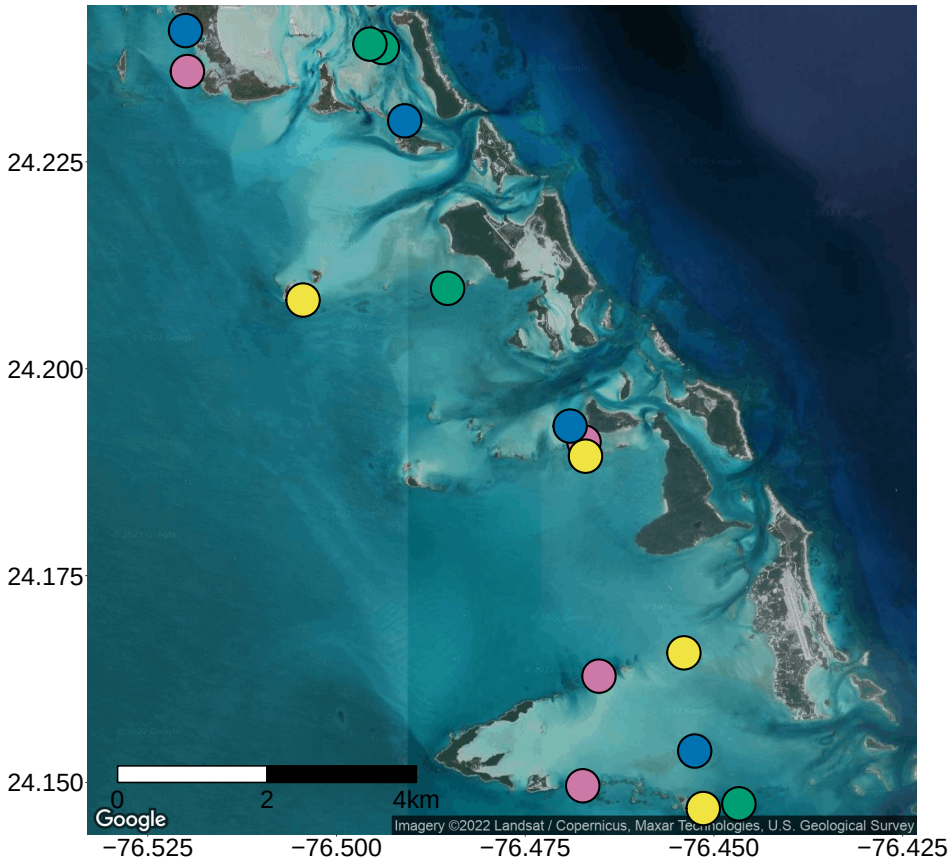

Figure S1: The 16 experimental islands near Staniel Cay in the Exuma Chain of the Bahamas. Islands are coloured by treatment type as in the main text: yellow, control islands; blue, predator islands; green, competitor islands; pink, all islands.

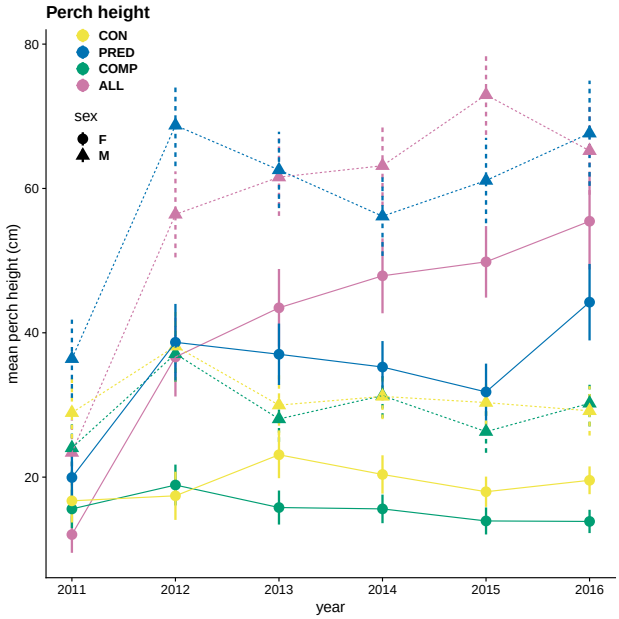

Figure S2: Change in perch height through time split by sexes. Males: triangles and dashed lines. Females: circles and solid lines. Points show treatment mean ± 2SE, pooling across islands.

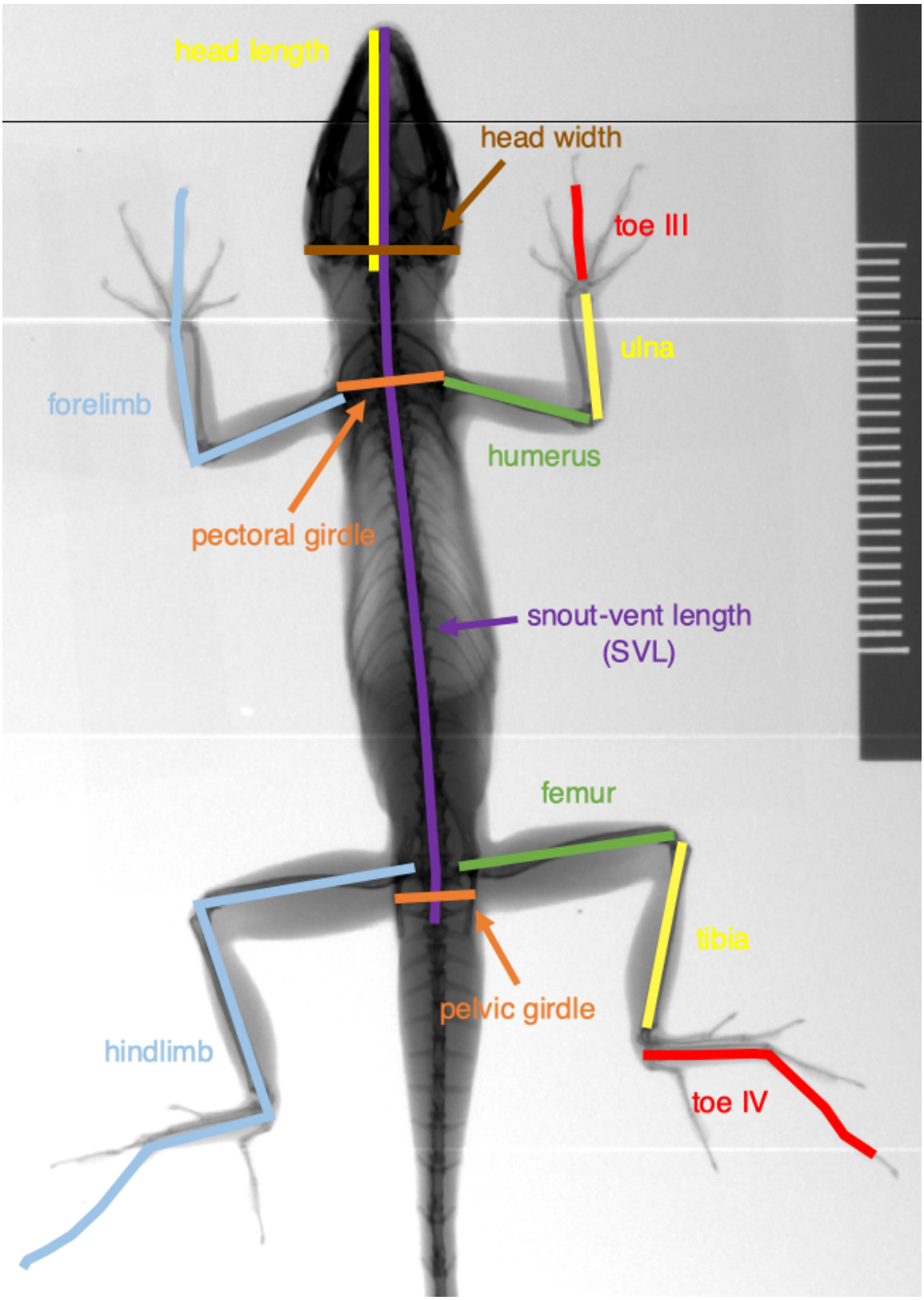

Figure S3: An example X-ray image of an *Anolis sagrei* male, labeled with the 13 skeletal traits we measured.

Supplemental Material (not copyedited or formatted) for: Timothy J. Thurman, Todd M. Palmer, Jason J. Kolbe, Arash M. Askary, Kiyoko M. Gotanda, Oriol Lapiedra, Tyler R. Kartzinel, Naomi Man in't Veld, Liam J. Revell, Johanna E. Wegener, Thomas W. Schoener, David A. Spiller, Jonathan B. Losos, Robert M. Pringle, Rowan D. H. Barrett. 2023. "The Difficulty of Predicting Evolutionary Change in Response to Novel Ecological Interactions: A Field Experiment with Anolis Lizards." The American Naturalist 201(4). DOI: https://doi.org/10.1086/723209.

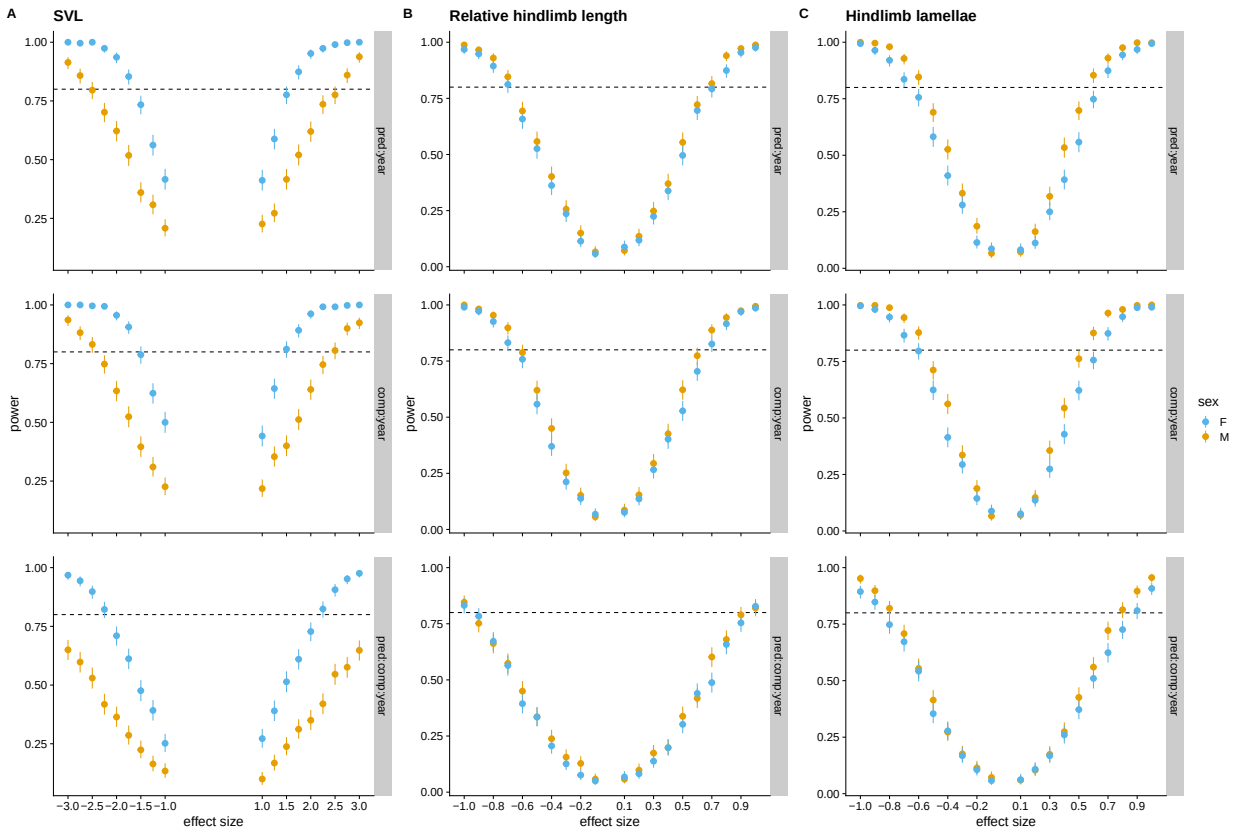

Figure S4: Power analysis of our linear mixed models. Power curves were calculated for a range of effect sizes for each statistical term of interest for each trait: **A**, SVL; **B**, relative hindlimb length; **C**, number of hindlimb lamellae. Points and lines show estimated power and 95% CIs for a given effect size. The dashed line shows the 80% power threshold.

Supplemental Material (not copyedited or formatted) for: Timothy J. Thurman, Todd M. Palmer, Jason J. Kolbe, Arash M. Askary, Kiyoko M. Gotanda, Oriol Lapiedra, Tyler R. Kartzinel, Naomi Man in't Veld, Liam J. Revell, Johanna E. Wegener, Thomas W. Schoener, David A. Spiller, Jonathan B. Losos, Robert M. Pringle, Rowan D. H. Barrett. 2023. "The Difficulty of Predicting Evolutionary Change in Response to Novel Ecological Interactions: A Field Experiment with Anolis Lizards." The American Naturalist 201(4). DOI: https://doi.org/10.1086/723209.

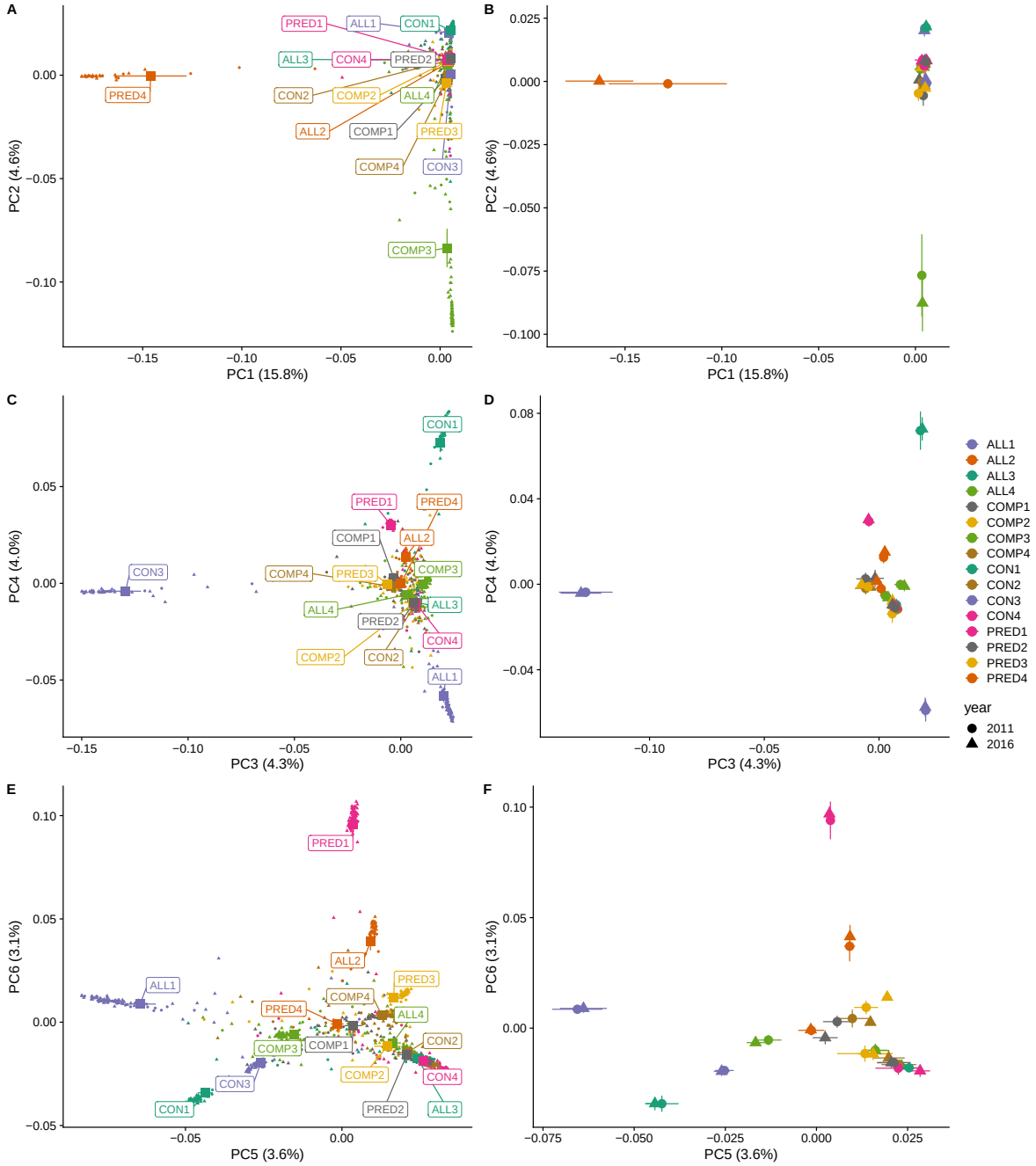

Figure S5: Principal component analysis of genome-wide population structure across islands. **A**, **C**, **E**: Individual (small circles and triangles) and per-island average (large squares) PC scores for the first 6 principal components of a genome-wide PCA. Vertical and horizontal lines around per-island averages show approximate 95% CIs (1.96 x SE). **B**, **D**, **F**: Per-island average PC scores at the beginning of the experiment in 2011 (circles) and after 5 generations in 2016 (triangles). Lines show approximate 95% CIs.

References

Bates, D., Mächler, M., Bolker, B., and Walker, S. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1–48.

Beuttell, K. and Losos, J. B. 1999. Ecological morphology of Caribbean anoles. Herpetological Monographs 13:1–28.

Butler, M. A. and Losos, J. B. 2002. Multivariate sexual dimorphism, sexual selection, and adaptation in Greater Antillean *Anolis* lizards. Ecological Monographs 72:541–559.

Calsbeek, R. 2009. Experimental evidence that competition and habitat use shape the individual fitness surface. Journal of Evolutionary Biology 22:97–108.

Calsbeek, R. and Bonneaud, C. 2008. Postcopulatory fertilization bias as a form of cryptic sexual selection. Evolution 62:1137–1148.

Calsbeek, R. and Smith, T. B. 2007. Probing the adaptive landscape using experimental islands: density-dependent natural selection on lizard body size. Evolution 61:1052–1061.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22:3124–3140.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. 2011. *Stacks*: building and genotyping loci *de novo* from short-read sequences. G3 1:171–182.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4:7.

Cox, R. M. and Calsbeek, R. 2010. Sex-specific selection and intraspecific variation in sexual size dimorphism. Evolution 64:798–809.

Danecek, P., Auton, A., Abecasis, G. R., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Danecek, P., Schiffels, S., and Durbin, R., 2014. Multiallelic calling model in bcftools (-m), http://samtools.github.io/bcftools/call-m.pdf. URL http://samtools.github.io/bcftools/call-m.pdf.

Donihue, C. M., Herrel, A., Fabre, A.-C., Kamath, A., Geneva, A. J., Schoener, T. W., Kolbe, J. J., and Losos, J. B. 2018. Hurricane-induced selection on the morphology of an island lizard. Nature 560:88–91.

Gelman, A. and Hill, J., 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York, USA.

Gelman, A. and Rubin, D. B. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7:457–472.

Glossip, D. and Losos, J. B. 1997. Ecological correlates of number of subdigital lamellae in Anoles. Herpetologica 53:192–199.

Green, P. and MacLeod, C. J. 2016. SIMR: an R package for power analysis of generalized linear mixed models by simulation. Methods in Ecology and Evolution 7:493–498.

Hecht, M. K. 1952. Natural selection in the lizard genus *Aristelliger*. Evolution 6:112–114.

Heckel, D. G. and Roughgarden, J. 1979. A technique for estimating the size of lizard populations. Ecology 60:966–975.

Hereford, J., Hansen, T. F., and Houle, D. 2004. Comparing strengths of directional selection: how strong is strong? Evolution 58:2133–2143.

Herrel, A., Joachim, R., Vanhooydonck, B., and Irschick, D. J. 2006. Ecological consequences of ontogenetic changes in head shape and bite performance in the jamaican lizard *Anolis lineatopus*. Biological Journal of the Linnean Society 89:443–454.

Herrel, A., Mcbrayer, L. D., and Larson, P. M. 2007. Functional basis for sexual differences in bite force in the lizard *Anolis carolinensis*. Biological Journal of the Linnean Society 91:111–119.

Jorde, P. E. and Ryman, N. 2007. Unbiased estimator for genetic drift and effective population size. Genetics 177:927–935.

Kim, S. Y., Lohmueller, K. E., Albrechtsen, A., Li, Y., Korneliussen, T., Tian, G., Grarup, N., Jiang, T., Andersen, G., Witte, D., Jørgensen, T., Hansen, T., Pedersen, O., Wang, J., and Nielsen, R. 2011. Estimation of allele frequency and association mapping using next-generation sequencing data. BMC Bioinformatics 12:231.

Kolbe, J. J., Leal, M., Schoener, T. W., Spiller, D. A., and Losos, J. B. 2012. Founder effects persist despite adaptive differentiation: a field experiment with lizards. Science 335:1086–1089.

Korneliussen, T. S., Albrechtsen, A., and Nielsen, R. 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics 15:356.

Les, A. M. and Powell, R. 2014. *Anolis smaragdinus*. Catalogue of American Amphibians and Reptiles 902:1–14.

Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993.

Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. R., Durbin, R., and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Losos, J. B. 1990*a*. Ecomorphology, performance capability, and scaling of West Indian *Anolis* lizards: an evolutionary analysis. Ecological Monographs 60:369–388.

Losos, J. B. 1990*b*. The evolution of form and function: morphology and locomotor performance in West Indian *Anolis* lizards. Evolution 44:1189–1203.

Losos, J. B. and Irschick, D. J. 1996. The effect of perch diameter on escape behaviour of *Anolis* lizards: laboratory predictions and field tests. Animal Behaviour 51:593–602.

Losos, J. B., Schoener, T. W., Langerhans, R. B., and Spiller, D. A. 2006. Rapid temporal reversal in predator-driven natural selection. Science 314:1111.

Losos, J. B., Schoener, T. W., and Spiller, D. A. 2004. Predator-induced behaviour shifts and natural selection in field-experimental lizard populations. Nature 432:505–508.

Losos, J. B. and Sinervo, B. 1989. The effects of morphology and perch diameter on sprint performance of *Anolis* lizards. Journal of Experimental Biology 145:23–30.

Losos, J. B. and Spiller, D. A. 1999. Differential colonization success and asymmetrical interactions between two lizard species. Ecology 80:252–258.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17.

McGlothlin, J. W., Cox, R. M., and Brodie, E. D. 2019. Sex-specific selection and the evolution of between-sex genetic covariance. Journal of Heredity 110:422–432.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135.

Pringle, R. M., Kartzinel, T. R., Palmer, T. M., Thurman, T. J., Fox-Dobbs, K., Xu, C. C. Y., Hutchinson, M. C., Coverdale, T. C., Daskin, J. H., Evangelista, D. A., Gotanda, K. M., Man in 't Veld, N. A., Wegener, J. E., Kolbe, J. J., Schoener, T. W., Spiller, D. A., Losos, J. B., and Barrett, R. D. H. 2019. Predator-induced collapse of niche structure and species coexistence. Nature 570:58–64.

Pringle, R. M., Kimuyu, D. M., Sensenig, R. L., Palmer, T. M., Riginos, C., Veblen, K. E., and Young, T. P. 2015. Synergistic effects of fire and elephants on arboreal animals in an African savanna. Journal of Animal Ecology 84:1637–1645.

Purcell, S. M. and Chang, C. C., 2017. Plink v1.90b4.6. www.cog-genomics.org/plink/1.9/.

Quinn, G. P. and Keough, M. J., 2002. Experimental design and data analysis for biologists. Cambridge University Press, Cambridge, UK.

Rohland, N. and Reich, D. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Research 22:939–946.

Schoener, T. W. and Schoener, A. 1983. The time to extinction of a colonizing propagule of lizards increases with island area. Nature 332:332–334.

Schoener, T. W., Slade, J. B., and Stinson, C. H. 1982. Diet and sexual dimorphism in the very catholic lizard genus, *Leiocephalus* of the Bahamas. Oecologia 53:160–169.

Shumate, A. and Salzberg, S. L. 2021. Liftoff: accurate mapping of gene annotations. Bioinformatics 37:1639–1643.

Stuart, Y. E., Campbell, T. S., Hohenlohe, P. A., Reynolds, R. G., Revell, L. J., and Losos, J. B. 2014. Rapid evolution of a native species following invasion by a congener. Science 346:463–466.

Vendrami, D. L. J., Telesca, L., Weigand, H., Weiss, M., Fawcett, K., Lehman, K., Clark, M. S., Leese, F., McMinn, C., Moore, H., and Hoffman, J. I. 2017. RAD sequencing resolves fine-scale population structure in a benthic invertebrate: implications for understanding phenotypic plasticity. Royal Society Open Science 4:160548–17.

Whitlock, M. C. and Schluter, D., 2009. The analysis of biological data. Roberts and Co. Publishers, Greenwood Village, CO.

Williams, E. E., 1972. The origin of faunas. Evolution of lizard congeners in a complex island fauna: a trial analysis *in* T. Dobzhansky, M. K. Hecht, and W. C. Steere, eds. Evolutionary Biology, volume 6. Springer, New York.

Winchell, K. M., Reynolds, R. G., Prado-Irwin, S. R., Puente-Rolón, A. R., and Revell, L. J. 2016. Phenotypic shifts in urban areas in the tropical lizard *Anolis cristatellus*. Evolution Pages 1–14.

Zhou, X., Carbonetto, P., and Stephens, M. 2013. Polygenic modeling with Bayesian sparse linear mixed models. PLoS Genetics 9:e1003264.

Zhou, X. and Stephens, M. 2012. Genome-wide efficient mixed-model analysis for association studies. Nature Ecology & Evolution 44:821–824.