Genomic heritability of song and condition in wild singing mice

Author names and addresses: Tracy T. Burkhard[a]*, Mikhail Matz[a], Steven M. Phelps[a]*

[a]Department of Integrative Biology, University of Texas at Austin, Austin, TX

***Corresponding authors**: tburk@utexas.edu, sphelps@utexas.edu

Key words: heritability, repeatability, *Scotinomys*, condition dependence, good genes, RAD-seq

Abstract

Courtship displays are dramatic examples of complex behaviors that vary within and among species. Evolutionary explanations for this diversity rely upon genetic variation, yet the heritability of complex phenotypes is seldom investigated in the field. Here, we estimate genomic heritability of advertisement song and body condition in a wild population of singing mice. The heritability of song exhibits a systematic pattern, with high heritability for spectral characteristics linked to vocal morphology, intermediate heritability for rhythmic patterns, and lower but significant heritability for measures of motivation, like song length and rate. Physiological measures of condition, like hormonal markers of adiposity, exhibited intermediate heritability. Among singing mice, song rate and body condition have a strong phenotypic correlation; our estimate suggests a comparable genetic correlation that merits further study. Our results illustrate how advances in genomics and quantitative genetics can be integrated in free-living species to address longstanding challenges in behavior and evolution.

Introduction

A fundamental challenge in evolutionary biology is to understand the origins of phenotypic diversity. The concept of heritability (h^2)—the proportion of phenotypic variance due to genetic differences in a population (Falconer, 1981; Visscher et al., 2008)—is crucial to this understanding because it predicts a trait's evolutionary response to selection (Falconer, 1981; Fisher, 1930; Lande, 1976; Price et al., 1991; Visscher et al., 2008). Traits of high heritability are more responsive to selection than traits of low heritability (Falconer, 1981; Lande, 1976), knowledge long exploited by livestock and agriculture breeders (Cassell et al., 2009; Dudley and Lambert, 2010; Hill and Kirkpatrick, 2010; Visscher et al., 2008) and famously illustrated by the rapid, repeated adaptation of highly heritable beak size in Darwin's finches (Grant and Grant, 1993, 1995).

Few traits pose such interesting and challenging examples of phenotypic evolution as male sexual signals. Advertisement displays are present across diverse taxonomic groups (Wiens and Tuschhoff, 2020), vary dramatically within and among species (within: Dubuc et al., 2014; Izzo and Tibbetts, 2015; Karin et al., 2018; between: Ingram et al., 2016; Vanhooydonck et al., 2009), and evolve under both natural and sexual selection (Andersson, 1981; Endler, 1992). Heritable variation is not merely a substrate for this diversity but underlies common explanations for exaggerated male traits. Models of signal evolution by female choice assume heritable variation of male display, for example, but a popular subset also predicts genetic correlations between sexual display and male fitness or condition—correlations that are interpreted to indicate that females use displays to gauge the heritable fitness of males (Andersson, 1986; Hamilton and Zuk, 1982; Kirkpatrick and Ryan, 1991; Rowe and Houle, 1996).

Of course, not all variation in displays is heritable: sexual displays are complex, encompassing both stable structural elements and the dynamic behaviors that animate them. Relative to stable morphological elements of display, behavioral components are often highly specific to social

and environmental contexts. For example, animals increase signaling effort (*e.g.* display rate or duration) in the presence of an audience, a widespread phenomenon that has been documented in vocalizing hyraxes (Demartsev et al., 2014), courting betta fish (Doutrelant et al., 2001), and many other taxa (reviewed in Zuberbühler, 2008). Behavioral components may also rely on male body condition (Cotton et al., 2004; Johnstone, 1993; Rowe and Houle, 1996), particularly if they are energetically expensive to produce (Chappell et al., 1995; Ophir et al., 2010) or incur the expense of fight or flight (Bachmann et al., 2017; Zuk and Kolluru, 1998). Different facets of display may thus reflect varying degrees of heritability. Even if selection pressures are consistent across these dimensions of display, sexual signals may evolve not as a single phenotype but as discrete components that undergo different rates of evolution (*e.g.* acoustic characteristics of frog calls, Cocroft and Ryan, 1995). Investigating the nature and extent of heritable variation underlying sexual signals is thus especially important to our broader understanding of how complex traits evolve.

Although heritability is essential to understanding evolution, it is challenging to study. By definition, heritability is highly population-specific – and its measures are sensitive to genetic variation and environmental circumstances (Falconer, 1981; Visscher et al., 2008; Waldmann, 2001; but see contrary evidence in Dochtermann et al., 2019; Waldmann, 2001). Despite its obvious value, our understanding of the quantitative genetics of wild populations has been limited due to the difficulties of ascertaining genetic relatedness in the field (Gienapp et al., 2017). One popular surrogate for heritability, especially in behavioral ecology, is repeatability, which eliminates the need for *a priori* knowledge of relatedness altogether (Bell et al., 2009; Boake, 1989; Dingemanse et al., 2002). Repeatability is facile and offers an upper bound estimate for heritability (Boake, 1989), but because direct extrapolation of repeatability to heritability can vary in reliability (*e.g.* can underestimate heritability, Dohm, 2002; St-Hilaire et al., 2017), its most appropriate use may be limited to hypothesis generation. Conversely,

rigorous pedigree analysis, long the standard for quantitative genetics, is labor-intensive and impractical for many species that are either long-lived or difficult to track (Gienapp et al., 2017).

Recent advances in genomics and quantitative genetics afford new opportunities for assessing heritability in non-model and wild populations. In traditional pedigree analysis, pedigrees are fitted into animal models, linear (or generalized linear) mixed models first developed for animal breeding programs (Henderson, 1950; Hill and Kirkpatrick, 2010) and later applied to evolutionary questions (Kruuk, 2004; Lynch et al., 1998). The advent of reduced representation next-generation sequencing techniques, such as RAD-seq (Peterson et al., 2012), allows researchers to estimate population genome-wide relatedness without need for pedigree, marking a significant turning point for researchers interested in the quantitative genetics of wild populations. Genomic relatedness matrices (GRM) are estimated by identity-by-descent inference (Mousseau et al., 1998; Ritland, 1996) and can then be fitted into mixed models in lieu of pedigrees, resulting in estimates of “SNP-based” or “genomic” heritability. Researchers in human genetics, biostatistics, and quantitative genetics have developed computational tools that accommodate complex genomic models, which has led to important insights into complex human traits (*e.g.* Yang et al., 2010). Researchers in ecology and evolution have recently begun to apply these methods in non-model species to investigate heritable variation in the wild (Gervais et al., 2020, 2019; Perrier et al., 2018).

Here, we apply SNP-based methods to investigate patterns of heritability in singing behavior and body condition in a Costa Rican population of Alston’s singing mice (*Scotinomys teguina*). Singing mice are small, diurnal, and insectivorous muroid rodents of the family Cricetidae, eponymously named for their stereotyped advertisement songs (Hooper, 1972; Fig. 1). *Scotinomys* song makes a particularly interesting model for the study of signal evolution, being both acoustically complex and ecologically important. By the standards of muroid rodent vocalizations, songs are loud and long (7-10 s; Hooper and Carleton, 1976), comprising a rapidly repeated series of stereotyped, tonal frequency sweeps notes that span both audible

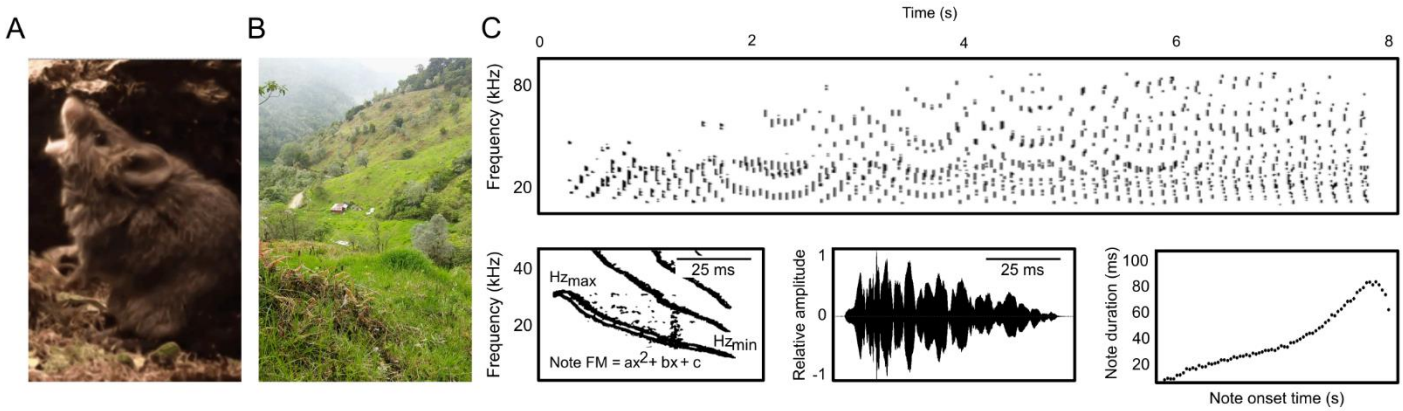

Figure 1: Singing behavior of a wild population of Alston's singing mice (*Scotinomys teguina*). *(A)* Still from a video of a singing male mouse. *(B)* Typical grassland and forest habitat characterizing San Gerardo de Dota, Costa Rica. *(C)* Example spectrogram of an advertisement song. Lower panels: *left* and *middle* depict a close-up of a single note at 5 s during the song. *Left,* spectrogram detail of note with minimum and maximum frequency labeled. A quadratic curve was fit to each note to measure note frequency modulation (FM). FMa describes note curvature, FMb describes slope, and FMc describes starting frequency. *Middle,* waveform detail of note with peak amplitude, peak amplitude within first quarter of note, and note duration labeled. *Right,* note duration changes as a function of song length. Photo credit: *(A)* B. Pasch, *(B)* T. Burkhard.

and ultrasonic (> 20kHz) frequencies. This suite of acoustic traits characterizes a notable departure from the short and simple ancestral ultrasonic vocalizations typical of most muroid rodents (Miller and Engstrom, 2007, 2012) and makes song detectable over long distances. Indeed, singing mouse song seems designed for that purpose, playing roles in both mate attraction and male-male competition (Fernández-Vargas et al., 2011; Hooper and Carleton, 1976; George, 2014; Pasch et al., 2011a, 2013). Interestingly, aspects of singing effort are predicted by male body condition (Burkhard et al., 2018; Pasch et al., 2011a), providing a putative mechanism for signaling decisions and female preferences.

Table 1: Trapping summary. Elevation taken at GPS waypoint.

Site	Latitude	Longitude	Elevation (m)	Sampled mice
Robles	9°33'7.13"N	83°47'44.57"W	2580	4
QERC	9°33'9.44"N	83°48'22.88"W	2227	134
PH	9°54'9.55"N	83°81'0.23"W	2221	6
PC	9°58'8.55"N	83°79'9.22"W	2577	1
MH	9°58'0.62"N	83°79'9.35"W	2521	20

To examine patterns of heritability within this species, we recorded and analyzed songs and singing behavior from 168 wild-caught mice from San Gerardo de Dota, Costa Rica (Fig. 1, Table 1), and we quantified acoustic repeatability within and variation among individuals. We measured a variety of condition phenotypes from these mice, including morphometric traits, plasma nutrients, and circulating hormone levels, and then examined the phenotypic relationships between condition and song (Fig. 2, Table 2). We generated genotype data for 157 mice using ddRAD-seq, and then deduced pairwise genetic covariance matrices. Finally, we used these data to fit animal models to explore patterns of heritability and co-heritability in song and condition phenotypes. Our study provides the first heritability estimates for complex signaling behaviors and their physiological substrates in a wild population and illustrates the utility of such methods to address evolutionary questions.

Results

Phenotypic sampling

We sampled a population of singing mice from San Gerardo de Dota, Costa Rica (Fig. 3, Table 1), taking a variety of measurements of song and condition from these wild-caught animals. Summary statistics, sample sizes, and breakdown by sex for all measured phenotypes are reported in Table 2.

We successfully recorded long songs and spontaneous song rate (*i.e.* number of songs in two hours) from 106 and 129 mice, respectively (Table 2). Mice with at least one song were included in principal components analysis (PCA) of acoustic variation, which revealed two latent variables, PC1 (“song effort”, 23.1% of overall variance), which explained individual variation in amplitude and duration, and PC2 (“spectral characteristics”, 19.1%), which explained individual variation in frequency characteristics of song, particularly note shape (Fig. 2, Table S1).

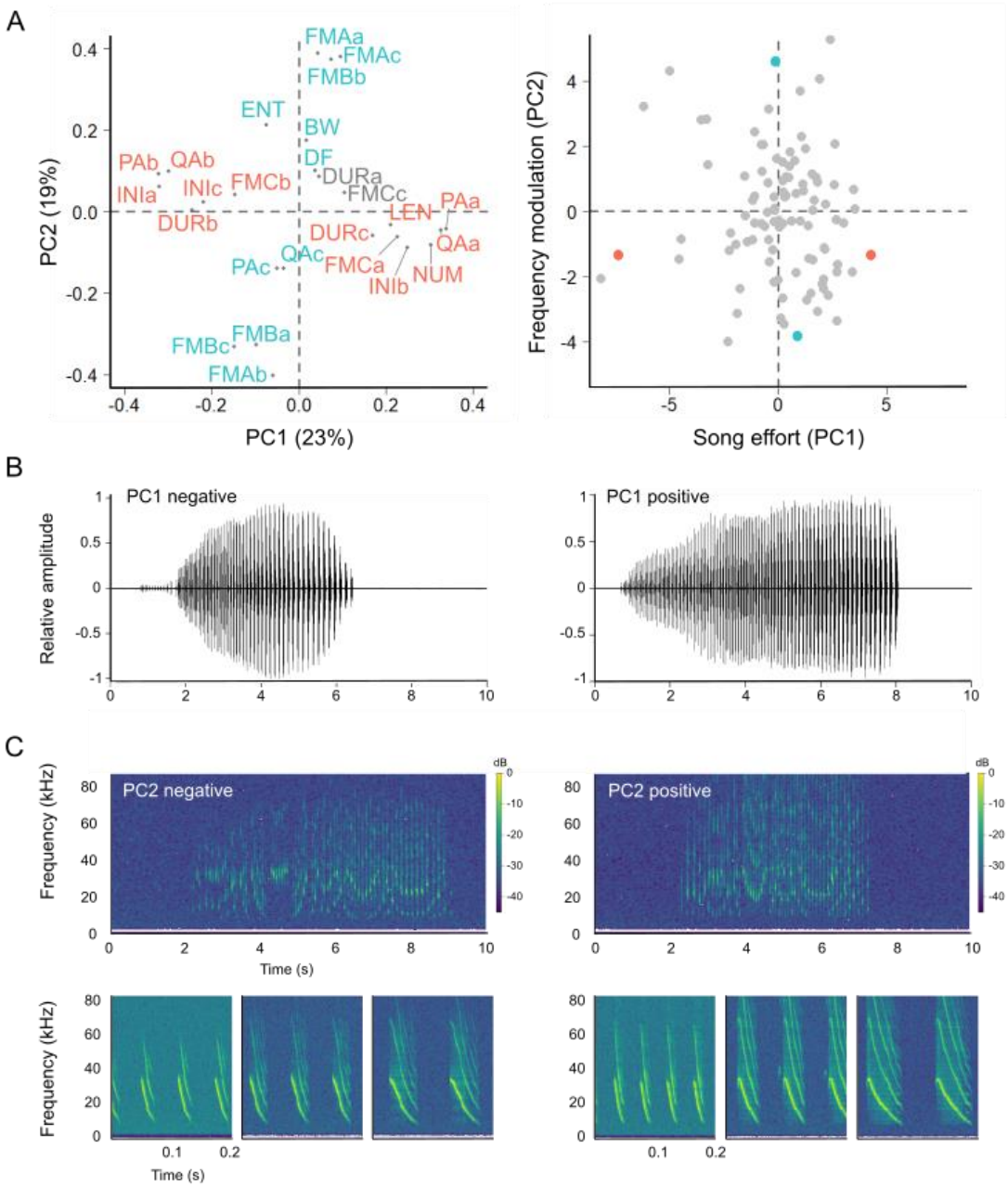

Figure 2: Acoustic variation in songs of wild singing mice. *(A)* Biplots from PCA of *Scotinomys* song. *Left,* biplot showing how variables loaded in PC space. In red, parameters loading strongly on PC1 ("song effort"). In blue, parameters loading strongly on PC2 ("spectral characteristics"). *Right,* individual singers plotted in PC space. Points colored red and blue are individuals that produced the songs plotted in B and C, respectively. *(B)* Variation in song effort. *Left,* song loading strongly and negatively on PC1. *Right,* song loading strongly and positively on PC1. *(C)* Variation in spectral characteristics. Top panels depict full spectrograms of songs with lower panels depicting 0.1 ms of song taken at 1 s, 3 s, and 5 s. *Left,* song loading strongly and negatively on PC2. *Right*, song loading strongly and positively on PC2.

We regressed body mass (RBM) as a common measure of body condition in 165 mice and measured anogenital distance (AGD) in 159 mice. We took plasma samples from each of the 165 mice, but only male plasma was processed; of these, 78 samples had enough plasma to complete at least one assay. All plasma measures fell within the range of values reported for laboratory mice except adiponectin, an adiposity hormone involved in glucose and fatty acid metabolism (*Mus*: 3000–15 000 ng ml^{-1}, Friedman and Halaas, 1998; *Scotinomys*: 50–350 ng

ml^{-1}, Burkhard et al., 2018). Fifty-two samples met the requirements to be included in PCA of body condition (see Methods), which revealed three latent variables. PC1 ("PC1 condition", 34.5% of total variation) described individual variation in glucose, fatty acids, and insulin, which we interpreted as metabolic responses to recent feeding. PC2 ("PC2 condition", 22.2%) described variation in residual body mass (RBM) and adiponectin, an adiposity hormone that modulates glucose and lipid metabolism; while PC3 ("PC3 condition", 17.2%) described variation in leptin, an adiposity hormone that helps regulate body weight, and adiponectin (Table S2). We interpreted these latter two composite variables as putatively stable indicators of long-term body composition and regulation of energy stores.

Table 2: Sample sizes and summary statistics for each song and condition phenotype recorded from 165 singing mice.

Phenotype	Units	N_{Total}	N (♂)	N (♀)	Mean ± SD	Range
Song rate	songs/2h	129	97	32	2.1 ± 2.5	0 – 14
					2.2 ± 2.3 (♂)	0 – 9 (♂)
					1.8 ± 2.8 (♀)	0 – 14 (♀)
Song length	s	106	84	22	6.3 ± 1.5	4.1 – 8.8
					6.7 ± 1.1 (♂)	4.1 – 8.8 (♂)
					6.1 ± 0.9 (♀)	4.4 – 7.7 (♀)
Trill rate	notes/s	106	84	22	6.7 ± 1.1	8.5 – 15.9
					11.8 ± 1.5 (♂)	8.5 – 15.9 (♂)
					12.1 ± 1.1 (♀)	9.1 – 13.6 (♀)
Note number	notes	106	84	22	76.5 ± 12.8	43.5 – 104
					77.5 ± 12.8 (♂)	43.5 – 104 (♂)
					73.0 ± 12.6 (♀)	55 – 104 (♀)
Dominant frequency	kHz	106	84	22	24670.8 ± 3782.8	18052 – 35272
					24596.1 ± 3856.7 (♂)	18052 – 35272 (♂)
					24942.5 ± 3573.2 (♀)	20079 – 32158 (♀)
Mean bandwidth	kHz	106	84	22	25064.6 ± 7799.7	3306 – 45949
					24439.3 ± 8209.9 (♂)	3306 – 45949 (♂)
					27338.3 ± 5667.5 (♀)	19599 – 44188 (♀)
Minimum frequency	kHz	106	84	22	13660.62 ± 13856.53	15872 – 92570
					14935.4 ± 15396.04 (♂)	15872 – 92570 (♂)
					9025.06 ± 1740 (♀)	36621 – 89264 (♀)
Maximum frequency	kHz	106	84	22	55035.0 ± 18821.1	15872 – 92570
					54309.9 ± 19123.3 (♂)	15872 – 92570 (♂)
					57671.6 ± 17851.8 (♀)	36621 – 89264 (♀)
Entropy	nat	106	84	22	-4.7 ± 0.4	-5.6 – -4.0

					-4.6 ± 0.35 (♂)	-5.6 – -4.0 (♂)
					-4.8 ± 0.4 (♀)	-5.3 – -4.0 (♀)
ddRAD	NA	157	128	29	NA	
PC condition	NA	52	52	0		
Mass	g	165	133	32	12.6 ± 1.5	8 – 16.4
Residual body mass	g	165	133	32	0 ± 1	-3.1 – 2.5
Hindfoot	mm	165	133	32	16.4 ± 0.6	15 – 18
Anogenital distance	mm	159	127	32	5.5 ± 1.3	2 – 8
					2.2 ± 2.3 (♂)	0 – 9 (♂)
					1.8 ± 2.8 (♀)	0 – 14 (♀)
Glucose	ml dl^{-1}	78	78	0	111.9 ± 66.6	14.71 – 378.5
Triglycerides	ml dl^{-1}	74	74	0	201.7 ± 119.5	52.2 – 718.3
Adiponectin	ng ml^{-1}	63	63	0	132.8 ± 52.2	53.4 – 324.2
Insulin	pg ml^{-1}	56	56	0	585.7 ± 502.7	69.0 – 2814.1
Leptin	pg ml^{-1}	56	56	0	2162.0 ± 1332.9	9.0 – 6711.0
Cholesterol	ml dl^{-1}	32	32	0	219.8 ± 64.6	18.5 – 340.7
Phospholipids	ml dl^{-1}	32	32	0	347.4 ± 106.4	41.9 – 526.4
Non-esterified fatty acids	ml dl^{-1}	30	30	0	0.7 ± 0.2	0.3 – 1.3
PC song + PC condition	NA	41	41	0	NA	
PC song + ddRAD	NA	101	81	20	NA	
PC condition+ ddRAD	NA	49	49	0	NA	
PC song + PC condition + ddRAD	NA	39	39	0	NA	

Variant calling

We used ddRAD-sequencing (Peterson et al., 2012) to genotype 162 mice. Five samples had poor coverage and were excluded from downstream analyses. We demultiplexed and trimmed reads using the *process_radtags* command from the Stacks pipeline (v1.46.0; Rochette and Catchen, 2017), then mapped trimmed reads to the *S. teguina* reference genome (annotated, unpublished) using Bowtie 2 (v2.3.2; Langmead and Salzberg, 2012). After pruning the remaining 157 genotypes for minor allele frequency (MAF) > 0.1 and for sex-linked markers using VCFTOOLS (v0.1.15; Danecek et al., 2011) and ANGSD (v0.929-13; Korneliussen et al., 2014), we retained 31,003 SNPs.

Table 3: Trapping site statistics. Geographic distance and elevational change calculated from GPS waypoints.

Site 1	Site 2	Geographic distance (m)	Elevation change (m)	FST
PH	Robles	1610	359	0.08
QERC	PH	538	6	0.05
QERC	Robles	1180	353	0.04
QERC	MH	3190	294	0.05
QERC	PC	4100	350	0.07
MH	PH	3650	300	0.07
MH	Robles	3220	59	0.08
MH	PC	870	56	0.06
PC	PH	4490	356	0.13
PC	Robles	4080	3	0.12

Population structure and individual relatedness

Because population structure can influence estimates of heritability, we first examined population genetic structure (Fig. 3). Permutational multivariate analysis of variance (PERMANOVA) revealed low but statistically significant genetic differentiation between trapping sites (R^2 = 0.027, P = 0.001), and principal coordinate analysis of the genotypic data revealed distinct but overlapping genetic clusters explained by trapping site (Fig. 3b). Consistent with this interpretation, we estimated generally low F_{ST} between sites (range: 5.7E-06 – 0.13) and found F_{ST} increased with geographic, though not elevational, distance (two-sided Mantel test; geographic distance: r = 0.68, P = 0.02; elevational distance: r = -0.05, P = 0.91; Table 3). We next estimated relatedness of individuals within our sample. Most pairwise genetic relationships were the equivalent of unrelated individuals, but we also uncovered stronger genetic relationships equivalent to half-sibling and full-sibling relatedness (Fig. 3c).

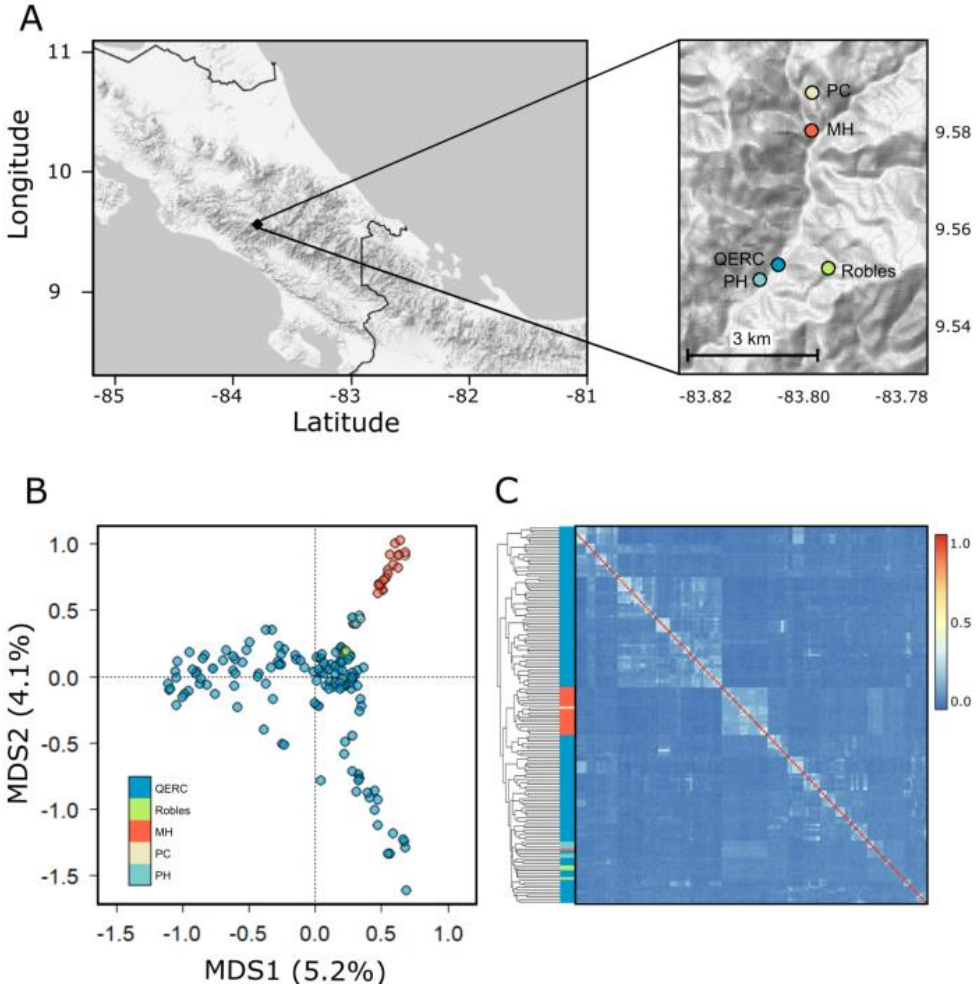

Figure 3: Geographic structure of relatedness in a wild population of singing mice. *(A)* Map of Costa Rica with zoomed-in panel of trapping sites within the valley of San Gerardo de Dota. *(B-C)* Colored by site from *A*. *(B)* Biplot of first two PCs from PCoA of genetic distance between individual mice. *(C)* Heatmap of genetic relatedness matrix (GRM) for 157 mice. Hierarchical cluster branches annotated by trapping site as in *A*. See Table 1 for elevation data and sample sizes.

Repeatability of acoustic variation

We estimated repeatability of acoustic characteristics by calculating intraclass correlation coefficients (ICC). Repeatability ranged from negligible (ICC < 0.2) to high (ICC ≥ 0.5; Table S3) and was generally greater for spectral song descriptors than for those related to song effort (spectral: mean *ICC* = 0.36, range = 0 – 0.76; effort: *ICC* = 0.21, 0 – 0.45). This pattern became more pronounced when considering only whole-song descriptors (*i.e.* song length, dominant frequency; Table S3), with whole-song spectral measures having greater mean repeatability than whole-song measures of effort (spectral: mean *ICC* = 0.32, range = 0.43 – 0.76; effort: *ICC* = 0.32, 0.30 – 0.33; Table S3, Fig. 4). The top three most repeatable aspects of song (*ICC* ≥ 0.6) were spectral characteristics: song mean bandwidth (*ICC* ± SE; BW: 0.76 ± 0.02), minimum frequency (minHz: 0.76 ± 0.03), and the starting frequency of the first note of a song (FMc_c:

Table 4: Heritability estimates for song and condition phenotypes. 95% HPDI = 95% highest posterior density credible interval.

					95% HDPI	
Type	Trait	N	V_G/V_P	SE	Lower	Upper
Condition	Hindfoot	154	0.20	0.13	0.00	0.52
	PC1 condition	49	0.07	0.07	0.00	0.26
	PC2 condition	49	0.34	0.24	0.00	0.94
	PC3 condition	49	0.45	0.25	0.00	1.00
	RBM	154	0.27	0.19	0.00	0.75
Development	AGD	148	0.25	0.17	0.00	0.65
Song	Entropy	101	0.48	0.25	0.00	0.97
	Max Hz	101	0.88	0.09	0.64	1.00
	Mean BW	101	0.90	0.10	0.62	1.00
	Min Hz	101	0.87	0.10	0.60	1.00
	PC2 spectral chars	97	0.61	0.25	0.02	0.99
	Song DF	101	0.55	0.25	0.02	1.00
	Fma_a	102	0.32	0.15	0.00	0.60
	Fma_b	102	0.48	0.22	0.00	0.84
	Fma_c	102	0.61	0.25	0.02	1.00
	Fmb_a	102	0.09	0.06	0.00	0.23
	Fmb_b	102	0.12	0.07	0.00	0.29
	Fmb_c	102	0.42	0.21	0.00	0.81
	Note num	101	0.03	0.03	0.00	0.10
	PC1 song effort	97	0.06	0.05	0.00	0.21
	Song length	101	0.02	0.01	0.00	0.06
	Song rate	119	0.59	0.25	0.00	0.98
	Trill rate	101	0.14	0.12	0.00	0.45

0.61 ± 0.04). Entropy, a whole-song description of tonality (0.59 ± 0.04), maximum frequency (maxHz: 0.56 ± 0.05), and FMb_c, starting note slope (0.52 ± 0.04) were the other measures with ICC ≥ 0.5. Twelve parameters, including dominant frequency and song length, were modestly repeatable (0.5 > *ICC* ≥ 0.2), while the remaining 13 parameters had negligible repeatability.

Heritability of song and condition

We estimated heritability by fitting animal models using maximum likelihood (ML) estimation implemented using standalone software *GCTA* (Yang et al., 2011) and using Bayesian Markov

chain Monte Carlo methods implemented with R-package *brms* (Bürkner, 2018, 2017). Heritability estimates derived from *GCTA* and *brms* were generally consistent for the same phenotypes with a notable exception, PC1 condition (*GCTA*: h^2 = 0.96 ± 0.59, *brms*: h^2 = 0.07 ± 0.07; Table 4, Table S4). Such disagreements result from how the two approaches deal with sparse data—for example, due to low levels of true heritability or small sample size. In these cases, ML estimation tends to be extremely variable and may report values that are either very large or zero (*e.g.* Beerli, 2006), while Bayesian approaches, given appropriate priors, offer more reliable results, as demonstrated by analyses of both empirical and simulated data (Beerli, 2006; Charmantier et al., 2011; Van de Schoot et al., 2015; Villemereuil et al., 2013). Credible intervals from Bayesian analysis are easily interpretable and provide distributional information indicating the most probable value and associated certainty (de Villemereuil, 2019; Kruschke and Liddell, 2018). For these reasons, we report here only heritability estimates (± SE) and 95% highest posterior density credible intervals (HPDI) from Bayesian models. For comparison, GCTA results are reported in Supplemental Files (Table S4).

Patterns of heritable variation in songs

Heritability estimates for whole-song measures related to song effort (h^2 range: 0.02 – 0.59, mean: 0.19) were generally lower than estimates for whole-song spectral measures (h^2 range: 0.48 – 0.90, mean: 0.57; Fig. 4, Table 4). PC composite variables followed this pattern, with PC1 having negligible heritability (h^2 = 0.06 ± 0.05) and PC2 having high heritability (h^2 = 0.61 ± 0.25). Mean song bandwidth, maximum song frequency, and minimum song frequency were the most heritable acoustic traits, with $h^2 \geq 0.8$, while song length and note number were the least heritable traits, with $h^2 \leq 0.05$. One notable exception to the general pattern was song rate, the number of songs produced in two hours, which had high heritability (h^2 = 0.59 ± 0.25).

We also examined the repeatability and heritability of several measures describing note shape (*i.e.* note curvature, FMa_X; note slope, FMb_x) and how shape changes over the course of a

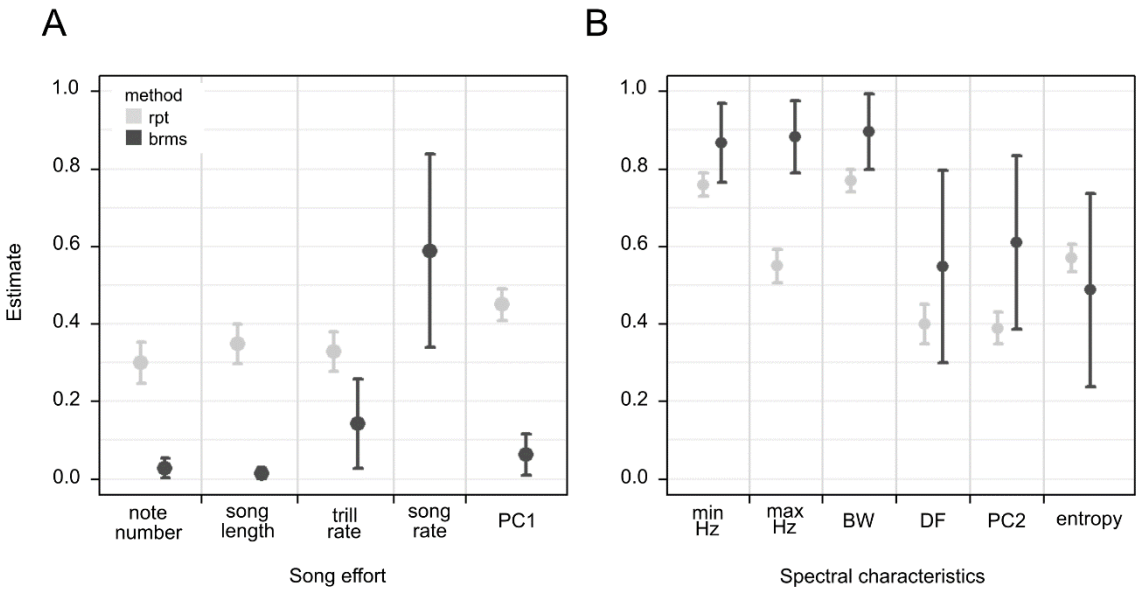

Figure 4: Heritable and repeatable variation in songs. *(A)* Repeatability (rpt) and heritability (brms) estimates from different models for song effort characteristics. *(B)* Estimates for spectral characteristics. Error bars indicate standard errors.

song (Fig 5); for example, FMa_A describes how rapidly note-to-note differences in note curvature happen over course of song, FMa_B describes the change in note curvature between first and second notes, and FMa_C describes the note curvature in first note in song (Campbell et al., 2010). We estimated intermediate to high repeatability and heritability for FMx_C terms; by contrast, we found low to intermediate repeatability and heritability for FMx_B terms and negligible to low repeatability and heritability for FMx_A terms (Tables 4, S3).

Patterns of heritable variation in body condition and morphology

Condition and morphological measures ranged in heritability from low to intermediate heritability (h^2 range: 0.07 – 0.45, Fig. 6a, Table 4). Anogenital distance (AGD), a morphological indicator of developmental exposure to androgens, had moderate heritability (h^2 = 0.25 ± 0.17). Hindfoot length and RBM also had moderate heritability (hindfoot: h^2 = 0.20 ± 0.13; RBM: h^2 = 0.27 ± 0.19). Measures of long-term body composition (*i.e.* PC2 and PC3 condition) were more heritable (PC2 condition: h^2 = 0.34 ± 0.24; PC3 condition: h^2 = 0.45 ± 0.25). Finally, short-term fluctuations in nutrients (PC1 condition) had low heritability (h^2 = 0.07 ± 0.07).

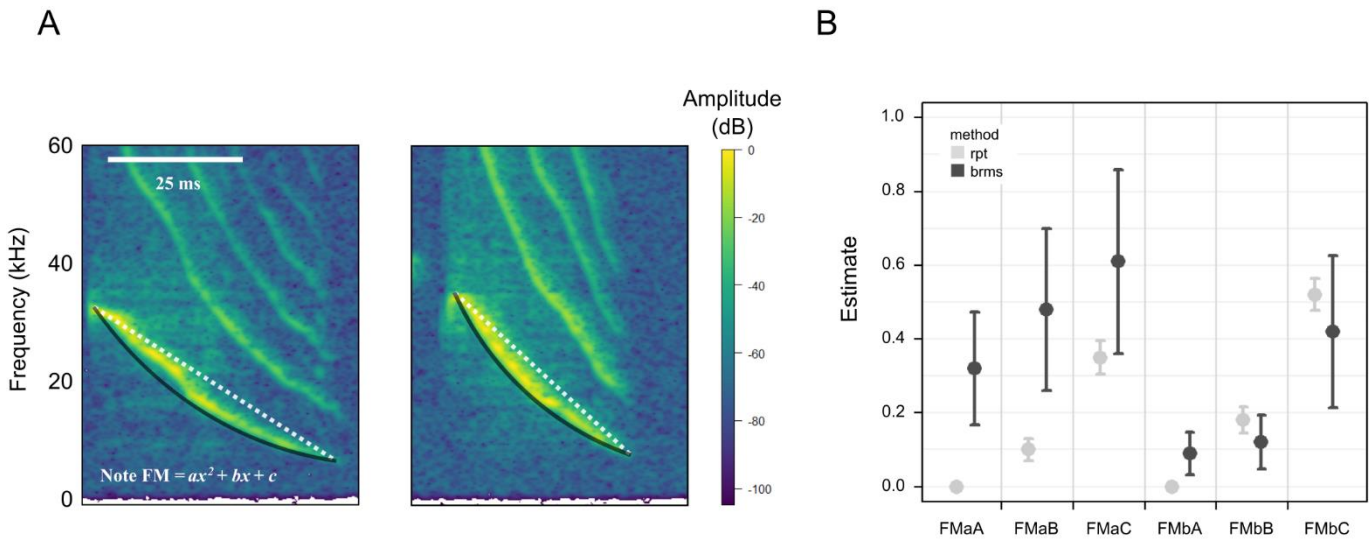

Figure 5: Heritability of note shape. *(A)* Close-ups of a single note over 50 ms. *Left*, note with low FMb and FMa (shallow note slope and curvature). *Right*, note with high FMb and FMa (steep note slope and curvature). Black lines trace note curvature; dotted white lines indicate note slope. *(B)* Repeatability (rpt) and heritability (brms) estimates from different models for song effort characteristics. Error bars indicate standard errors.

Correlations between body condition and singing behavior

Phenotypic correlations of condition and song

We next examined phenotypic correlations of condition and song by fitting mixed models. Forty-one male mice had both song and PC condition data and could be included in linear or generalized linear mixed models for condition-dependence (Tables 2, S5). Acoustic characteristics related to song effort were predicted by both body size and adiposity measures of condition: males with high PC2 condition scores sang longer songs with more notes than individuals with low PC2 condition scores (song length: β_{PC2con} = 0.32, *SE* = 0.10, *P* = 0.002; note number: β_{PC2con} = 0.32, *SE* = 0.10, *P* = 0.001, Fig. 6b). Similarly, males with high PC3 condition scores (high leptin, low adiponectin) produced more songs in the span of two hours than males with low PC3 condition scores (song rate: β_{PC3con} = 0.21, *SE* = 0.10, *P* = 0.04, Fig. 6c). Interestingly, we did not find a statistically clear effect of body size or adiposity measures on PC1 song effort, but we did find a clear relationship between song effort and AGD, an indicator of phenotypic masculinity (song effort: β_{agd} = 0.76, *SE* = 0.36, *P* = 0.04).

While increases in PC3 condition scores also predicted modest increases in song maximum frequency (max Hz: β_{PC3con} = 0.07, *SE* = 0.13, *P* = 0.007) and modest decreases in song minimum frequency (min Hz: β_{PC3con} = -0.04, *SE* = 0.01, *P* < 0.001), other spectral characteristics were not strongly predicted by any other measure of body condition (Table S5). Finally, PC1 condition, variation in immediate metabolic responses to recent feeding, did not predict any measure of song (Table S5).

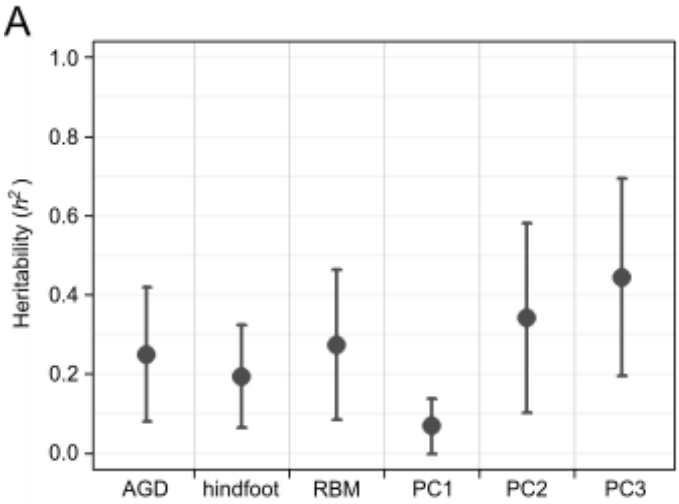

Figure 6: Heritability of condition. Error bars indicate standard errors.

Co-heritability of condition and song

We estimated genetic correlations for combinations of song effort characteristics (song length, note number, song rate) and condition measures (PC2, PC3) that were substantially phenotypically correlated (Table S5). We also investigated genetic correlations between song effort characteristics and RBM; this comparison allowed us to examine the utility of the traditional body condition metric and afforded us greater statistical power. Correlations ranged from -0.19 to 0.61 and were generally larger in magnitude for traits that were more strongly phenotypically correlated (Tables 5, S5). Total songs and PC3 of condition had the largest genetic correlation (r = 0.61 ± .29); this was unsurprising given the intermediate to high heritability estimates for each trait. In all cases, however, the 95% HPDI included the region of practical equivalence (ROPE) around 0 (Kruschke and Liddell, 2018); thus, we did not find clear statistical evidence for significant genetic correlations in our study.

Table 5: Genetic correlations (r_G) between condition and song. 95% HPDI = highest posterior density credible intervals.

Trait 1	Trait 2	N	r_G	SE	95% HPDI	
					lower	upper
call length	RBM	80	0.14	0.15	-0.17	0.43
note number	RBM	80	-0.01	0.34	-0.65	0.66
total songs	RBM	118	0.19	0.35	-0.7	0.87
call length	PC2 condition	31	0.35	0.2	-0.07	0.69
note number	PC2 condition	31	0.02	0.34	-0.63	0.64
total songs	PC3 condition	49	0.61	0.29	-0.12	0.98

Discussion

In this study, we used genome-wide data to investigate heritable variation in a variety of phenotypes related to advertisement display and body condition in a Costa Rican population of Alston's singing mouse (*Scotinomys teguina*). We obtained 31,000 SNPs from ddRAD-seq that we then used to estimate genome-wide relatedness, finding genetic relationships ranging from unrelated (rel ≅ 0) to the equivalent of full siblings (rel ≅ 0.5). We next calculated SNP-based heritability for various characteristics of singing behavior and measures of body condition, confirming results with both frequentist and Bayesian modeling. We found that measures of repeatability, a common surrogate for heritability (Boake, 1989), accurately predicted patterns of heritability: characteristics related to song effort, which were modestly repeatable, had low to moderate heritability, whereas spectral characteristics were often highly repeatable and heritable. We estimated intermediate heritability for patterns of body adiposity, while the heritability of insulin and circulating nutrients was negligible. Adiposity and song effort exhibited a strong phenotypic correlation, which was accompanied by a surprisingly high estimate of their co-heritability (r = 0.61). We now explore these findings in more depth.

Heritability and repeatability of singing behavior

Scotinomys songs vary in a wide range of acoustic attributes, and a principal components analysis of acoustic features suggested two major dimensions of song variation: those related to duration and amplitude, which we consider “song effort”; and those describing spectral characteristics, specifically, the frequency modulation that occurs over the course of a song (Fig 2, 4). Although heritability varied greatly from feature to feature, ranging from values near 0.0 to above 0.8, there were systematic patterns to the heritability of songs and their attributes (Table 5).

Among our various measures, whole-song spectral variation (dominant, maximum, and minimum frequency; mean bandwidth, entropy) was both highly repeatable and heritable. We hypothesize that this heritability reflects anatomical variation or low-level motor control. For example, dominant frequency (h^2 = 0.55 ± 0.25; *ICC* = 0.43 ± 0.05) is a property of both the vocal tract and adjustments to its length or shape (Fitch, 2006; *e.g.* via lip contractions in humans, Story et al., 2001; tongue placement or bill gape in songbirds, Langin et al., 2017, Nelson et al., 2005; laryngeal descent in humans and deer, Fitch and Reby, 2001). Maximum frequency (h^2 = 0.88 ± 0.1; *ICC* = 0.56) is likely a function of the cricothyroid, a muscle controlling the length and tension of vocal folds (Jürgens, 2002; Riede and Brown, 2013). In muroid mice that produce laryngeal whistles, variation in structures like the vocal folds and the ventral pouch are proposed to shape interspecific variation in dominant frequency (Riede et al., 2017; Riede and Pasch, 2020; Smith et al., 2020). At a phylogenetic scale, spectral properties are often constrained by body size, a relationship generally mediated by scaling of vocal structures (Gillooly and Ophir, 2010; Ophir et al., 2010; Ryan and Brenowitz, 1985). The relationship between body size and frequency is often observed within species as well, such as in red deer (Reby and McComb, 2003), anurans (Gerhardt, 1982), songbirds (Ryan and Brenowitz, 1985), loons (Mager et al., 2007), and humans (Pisanski et al., 2014). It is by no

means universal, however (*e.g.* Martin et al., 2017), and we did not observe a relationship in our sample.

While whole-song spectral characteristics were highly heritable, we also found substantial heritability in frequency modulation. Six related variables strongly loaded on PC2; together, these variables describe the frequency modulation that characterizes a note and how it changes over the course of a song (Figs. 2, 5). To appreciate how these are related, imagine a single note composed of a downward frequency sweep that begins and ends at characteristic frequencies defined by the performance of the vocal tract. A given note may proceed linearly from the beginning frequency to the end frequency with a relatively shallow slope, or may change rapidly in frequency initially, and then flatten out (Fig. 5a). These two parameters are the starting slope and curvature of a note (*i.e.* FMb, and FMa, the linear and quadratic terms in the equation: $Frequency = FMc + FMb \times time + FMa \times time^2$; Figs. 1, 5). We find that the slope and curvature of notes are strongly correlated with one another, and change systematically over the course of the song. Starting note shape (FMb_C, FMa_C) exhibits high repeatability and heritability (Tables 4, S3). Unlike many other spectral features of the song, note shape reflects the temporal dynamics of muscle activity over the course of a note. Such timing must reside in central pattern generators (CPGs), such as those vocal, orofacial and respiratory CPGs thought to reside in the brainstem (Zheng, 2020). In singing mice, laryngeal and jaw muscles important for vocalization seem to be coordinated by a putative CPG within the reticular formation (Zheng et al., *manuscript*).

The fact that note shapes vary by individual and are highly heritable is also consistent with prior suggestions that they contribute to identity signaling in this species (Burkhard et al., 2018). Other studies have found that ecological contexts can favor individually distinctive social signals (Burt and Beecher, 2008; Sheehan and Tibbetts, 2010; Stoddard and Beecher, 1983). If this were true, it would predict that not only could animals use note shape and other spectral features to recognize individuals, but that balancing selection may actively promote diversity of

neural activity patterns within CPGs. These results highlight a need to examine whether note shapes are in fact used in individual recognition.

In contrast to spectral features, we found low heritability for most song-effort traits. Our PC1 of song variation, which we labelled song effort, was low both in its repeatability and heritability (ICC = 0.21, h^2 = 0.06). Measures like song length, note number, and trill rate (*ICC* < 0.35) all had negligible heritability (h^2 < 0.1). One explanation for this pattern is that highly plastic behaviors, which are shaped by social context and experience, may be less heritable than traits constrained by morphology (Boake, 1989; Mousseau and Roff, 1987). This phenomenon has been documented in other natural populations (*e.g.* three-spined stickleback: Dingemanse et al., 2009; island-adapted deer mice: Baier et al., 2019) and confirmed by meta-analyses of behavioral and morphological datasets (Mousseau and Roff, 1987; Stirling et al., 2002; but see Jensen et al., 2003; Dochtermann et al., 2019). This interpretation is consistent with evidence from singing mice suggesting that song effort is highly sensitive to social context and motivation. Male singing mice increase song rate in response to winning fights (George, 2014), encounters with females (Fernández-Vargas et al., 2011; Hooper and Carleton, 1976), and simply hearing another animal vocalize (Hooper and Carleton, 1976; Pasch et al., 2013). Singing mice also exhibit dynamic “turn-taking” counter-singing behavior, rapidly modifying the starts and stops of songs during social interactions (Okobi et al., 2019). Similar relationships between vocal effort and social experience are well documented in other taxa, particularly in avian (Catchpole and Slater, 2008; Podos and Cohn-Haft, 2019) and anuran species (Bernal et al., 2009; Gerhardt, 1991; Ryan, 1985), in wild muroid rodents (Fernández-Vargas, 2018 a; Fernández-Vargas, 2018 b), and in other mammals (Briefer, 2012; Gouzoules and Gouzoules, 2000), including humans (Bachorowski, 1999). Male music frogs dynamically adjust calling efforts based on perceived sexual attractiveness of rivals (Fang et al., 2014), comparable to male responses to competition in gray treefrogs (Gerhardt, 1991) and túngara frogs (Bernal et al., 2009; Ryan, 1985). Likewise, pair-bonded California mice increase their use of aggressive “bark”

vocalizations towards recently unfaithful partners (Pultorak et al., 2018), a behavioral response with uncanny parallels to non-verbal human expressions of anger (Sauter et al., 2010).

Taken together, our data suggest that as we move from the peripheral vocal organs, through low-level motor control, and finally to the higher-level neural centers that govern motivational state and dynamic social interactions, aspects of song become progressively less heritable. This seems to be a major finding of our analysis, but there are two anomalous aspects of our data that merit some further discussion. The first is that stable differences in song duration, trill rate, and related characteristics that seem negligibly heritable in our sample are nevertheless distinct among populations of singing mice (Campbell et al., 2010; Hooper and Carleton, 1976); these differences persist in laboratory-reared mice (Campbell et al., 2014), suggesting a heritable basis for lineage differences. One possibility is that meaningful heritability is present in these measures, but our power is insufficient to detect it. A second possibility is that low heritability in contemporary populations may be a by-product of past selection. Local adaptation, for example, would preserve between-lineage heritability while reducing within-population genetic variance (Falconer, 1981; Mousseau and Roff, 1987). Future work will investigate the relationship between acoustic and genetic variation across populations of singing mice. The second anomaly in our findings is that song rate, which we interpret as a motivational aspect of song, is surprisingly heritable ($h^2 = 0.59 \pm 0.25$). We recently reported that manipulating perceived body condition by injecting leptin increases song rate, but not song duration. Perhaps song rate is particularly strongly influenced by trade-offs associated with body condition – a hypothesis that is supported by phenotypic correlations between the two dimensions (discussed below).

Heritability of condition

Body condition holds evolutionary significance for two major reasons: it is thought to be tightly linked to fitness, and it is integral to many models of signal evolution. As a result, its heritability is also of evolutionary interest. While fitness-related traits are generally expected to have low

heritability (Falconer, 1981; Mousseau and Roff, 1987), it is not immediately obvious whether condition should follow this pattern. First, certain indices of condition may not be as predictive of evolutionary fitness as others, or may predict different aspects of fitness; for instance, in damselflies, body weight is negatively correlated with parasite-mediated mortality but has no relationship with reproductive success (Rolff et al., 2002). Further, condition may be variably affected by contributions of environmental and genetic variation underlying its many physiological and morphological constituents. Lastly, if condition is highly polygenic due to its dependence on a large number of other aspects of phenotype, the high “mutational target size” of the trait could allow heritable variation to persist in the face of selection – an argument that has been used to suggest that elaborate displays can serve as indicators of underlying genetic viability (Rowe and Houle, 1996).

In the present study, we described body condition with measures of residual body mass and various circulating nutrients and metabolic hormones, using PCA to define major dimensions of variation. PC1 described variation in glucose, fatty acids, and insulin, which we interpreted as immediate metabolic responses to recent feeding or fasting. PC2 and PC3 each reflected patterns of adiposity: PC2 described variation in body size and adiponectin, an adiposity hormone involved in glucose and fat digestion, while PC3 described variation in adiponectin and leptin, an adiposity hormone regulating body weight. Thus, we interpreted both PC2 and PC3 of condition as putatively stable indicators of body composition.

We found low heritability for circulating nutrients and insulin (PC1 condition: $h^2 = 0.07$), a result that is consistent with these measures as dynamic responses to recent feeding. Relative to the moderate heritability of RBM, stable indicators of body composition (PC2 and PC3 condition) were higher in heritability (Table 4, Fig. 6) but were lower than expected. Evidence from humans and laboratory rodents suggests a high heritability of adiposity- and metabolism-related traits, including body fat distribution (Schleinitz et al., 2014), percent body fat (Wuschke et al., 2007), fat type (Ferrannini et al., 2016), and metabolic rate (Pettersen et al., 2018); indeed, twin studies

report higher heritability of human body fat (h^2 = 0.59-0.63, Schousboe et al., 2004; h^2 = 0.71-0.75, Lehtovirta et al., 2010) and body mass index (BMI; h^2 = 0.58-0.63, Schousboe et al., 2004; h^2 = 0.63-0.85, Allison et al., 1996) than values we report here. However, an important distinction between human or lab rodent work and this study is that the former significantly reduces environmental variance by design whereas our focal population has been subject to natural environmental conditions.

Condition dependence

Elaborate displays can be costly to express, and are thought to rely on body condition (Bradbury and Vehrencamp, 1998; Cotton et al., 2004; Johnstone, 1993). Condition dependence thus plays an important role in individual signaling decisions, determining the expression and intensity of labile signals, in addition to serving as a key mechanism for many models of signal evolution (Grafen, 1990; Kirkpatrick and Ryan, 1991; Pomiankowski, 1987; Zahavi, 1975). Condition-dependence could also lead to genetic covariance between loci affecting condition and display, a critical prediction of good-genes models of signal evolution (Kirkpatrick and Ryan, 1991; Rowe and Houle, 1996; Tomkins et al., 2004).

We tested whether major dimensions of song were phenotypically corelated with measures of condition. As expected, stable spectral characteristics, putatively constrained by variation in vocal structures and morphology, lacked strong correlations with measures of body condition. Conversely, we found significant relationships between elements of song effort, including song length, vocalization rate, and note number, and stable measures of body composition (PC2 and PC3 of condition). These findings are consistent with previously reported relationships between body condition and song in singing mice (Burkhard et al., 2018; Pasch et al., 2011b) and with general patterns described in other taxa (*e.g.* birds: Houtman, 1992, Lampe et al., 1994; crickets: Thomson et al., 2014). For example, Thomas and Cuthill (2002) demonstrated that European robins (*Erithacus rubecula*) song rate was positively correlated with body mass, such

that increases in body mass (*i.e.* increased fat reserves) predict more frequent singing behavior (Thomas and Cuthill, 2002). Likewise, greater body mass predicts more elaborate repertoires and better performance in Java sparrows (*Lonchura oryzibora*) (Kagawa and Soma, 2013). Possible explanations for this general phenomenon include the hypothesis that conspicuous song, along with other labile displays, may be energetically costly to produce (Barske et al., 2013; Ward et al., 2003) or may elicit the energetic costs of predator evasion (Zuk and Kolluru, 1998). Signaling decisions are thus likely modulated not only by social context but by self-assessment of body condition (Santori et al., 2020), perhaps through interoceptive cues, like adiponectin or leptin, that signal status of energetic reserves (Havel, 2001).

Both song rate and PC3 of condition (the adiposity hormones leptin and adiponectin) were remarkably heritable (Table 5), and our estimate for the genetic correlation between the two suggests that roughly 60% of the heritability in song rate could be predicted by this measure of condition (Table 6). This estimate is surprisingly large – it is comparable, for example, to published genetic correlations between closely related psychiatric conditions like schizophrenia and bipolar disorder (r_g = 0.64, Lee et al., 2013). An important caveat, however, is that the credible intervals overlap the region of practical equivalence around 0 (Kruschke and Liddell, 2018), indicating that we lack the power to precisely estimate the genetic correlation. Our findings indicate, however, that it would be worthwhile to generate a more precise estimate of this genetic correlation. Such studies should focus on hormonal indicators of condition, and should include a substantially larger number of animals. What constitutes a reasonable sampling effort will vary by study system, its demography, and its natural history (eg Bérénos et al., 2014; Villemereuil et al., 2013).

Finding a genetic correlation is necessary for the good genes hypothesis, but it is not sufficient. The idea that an expensive trait could be an index for heritable variation in condition is plausible, but we would expect genetic correlations to be present even if female choice plays no role in the evolution of effort—because males must make these decisions in order to balance risks and

rewards in general. Our approach, however, provides a way to generate meaningful measures of body condition in the wild, as well as meaningful measures of heritability and co-heritability. Such data will be essential to disentangling the role of direct benefits – such as males maximizing the outcomes of trade-offs, and females avoiding parasites – and indirect benefits, such as minimizing the genetic burden of deleterious mutations passed on to offspring.

Conclusions

For decades, the challenges of studying heritability in non-model and wild species has compelled many behavioral ecologists and field biologists to either strategically ignore the genetic mechanisms underlying traits (the “phenotypic gambit”; Grafen, 1984; Rittschof and Robinson, 2014) or focus instead on more tractable but less informative alternative measures, like repeatability (Boake, 1989). Despite the obvious logistical advantages of lab estimates of heritability, field estimates are more ecologically meaningful. Lab conditions can reduce genetic variation and environmental variation, which may lead to under- or overestimation of heritability (Orengo and Prevosti, 1999; Roff and Simons, 1997; Schneider et al., 2011; Simons and Roff, 1994; Stirling et al., 2002). Lab conditions can also erode ecologically interesting phenotypic variation (*e.g.* the relationship of male song and aggression in crickets; Hedrick and Bunting, 2014), rendering further examination of phenotypic and genetic correlations meaningless. Finally, heritability is not a stable characteristic of a trait—it is contingent on a population’s history of selection, gene flow, and environmental heterogeneity (Visscher et al., 2008). For these reasons, heritability is best assessed in wild populations (though see Dochtermann et al., 2019).

Today, advances in genomics and bioinformatics provide novel opportunities for the estimation of heritability and other quantitative genetics parameters in the field. Our study—in which we collected phenotypic data and DNA in the field, generated genome-wide relatedness estimates, and related these datasets to one another to estimate heritability and co-heritability of condition

and song—demonstrates the feasibility of implementing these methods to study both wild, non-model species and the complex behaviors and traits often neglected in traditional genetics analyses. Finally, our data highlight the potential for using such methods to address longstanding evolutionary problems. Harnessing the potential of these techniques and their applications will expand the role of genetics in evolutionary and ecological studies and allow us to shed the confines of the phenotypic gambit.

Materials and Methods

Sampling and data collection

We focused on a population of singing mice inhabiting the San Gerardo de Dota valley of Costa Rica, a high-elevation region in the Talamanca Mountain range (>2200 m) characterized by frequent rainfall and its lush cloud forests, grasslands, and paramo (Fig. 1a). Within this region, we sampled mice at five trapping sites over three years (June – August 2014, 2015, and 2016). Trapping sites were located within 6 km^2 of one another and occurred at elevations between 2200 m and 2600 m (Figure 3a, Table 1). Two of the sites comprised primary and secondary forests (PC and Robles) while the other two comprised grassy pastures (QERC, MH).

We captured a total of 168 adult mice and brought them to the Quetzal Education and Research Center (QERC) for processing (*N* = 32 females and 134 males). This number was male-biased due to the needs of concurrent projects. We recorded the hindfoot length, anogenital distance, and mass of each focal mouse. Mice were housed singly in 28 x 28 x 28 cm^3 PVC-coated wire mesh cages and provided with food and water *ad libitum*. Each mouse was given at least one day of acclimation before recording sessions.

Focal animals were moved in their home cage into a 42 x 42 x 39 cm^3 acoustic isolation chamber the night before recording. Recording sessions took place between 500h and 1100 h in the morning. We have previously described lab methods for obtaining and recording singing

mouse songs (Burkhard et al., 2018; Campbell et al., 2010); briefly, we recorded both songs produced in response to playback stimuli and songs spontaneously produced in two hours of silence. Vocalizations were recorded using ACO Pacific microphones on Tucker-Davis RX6 hardware at 32-bit, 195.3 kHz resolution. We visualized the spectrogram of each recording in MATLAB and excluded poor-quality recordings from analysis.

Between 1400 and 1700 h on the day of recording, we euthanized focal animals. Trunk blood was collected within two minutes of sacrifice and immediately centrifuged for plasma extraction. We then harvested fresh liver tissue and counted the number of nematode endoparasites present in the stomach and body cavity. Plasma and liver tissue were stored at -20ºC until analysis.

Song analysis

Acoustic measurement

We successfully obtained 467 vocalizations from 111 mice (*N* = 24 females, 87 males). Vocalizations were bimodally distributed, ranging in length from 0.3 s to 9.9 s long (mean ± s.d. = 6.3 ± 1.8 s). We operationally define long songs as those four seconds or longer and retained long songs from 108 mice (*N* = 24 females, 84 males, Table 2). We assessed acoustic properties of each song using custom script in MATLAB. We measured 21 parameters describing changes to spectral, temporal, and amplitude characteristics throughout a song and 5 parameters describing whole-song characteristics (mean bandwidth, song length, note number, entropy, and dominant frequency; Table S1) (Burkhard et al., 2018; Campbell et al., 2010). To these 26 parameters, we added three additional whole-song measures, trill rate (notes/s), and minimum and maximum frequency of song.

Multivariate analysis of acoustic variation

To characterize dimensions of acoustic variation, we performed principal components analysis (PCA) on 26 acoustic parameters (Table S1) in R version 3.5.3 (R Core Team, 2016). These

variables were selected to accommodate comparison of unpublished data to previously reported data (Burkhard et al., 2018). The first two principal components (PCs) had λ > 4.9 and explained 42.2% of overall variance (Figure 2b). To aid interpretation, we multiplied PC1 scores by -1 (Jolliffe, 2002). Songs scoring strongly and positively on PC1 were longer, had longer beginning notes, and had shorter inter-note intervals than those scoring strongly and negatively on PC1. Songs that loaded strongly and positively on PC2 had a lower dominant frequency, greater mean bandwidth, began at lower frequencies, and ended on higher frequencies than those scoring strongly and negatively on PC2. As these results resembled previously published work (Burkhard et al., 2018), we interpreted PC1 as individual variation in "song effort" and PC2 as variation in "frequency modulation" and retained these as composite acoustic variables for subsequent models (Figure 2c).

Repeatability

Because repeatability describes the proportion of total phenotypic variation due to differences between individuals, the metric is often used to estimate the upper limits to a trait's heritability (Bell et al., 2009; Boake, 1989, though see Dohm, 2002). We estimated the repeatability for each of the eight whole-song measures and for PC1 and PC2 of song. Repeatability is often estimated as the intraclass correlation coefficient (RA., 1954; Wolak et al., 2012), such that:

$$ICC = \frac{V_A}{V_A + V_W} \qquad \text{(equation 1)}$$

where V_A is the variance among individuals, V_W is the variance within individuals, and total phenotypic variation $V_P = V_A + V_W$. Intraclass correlation coefficients were calculated by fitting mixed-effect models using the *rptR* R-package (v0.9.22, Stoffel et al., 2017) with animal ID as a random effect. Confidence intervals were estimated with parametric bootstrapping for 1000 iterations, and p-values were calculated with likelihood ratio tests.

Assessment of body condition

Morphometric evaluation

Residual body mass (RBM) was determined by regressing mass on skeletal length (*i.e.* hindfoot length) and calculating the residual error in grams for each animal. RBM, a traditional proxy for total energy reserves, was included in subsequent analyses of body condition.

We also measured anogenital distance (AGD), a morphological cue of prenatal exposure to androgens. We performed Z-score normalization on AGD (AGD_z); because AGD was sexually dimorphic (Welch two sample t-test; T= -12.5, P < 2.2e-16), it was transformed for each sex separately. Thus, female AGD_z > 1.5 corresponds to females with AGD ≥ 5.0 mm whereas male AGD_z > 1.5 corresponds to males with AGD ≥ 7.5 mm. Because variation in AGD predicts individual differences in "masculinity" but not body condition itself, we considered AGD as a covariate separate from body condition metrics in subsequent analyses.

Plasma assays

Plasma samples collected in 2015 were assayed by the Mouse Metabolic Phenotyping Center at the University of Cincinnati, Ohio. Those collected in 2016 were assayed by the Mouse Metabolic and Phenotypic Core at the Baylor College of Medicine. Only male samples were processed. We selected assays for cholesterol, phospholipids, non-esterified fatty acids (NEFA), glucose, and triglycerides (TG) and the hormones adiponectin, insulin, and leptin. Not every male had enough plasma for all assays, so assays were assigned an order of priority. Assays for 2015 samples were completed in the following order: 1) glucose, 2) lipid profile (TG, cholesterol, phospholipids, NEFA), and 3) multiplex ELISA assay (insulin, leptin, adiponectin). As results from these first analyses suggested a strong relationship between adiposity hormones and singing behavior (Burkhard et al., 2018), we modified the order for 2016 assays to the following: 1) multiplex ELISA, 2) glucose, and 3) lipid profile. As a result, sample sizes

vary across plasma measures (Table 2). Values for plasma measures did not significantly differ between years and were thus pooled for further analysis.

Multivariate analysis of variation in body condition

We performed exploratory PCA and assessed a correlation matrix of all body condition parameters (*i.e.* RBM and plasma measures) to examine how body condition varied among individuals. To maximize our statistical power, we ultimately focused on the subset of individuals (N = 52) with each of the following parameters: RBM, glucose, leptin, insulin, adiponectin, and triglycerides. We performed PCA on these condition measures in R to create composite variables for downstream analyses. Three PCs had $\lambda > 1.0$, explaining 34.5%, 22.2%, and 17.2% of total variance, respectively. Glucose, insulin, and triglycerides strongly loaded on PC1; RBM and adiponectin contributed most to PC2; and leptin and adiponectin contributed strongly and in different directions to PC3 (Table S2). We retained these three components as composite variables in downstream analyses (*i.e.* PC1con, PC2con, PC3con).

Library preparation and variant calling

We extracted DNA from thawed liver tissue using Qiagen DNeasy Blood and Tissue kits. DNA was first quality-checked by gel electrophoresis and then quantified with a PicoGreen fluorometer quantitator and Quant-iT™ PicoGreen™ dsDNA Assay Kit (Invitrogen™, ThermoFisher Scientific). The Genomic Sequencing and Analysis Facility (GSAF; University of Texas, Austin) performed double-digest restriction site associated DNA sequencing (ddRAD-seq; Peterson et al., 2012) library preparation on normalized DNA concentrations. The restriction enzymes *EcoR1* and *MspI* (New England BioLabs) were selected to shear DNA with Pippen size selection for fragments of 335-435 bp. Each sample was tagged with a unique five-nucleotide-long barcode and then pooled in groups of 24 for sequencing. Samples included biological replicates to check for procedural error. Multiplexed samples were then sequenced on

one lane of either the HiSeq 2500 or NextSeq 500 (Illumina, Inc.), depending on availability. Lanes also included samples for studies not described here.

After inspecting the quality of raw reads with FastQC (Andrews, 2010), we demultiplexed and trimmed reads to 125 bps using the *process_radtags* command from the Stacks pipeline (v1.46.0; Rochette and Catchen, 2017). Trimmed reads were mapped to the *S. teguina* reference genome (unpublished) using Bowtie 2 (v2.3.2; Langmead and Salzberg, 2012), and resulting *sam* files were converted to *bam* files using SAMtools (v1.10; Li et al., 2009). We used ANGSD (v0.929-13; Korneliussen et al., 2014) to assess base qualities and coverage depth of resulting bam files, excluding poor-quality and poor-coverage samples from downstream analyses.

Samples that passed these initial filtering steps were then further filtered with ANGSD to retain only loci with minor allele frequency (MAF) above 0.1 and which were present in at least 80% of all individuals. We detected sex-linked markers using VCFTOOLS (v0.1.15; Danecek et al., 2011) and then excluded them with the -rf option in ANGSD. Finally, we used ANGSD to call variants and convert genotype data into various formats, including identity-by-state matrices (ibsmat), mutation annotation format files (maf), variant calling files (vcf), and genetic covariance matrices (covmat).

Population structure and genomewide relatedness

Pairwise genetic distance between trapping locations (F_{ST}) was estimated using VCFTOOLS. We also used the R-package vegan (v.2.5.6 ; Oksanen et al., 2019) to perform principal coordinates analysis (PCoA) to help resolve genetic distance. Permutational multivariate analysis of genetic distance (PERMANOVA) explained by trapping site was performed with the *adonis* function in vegan.

There are many methods available to calculate genetic relatedness which produce a variety of data formats (Eu-ahsunthornwattana et al., 2014). Certain combinations of data formats with heritability estimation method are more compatible and straightforward than other combinations, however. For use with *GCTA* software (see Inference of SNP-based heritability), we produced a population genomic relationship matrix (GRM) in PLINK (v1.90b6.7; Chang et al., 2015), which uses a method-of-moments approach to estimate the proportion of the genome at which two individuals share 0, 1, or 2 alleles identical by descent (IBD; Purcell et al., 2007). PLINK operates on 'hard-call' genotypes and is thus best suited for high-coverage, high-quality reads. To check the accuracy of PLINK relatedness estimation on our data, we also estimated relatedness using NgsRelate software (v2; Korneliussen and Moltke, 2015), which uses a maximum likelihood approach on genotype likelihoods to estimate the proportion of alleles two individuals share IBD. Relatedness estimates from NgsRelate, a method based on genotype likelihoods, were strongly correlated with estimates from PLINK (R^2 = 0.90, P < 2.2e-16; Fig. S1). Finally, for use with *brms* (see Inference of SNP-based heritability), we produced a population genetic covariance matrix using ANGSD, which uses genotype likelihoods rather than hard-called genotypes.

Phenotypic correlations between condition and song

To ask how condition phenotypes related to singing behaviors, we fitted mixed models using the R-package *glmmTMB* (v1.0.2.1; Brooks et al., 2017). We focused on the eight whole-song measures, song rate, and the first two song PCs as response phenotypes. Whole-song measures were transformed by Z-score normalization before model fitting. Whole-song measures and song PCs were fitted assuming Gaussian distributions, while song rate was fitted assuming a negative binomial distribution with a logit link function. Each response variable was initially fitted with a full model, which included PC1, PC2, and PC3 of condition and AGD_z as covariates. Composite variables, rather than individual parameters, were included to reduce multicollinearity of parameters and to mitigate interpretation of results. Trapping site (Site) and

individual identity (1|ID) were also included as a fixed factor and a random effect, respectively. Because only male plasma samples were processed, we did not include sex in the models. Finally, we specified the number of days spent in acclimation (Days) as an observation-level dispersion factor to account for data overdispersion (Harrison, 2014). After fitting the full model, we used R-package *buildmer* to perform a backwards stepwise elimination procedure based on the change in AIC (v1.6; Voeten, 2020). We report the estimates, SEs, and p-values for all terms in the final models and the change in AIC between initial and final models. Model results were visualized using the *effects* package (v4.1.4; Fox and Hong, 2009; Fox and Weisberg, 2018), and the *emmeans* (v1.5.0; Lenth, 2020) and *boot* packages were used to help interpret results (v1.3.25; Canty and Ripley, 2020).

Quantitative genetics

Inference of SNP-based heritability

We estimated h2 values for the eight whole-song measures, PC1 and PC2 song, RBM, AGDz, and PC1, PC2, and PC3 condition using animal models implemented with both a maximum likelihood estimation (MLE) approach and a Bayesian Hamiltonian Monte Carlo approach. The MLE method was implemented with *GCTA* standalone software (v1.92.1; Yang et al., 2011), and the Bayesian approach was implemented using the *brms* R-package (v2.13.5; Bürkner, 2018, 2017). While easy to implement and fast to run, GCTA and many related frequentist algorithms cannot accommodate more complex models or non-parametric assumptions (*e.g.* non-normally distributed phenotypes; *e.g.* de Villemereuil, 2019). Conversely, linear (or generalized linear) mixed models (GLMM) implemented with Bayesian approaches are less user-friendly and more computationally intensive but are more robust to scarce data and complex models and are easy to interpret (de Villemereuil, 2019; Morrissey et al., 2014; Villemereuil et al., 2013).

Animal model approaches rely on estimating heritability by correlation between genetic similarity and phenotypic similarity (Wilson et al., 2010).The most basic animal model assumes a normally distributed phenotype, where a trait (*y*) in an individual (*i*) can be described as:

$$y_i = \mu + a_i + e_i \qquad \text{(equation 2)}$$

where μ is the population mean; a_i is the breeding value, or effect of individual *i*'s genotype on the trait; and e_i is a residual term. This equation is specified as a mixed model containing both fixed and random effects with individual genotype treated as a random effect such that:

$$Y = X\beta + Z\gamma + \varepsilon \qquad \text{(equation 3)}$$

where *Y* is a vector of phenotypic values, *X* is a fixed variable, *β* is the fixed variable effect size, *Z* is the GRM, γ is a vector of random effects, and ε is a residual random effect (non-genetic).

Narrow-sense heritability was calculated as

$$h^2 = \frac{V_A}{V_A + V_R} \qquad \text{(equation 4)}$$

, where V_A is additive genetic variance, V_R is residual variance, and V_P, is total phenotypic variance, = $V_A + V_R$.

GCTA models

Models were fitted with *GCTA* with sex (Sex) and trapping site (Site) included as fixed effects and PLINK-derived GRM as a random effect.

Brms models

Heritability estimates for each trait were estimated with models including ANGSD-derived genetic covariance matrix as a random effect, and Sex (excluding models involving only one sex, *i.e.* heritability of PC condition variables) and Site as fixed effects. Where applicable, response variables were log-transformed and scaled to address skew. We specified gaussian family distributions for all variables except song rate, which we fitted assuming a negative binomial distribution. Each model was run with 4 chains of 5000 iterations, with burn-in of 1000, and thin of 1 and provided with weakly informative priors. Model output was checked for proper mixing and convergence by inspection of autocorrelation and diagnostic plots. We took the mean of values from the posterior distribution to calculate heritability and report 95% highest posterior density credible intervals (HPDI) in Table 4. We calculated heritability following equation 4; residual variance for negative binomial models was calculating following (Matos et al., 1997). Credible intervals contain the most probable values (*i.e.* 95%, in this case) of the estimated parameter. Unlike a confidence interval, credible intervals have distributional information such that the most probable values are closest to the point estimate and the width indicates certainty.

Genetic correlation of song and condition

The proportion of covariance between two traits explained by genetics is calculated by the following equation:

$$r_g = \frac{cov_g}{\sqrt{V_{g1}V_{g2}}} \quad \text{(equation 5)}$$

Where cov_g is the genetic covariance, V_{g1} is the additive genetic variance for trait 1, and V_{g2} is the additive genetic variance for trait 2. To determine whether traits were genetically correlated, we ran bivariate models in *brms* between all pairwise comparisons of condition and song variables that were found to have statistically clear phenotypic correlations (see Phenotypic

correlations between condition and song). These models used the same data and parameters as the univariate models described above for *brms* models. Genetic correlations were considered significant in magnitude if the 95% highest posterior density credible intervals (HPDI) excluded zero.

Acknowledgments

We thank R. Westwick, K. Garner, W. Betty, and the QERC staff for their assistance in the field. We also thank J. Podnar for support with RAD-seq; B. Goetz and J. Sardell for advice on bioinformatics analysis; P. Vullioud, P. de Villemereuil, and M. Modrák for advice on statistical analysis; G. Otto, P. Saha, and P. Tso for guidance on bloodwork; and the Phelps lab and M. Ryan for comments on drafts of this manuscript.

Ethics

All animal procedures were approved by the Institutional Animal Care and Use Committee at the University of Texas and the Costa Rican Ministerio de Ambiente y Energía.

References

Allison D, Kaprio J, Korkeila M, Koskenvuo M, Neale M, Hayakawa K. 1996. The heritability of body mass index among an international sample of monozygotic twins reared apart. *Int J Obes* **20**:501–506.

Andersson M. 1986. Evolution of Condition-Dependent Sex Ornaments and Mating Preferences: sexual selection based on viability differences **40**:804–816.

Andersson M. 1981. Sexual selection, natural selection and quality advertisement. *Biol J Linn Soc* **17**:375–393. doi:10.4324/9781315129266-5

Andrews S. 2010. FastQC: A quality control tool for high throughput sequence data.

Bachmann JC, Cortesi F, Hall MD, Marshall NJ, Salzburger W, Gante HF. 2017. Real-time social selection maintains honesty of a dynamic visual signal in cooperative fish 269–278. doi:10.1002/evl3.24

Bachorowski J. 1999. Vocal expression and perception of emotion. *Curr Dir Psychol Sci* **8**:53–57.

Baier F, Hoekstra HE. 2019. The genetics of morphological and behavioural island traits in deer mice. *Proc R Soc B Biol Sci* **286**. doi:10.1098/rspb.2019.1697

Barske J, Fusani L, Wikelski M, Feng NY, Santos M, Schlinger B a. 2013. Energetics of the acrobatic courtship in male golden-collared manakins (Manacus vitellinus). *Proc R Soc B Biol Sci* **281**:20132482–20132482. doi:10.1098/rspb.2013.2482

Beerli P. 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* **22**:341–345. doi:10.1093/bioinformatics/bti803

Bell AM, Hankison SJ, Laskowski KL. 2009. The repeatability of behaviour: a meta-analysis. *Anim Behav* **77**:771–783. doi:10.1016/j.anbehav.2008.12.022

Bérénos C, Ellis PA, Pilkington JG, Pemberton JM. 2014. Estimating quantitative genetic parameters in wild populations: A comparison of pedigree and genomic approaches. *Mol Ecol* **23**:3434–3451. doi:10.1111/mec.12827

Bernal XE, Akre KL, Baugh AT, Rand AS, Ryan MJ. 2009. Female and male behavioral response to advertisement calls of graded complexity in túngara frogs, Physalaemus pustulosus. *Behav Ecol Sociobiol* **63**:1269–1279. doi:10.1007/s00265-009-0795-5

Boake CRB. 1989. Repeatability: Its role in evolutionary studies of mating behavior. *Evol Ecol* **3**:173–182. doi:10.1007/BF02270919

Bradbury JW, Vehrencamp S. 1998. Costs and ConstraintsPrinciples of Animal Communication.

Sunderland, Massachusetts: Sinauer Associates, Inc.

Briefer EF. 2012. Vocal expression of emotions in mammals: Mechanisms of production and evidence. *J Zool* **288**:1–20. doi:10.1111/j.1469-7998.2012.00920.x

Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Maechler M, Bolker BM. 2017. Modeling Zero-Inflated Count Data With glmmTMB. *bioRxiv* 132753. doi:10.1101/132753

Burkhard TT, Westwick RR, Phelps SM. 2018. Adiposity signals predict vocal effort in Alston's singing mice. *Proc R Soc B Biol Sci* **285**. doi:10.1098/rspb.2018.0090

Bürkner PC. 2018. Advanced Bayesian multilevel modeling with the R package brms. *R J* **10**:395–411. doi:10.32614/rj-2018-017

Bürkner PC. 2017. brms: An R package for Bayesian multilevel models using Stan. *J Stat Softw* **80**. doi:10.18637/jss.v080.i01

Burt JM, Beecher MD. 2008. The social interaction role of song in song sparrows: implications for signal design. *Comp Cogn Behav Rev* **3**:86–98. doi:10.3819/ccbr.2008.30005

Campbell P, Pasch B, Pino JL, Crino OL, Phillips M, Phelps SM. 2010. Geographic variation in the songs of neotropical singing mice: testing the relative importance of drift and local adaptation. *Evolution* **64**:1955–72. doi:10.1111/j.1558-5646.2010.00962.x

Campbell P, Pasch B, Warren AL, Phelps SM. 2014. Vocal ontogeny in neotropical singing mice (scotinomys). *PLoS One* **9**:e113628. doi:10.1371/journal.pone.0113628

Canty A, Ripley B. 2020. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-25.

Cassell B, Scientist ED, Tech V. 2009. Using Heritability for Genetic Improvement. *Virginia Coop Ext*.

Catchpole CK, Slater PJ. 2008. Bird song: biological themes and variations. New York:

Cambridge University Press. doi:10.1017/CBO9781107415324.004

Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**:1–22. doi:10.1186/s13742-015-0047-8

Chappell M a., Zuk M, Kwan TH, Johnsen TS. 1995. Energy cost of an avian vocal display: crowing in red junglefowl. *Anim Behav* **49**:255–257. doi:10.1016/0003-3472(95)80177-4

Charmantier A, Buoro M, Gimenez O, Weimerskirch H. 2011. Heritability of short-scale natal dispersal in a large-scale foraging bird, the wandering albatross. *J Evol Biol* **24**:1487–1496. doi:10.1111/j.1420-9101.2011.02281.x

Cocroft RB, Ryan MJ. 1995. Patterns of advertisement call evolution in toads and chorus frogs. *Anim Behav* **49**:283–303. doi:10.1006/anbe.1995.0043

Cotton S, Fowler K, Pomiankowski A. 2004. Do sexual ornaments demonstrate heightened condition-dependent expression as predicted by the handicap hypothesis? *Proc Biol Sci* **271**:771–783. doi:10.1098/rspb.2004.2688

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 2011. The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158. doi:10.1093/bioinformatics/btr330

de Villemereuil P. 2019. On the relevance of Bayesian statistics and MCMC for animal models. *J Anim Breed Genet* **136**:339–340. doi:10.1111/jbg.12426

Demartsev V, Kershenbaum A, Ilany A, Barocas A, Ziv EB, Koren L, Geffen E. 2014. Male hyraxes increase song complexity and duration in the presence of alert individuals. *Behav Ecol* **25**:1451–1458. doi:10.1093/beheco/aru155

Dingemanse NJ, Both C, Drent PJ, Van Oers K, Van Noordwijk AJ. 2002. Repeatability and

heritability of exploratory behaviour in great tits from the wild. *Anim Behav* **64**:929–938. doi:10.1006/anbe.2002.2006

Dingemanse NJ, Van Der Plas F, Wright J, Réale D, Schrama M, Roff DA, Van Der Zee E, Barber I. 2009. Individual experience and evolutionary history of predation affect expression of heritable variation in fish personality and morphology. *Proc R Soc B Biol Sci* **276**:1285–1293. doi:10.1098/rspb.2008.1555

Dochtermann NA, Schwab T, Anderson Berdal M, Dalos J, Royauté R. 2019. The Heritability of Behavior: A Meta-analysis. *J Hered* **110**:403–410. doi:10.1093/jhered/esz023

Dohm MR. 2002. Repeatability estimates do not always set an upper limit to heritability. *Funct Ecol* **16**:273–280. doi:10.1046/j.1365-2435.2002.00621.x

Doutrelant C, Mcgregor PK, Oliveira RF. 2001. The effect of an audience on intrasexual communication in male Siamese fighting fish, Betta splendens. *Behav Ecol* **12**:283–286. doi:10.1093/beheco/12.3.283

Dubuc C, Winters S, Allen WL, Brent LJN, Cascio J, Maestripieri D, Ruiz-Lambides A V, Widdig A, Higham JP. 2014. Sexually selected skin colour is heritable and related to fecundity in a non-human primate. *Proc Biol Sci* **281**:20141602. doi:10.1098/rspb.2014.1602

Dudley JW, Lambert RJ. 2010. 100 Generations of Selection for Oil and Protein in Corn, Plant Breeding Reviews. doi:10.1002/9780470650240.ch5

Endler J. 1992. Signals, signal conditions, and the direction of evolution. *Am Nat* **139**.

Eu-ahsunthornwattana J, Miller EN, Fakiola M, Jeronimo SMB, Blackwell JM, Cordell HJ. 2014. Comparison of Methods to Account for Relatedness in Genome-Wide Association Studies with Family-Based Data. *PLoS Genet* **10**. doi:10.1371/journal.pgen.1004445

Falconer DS. 1981. Introduction to Quantitative Genetics.

Fang G, Jiang F, Yang P, Cui J, Brauth SE, Tang Y. 2014. Male vocal competition is dynamic and strongly affected by social contexts in music frogs. *Anim Cogn* **17**:483–494. doi:10.1007/s10071-013-0680-5

Fernández-Vargas M. 2018a. Vocal Signals of Sexual Motivation in Male and Female Rodents. *Curr Sex Heal Reports* **10**:315–328. doi:10.1007/s11930-018-0179-9

Fernández-Vargas M. 2018b. Presence of a potential competitor and its individual identity modulate ultrasonic vocalizations in male hamsters. *Anim Behav* **145**:11–27. doi:10.1016/j.anbehav.2018.08.014

Fernández-Vargas M, Phelps SMM, Tang-Martínez Z. 2011. Singing, allogrooming, and allomarking behaviour during inter- and intra-sexual encounters in the Neotropical short-tailed singing mouse (Scotinomys teguina). *Behaviour* **148**:945–965. doi:10.1163/000579511X584591

Ferrannini G, Namwanje M, Fang B, Damle M, Li D, Liu Q, Lazar MA, Qiang L. 2016. Genetic backgrounds determine brown remodeling of white fat in rodents. *Mol Metab* **5**:948–958. doi:10.1016/j.molmet.2016.08.013

Fisher R a. 1930. The Genetical Theory of Natural Selection. Oxford: Clarendon Press. doi:10.1038/158453a0

Fitch T. 2006. Production of Vocalizations in MammalsEncyclopedia of Language & Linguistics. pp. 115–121. doi:10.1016/b0-08-044854-2/00821-x

Fitch WT, Reby D. 2001. The descended larynx is not uniquely human. *Proc R Soc B Biol Sci* **268**:1669–1675. doi:10.1098/rspb.2001.1704

Fox J, Hong J. 2009. Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. *J Stat Softw* **32**:1–24. doi:10.18637/jss.v032.i01

Fox J, Weisberg S. 2018. Visualizing fit and lack of fit in complex regression models with predictor effect plots and partial residuals. *J Stat Softw* **87**. doi:10.18637/jss.v087.i09

Friedman JM, Halaas JL. 1998. Leptin and the regulation of body weight in mammals. *Nature* **395**:763–770. doi:10.1038/27376

George AS. 2014. Songs of Aggression: The Singing Mouse Model. *Master's thesis, University of Texas at Austin.*

Gerhardt HC. 1991. Female mate choice in treefrogs: static and dynamic acoustic criteria. *Anim Behav* **42**:615–635. doi:10.1016/S0003-3472(05)80245-3

Gerhardt HC. 1982. Sound pattern recognition in some North American treefrogs (anura: Hylidae): Implications for mate choice. *Integr Comp Biol* **22**:581–595. doi:10.1093/icb/22.3.581

Gervais L, Hewison AJM, Morellet N, Bernard M, Merlet J, Cargnelutti B, Chaval Y, Pujol B, Quéméré E. 2020. Pedigree-free quantitative genetic approach provides evidence for heritability of movement tactics in wild roe deer. *J Evol Biol* **33**:595–607. doi:10.1111/jeb.13594

Gervais L, Perrier C, Bernard M, Merlet J, Quéméré E, Pemberton JM, Pujol B, Quéméré E, Pemberton JM, Pujol B. 2019. RAD-sequencing for estimating genomic relatedness matrix-based heritability in the wild: A case study in roe deer. *Mol Ecol Resour* 1205–1217. doi:10.1111/1755-0998.13031

Gienapp P, Fior S, Guillaume F, Lasky JR, Sork VL, Csilléry K. 2017. Genomic Quantitative Genetics to Study Evolution in the Wild. *Trends Ecol Evol* **32**:897–908. doi:10.1016/j.tree.2017.09.004

Gillooly JF, Ophir AG. 2010. The energetic basis of acoustic communication 1325–1331. doi:10.1098/rspb.2009.2134

Gouzoules H, Gouzoules S. 2000. Agonistic screams differ among four species of macaques: The significance of motivation-structural rules. *Anim Behav* **59**:501–512. doi:10.1006/anbe.1999.1318

Grafen A. 1990. Sexual selection unhandicapped by the Fisher process. *J Theor Biol* **144**:473–516. doi:10.1016/S0022-5193(05)80087-6

Grafen A. 1984. Natural selection, kin selection and group selection In: Krebs JR DN, editor. Behavioural Ecology: An Evolutionary Approach. Oxford: Blackwell Scientific Publications. pp. 62–84.

Grant BR, Grant PR. 1993. Evolution of Darwin's finches caused by a rare climatic event. *Proc R Soc B Biol Sci* **251**:111–117. doi:10.1098/rspb.1993.0016

Grant PR, Grant BR. 1995. Predicting microevolutionary responses to directional selection on heritable variation. *Evolution (N Y)* **49**:241. doi:10.2307/2410334

Hamilton WD, Zuk M. 1982. Heritable true fitness and bright birds: a role for parasites? *Science* **218**:384–7.

Harrison XA. 2014. Using observation-level random e ff ects to model overdispersion in count data in ecology and evolution. doi:10.7717/peerj.616

Havel PJ. 2001. Peripheral signals conveying metabolic information to the brain: short-term and long-term regulation of food intake and energy homeostasis. *Exp Biol Med (Maywood)* **226**:963–977.

Hedrick A, Bunting J. 2014. An attractive male trait and aggressiveness are negatively correlated in wild field crickets, but uncorrelated in lab-reared crickets. *Behav Ecol Sociobiol* **68**:233–238. doi:10.1007/s00265-013-1638-y

Henderson C. 1950. Estimation of genetic parameters. *Ann Math Stat* **21**:309–210.

Hill WG, Kirkpatrick M. 2010. What Animal Breeding Has Taught Us about Evolution. *Annu Rev Ecol Evol Syst* **41**:1–19. doi:10.1146/annurev-ecolsys-102209-144728

Hooper ET. 1972. A synopsis of the rodent genus Scotinomys. *Occas Pap Museum Zool Univ Michigan* 1–32.

Hooper ET, Carleton MDM. 1976. Reproduction, growth and development in two contiguously allopatric rodent species, genus Scotinomys. *Misc Publ Museum Zool Univ Michigan* **151**:1–52.

Houtman AM. 1992. Female zebra finches choose extra-pair copulations with genetically attractive males. *Proc R Soc B Biol Sci* **249**:3–6. doi:10.1098/rspb.1992.0075

Ingram T, Harrison A, Mahler DL, Castañeda MDR, Glor RE, Herrel A, Stuart YE, Losos JB. 2016. Comparative tests of the role of dewlap size in Anolis lizard speciation. *Proc R Soc B Biol Sci* **283**. doi:10.1098/rspb.2016.2199

Izzo A, Tibbetts EA. 2015. Heightened Condition Dependence of a Sexually Selected Signal in Male Polistes dominulus Paper Wasps. *Ethology* **121**:586–592. doi:10.1111/eth.12371

Jensen H, Sæther BE, Ringsby TH, Tufto J, Griffith SC, Ellegren H. 2003. Sexual variation in heritability and genetic correlations of morphological traits in house sparrow (Passer domesticus). *J Evol Biol* **16**:1296–1307. doi:10.1046/j.1420-9101.2003.00614.x

Johnstone RA. 1993. Chapter 7 The Evolution of Animal Signals. *Behav Ecol An Evol approach* 155–78.

Jolliffe IT. 2002. Principal Component Analysis, Second Edition, Encyclopedia of Statistics in Behavioral Science. doi:10.2307/1270093

Jürgens U. 2002. Neural pathways underlying vocal control. *Neurosci Biobehav Rev* **26**:235–258.

Kagawa H, Soma M. 2013. Song performance and elaboration as potential indicators of male quality in Java sparrows. *Behav Processes* **99**:138–144. doi:10.1016/j.beproc.2013.07.012

Karin BR, Cicero C, Koo MS, Bowie RCK. 2018. The role of history and ecology as drivers of song divergence in Bell's and Sagebrush sparrows (Artemisiospiza, Aves: Passerellidae). *Biol J Linn Soc* **125**:1–20. doi:10.1093/biolinnean/bly090

Kirkpatrick M, Ryan M. 1991. The evolution of mating preferences and the paradox of the lek. *Nature*.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**. doi:10.1007/978-1-4614-1174-1_8

Korneliussen TS, Moltke I. 2015. NgsRelate: A software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics* **31**:4009–4011. doi:10.1093/bioinformatics/btv509

Kruschke JK, Liddell TM. 2018. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychon Bull Rev* **25**:178–206. doi:10.3758/s13423-016-1221-4

Kruuk LEB. 2004. Estimating genetic parameters in natural populations using the "animal model." *Philos Trans R Soc B Biol Sci* **359**:873–890. doi:10.1098/rstb.2003.1437

Lampe HMH, Espmark YYO. 1994. Song structure reflects male quality in pied flycatchers, Ficedula hypoleuca. *Anim Behav*. doi:10.1006/anbe.1994.1118

Lande R. 1976. Natural Selection and Random Genetic Drift in Phenotypic Evolution. *Evolution (N Y)* **30**:314. doi:10.2307/2407703

Langin KM, Sillett TS, Morrison SA, Ghalambor CK. 2017. Bill morphology and neutral genetic structure both predict variation in acoustic signals within a bird population. *Behav Ecol*

28:866–873. doi:10.1093/beheco/arx051

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357–359. doi:10.1038/nmeth.1923

Lee SH, Ripke S, Neale BM, Faraone S V., Purcell SM, Perlis RH, et al. 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* **45**:984–994. doi:10.1038/ng.2711

Lehtovirta M, Pietiläinen KH, Levälahti E, Heikkilä K, Groop L, Silventoinen K, Koskenvuo M, Kaprio J. 2010. Evidence that BMI and type 2 diabetes share only a minor fraction of genetic variance: A follow-up study of 23,585 monozygotic and dizygotic twins from the Finnish twin cohort Study. *Diabetologia* **53**:1314–1321. doi:10.1007/s00125-010-1746-4

Lenth R. 2020. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.8.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. doi:10.1093/bioinformatics/btp352

Lynch M, Walsh B, B. LM& W. 1998. Genetics and Analysis of Quantitative Traits. Sunderland; MA: Sinauer Associates. doi:10.1086/318209

Mager JN, Walcott C, Piper WH. 2007. Male common loons, Gavia immer, communicate body mass and condition through dominant frequencies of territorial yodels. *Anim Behav* **73**:683–690. doi:10.1016/j.anbehav.2006.10.009

Martin K, Tucker MA, Rogers TL. 2017. Does size matter? Examining the drivers of mammalian vocalizations. *Evolution (N Y)* **71**:249–260. doi:10.1111/evo.13128

Matos CAP, Thomas DL, Gianola D, Tempelman RJ, Young LD. 1997. Genetic Analysis of

Discrete Reproductive Traits in Sheep Using Linear and Nonlinear Models: I. Estimation of Genetic Parameters. *J Anim Sci* **75**:76–87. doi:10.2527/1997.75176x

Miller JJRJ, Engstrom MMD. 2007. Vocal Stereotypy and Singing Behavior in Baiomyine Mice. *J Mammal* **88**:1447–1465. doi:10.1644/06-MAMM-A-386R.1

Miller JR, Engstrom MD. 2012. Vocal stereotypy in the rodent genera Peromyscus and Onychomys (Neotominae): taxonomic signature and call design. *Bioacoustics* **21**:193–213. doi:10.1080/09524622.2012.675176

Morrissey MB, de Villemereuil P, Doligez B, Gimenez O. 2014. Bayesian approaches to the quantitative genetic analysis of natural populations. *Quant Genet Wild* 228–253. doi:10.1093/acprof:oso/9780199674237.003.0014

Mousseau T, Ritland K, Heath D. 1998. A novel method for estimating heritability using molecular markers. *Heredity (Edinb)* **80**:218–224.

Mousseau TA, Roff DA. 1987. Natural selection and the heritability of fitness components. *Heredity (Edinb)* **59**:181–197. doi:10.1038/hdy.1987.113

Nelson BS, Beckers GJL, Suthers RA. 2005. Vocal tract filtering and sound radiation in a songbird. *J Exp Biol* **208**:297–308. doi:10.1242/jeb.01378

Okobi DE, Banerjee A, Matheson AMMM, Phelps SM, Long MA. 2019. Motor cortical control of vocal interaction in neotropical singing mice. *Science (80-)* **363**:983–988. doi:10.1126/science.aau9480

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, Mcglinn D, Minchin PR, O'hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Szoecs E, Maintainer HW. 2019. Package "vegan" Title Community Ecology Package.

Ophir a. G, Schrader SB, Gillooly JF. 2010. Energetic cost of calling: general constraints and

species-specific differences. *J Evol Biol* **23**:1564–1569. doi:10.1111/j.1420-9101.2010.02005.x

Orengo DJ, Prevosti A. 1999. Wing-size heritability in a natural population of Drosophila subobscura. *Heredity (Edinb)* **82**:100–106. doi:10.1038/sj.hdy.6884390

Pasch B, Bolker BM, Phelps SM. 2013. Interspecific dominance via vocal interactions mediates altitudinal zonation in neotropical singing mice. *Am Nat* **182**:E161-73. doi:10.1086/673263

Pasch B, George AS, Campbell P, Phelps SM. 2011a. Androgen-dependent male vocal performance influences female preference in Neotropical singing mice. *Anim Behav* **82**:177–183. doi:10.1016/j.anbehav.2011.04.018

Pasch B, George AS, Hamlin HJ, Guillette LJ, Phelps SM. 2011b. Androgens modulate song effort and aggression in Neotropical singing mice. *Horm Behav* **59**:90–97. doi:10.1016/j.yhbeh.2010.10.011

Perrier C, Delahaie B, Charmantier A. 2018. Heritability estimates from genomewide relatedness matrices in wild populations: Application to a passerine, using a small sample size. *Mol Ecol Resour* **18**:838–853. doi:10.1111/1755-0998.12886

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**:e37135. doi:10.1371/journal.pone.0037135

Pettersen AK, Marshall DJ, White CR. 2018. Understanding variation in metabolic rate. *J Exp Biol* **221**. doi:10.1242/jeb.166876

Pisanski K, Fraccaro PJ, Tigue CC, O'Connor JJM, Röder S, Andrews PW, Fink B, DeBruine LM, Jones BC, Feinberg DR. 2014. Vocal indicators of body size in men and women: A meta-analysis. *Anim Behav* **95**:89–99. doi:10.1016/j.anbehav.2014.06.011

Podos J, Cohn-Haft M. 2019. Extremely loud mating songs at close range in white bellbirds. *Curr Biol* **29**:R1068–R1069. doi:10.1016/j.cub.2019.09.028

Pomiankowski a. 1987. Sexual Selection: The Handicap Principle Does Work -- Sometimes. *Proc R Soc B Biol Sci* **231**:123–145. doi:10.1098/rspb.1987.0038

Price T, Schluter D, Jolla L, Group TE. 1991. On the low heritability of life-history traits. *Evolution (N Y)* **45**:853–861.

Pultorak JD, Alger SJ, Loria SO, Johnson AM, Marler CA. 2018. Changes in behavior and ultrasonic vocalizations during pair bonding and in response to an infidelity challenge in monogamous California mice. *Front Ecol Evol* **6**:1–15. doi:10.3389/fevo.2018.00125

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**:559–575. doi:10.1086/519795

RA. F. 1954. Statistical methods for research. Edinburgh.

Reby D, McComb K. 2003. Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags. *Anim Behav* **65**:519–530. doi:10.1006/anbe.2003.2078

Riede T, Borgard HL, Pasch B. 2017. Laryngeal airway reconstruction indicates that rodent ultrasonic vocalizations are produced by an edge-tone mechanism. *R Soc Open Sci* **4**. doi:10.1098/rsos.170976

Riede T, Brown C. 2013. Body Size , Vocal Fold Length, and Fundamental Frequency – Implications for Mammal Vocal Communication. *Nov Acta Leopoldina NF 111* **380**:295–314.

Riede T, Pasch B. 2020. Pygmy mouse songs reveal anatomical innovations underlying acoustic signal elaboration in rodents. *J Exp Biol* **223**:jeb223925. doi:10.1242/jeb.223925

Ritland K. 1996. Marker-based method for inferences about quantitative inheritance in natural populations. *Evolution* **50**:1062–1073. doi:10.2307/2410647

Rittschof CC, Robinson GE. 2014. Genomics: moving behavioural ecology beyond the phenotypic gambit. *Anim Behav* **92**:263–270. doi:10.1016/j.anbehav.2014.02.028

Rochette NC, Catchen JM. 2017. Deriving genotypes from RAD-seq short-read data using Stacks. *Nat Protoc* **12**:2640–2659. doi:10.1038/nprot.2017.123

Roff DA, Simons AM. 1997. The quantitative genetics of wing dimorphism under laboratory and “field” conditions in the cricket Gryllus pennsylvanicus. *Heredity (Edinb)* **78**:235–240. doi:10.1038/hdy.1997.37

Rolff J, Joop G. 2002. Estimating condition: Pitfalls of using weight as a fitness correlate. *Evol Ecol Res* **4**:931–935.

Rowe L, Houle D. 1996. The lek paradox and the capture of genetic variance by condition dependent traits. *Proc R Soc B Biol Sci* **263**:1415–1421. doi:10.1098/rspb.1996.0207

Ryan M. 1985. The Tungara Frog. A Study in Sexual Selection and Communication. University of Chicago Press. doi:10.1086/415117

Ryan MJ, Brenowitz EA. 1985. The role of body size, phylogeny, and ambient noise in the evolution of bird song. *Am Nat* **126**:87. doi:10.1086/284398

Santori C, Bussière LF, Houslay TM. 2020. Heightened perception of competition hastens courtship. *Behav Ecol* **31**:239–246. doi:10.1093/beheco/arz183

Sauter DA, Eisner F, Calder AJ, Scott SK. 2010. Perceptual cues in nonverbal vocal expressions of emotion. *Q J Exp Psychol* **63**:2251–2272. doi:10.1080/17470211003721642

Schleinitz D, Böttcher Y, Blüher M, Kovacs P. 2014. The genetics of fat distribution. *Diabetologia* **57**:1276–1286. doi:10.1007/s00125-014-3214-z

Schneider JR, Chadee DD, Mori A, Romero-Severson J, Severson DW. 2011. Heritability and adaptive phenotypic plasticity of adult body size in the mosquito Aedes aegypti with implications for dengue vector competence. *Infect Genet Evol*. doi:10.1038/jid.2014.371

Schousboe K, Visscher PM, Erbas B, Kyvik KO, Hopper JL, Henriksen JE, Heitmann BL, Sørensen TIA. 2004. Twin study of genetic and environmental influences on adult body size, shape, and composition. *Int J Obes* **28**:39–48. doi:10.1038/sj.ijo.0802524

Schweiger R, Fisher E, Weissbrod O, Rahmani E, Müller-Nurasyid M, Kunze S, Gieger C, Waldenberger M, Rosset S, Halperin E. 2018. Detecting heritable phenotypes without a model using fast permutation testing for heritability and set-tests. *Nat Commun* **9**. doi:10.1038/s41467-018-07276-w

Sheehan MJ, Tibbetts EA. 2010. Selection for individual recognition and the evolution of polymorphic identity signals in polistes paper wasps. *J Evol Biol* **23**:570–577. doi:10.1111/j.1420-9101.2009.01923.x

Simons AM, Roff DA. 1994. The effect of environmental variability on the heritabilities of traits of a field cricket. *Evolution (N Y)* **48**:1637–1649. doi:10.1111/j.1558-5646.1994.tb02201.x

Smith SK, Burkhard TT, Phelps SM. 2020. A comparative characterization of laryngeal anatomy in the singing mouse. *J Anat* 1–13. doi:10.1111/joa.13315

St-Hilaire É, Réale D, Garant D. 2017. Determinants, selection and heritability of docility in wild eastern chipmunks (Tamias striatus). *Behav Ecol Sociobiol* **71**. doi:10.1007/s00265-017-2320-6

Stirling DG, Réale D, Roff DA. 2002. Selection, structure and the heritability of behaviour. *J Evol Biol* **15**:277–289. doi:10.1046/j.1420-9101.2002.00389.x

Stoddard PK, Beecher MD. 1983. Parental Recognition of Offspring in the Cliff Swallow. *Auk* **100**:795–799. doi:10.1093/auk/100.4.795

Stoffel MA, Nakagawa S, Schielzeth H. 2017. rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol Evol* **8**:1639–1644. doi:10.1111/2041-210X.12797

Story BH, Titze IR, Hoffman EA. 2001. The relationship of vocal tract shape to three voice qualities. *J Acoust Soc Am* **109**:1651–1667. doi:10.1121/1.1352085

Thomas RJ, Cuthill IC. 2002. Body mass regulation and the daily singing routines of European robins. *Anim Behav* **63**:285–295. doi:10.1006/anbe.2001.1926

Thomson IR, Darveau CA, Bertram SM. 2014. Body morphology, energy stores, and muscle enzyme activity explain cricket acoustic mate attraction signaling variation. *PLoS One* **9**:1–16. doi:10.1371/journal.pone.0090409

Tomkins JL, Radwan J, Kotiaho JS, Tregenza T. 2004. Genic capture and resolving the lek paradox. *Trends Ecol Evol* **19**:323–8. doi:10.1016/j.tree.2004.03.029

Van de Schoot R, Broere JJ, Perryck KH, Loey NE Van. 2015. Analayzing small data sets using Bayesian estimation. *Eur J Psychotraumatol* **1**:1–13.

Vanhooydonck B, Herrel A, Meyers JJ, Irschick DJ. 2009. What determines dewlap diversity in Anolis lizards? An among-island comparison. *J Evol Biol* **22**:293–305. doi:10.1111/j.1420-9101.2008.01643.x

Villemereuil P De, Gimenez O, Doligez B, Lyon D, De Villemereuil P, Gimenez O, Doligez B. 2013. Comparing parent-offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: A simulation study for Gaussian and binary traits. *Methods Ecol Evol* **4**:260–275. doi:10.1111/2041-210X.12011

Visscher PM, Hill WG, Wray NR. 2008. Heritability in the genomics era — concepts and misconceptions. *Nat Rev Genet* **9**:255–266. doi:10.1038/nrg2322

Voeten CC. 2020. buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression. R package version 1.6.

Waldmann P. 2001. Additive and non-additive genetic architecture of two different-sized populations of Scabiosa canescens. *Heredity (Edinb)* **86**:648–657. doi:10.1046/j.1365-2540.2001.00873.x

Ward S, Speakman JR, Slater PJB. 2003. The energy cost of song in the canary, Serinus canaria. *Anim Behav* **66**:893–902. doi:10.1006/anbe.2003.2250

Wiens JJ, Tuschhoff E. 2020. Songs versus colours versus horns: what explains the diversity of sexually selected traits? . *Biol Rev* **7**. doi:10.1111/brv.12593

Wilson AJ, Réale D, Clements MN, Morrissey MM, Postma E, Walling CA, Kruuk LEBB, Nussey DH. 2010. An ecologist's guide to the animal model. *J Anim Ecol* **79**:13–26. doi:10.1111/j.1365-2656.2009.01639.x

Wolak ME, Fairbairn DJ, Paulsen YR. 2012. Guidelines for estimating repeatability. *Methods Ecol Evol* **3**:129–137. doi:10.1111/j.2041-210X.2011.00125.x

Wuschke S, Dahm S, Schmidt C, Joost HG, Al-Hasani H. 2007. A meta-analysis of quantitative trait loci associated with body weight and adiposity in mice. *Int J Obes* **31**:829–841. doi:10.1038/sj.ijo.0803473

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Others. 2010. Common {SNPs} explain a large proportion of the heritability for human height. *Nat Gen* **42**:565–569. doi:10.1038/ng.608.Common

Yang J, Lee SH, Goddard MEM, Visscher PM, Analysis CT, Yang J, Lee H, Goddard MEM,

Visscher PM, Lee SH, Goddard MEM, Visscher PM. 2011. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**:76–82. doi:10.1016/j.ajhg.2010.11.011

Zahavi A. 1975. Mate selection-A selection for a handicap. *J Theor Biol* **53**:205–214. doi:10.1016/0022-5193(75)90111-3

Zheng D. 2020. Androgen modulation of vocal circuits in Alston's singing mouse. *Doctoral thesis, University of Texas at Austin*

Zuberbühler K. 2008. Audience effects. *Curr Biol*. doi:10.1016/j.cub.2007.12.041

Zuk M, Kolluru GR. 1998. Exploitation of Sexual Signals by Predators and Parasitoids. *Q Rev Biol* **73**:415–438. doi:10.1086/420412