Received: 15 November 2018 | Revised: 30 January 2019 | Accepted: 30 January 2019

DOI: 10.1111/mec.15042

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Is genomic diversity a useful proxy for census population size? Evidence from a species-rich community of desert lizards

Maggie R. Grundler[1,2] | Sonal Singhal[1,3] | Mark A. Cowan[4] | Daniel L. Rabosky[1]

[1]Museum of Zoology and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan

[2]Department of Environmental Science, Policy, & Management, University of California, Berkeley, Berkeley, California

[3]Department of Biology, CSU Dominguez Hills, Carson, California

[4]Department of Biodiversity, Conservation and Attractions, Kensington, Western Australia, Australia

Correspondence
Maggie R. Grundler, Department of Environmental Science, Policy, & Management, University of California, Berkeley, Berkeley, CA.
Email: mgrundler@berkeley.edu

Funding information
National Science Foundation, Grant/Award Number: DEB-0814277, DEB-1519732, DEB-1754398 and OSIE-0612855; David and Lucile Packard Foundation

Abstract

Species abundance data are critical for testing ecological theory, but obtaining accurate empirical estimates for many taxa is challenging. Proxies for species abundance can help researchers circumvent time and cost constraints that are prohibitive for long-term sampling. Under simple demographic models, genetic diversity is expected to correlate with census size, such that genome-wide heterozygosity may provide a surrogate measure of species abundance. We tested whether nucleotide diversity is correlated with long-term estimates of abundance, occupancy and degree of ecological specialization in a diverse lizard community from arid Australia. Using targeted sequence capture, we obtained estimates of genomic diversity from 30 species of lizards, recovering an average of 5,066 loci covering 3.6 Mb of DNA sequence per individual. We compared measures of individual heterozygosity to a metric of habitat specialization to investigate whether ecological preference exerts a measurable effect on genetic diversity. We find that heterozygosity is significantly correlated with species abundance and occupancy, but not habitat specialization. Demonstrating the power of genomic sampling, the correlation between heterozygosity and abundance/occupancy emerged from considering just one or two individuals per species. However, genetic diversity does no better at predicting abundance than a single day of traditional sampling in this community. We conclude that genetic diversity is a useful proxy for regional-scale species abundance and occupancy, but a large amount of unexplained variation in heterozygosity suggests additional constraints or a failure of ecological sampling to adequately capture variation in true population size.

KEYWORDS

heterozygosity, Lewontin's paradox, species abundance, squamates, target capture

1 | INTRODUCTION

Species abundance distributions represent one of the most basic descriptions of a community, and are the foundation of many ecological theories and conservation management practices (He & Gaston, 2000; McGill et al., 2007). Relative abundance is typically estimated through repeated community sampling efforts, with study durations that frequently span multiple years (Gotelli & Chao, 2013; Magurran et al., 2010; Meyer et al., 2010; Pianka, 2014). An obvious difficulty is that such "brute-force" sampling requires considerable time and funding in order to produce reliable estimates (Pearce & Ferrier, 2001; Yin & He, 2014). Consequently, there has been widespread interest in developing statistical methods for estimating relative species abundance from imperfect survey data as well as from proxy variables, including point occurrences, spatial distributions, and environmental suitability (He & Gaston, 2000; VanDerWal, Shoo, Johnson, & Williams, 2009; Yin & He, 2014). Although these methods perform well at small spatial scales, predicting regional abundance remains challenging, in part because of the paucity of data available to validate model-predicted abundances (Pearce & Ferrier, 2001; Yin & He, 2014).

Neutral theory predicts that genetic diversity should correlate with population census size (Leffler et al., 2012; Tallmon et al., 2010; Wright, 1931), and intraspecific genetic variation is thus an attractive proxy for species abundance in natural populations. Genetic material is both easy and inexpensive to sample (Schwartz, Luikart, & Waples, 2007), and recent advances in sequencing technology and bioinformatics allow researchers to harvest information from across the genome at low cost. Furthermore, reliable estimates of both genetic diversity and past population size have been recovered from sample sizes as small as a single individual (Li & Durbin, 2011; Nazareno, Bemmels, Dick, & Lohmann, 2017). However, historical signals of demographic processes can weaken the relationship between census and effective population size, and thus the relationship between abundance and heterozygosity (Frankham, 1995). For example, range expansions or population bottlenecks can lead to reduced genetic diversity in marginal or founder populations, and consequently to decoupling from contemporary census population size (Banks et al., 2013; Charlesworth, 2009; Dalongeville, Andrello, Mouillot, Albouy, & Manel, 2016; Excoffier & Ray, 2008).

Previous studies have found positive relationships between heterozygosity and proxies for species abundance, including population size estimated from calculations of density and acreage (Patton & Yang, 1977), extrapolations calculated from active social groups (Stangel, Lennartz, & Smith, 1992), categorical estimates (“large,” “small”) of population size (Godt, Johnson, & Hamrick, 1996; Hague & Routman, 2016) and museum occurrence records (Singhal, Huang, et al., 2017). Studies that more directly compare intraspecific genetic diversity and abundance have reported positive associations, but these studies have generally focused on single species or paired species comparisons (Devillard, Santin-Janin, Say, & Pontier, 2011; Lozier, 2014; Ortego, Aparicio, Cordero, & Calabuig, 2008; Sun, 1996), or on many species sampled at a broad geographical scale (Bazin, Glémin, & Galtier, 2006; Leimu, Mutikainen, Koricheva, & Fischer, 2006; McCusker & Bentzen, 2010; Pinsky & Palumbi, 2014). Few studies have directly assessed the relationship between abundance and genetic diversity within species-rich communities of potentially interacting species, despite the importance of such communities for biodiversity monitoring and studies of ecoevolutionary dynamics.

In addition to these rather practical motivations, the ecological analysis of genetic variation may help us to understand why the range of genetic diversity among species is orders of magnitude smaller than that of census population size. This observation represents a long-standing but unresolved puzzle for evolutionary biology known as “Lewontin's paradox” (Leffler et al., 2012; Lewontin, 1972). Addressing this knowledge gap may be especially critical for conservation efforts, which would benefit from a comprehensive understanding of which factors constrain genetic diversity and influence effective population size in wild populations (Leffler et al., 2012; Palstra & Ruzzante, 2008).

Foundational studies based on results from a small number of loci, including microsatellites characterized by high variability in mutation rate, found few discernible links between genetic polymorphism and ecological traits (Ellegren & Galtier, 2016). Genome-wide data increase our power to recover associations between genetic diversity, ecological correlates, and evolutionary processes acting over shorter time periods (Faircloth et al., 2012; Harvey, Aleixo, Ribas, & Brumfield, 2017). Recent studies have found strong relationships between ecological traits and genetic diversity at broad phylogenetic scales (e.g., across phyla; Romiguier et al., 2014); whether such patterns are also exhibited by more closely related species with similar life history traits remains equivocal (Ellegren & Galtier, 2016; Romiguier et al., 2014; Singhal, Huang, et al., 2017). Additionally, there is some evidence that microhabitat preference can limit dispersal and reduce within-population genetic diversity (Brouat, Chevallier, Meusnier, Noblecourt, & Rasplus, 2004; Brouat, Sennedot, Audiot, Leblois, & Rasplus, 2003; Dalongeville et al., 2016; Khimoun et al., 2016; Pianka, 1986, 2014; Rabosky, Cowan, Talaba, & Lovette, 2011), but studies combining genome-wide diversity estimates with ecological traits at the community level are rare.

In this study, we test whether genome-wide estimates of heterozygosity are correlated with species abundance, landscape-level occupancy, and habitat specialization. We focus on a species-rich community of lizards from the western Australian arid zone that has been subject to multiyear demographic study and is characterized by variation in abundance and other ecological traits (Grundler, Pianka, Pelegrin, Cowan, & Rabosky, 2017; Pianka, 1986, 2014; Rabosky et al., 2011). Despite ecological variation among clades, species in this community are generally similar in major life history characteristics (Mesquita et al., 2016) and share a common geological and climatic history, potentially minimizing the confounding effects of variation in environment, demographic history, and phylogeny.

We also explore relationships with additional ecological traits that may relate to abundance, including habitat preference and body size. Based on classic theory and previous work, we predict that (i) greater species abundance will correlate with greater levels of nucleotide diversity; and (ii) increasing habitat specialization will correlate with reduced levels of nucleotide diversity, due to reduced gene flow between populations restricted by narrow ecological preference. We construct a multipredictor model to assess the relative importance of each of these ecological traits in explaining variation in heterozygosity. This framework provides a means of better understanding what ecological processes influence genetic diversity in light of Lewontin's paradox.

2 | METHODS

2.1 | Sample and ecological data collection

Tissue samples from 30 species of lizards were collected by Rabosky et al. (2011) as part of a long-term monitoring project at the former pastoral station of Lorna Glen in the western Australian arid zone. This area is now known by its traditional name Matuwa, and hereafter will be referred to by this name. The Matuwa region—and the spinifex desert of the western Australian arid zone more generally—harbours the most diverse lizard communities on Earth (James & Shine, 2000; Morton & James, 1988; Pianka, 1972; Roll et al., 2017), with potentially more than 50 species co-occurring at single sites from spinifex sandplain habitats (Pianka, 1986; Rabosky et al., 2011). Species from which genetic data were obtained included representatives of at least four Australian lizard radiations, including sphenomorphine and nonsphenomorphine skinks, agamids, and geckos. Ecological data including long-term cumulative measures of species abundance, species habitat preference, and body size were collected as part of the same study. The data presented here include genetic and ecological data for an additional 10 species that were not included in Rabosky et al. (2011). Briefly, the study entailed multiyear pitfall trapping of lizard communities at 24 sites at Matuwa, with sites selected to encompass the majority of habitats in the study region. Each site comprised two lines of six 20-L buckets buried in the substrate and connected by a continuous barrier of drift fencing. Sites were separated by approximately 3–10 km, and sampled for 21–28 days per year between 2002 and 2008. During each sampling period, traps on all sites were kept open for the same number of days, ensuring that sampling was standardized across the landscape. Further details on study design are available in Rabosky, Reid, Cowan, and Foulkes (2007), Rabosky et al. (2011).

Abundance for each species was calculated as the sum of each annual survey total across sites. While even the most rigorous sampling methods cannot capture true population size across a region, the consistent effort applied to the Matuwa lizard community represents one of the most direct assessments of abundance feasible. Pitfall traps combined with drift fencing are an effective method for capturing arid Australian lizards (Morton, Gillam, Jones, & Fleming, 1988). They are superior to alternative methods tested in this region (Cowan, Edinger, & Coate, 2017), and demonstrate low capture bias for the small-bodied taxa of this study, based on visual surveys by the authors.

To approximate habitat preference, 14 habitat variables were measured for each pitfall trap included in the survey, accounting for variation in nearby vegetation type, substrate type, soil compaction and shear strength, woody debris, and distance to and diameter of nearest sheltering vegetation. Each individual lizard was associated with the habitat variables of the trap in which it was captured. Habitat variables were log-transformed and *z*-score-standardized following methods reported by Rabosky et al. (2011). For each species composed of *n* individuals, we calculated the Euclidean distance between habitat variables for all pairwise comparisons of individuals, resulting in an $[n \times n]$ distance matrix. A simple index of habitat specialization was computed by taking the average of the distance matrix, excluding the diagonal. This final value provides an approximation of the average distance between two individuals in the habitat space occupied by a species, and is robust to variation in sample size between species. For example, individuals from a specialized species are expected to be associated with similar habitat variables (demonstrating adherence to a restricted set of environmental attributes) and therefore will generate a smaller average distance in this calculation, relative to generalist species.

We additionally compare genetic diversity to body size, a traditional proxy for species abundance whereby smaller species are expected to be more abundant (White, Ernest, Kerkhoff, & Enquist, 2007). As one of the most fundamental properties of an organism, body size is also thought to correlate with multiple aspects of trophic and other ecological and life history traits (Woodward et al., 2005). We therefore include body size in the multipredictor model in the interest of understanding what factors contribute to variation in genetic diversity. The distributions of snout–vent length (SVL) within species were often multimodal, reflecting contamination by several distinct age cohorts of lizards (e.g., first-year juveniles; subadults) during our sampling periods (Rabosky et al., 2007). We thus used kernel density estimation (KDE) to estimate adult body size. This method is nonparametric, as the shape of the estimated density function is determined by the data without assuming an underlying distribution, and has been used for a variety of ecological applications (Manly, 1996; Rabosky et al., 2011; Seaman & Powell, 1996). Following Rabosky et al. (2007), we took the upper mode of the empirical probability density function for SVL for each species as representative of the "typical" adult body size.

2.2 | Genomic data collection

Methods for genomic data collection are identical to those described in greater detail by Singhal, Grundler, Colli, and Rabosky (2017). Using the high-salt DNA extraction method (Aljanabi & Martinez, 1997), we collected high-molecular-weight DNA from one individual per species for 19 of the 30 species collected at Matuwa, and two individuals per species for 11 of the 30 species. Dual-barcoded libraries were produced for each sample by Arbor Biosciences. Arbor Biosciences also designed probes at 2 × tiling density across the 5,462 unique SqCL target loci identified by Singhal, Grundler, et al. (2017), including anchored hybrid enrichment (AHE) loci, ultraconserved elements (UCEs) and traditional genes used in squamate phylogenetics (other loci). We refer to this set of loci as the SqCL marker set (Singhal, Grundler, et al., 2017). Target capture reactions were performed on size-selected, amplified, and cleaned libraries following a modified MYbaits protocol described by Singhal, Grundler, et al. (2017), and sequenced by Hudson Alpha on one 100 paired-end run of a HiSeq 2,500 version 4.

2.3 | Data analysis

To obtain estimates of nucleotide diversity, raw sequencing reads were analysed following the bioinformatics pipeline provided for SqCL, available at https://github.com/singhal/SqCL along with explanatory documentation. Methods for the present study were modified to include error correction of cleaned reads using BLESS-EC2 before targets were matched to probes (Heo, Ramachandran, Hwu, Ma, & Chen, 2016). To perform read error correction, an estimated k-mer size was calculated using KMERGENIE with a default maximum length of 121, resulting in an optimal k-mer length of 31.

Within-population diversity was estimated by calculating the average pairwise difference (π) across all loci for each individual (Begun et al., 2007; Tajima, 1983). With greater than 5,000 loci, this is equivalent to estimating population diversity by sampling a few loci for many individuals (Harvey et al., 2017; Willing, Dreyer, & Oosterhout, 2012). For species that had two sampled individuals, we calculated π for each individual and then averaged the two measurements. All references to nucleotide diversity below refer to the average nucleotide diversity within a single individual, averaged across individuals within species for the 11 species for which multiple individuals were sampled. Additionally, we calculated the Pearson correlation between estimates of nucleotide diversity for individuals from the same species and used ANOVA to estimate the variance explained within and between species. For each individual, nucleotide diversity was also calculated for each locus in order to produce bootstrapped estimates of genetic diversity with variation in the number of loci sampled.

To test the relationship between nucleotide diversity and ecological predictors, we performed a pairwise correlation analysis as well as phylogenetically informed model selection, using the phylogeny from Tonini, Beard, Ferreira, Jetz, and Pyron (2016) (for this and all subsequent phylogenetic analyses), to estimate the importance of each predictor variable. We first computed the pairwise correlation between individual-level nucleotide diversity and the following ecological attributes for each species, using an expanded version of the Rabosky et al. (2011) data set: species abundance, computed as the total number of individuals captured during the 7-year survey period; and species occupancy, computed as the total number of sites where a species was detected. We note that our estimates of abundance are not based on an explicit capture–mark–recapture study, and we view the total number of individuals captured per species as a proxy for true total abundance (see Rabosky et al., 2011, for discussion). Because all sites were sampled for an identical number of days, we also note that all results reported below will be identical regardless of whether we analyse total abundance, mean annual abundance, or relative (fractional) abundance. We also included estimates of SVL and the proxy for species habitat specialization described above. We report *p*-values as the significance of each variable after accounting for phylogeny, calculated from phylogenetic linear models including only the predictor and genetic diversity, using Pagel's λ model for the covariance (Ho et al., 2018). Because our primary hypothesis is that genetic diversity should be positively correlated with species abundance, we did not correct for multiple comparisons despite statistical tests including additional predictors. These analyses were conducted to facilitate interpretation of data concerning our primary hypothesis, and thus we do not believe multiple comparison correction to be appropriate for our study design.

To determine the relative importance of each predictor, we constructed phylogenetic generalized least squares (PGLS) models for the full model (four predictors, not including interactions) and for each possible submodel, and we computed Akaike's information criterion (AIC) weights using the full set of fitted models. Because ordinary least squares models assume unequal variance in error among dependent and independent variables, we confirmed homoscedasticity of the residuals of each submodel using a Breusch–Pagan test with a significance level of $\alpha = 0.05$. These tests verified the consistency of standard errors in all models; however, we note that error in the independent variables would bias slopes toward zero, thus making these models a conservative approach.

We then calculated the relative importance of each variable by summing the AIC weights of the models in which the variable appears and dividing this by the sum of the AIC weights of all models (Anderson & Burnham, 2002; Kisel & Barraclough, 2010). We constructed a final model including all variables with a relative importance greater than 0.6 to calculate coefficients and values of significance for the regression (Wagner, Harmon, & Seehausen, 2012).

Finally, to quantify the effectiveness of genetic diversity as a proxy for species abundance and to facilitate comparisons with other methods, we computed the relative root mean squared error (rRMSE) of species abundance estimates predicted from heterozygosity, using the following equation from Yin and He (2014):

$$\mathrm{rRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - o_i}{o_i}\right)^2}$$

where x_i is the predicted log abundance for species i based on a linear model with heterozygosity and empirical estimates of species log abundance; o_i is the observed abundance of species i; and n is the total number of species sampled. We additionally calculated Pearson's product-moment correlation and R^2 values to quantify the significance of the correlation between observed and predicted abundance estimates, and compare these to alternative statistical models reported in Yin and He (2014) for predicting abundance.

As a second approach to evaluating the strength of genetic diversity as an abundance proxy, we performed a sliding-window resampling analysis to determine how many days of standardized community sampling are required to predict overall species abundance with the same correlation obtained from genetic diversity. In other words, how many consecutive days of sampling would have been required, on average, to recover a similar correlation between total abundance (across the full survey period) as that which we obtained from genetic data alone? We regenerated abundance and occupancy estimates beginning with a single day of sampling drawn from every unique date in the sampling period, increasing the window 1 day at a time from each starting point and averaging the results across dates within each window. For example, the mean correlation obtained for a window size of 5 days corresponds to the expected correlation between overall (multiyear) abundance and a much shorter subsurvey of just five sequential survey days.

3 | RESULTS

We recovered an average of 4,728 UCEs, 309 AHEs, and 27 additional loci traditionally used in squamate phylogenetics for each

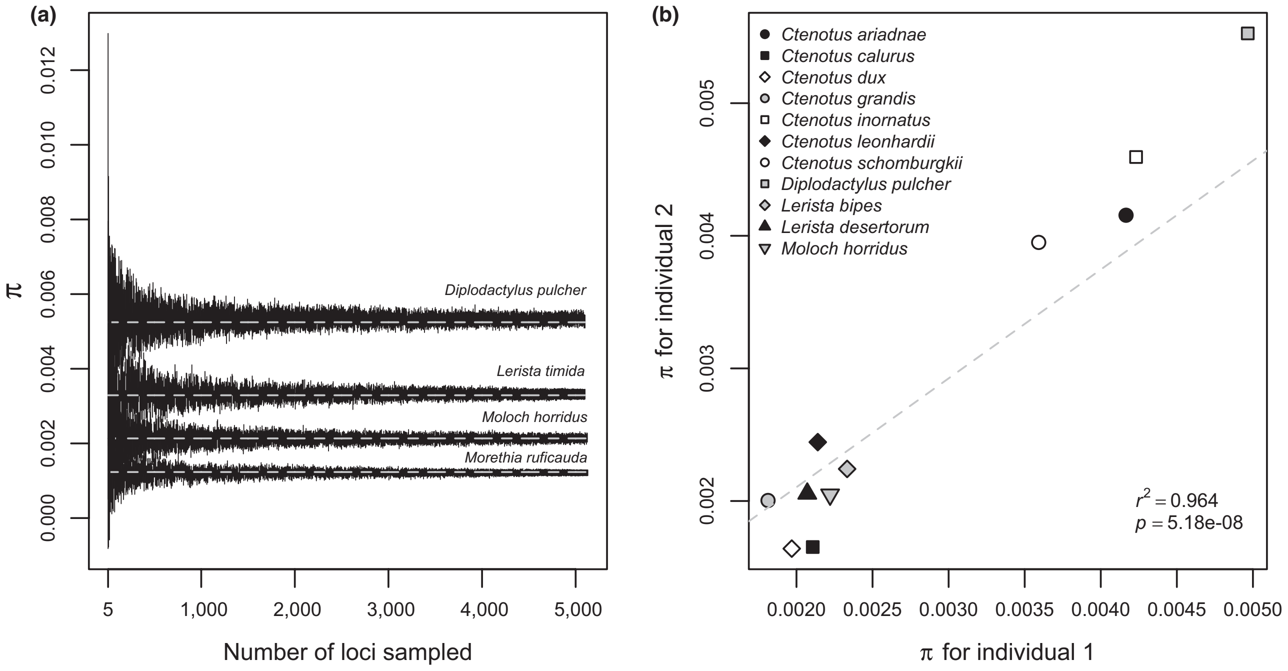

FIGURE 1 (a) Bootstrapped estimates of heterozygosity across four species of Matuwa lizards, as a function of the number of loci used to compute the estimate. Genome-wide heterozygosity converges rapidly to a limiting value and can be robustly estimated from samples of 500 or more loci. Species illustrated are *Diplodactylus pulcher*, *Lerista timida*, *Moloch horridus*, and *Morethia ruficauda*. Also drawn are the 95% confidence intervals around the estimated mean heterozygosity for each species, shown as a grey dotted line, as measured across all loci. (b) Genome-wide estimates of heterozygosity are highly correlated across conspecifics. Each point represents the pairwise relationship between heterozygosity estimates from two individual lizards of a given species. Dotted line illustrates isometric scaling relationship

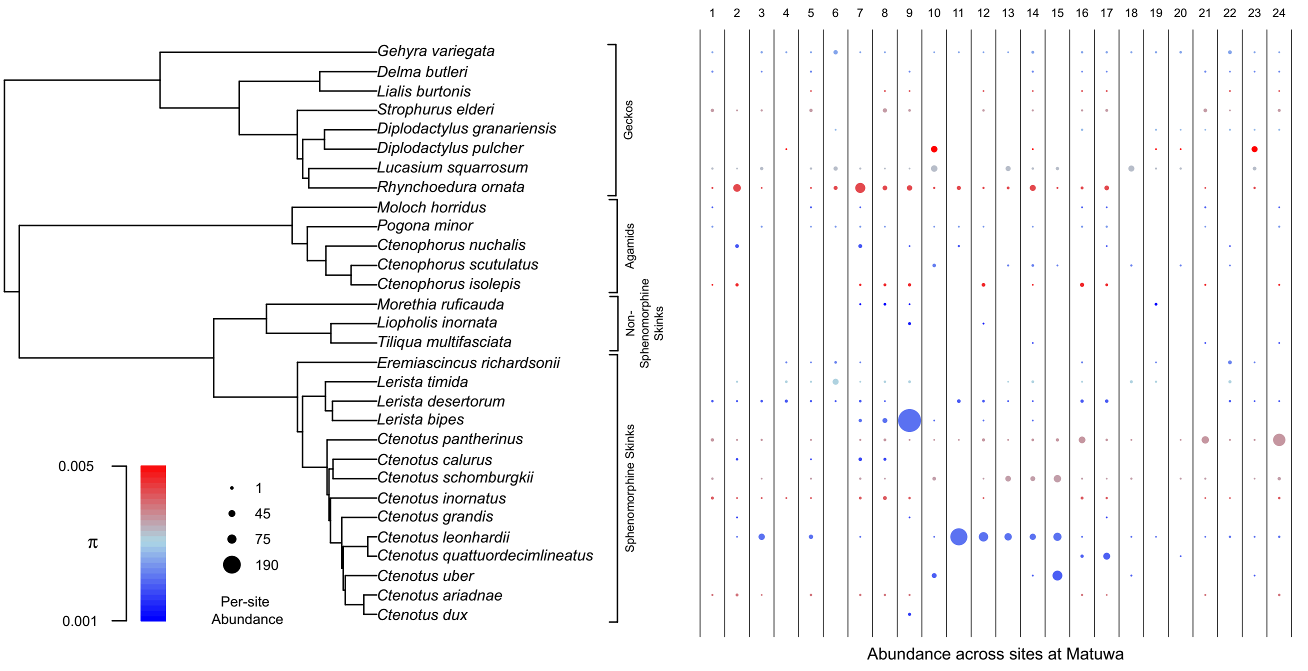

FIGURE 2 Phylogenetic distribution of occupancy, abundance, and heterozygosity across 30 species of lizards from Matuwa. Abundance at each of 24 sites is represented by the size of the point, and within-species genetic diversity is represented by colour. Occupancy is simply the number of sites at which a species was detected (e.g., empty “cells” denote absence). Corresponding abundance data can be found in Supporting Information Table S1

individual, resulting in an average of 5,066 loci per individual with a total of 2,946 loci in common between all individuals. We recovered an average of 3.6 Mb for each individual, with an average of 3.16 Mb ≥ 10× coverage. Measuring diversity from thousands of loci provides more reliable estimates of individual-level heterozygosity among species, as demonstrated by wider variability in bootstrapped estimates of diversity from fewer loci (Figure 1a). Eleven of 30 species were represented by two individuals, and nucleotide diversity between conspecifics was highly correlated (Pearson's $r = 0.984$; $p = 5.178 \times 10^{-8}$; Figure 1b). Using ANOVA on this set of individuals, we estimate that intraspecific variation accounts for ~1.53% of the total variation in individual-level nucleotide diversity across our data set. These results suggest that, with genome-wide sampling, even single individuals contain sufficient information to estimate “average” levels of within-population genetic variation.

Both log abundance and occupancy exhibited a positive and significant correlation with genetic diversity ($r = 0.43$, $p = 0.02$; $r = 0.50$, $p = 0.005$, respectively; correlation coefficients are not phylogenetically corrected) (Figures 2 and 3, Supporting Information Table S1). In line with Lewontin's paradox, our estimates of abundance varied across two orders of magnitude while genetic diversity varied across less than one. Abundance and occupancy are also highly correlated ($r = 0.70$, $p = 1.5 \times 10^{-5}$; Figure 3). There was no relationship between genetic diversity and log SVL ($r = -0.10$, $p = 0.60$), but log SVL was negatively and significantly correlated with abundance ($r = -0.49$, $p = 0.006$; Figure 3).

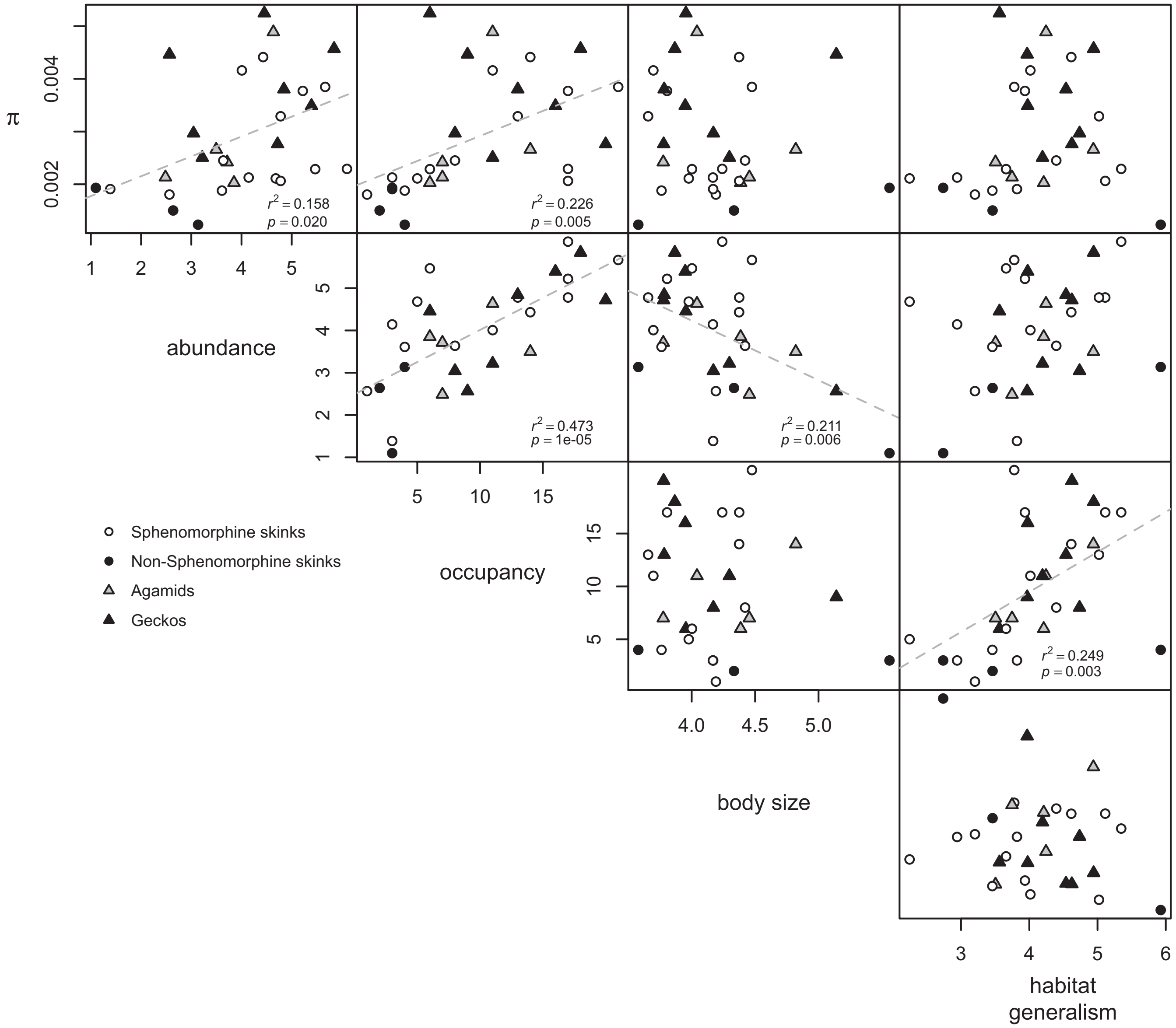

FIGURE 3 Correlation matrix showing relationships between genetic diversity and four predictor variables: log abundance, occupancy, body size as measured by snout–vent length, and a metric of habitat specialization (habitat distance) wherein larger values indicate generalist species and vice versa. Regression lines and phylogenetic p-values are shown for significant relationships, and points are coloured in greyscale by clade

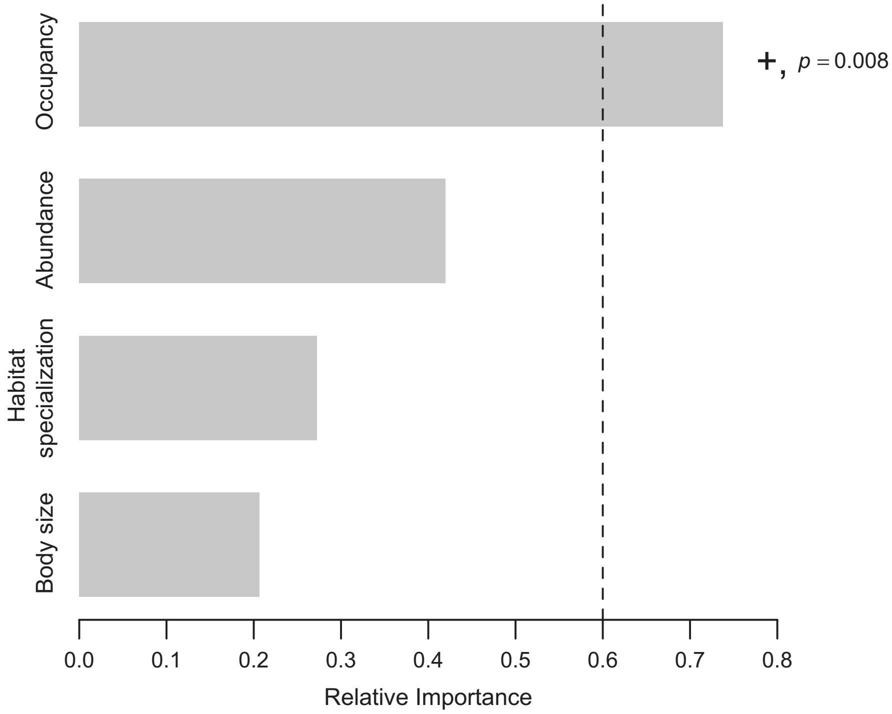

FIGURE 4 Results from model averaging, showing the relative importance of each independent variable for predicting within-species genetic diversity. Also shown are the coefficient sign and significance from the final phylogenetic linear model including occupancy, the only variable to exceed the cutoff value of cutoff value of relative importance ≥ 0.6

While there was a significant and positive relationship between occupancy and habitat specialization as approximated by the habitat distance metric ($r = 0.52$, $p = 0.003$), there was no correlation between genetic diversity and habitat distance ($r = 0.13$, $p = 0.500$; Figure 3).

The predictor variable of most importance after model averaging and the only variable to exceed the cutoff of 0.6 was occupancy, with a relative importance of 0.74 (Figure 4). Using only occupancy as a predictor in our final model results in $p = 0.008$ with coefficient = 8.7×10^{-5}. This small coefficient probably reflects the difference in scale between heterozygosity and occupancy, rather than a minor effect size; scaling genetic diversity to the same order of magnitude as the occupancy data results in a corresponding increase in the response of genetic diversity to changes in occupancy (in other words, an increase in the regression coefficient; coefficient after scaling genetic diversity = 0.870).

The rRMSE calculated for the prediction of species abundance based on heterozygosity was 0.55, with an $r = 0.4$ and $p = 0.02$. These values are similar to those reported in Yin and He (2014), but suggest a reduced predictive power for genetic diversity compared to statistical models based on occupancy and spatial distribution. However, the data used for validation of the models summarized in Yin and He (2014) are from an area of 1 km^2, whereas the current study is still able to recover a significant correlation between observed and predicted abundance estimates from a study region of 2,350 km^2 (maximum distance between sites = 38.4 km). Nonetheless, a sliding-window resampling analysis of abundance data demonstrates that a single day of sampling can generate a stronger correlation with long-term abundance than genetic diversity (Figure 5a). Moreover, only 3 days of sampling are required to produce a stronger correlation with occupancy (Figure 5b).

4 | DISCUSSION

We recovered a significant positive relationship between genetic diversity and empirical estimates of species abundance and occupancy, with occupancy being the most significant predictor of nucleotide diversity. While occupancy is correlated with patterns of habitat use among species, we find that direct measures of habitat specialization were not significant predictors of genetic diversity. Importantly, our results suggest that genetic diversity is at least a weak proxy for species abundance in the regional community. Additionally, we show a strong, positive relationship between occupancy and abundance, lending further support to a classic ecological relationship recovered for many taxa but rarely examined in squamates (Freckleton, Noble, & Webb, 2005; Gaston et al., 2002; Gaston, Blackburn, & Lawton, 1997). The pervasiveness of this relationship in macroecology suggests a mechanism linking changes in population dynamics at local and regional scales (Freckleton et al., 2005).

Our study reveals that a species-rich community of related taxa displays the same positive diversity–abundance association found by other studies at both narrower and broader phylogenetic and geographical scales. The proportion of variance in genetic diversity that was explained by local-scale abundance in the present study is similar to that explained by museum occurrence records (a proxy for global population size) in a recent study of genetic diversity in the lizard genus *Ctenotus*, a member of the Sphenomorphine clade that is represented in the current data set (Singhal, Huang et al., 2017). Similarly, Pearson correlations between abundance and diversity recovered from broader phylogenetic sampling and meta-analyses in other taxa are also comparable to the correlation recovered here ($r = 0.4$; Leimu et al., 2006; McCusker & Bentzen, 2010). It is useful

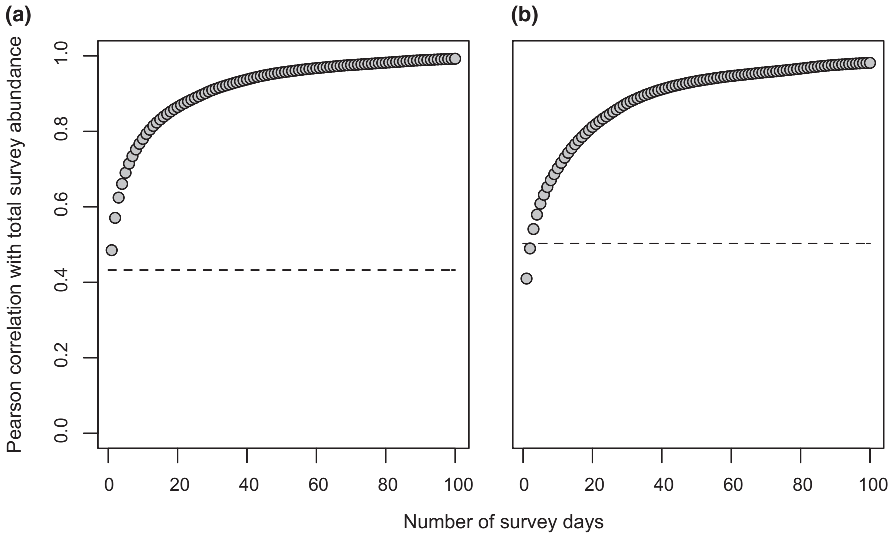

FIGURE 5 Expected correlation in abundance (a) and occupancy (b) between cumulative survey totals and a specified survey duration (*x*-axis). For example, a value of $x = 20$ gives the expected correlation between two estimates of the species abundance distributions: the cumulative (133-day) survey totals, versus a shorter interval of 20 consecutive survey days. A value of $x = 1$ indicates the extent to which a single day of sampling would have estimated cumulative abundance and occupancy totals. The observed correlation of genetic diversity with abundance and occupancy is illustrated by the corresponding horizontal dotted lines in each panel. Genetic diversity predicts overall species abundance with a correlation slightly below that obtained from 1 day of sampling, and it predicts species occupancy with a correlation slightly below that obtained from 3 days of sampling

to discover that genetic diversity retains at least some predictive power for population size across these different scales.

However, much of the variation in genetic diversity among species is unexplained by abundance and occupancy. With respect to predicting total abundance (e.g., pooled across all survey years), we found that genetic variation was equivalent to just a single "average" day of sampling at Matuwa (Figure 5). These results suggest that nucleotide diversity is heavily influenced by factors other than census population size, or that ecological sampling in this community is unable to capture true population size. Interpreting results from ecological analyses may help to clarify this issue. For example, habitat specialists may be locally abundant given nonrandom site selection, with low heterozygosity indicative of lower regional abundance. Although the most abundant species in this community also tend to be the most widespread, some high- and midabundance species of the sphenomorphine clade exhibit relatively low occupancy, in addition to a higher degree of habitat specialization. Assuming occupancy at the chosen sites is to some degree correlated with population connectivity, these observations suggest that gene flow could be mediated by habitat preference, restricting individuals to certain microhabitats and preventing movement through suboptimal patches across large areas of landscape (Wang & Bradburd, 2014). Concordantly, Pianka (2014) found that many of the most abundant species, as measured over 42 years in broadly the same region as the current study, are also the most restricted in dietary and microhabitat niche breadth. However, because occupancy at Matuwa is strongly correlated with both abundance and habitat generalism, the wide range of heterozygosity values exhibited by species at the upper range of occupancy suggests that other processes not evaluated by this study may constrain heterozygosity in these taxa. For example, the relationship between abundance and genetic diversity can be weakened by past demographic processes such as bottlenecks or range expansions not reflected by current population size (Banks et al., 2013; Dalongeville et al., 2016; Excoffier & Ray, 2008).

Similarly, because the chosen sample sites are not strictly a random draw from the landscape, there is no guarantee that ecological sampling at the community level is sufficient to accurately track true variation in population size. It may therefore be the case that occupancy is a more reliable estimate of long-term abundance than the pooled abundance measurements reported here. Population size will fluctuate through time, and abundance data are likely to be susceptible to noise generated by detection bias or recapture of individuals. While this is also true for occupancy data, occupancy provides a more coarse-grained metric that could be more sensitive to increases in relative species abundance or to the number of subpopulations in the region. Occupancy in a regional and community context may thereby provide a more robust proxy for true population size relative to other species in a comparative analysis.

5 | CONCLUSION

The results of this study support our initial prediction that greater abundance would correlate with higher levels of genetic diversity, and further suggest that factors associated with landscape occupancy contribute to observed variation in heterozygosity. Our study provides evidence that genetic diversity is at least weakly informative regarding demographic processes occurring at phylogenetic scales broader than conspecific populations. However, genetic diversity does no better at predicting species abundance, as measured in the community at Matuwa, than a single "average" day of standardized sampling, and does only marginally better at predicting occupancy. Whether these results reflect a failure of genetic diversity or ecological sampling to adequately capture variation in true census population size remains unknown. In conclusion, patterns of intraspecific genetic diversity potentially can serve as proxies for regional-scale species abundance, but further evidence from diverse communities is needed, ideally incorporating information on species abundance through time (Magurran, 2007). More broadly, our study provides further evidence for Lewontin's paradox and suggests that resolving the paradox may require further understanding and consideration of the relationship between historical demography and present-day census size.

ACKNOWLEDGEMENTS

We thank C. Dardia, A. Davis Rabosky, P. Doughty, I. Kealley, G. Schneider, D. Stefoni, P. Title and the staff of the Kalgoorlie office of the Department of Biodiversity, Conservation and Attractions for logistical assistance associated with fieldwork and sample acquisition. We gratefully acknowledge the Martu, the Traditional Owners of Matuwa, for access to their lands and for supporting our fieldwork. For technical advice and logistical support we thank Allison Devault and Jake Enk from MYcroarray, R. Marchan Rivadneira and R. Murrell. This work was supported by NSF grants OSIE-0612855, DEB-0814277 and DEB-1754398, and by a fellowship to D.L.R. from the David and Lucile Packard Foundation. S.S. was supported by NSF DEB-1519732. Field sampling at Matuwa was conducted under permit SF0004654 to D.L.R. (Western Australian Department of Parks and Wildlife).

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

AUTHOR CONTRIBUTIONS

M.R.G. and D.L.R. designed the study, analysed data and wrote the paper. M.R.G. and S.S. conducted lab work. S.S. contributed analytical tools. M.A.C. and D.L.R. conducted fieldwork.

DATA ACCESSIBILITY

Raw reads are associated with BioProject PRJNA518083, and additional data are available in a dryad digital repository (https://doi.org/10.5061/dryad.3b1rj70).

ORCID

Maggie R. Grundler https://orcid.org/0000-0002-7104-0494
Sonal Singhal https://orcid.org/0000-0001-5407-5567

REFERENCES

Aljanabi, S. M., & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, *25*(22), 4692–4693. https://doi.org/10.1093/nar/25.22.4692

Anderson, D. R., & Burnham, K. P. (2002). Avoiding pitfalls when using information-theoretic methods. *The Journal of Wildlife Management*, *66*(3), 912–918. https://doi.org/10.2307/3803155

Banks, S. C., Cary, G. J., Smith, A. L., Davies, I. D., Driscoll, D. A., Gill, A. M., … Peakall, R. (2013). How does ecological disturbance influence genetic diversity? *Trends in Ecology & Evolution*, *28*(11), 670–679.

Bazin, E., Glémin, S., & Galtier, N. (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science*, *312*(5773), 570–572. https://doi.org/10.1126/science.1122033

Begun, D. J., Holloway, A. K., Stevens, K., Hillier, L. W., Poh, Y.-P., Hahn, M. W., … Langley, C. H. (2007). Population genomics: Whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biology*, *5*(11), e310. https://doi.org/10.1371/journal.pbio.0050310

Brouat, C., Chevallier, H., Meusnier, S., Noblecourt, T., & Rasplus, J. Y. (2004). Specialization and habitat: Spatial and environmental effects on abundance and genetic diversity of forest generalist and specialist *Carabus* species. *Molecular Ecology*, *13*(7), 1815–1826. https://doi.org/10.1111/j.1365-294X.2004.02206.x

Brouat, C., Sennedot, F., Audiot, P., Leblois, R., & Rasplus, J. Y. (2003). Fine-scale genetic structure of two carabid species with contrasted levels of habitat specialization. *Molecular Ecology*, *12*(7), 1731–1745. https://doi.org/10.1046/j.1365-294X.2003.01861.x

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, *10*, 195–205. https://doi.org/10.1038/nrg2526

Cowan, M. A., Edinger, D., & Coate, K. (2017). Biodiversity in the southern rangelands: Variation in biota over time and space on the Black Range and Lake Mason stations, Murchison Bioregion, Western Australia. *Conservation Science Western Australia*, *12*(1), 1–40.

Dalongeville, A., Andrello, M., Mouillot, D., Albouy, C., & Manel, S. (2016). Ecological traits shape genetic diversity patterns across the Mediterranean Sea: A quantitative review on fishes. *Journal of Biogeography*, *43*(4), 845–857. https://doi.org/10.1111/jbi.12669

Devillard, S., Santin-Janin, H., Say, L., & Pontier, D. (2011). Linking genetic diversity and temporal fluctuations in population abundance of the introduced feral cat (*Felis silvestris catus*) on the Kerguelen archipelago. *Molecular Ecology*, *20*(24), 5141–5153. https://doi.org/10.1111/j.1365-294X.2011.05329.x

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, *17*(7), 422–433. https://doi.org/10.1038/nrg.2016.58

Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution*, *23*(7), 347–351. https://doi.org/10.1016/j.tree.2008.04.004

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, *61*(5), 717–726. https://doi.org/10.1093/sysbio/sys004

Frankham, R. (1995). Effective population size/adult population size ratios in wildlife: A review. *Genetics Research*, *66*(2), 95–107. https://doi.org/10.1017/S0016672300034455

Freckleton, R. P., Noble, D., & Webb, T. J. (2005). Distributions of habitat suitability and the abundance–occupancy relationship. *The American Naturalist*, *167*(2), 260–275.

Gaston, K. J., Blackburn, T. M., Greenwood, J. J. D., Gregory, R. D., Quinn, R. M., & Lawton, J. H. (2002). Abundance–occupancy relationships. *Journal of Applied Ecology*, *37*(s1), 39–59.

Gaston, K. J., Blackburn, T. M., & Lawton, J. H. (1997). Interspecific abundance-range size relationships: An appraisal of mechanisms. *Journal of Animal Ecology*, *66*(4), 579–601. https://doi.org/10.2307/5951

Godt, M. J. W., Johnson, B. R., & Hamrick, J. L. (1996). Genetic diversity and population size in four rare southern Appalachian plant species. *Conservation Biology*, *10*(3), 796–805. https://doi.org/10.1046/j.1523-1739.1996.10030796.x

Gotelli, N. J., & Chao, A. (2013). Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. In S. A. Levin (Ed.), *The encyclopedia of biodiversity* (Vol. 5, 2nd ed, pp. 195–211). Waltham, MA: Academic Press.

Grundler, M. R., Pianka, E. R., Pelegrin, N., Cowan, M. A., & Rabosky, D. L. (2017). Stable isotope ecology of a hyper-diverse community of scincid lizards from arid Australia. *PLoS ONE*, *12*(2), e0172879. https://doi.org/10.1371/journal.pone.0172879

Hague, M. T. J., & Routman, E. J. (2016). Does population size affect genetic diversity? A test with sympatric lizard species. *Heredity*, *116*(1), 92–98. https://doi.org/10.1038/hdy.2015.76

Harvey, M. G., Aleixo, A., Ribas, C. C., & Brumfield, R. T. (2017). Habitat association predicts genetic diversity and population divergence in Amazonian birds. *The American Naturalist*, *190*(5), 631–648. https://doi.org/10.1086/693856

He, F., & Gaston, K. J. (2000). Estimating species abundance from occurrence. *The American Naturalist*, *156*(5), 553–559. https://doi.org/10.1086/303403

Heo, Y., Ramachandran, A., Hwu, W.-M., Ma, J., & Chen, D. (2016). BLESS 2: Accurate, memory-efficient and fast error correction method. *Bioinformatics*, *32*(15), 2369–2371. https://doi.org/10.1093/bioinformatics/btw146

Ho, L. S. T., Ane, C., Lachlan, R., Tarpinian, K., Feldman, R., Yu, Q., …Ho, M. L. S. T. (2018). *Package 'phylolm'*. Retrieved from https://cran.r-project.org/web/packages/phylolm/index.html

James, C. D., & Shine, R. (2000). Why are there so many coexisting species of lizards in Australian deserts? *Oecologia*, *125*(1), 127–141.

Khimoun, A., Eraud, C., Ollivier, A., Arnoux, E., Rocheteau, V., Bely, M., … Garnier, S. (2016). Habitat specialization predicts genetic response to fragmentation in tropical birds. *Molecular Ecology*, *25*(16), 3831–3844. https://doi.org/10.1111/mec.13733

Kisel, Y., & Barraclough, T. G. (2010). Speciation has a spatial scale that depends on levels of gene flow. *The American Naturalist*, *175*(3), 316–334. https://doi.org/10.1086/650369

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., … Przeworski, M. (2012). Revisiting an old riddle: What determines genetic diversity levels within species? *PLoS Biology*, *10*(9), e1001388.

Leimu, R., Mutikainen, P., Koricheva, J., & Fischer, M. (2006). How general are positive relationships between plant population size, fitness and genetic variation? *Journal of Ecology*, *94*(5), 942–952.

Lewontin, R. C. (1972). The apportionment of human diversity. In *Evolutionary biology* (6th edn, pp. 381–398). Boston, MA: Springer.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496. https://doi.org/10.1038/nature10231

Lozier, J. D. (2014). Revisiting comparisons of genetic diversity in stable and declining species: Assessing genome-wide polymorphism in North American bumble bees using RAD sequencing. *Molecular Ecology*, *23*(4), 788–801. https://doi.org/10.1111/mec.12636

Magurran, A. E. (2007). Species abundance distributions over time. *Ecology Letters*, *10*, 347–354. https://doi.org/10.1111/j.1461-0248.2007.01024.x

Magurran, A. E., Baillie, S. R., Buckland, S. T., Dick, J. M., Elston, D. A., Scott, E. M., … Watt, A. D. (2010). Long-term datasets in biodiversity research and monitoring: Assessing change in ecological communities through time. *Trends in Ecology & Evolution*, *25*(10), 574–582. https://doi.org/10.1016/j.tree.2010.06.016

Manly, B. F. J. (1996). Are there clumps in body-size distributions? *Ecology*, *77*(1), 81–86.

McCusker, M. R., & Bentzen, P. (2010). Positive relationships between genetic diversity and abundance in fishes. *Molecular Ecology*, *19*(22), 4852–4862. https://doi.org/10.1111/j.1365-294X.2010.04822.x

McGill, B. J., Etienne, R. S., Gray, J. S., Alonso, D., Anderson, M. J., Benecha, H. K., … White, E. P. (2007). Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, *10*(10), 995–1015. https://doi.org/10.1111/j.1461-0248.2007.01094.x

Mesquita, D. O., Costa, G. C., Colli, G. R., Costa, T. B., Shepard, D. B., Vitt, L. J., & Pianka, E. R. (2016). Life-history patterns of lizards of the world. *The American Naturalist*, *187*(6), 689–705. https://doi.org/10.1086/686055

Meyer, C. F. J., Aguiar, L. M. S., Aguirre, L. F., Baumgarten, J., Clarke, F. M., Cosson, J.-F., … Kalko, E. K. V. (2010). Long-term monitoring of tropical bats for anthropogenic impact assessment: Gauging the statistical power to detect population change. *Biological Conservation*, *143*(11), 2797–2807. https://doi.org/10.1016/j.biocon.2010.07.029

Morton, S. R., Gillam, M. W., Jones, K. R., & Fleming, M. R. (1988). Relative efficiency of different pit-trap systems for sampling reptiles in spinifex grasslands. *Wildlife Research*, *15*(5), 571–577.

Morton, S. R., & James, C. D. (1988). The diversity and abundance of lizards in arid Australia: A new hypothesis. *The American Naturalist*, *132*(2), 237–256. https://doi.org/10.1086/284847

Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. (2017). Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. *Molecular Ecology Resources*, *17*, 1136–1147. https://doi.org/10.1111/1755-0998.12654

Ortego, J., Aparicio, J. M., Cordero, P. J., & Calabuig, G. (2008). Individual genetic diversity correlates with the size and spatial isolation of natal colonies in a bird metapopulation. *Proceedings of the Royal Society of London B: Biological Sciences*, *275*(1646), 2039–2047.

Palstra, F. P., & Ruzzante, D. E. (2008). Genetic estimates of contemporary effective population size: What can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular Ecology*, *17*, 3428–3447.

Patton, J. L., & Yang, S. Y. (1977). Genetic variation in *Thomomys bottae* pocket gophers: Macrogeographic patterns. *Evolution*, *31*(4), 697–720.

Pearce, J., & Ferrier, S. (2001). The practical value of modelling relative abundance of species for regional conservation planning: A case study. *Biological Conservation*, *98*(1), 33–43. https://doi.org/10.1016/S0006-3207(00)00139-7

Pianka, E. R. (1972). Zoogeography and speciation of Australian desert lizards: An ecological perspective. *Copeia*, *8*, 127–145. https://doi.org/10.2307/1442789

Pianka, E. R. (1986). *Ecology and natural history of desert lizards: Analyses of the ecological niche and community structure*. Princeton, NJ: Princeton University Press.

Pianka, E. R. (2014). Rarity in Australian desert lizards. *Austral Ecology*, *39*(2), 214–224. https://doi.org/10.1111/aec.12061

Pinsky, M. L., & Palumbi, S. R. (2014). Meta-analysis reveals lower genetic diversity in overfished populations. *Molecular Ecology*, *23*(1), 29–39. https://doi.org/10.1111/mec.12509

Rabosky, D. L., Cowan, M. A., Talaba, A. L., & Lovette, I. J. (2011). Species interactions mediate phylogenetic community structure in a hyperdiverse lizard assemblage from arid Australia. *The American Naturalist*, *178*(5), 579–595. https://doi.org/10.1086/662162

Rabosky, D. L., Reid, J., Cowan, M. A., & Foulkes, J. (2007). Overdispersion of body size in Australian desert lizard communities at local scales only: No evidence for the Narcissus effect. *Oecologia*, *154*(3), 561–570. https://doi.org/10.1007/s00442-007-0849-1

Roll, U., Feldman, A., Novosolov, M., Allison, A., Bauer, A. M., Bernard, R., … Meiri, S. (2017). The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nature Ecology & Evolution*, *1*(11), 1677–1682. https://doi.org/10.1038/s41559-017-0332-2

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., … Galtier, N. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, *515*(7526), 261–263.

Schwartz, M. K., Luikart, G., & Waples, R. S. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, *22*(1), 25–33. https://doi.org/10.1016/j.tree.2006.08.009

Seaman, D. E., & Powell, R. A. (1996). An evaluation of the accuracy of kernel density estimators for home range analysis. *Ecology*, *77*(7), 2075–2085. https://doi.org/10.2307/2265701

Singhal, S., Grundler, M., Colli, G., & Rabosky, D. L. (2017). Squamate Conserved Loci (Sq CL): A unified set of conserved loci for phylogenomics and population genetics of squamate reptiles. *Molecular Ecology Resources*, *17*(6), e12–e24.

Singhal, S., Huang, H., Title, P. O., Donnellan, S. C., Holmes, I., & Rabosky, D. L. (2017). Genetic diversity is largely unpredictable but scales with museum occurrences in a species-rich clade of Australian lizards. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1854), 20162588.

Stangel, P. W., Lennartz, M. R., & Smith, M. H. (1992). Genetic variation and population structure of red-cockaded woodpeckers. *Conservation Biology*, *6*(2), 283–292. https://doi.org/10.1046/j.1523-1739.1992.620283.x

Sun, M. (1996). Effects of population size, mating system, and evolutionary origin on genetic diversity in *Spiranthes sinensis* and *S. hongkongensis*. *Conservation Biology*, *10*(3), 785–795. https://doi.org/10.1046/j.1523-1739.1996.10030785.x

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, *105*(2), 437–460.

Tallmon, D. A., Gregovich, D., Waples, R. S., Baker, C. S., Jackson, J., Taylor, B. L., … Schwartz, M. K. (2010). When are genetic methods useful for estimating contemporary abundance and detecting population trends? *Molecular Ecology Resources*, *10*(4), 684–692.

Tonini, J. F. R., Beard, K. H., Ferreira, R. B., Jetz, W., & Pyron, R. A. (2016). Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. *Biological Conservation*, *204*, 23–31. https://doi.org/10.1016/j.biocon.2016.03.039

VanDerWal, J., Shoo, L. P., Johnson, C. N., & Williams, S. E. (2009). Abundance and the environmental niche: Environmental suitability estimated from niche models predicts the upper limit of local abundance. *The American Naturalist*, *174*(2), 282–291. https://doi.org/10.1086/600087

Wagner, C. E., Harmon, L. J., & Seehausen, O. (2012). Ecological opportunity and sexual selection together predict adaptive radiation. *Nature*, *487*(7407), 366. https://doi.org/10.1038/nature11144

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662. https://doi.org/10.1111/mec.12938

White, E. P., Ernest, S. M., Kerkhoff, A. J., & Enquist, B. J. (2007). Relationships between body size and abundance in ecology. *Trends in Ecology & Evolution*, *22*(6), 323–330. https://doi.org/10.1016/j.tree.2007.03.007

Willing, E.-M., Dreyer, C., & Van Oosterhout, C. (2012). Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE*, *7*(8), e42649. https://doi.org/10.1371/journal.pone.0042649

Woodward, G., Ebenman, B., Emmerson, M., Montoya, J. M., Olesen, J. M., Valido, A., & Warren, P. H. (2005). Body size in ecological networks. *Trends in Ecology & Evolution*, *20*(7), 402–409. https://doi.org/10.1016/j.tree.2005.04.005

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97.

Yin, D., & He, F. (2014). A simple method for estimating species abundance from occurrence maps. *Methods in Ecology and Evolution*, *5*(4), 336–343. https://doi.org/10.1111/2041-210X.12159

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Grundler MR, Singhal S, Cowan MA, Rabosky DL. Is genomic diversity a useful proxy for census population size? Evidence from a species-rich community of desert lizards. *Mol Ecol*. 2019;00:1–11. https://doi.org/10.1111/mec.15042