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

Authors: Maggie R. Grundler1,2,*, Sonal Singhal1,3, Mark A. Cowan4, Daniel L. Rabosky1

1Museum of Zoology and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109
2Department of Environmental Science, Policy, & Management, University of California, Berkeley, Berkeley, CA 94720
3Department of Biology, CSU Dominguez Hills, Carson, CA 90747
4Department of Biodiversity, Conservation and Attractions, Kensington, WA, Australia 6151

* Corresponding Author
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 ask 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, species abundance, Lewontin’s paradox, target capture, squamates

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 and Gaston 2000; McGill et al. 2007). Relative abundance is typically estimated through repeated community sampling efforts, with study durations that frequently span multiple years (Magurran et al. 2010; Meyer et al. 2010; Gotelli and Chao 2013; Pianka 2014). An
obvious difficulty is that such "brute-force" sampling requires considerable time and funding in
order to produce reliable estimates (Pearce and Ferrier 2001; Yin and 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 and Gaston 2000;
Jeremy VanDerWal et al. 2009; Yin and 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 and Ferrier 2001; Yin
and He 2014).

Neutral theory predicts that genetic diversity should correlate with population census size
(Tallmon et al. 2010; Wright 1931; Leffler et al. 2012), 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 et al. 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 samples sizes as small as a single individual (Li and Durbin 2011; Nazareno et
al. 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 consequent decoupling
from contemporary census population size (Excoffier and Ray 2008; Charlesworth 2009; Banks

Previous studies have found positive relationships between heterozygosity and proxies
for species abundance, including population size estimated from calculations of density and
acreage (Patton and Yang 1977), extrapolations calculated from active social groups (Stangel et
al. 1992), categorical estimates ("large", "small") of population size (Godt et al. 1996; Hague and
Routman 2016), and museum occurrence records (Singhal et al. 2017b). 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 (Lozier 2014; Sun 1996; Ortego et al. 2008; Devillard et al. 2011), or on many
species sampled at a broad geographic scale (Bazin et al. 2006; Leimu et al. 2006; McCusker and
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 eco-evolutionary 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” (Lewontin 1972; Leffler et al. 2012). 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 (Palstra and Ruzzante 2008; Leffler et al. 2012).

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 and 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 et al. 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 (Romiguier et al. 2014; Ellegren and Galtier 2016; Singhal et al. 2017b). Additionally, there is some evidence that microhabitat preference can limit dispersal and reduce within-population genetic diversity (Pianka 1986; Brouat et al. 2003; Brouat et al. 2004; Rabosky et al. 2011; Pianka 2014; Dalongeville et al. 2016; Khimoun et al. 2016), 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 multi-year demographic study and is characterized by variation in abundance and other ecological traits (Pianka 1986; Rabosky et al. 2011; Pianka 2014; Grundler et al. 2017). 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 (1) greater species abundance will correlate with greater levels of
nucleotide diversity; and (2) that 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.

METHODS

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 – harbors the most diverse lizard communities on Earth
(Pianka 1972; Morton and James 1988; James and Shine 2000; 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 non-sphenomorphine
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 multi-year pitfall

This article is protected by copyright. All rights reserved
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 et al. (2007; 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 et al. 1988). They are superior to
alternative methods tested in this region (Cowan et al. 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, fourteen 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 et al. 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 non-parametric, 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; Seaman and Powell 1996; Rabosky et al. 2011). 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.

Genomic Data Collection

Methods for genomic data collection are identical to those described in greater detail by Singhal et al. (2017a). Using the high-salt DNA extraction method (Aljanabi and 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 (Ann Arbor, MI). Arbor Biosciences also designed probes at 2x tiling density across the 5,462 unique SqCL target loci identified by Singhal et al. (2017a), 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 et al. 2017a). Target capture reactions were performed on size-selected, amplified, and cleaned libraries following a modified MYbaits protocol described by Singhal et al. (2017a), and sequenced by Hudson Alpha on one 100 paired-end run of a HiSeq 2500 v4.

Data Analysis

To obtain estimates of nucleotide diversity, raw sequencing reads were analyzed 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 et al. 2016). To perform read error correction, an estimated
k-mer size was calculated using kmergenie with a default max length of 121, resulting in an optimal k-mer length of 31.

Within-population diversity was estimated by calculating the average pairwise difference (\(\pi\)) across all loci for each individual (Tajima 1983; Begun et al. 2007). With greater than 5000 loci, this is equivalent to estimating population diversity by sampling a few loci for many individuals (Willing et al. 2012; Harvey et al. 2017). For species that had two sampled individuals, we calculated \(\pi\) 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 where 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 et al. (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) dataset: species abundance, computed as the total number of individuals captured during the seven 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 analyze 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 \(\lambda\) 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 (4 predictors, not including interactions) and for each possible submodel, and we computed 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 and Burnham 2002; Kisel and 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 et al. 2012).

Finally, to quantify the effectiveness of genetic diversity as a proxy for species abundance and 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):

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

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 one 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 five days corresponds to the expected correlation between overall (multi-year) abundance and a much shorter sub-survey of just five sequential survey days.

RESULTS

We recovered an average of 4,728 UCEs, 309 AHEs, and 27 additional loci traditionally used in squamate phylogenetics for each 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 ≥ 10x 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 (Fig. 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 x 10^{-8}) (Fig. 1B). Using ANOVA on this set of individuals, we estimate that intraspecific variation accounts for approximately 1.53% of the total variation in individual-level nucleotide diversity across our dataset. 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) (Fig. 2,3, 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 x 10^{-7})
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 ($r = -0.49$, $p = 0.006$) (Fig. 3).

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 ($r = 0.13$, $p = 0.500$) (Fig. 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 (Fig. 4). Using only occupancy as a predictor in our final model results in $p = 0.008$ with coefficient $= 8.7 \times 10^{-5}$. This small coefficient likely reflects difference in scale between heterozygosity and occupancy, rather than 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 somewhat comparable 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 2350 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 (Fig. 5A). Moreover, only three days of sampling are required to produce a stronger correlation with occupancy (Fig. 5B).

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 (Gaston et al. 2002; Gaston et al. 1997; Freckleton et al. 2005). 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 geographic 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 dataset (Singhal et al. 2017b). 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 and Bentzen 2010). It is useful 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 (Fig. 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 non-random 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 mid-abundance 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 sub-optimal patches across large areas of landscape (Wang and 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 suggest 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 (Excoffier and Ray

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.

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 about 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

This article is protected by copyright. All rights reserved
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, DEB-1754398, and by a fellowship to DLR from the David and Lucile Packard Foundation. SS was supported by a NSF DEB-1519732.

Field sampling at Matuwa was conducted under permit SF0004654 to DLR (Western Australian Department of Parks and Wildlife).

DATA ACCESSIBILITY

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

AUTHOR CONTRIBUTIONS

MRG and DLR designed the study, analyzed data, and wrote the paper. MRG and SS conducted lab work. SS contributed analytical tools. MAC and DLR conducted fieldwork.
CONFLICT OF INTEREST
The authors have no conflicts of interest to declare.

FIGURE LEGENDS
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 include 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 color. 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 Table S1.

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 colored 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 0.6 relative importance.
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 one day of sampling, and it predicts species occupancy with a correlation slightly below that obtained from three days of sampling.

Table S1. Abundance data for each species across sites at Matuwa, corresponding to Figure 2.

REFERENCES


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved


The diagram shows the relationship between the number of loci sampled and the diversity index $\pi$ for different species. The species include Diplodactylus pulcher, Lerista timida, Moloch horridus, and Morethia ruficauda. The $\pi$ values are plotted against the number of loci sampled, with a clear trend indicating an increase in diversity as the number of loci increases.

The scatter plot on the right shows the pairwise diversity index $\pi$ for individual 1 and individual 2 for the species Ctenotus ariadnae, Ctenotus calurus, Ctenotus dux, Ctenotus grandis, Ctenotus inornatus, Ctenotus leonhardii, Ctenotus schomburgkii, Ctenotus schomburgkii, Diplodactylus pulcher, Lerista bipes, Lerista desertorum, and Moloch horridus.

The correlation coefficient $r^2$ is 0.964, and the p-value $p = 5.18 \times 10^{-08}$, indicating a strong positive correlation between the two variables.
Moloch horridus
Delma butleri
Lialis burtonis
Strophurus elderi
Diplodactylus granariensis
Diplodactylus pulcher
Lucasium squarrosum
Rhynchoedura ornata
Moloch horridus
Pogona minor
Ctenophorus nuchalis
Ctenophorus scutulatus
Ctenophorus isolepis
Morethia ruficauda
Liopholis inornata
Tiliqua multifasciata
Eremiascincus richardsonii
Lerista timida
Lerista desertorum
Lerista bipes
Ctenotus pantherinus
Ctenotus calurus
Ctenotus schomburgkii
Ctenotus inornatus
Ctenotus grandis
Ctenotus leonhardii
Ctenotus quattuordecimlineatus
Ctenotus uber
Ctenotus ariadnae
Ctenotus dux

Abundance across sites at Matuwa
This article is protected by copyright. All rights reserved.