Converting quadratic entropy to diversity: both animals and alleles are diverse, but some are more diverse than others

Converting quadratic entropy to diversity: both animals and alleles are diverse, but some are more diverse than others

Peter E. Smouse [1¶], Sam C. Banks [2¶*], Rod Peakall [3¶]

[1] Department of Ecology, Evolution & Natural Resources, Rutgers University, New Brunswick, New Jersey, United States of America

[2] The Fenner School of Environment and Society, The Australian National University, Acton, ACT, Australia

[3] Research School of Biology, The Australian National University, Acton, ACT, Australia

Short title: Some are more diverse than others

[*]Corresponding author

Email: Sam.Banks@anu.edu.au

[¶] All authors contributed equally to this work

Abstract

The use of diversity metrics has a long history in population ecology, while population genetic work has been dominated by variance-derived metrics instead, a technical gap that has slowed cross-communication between the fields. Interestingly, Rao's Quadratic Entropy (RQE), comparing elements for 'degrees of divergence', was originally developed for population ecology, but has recently been deployed for evolutionary studies. We here translate RQE into a continuous diversity analogue, and then construct a multiply nested diversity partition for alleles, individuals, populations, and species, each component of which exhibits the behavior of proper diversity metrics, and then translate these components into [0,1] - scaled form. We also deploy non-parametric statistical tests of the among-stratum components and novel tests of the homogeneity of within-stratum diversity components at any hierarchical level. We then illustrate this new analysis with eight nSSR loci and a pair of close Australian marsupial (*Antechinus*) congeners, using both 'different is different' and 'degree of difference' distance metrics. The total diversity in the collection is larger than that within either species, but most of the within-species diversity is resident within single populations. The combined *A. agilis* collection exhibits more diversity than does the combined *A. stuartii* collection, possibly attributable to localized differences in either local ecological disturbance regimes or differential levels of population isolation. Beyond exhibiting different allelic compositions, the two congeners are becoming more divergent for the arrays of allele sizes they possess.

Key Words: Rao's quadratic entropy, Amova, Permanova, Gamova, Hill numbers, phylogenetic diversity, scaled diversity, *Antechinus*

Introduction

The use of genetic distance matrices to estimate genetic diversity within and among populations offers a number of benefits, including the ability to accommodate different genetic distance coding schemes, and computational tractability for large datasets. Here, we elaborate Rao's Quadratic Entropy to quantify and statistically evaluate patterns of genetic diversity, both within and among strata of a multiply nested taxonomic hierarchy, which can be used for diverse types of genetic data. This approach is related to, but exhibits a variety of innovations, relative to the traditional variance-based criteria commonly applied in population genetics. Quadratic $(q = 2)$ diversity metrics of several different types, originally developed for community ecology, have begun to infiltrate population genetic analysis, traditionally dominated by variance-based, least squares analyses [1-7]. To date, most such metrics have deployed 'different is different' coding of genetic markers, though sometimes measured along spanning networks [8-12], reflecting evolutionary separation. The use of 'degree of difference' as an evolutionary metric traces to early work [13-19], and has spawned recent efforts to elaborate diversity theory in that same vein [20-23].

In that context, Rao's Quadratic Entropy (henceforth *Q*) has drawn some attention [24-29], because conversion into inverted Gini-Simpson $1/(1 - Q)$ form yields a well-behaved diversity metric, provided that certain conditions are met [30-35]. Our object here is to elaborate *Q*, incorporating the 'degree of difference' between pairs of individual genets into a well-behaved diversity metric. We can translate a considerable array of paired-individual Euclidean distance matrices, as deployed for Amova [8, 36-38], Permanova [39-41], or Gamova [42], into *Q*, and can then convert *Q* into diversity analogue that may prove evolutionarily and/or ecologically informative.

Conversion of Q into well-behaved diversity metric is only possible if $[\, 0 \leq Q < 1 \,]$. There are three practical issues that must be dealt with in that translation. (1) Since quadratic genetic distance increases rapidly with the 'degree of difference', how are we to ensure that Q is properly bounded, given the wide array of quantitative divergence measures one could imagine for pairs of genets? (2) Can we estimate a well-behaved (and multiple level) partition of that total diversity, given the limited and typically unbalanced sampling routinely available from field studies? (3) Can we use this novel treatment for useful statistical evaluation of among-stratum diversification, as well as for evaluation of homo/heterogeneity of within-stratum diversity components? To illustrate both the formalisms and the utility of diversity translation, we employ a pair of Australian marsupial (*Antechinus*) congeners, sampled from contiguous Australian regions in New South Wales and Victoria, presenting evolutionary / geographic / environmental contrasts. We address a trio of additional questions: (4) How has evolutionary divergence within the complex been translated into genetic diversification within and between the two taxa? (5) Do responses to geographic or ecological challenges align with divergent patterns of diversity within the two organisms? (6) Do 'different is different' and 'degree of difference' treatments yield similar or disparate patterns of diversity within and between these close congeners?

Mathematical and computational methods

Rao's quadratic entropy

We start with Rao's quadratic entropy (henceforth Q), defined in terms of multi-locus genotypic arrays for the grand total collection of N diploid individuals. For any single genetic locus, we compute a squared allelic-pair distance as '0' (if identical) or '1' (if different) for all allelic pairs. Given a set of $(j, k = 1, \cdots, J)$ different alleles for that locus, we define Q for the total collection of N individuals ($2N$ alleles) as:

$$Q = \sum_j \sum_k p_j \cdot p_k \cdot d_{jk}^2 \,, \qquad [1]$$

where the relative frequencies of the j^{th} and k^{th} alleles are p_j andp_k. Eq. [1] can be rewritten in matrix form, using a matrix (**D**) of squared distances between all $(4N^2)$ allelic-pairs, and as a vector, $\boldsymbol{P}' = [\,(1/2N), \cdots, (1/2N)\,]$, one entry per sampled allele

$$Q = \boldsymbol{P}'\mathbf{D}\boldsymbol{P} = (sum\mathbf{D}/4N^2) \,, \qquad [2]$$

where $sum\mathbf{D}$ is the total of $(4N^2)$ squared distances within matrix **D**, and *Q* is the average.

Many current genetic surveys deploy microsatellite or simple sequence repeat (SSR) markers, for which we routinely use 'different is different' coding, $\left(d_{jk}^2 = 0,\ \text{same}\right)$ or $\left(d_{jk}^2 = 1, \text{different}\right)$, to construct a multi-allelic distance matrix $\mathbf{D}^{\mathrm{C}}$ [8, 36-38]. For SSR markers, we can also translate differences in numbers of repeat units into a 'degree of difference' distance metric [43-45]. If the j^{th} and k^{th} alleles have r_j and r_k repeat units, respectively, then $d_{jk}^2 = \left(r_j - r_k\right)^2$, we can pack those squared repeat-number distances into a $\mathbf{D}^{\mathrm{R}}$ matrix, connected by the assumption of single step mutation (SSM) models of evolution [14, 18-19]. We can use any quadratic Euclidean distance metric that makes sense for the problem at hand [46].

Translating quadratic entropy into diversity

Provided that $[\,0 \leq Q < 1\,]$ for the total collection of all individuals, we can convert *Q* into a measure of total diversity within the collection, using inverted Gini-Simpson translation; for the *N* individuals (2*N* alleles), that translation takes the simple form

$$\gamma = 1/(1 - Q) \,, \qquad [3]$$

where (γ) estimates the 'effective number' of equally frequent and equally different alleles within the collection. For any single locus of the 'different is different' $\mathbf{D}^{\mathrm{C}}$ matrix, *Q* is properly bounded by virtue of its (0 vs 1) construction. For the 'degree of difference' $\mathbf{D}^{\mathrm{R}}$ matrix, Eq. [1]

employs the squared difference in repeat units for the two alleles in question. Many individual elements in the matrix exceed '1', and it is customary [31] to scale the pairwise distances to ensure that [$0 \leq Q < 1$]. If the smallest allele for our single SSR locus has a count of (r_{min}) repeat units and the largest has (r_{max}) repeat units, each element of the single-locus $\mathbf{D}^{\mathrm{R}}$ matrix should be divided by the maximum squared distance for that locus, $d^2_{max} = (r_{max} - r_{min})^2$, thus ensuring that [$0 \leq Q < 1$] . To complete our scaling, we sum the squared distance values over loci, for either $\mathbf{D}^{\mathrm{C}}$ or $\mathbf{D}^{\mathrm{R}}$ loci, and then divide each multi-locus element by the number of loci (L) scored. The best scaling will depend on the genetic markers in question, of course, but by insuring that [$0 \leq Q < 1$], we ensure that $[1 \leq \gamma < 2N]$ for diploids. We assume that our **D** matrices have been appropriately scaled, from the outset. The essential point is that the diversity translations of different distance matrices may shed useful light on the ecological and/or evolutionary reality we have sampled.

The diversity partition

We are generally interested in a partition of that diversity across space, ecological context and/or taxonomic subdivision. To illustrate the partition of the total genetic diversity (γ) into separate within and among stratum levels of hierarchical sampling, consider a pair of congeneric species (S_A and S_B). Let the numbers of sampled alleles within the respective species be $(2N_{\mathrm{A}} = 6, 2N_{\mathrm{B}} = 4)$, and start with the example (five-allele array) illustrated in (Table 1). The average of all $(4N^2 = 100)$ elements in $(\mathbf{D}^{\mathrm{C}})$ is $Q = (sum\mathbf{D}^{\mathrm{C}}/4N^2) = (78/100) = 0.78$, from which Eq. [3] yields $(\gamma = 4.545)$ 'effective (equi-frequent, equi-different) alleles'. The two species are not equally replicated within our sample, and we need to account for that imbalance with our estimation protocols. We compute a separate *Q*-value within each of the species,

$Q_{\mathrm{WA}} = (22/36)$ and $Q_{\mathrm{WB}} = (10/16)$ and then convert the Q-values into separate diversity estimates within each of the species,

$$\sigma_{\mathrm{WA}} = 1/(1 - Q_{\mathrm{WA}}) = 2.571 \quad \text{and} \quad \sigma_{\mathrm{WB}} = 1/(1 - Q_{\mathrm{WB}}) = 2.667\,, \qquad [4]$$

within species A and B, respectively. We next compute a weighted average estimate of the within-species Q_{WS}-value; for the sample entries in (Table 1), the sample size weights are

$$f_{\mathrm{WA}} = 4N_{\mathrm{WA}}^2/(4N_{\mathrm{WA}}^2 + 4N_{\mathrm{WB}}^2) \qquad f_{\mathrm{WB}} = 4N_{\mathrm{WB}}^2/(4N_{\mathrm{WA}}^2 + 4N_{\mathrm{WB}}^2)\,, \qquad [5]$$

yielding weighted average within-species (Q_{WS}) and diversity (σ_{WS}) values of the form

$$Q_{\mathrm{WS}} = [f_{\mathrm{WA}} \cdot Q_{\mathrm{WA}} + f_{\mathrm{WB}} \cdot Q_{\mathrm{WB}}] = (8/13) \qquad \sigma_{\mathrm{WS}} = 1/(1 - Q_{\mathrm{WS}}) = 2.6\,. \qquad [6]$$

The extension to multiple, unequally sampled species, is obvious. We can also compute an among-species Q_{AS} value and (derivative) diversity (δ_{AS}) estimate, which is the 'effective (equi-frequent, equi-different) number' of species with no cross-species allelic sharing. To illustrate that extraction, in the context of (Table 1), we compute

$$\delta_{\mathrm{AS}} = (\gamma/\sigma_{\mathrm{WS}}) = 1/(1 - Q_{\mathrm{AS}}) = (1 - Q_{\mathrm{WS}})/(1 - Q) = 1.748\,, \qquad [7]$$

which we can back-translate into an equivalent 'among-species' (Q_{AS}) value,

$$Q_{\mathrm{AS}} = (Q - Q_{\mathrm{WS}})/(1 - Q_{\mathrm{WS}}) = (\delta_{\mathrm{AS}} - 1)/\delta_{\mathrm{AS}} = 0.428\,. \qquad [8]$$

By construction, $\gamma = (\delta_{\mathrm{AS}} \cdot \sigma_{\mathrm{WS}}) = (1.748) \cdot (2.6) = 4.545$ for the example (Table 1).

<u>Table 1</u>: Illustrative distance matrix (D^C) for two species: (A, three diploid individuals) and (B, two diploid individuals): five different alleles (g-1) · · · (g-5), with squared distances $(d_{jk}^2 = 0)$ if alleles are identical but $(d_{jk}^2 = 1)$ if different.

	Species A						Species B				
Allele	g-1	g-1	g-1	g-2	g-2	g-3	g-3	g-4	g-4	g-5	Allele
g-1	0	0	0	1	1	1	1	1	1	1	g-1
g-1	0	0	0	1	1	1	1	1	1	1	g-1
g-1	0	0	0	1	1	1	1	1	1	1	g-1
g-2	1	1	1	0	0	1	1	1	1	1	g-2
g-2	1	1	1	0	0	1	1	1	1	1	g-2
g-3	1	1	1	1	1	0	0	1	1	1	g-3
g-3	1	1	1	1	1	0	0	1	1	1	g-3
g-4	1	1	1	1	1	1	1	0	0	1	g-4
g-4	1	1	1	1	1	1	1	0	0	1	g-4
g-5	1	1	1	1	1	1	1	1	1	0	g-5
Allele	g-1	g-1	g-1	g-2	g-2	g-3	g-3	g-4	g-4	g-5	Allele
	Species A						Species B				

Similar treatment of the (sample frame determined) max-diversity dataset (Table 2) yields maximum achievable (Q^*) values and their derivative diversity maxima, using the same computational stream as used in Eqq. [3] - [8]. Both the observed (Table 1) *Q*-values and their sample-frame dependent (Table 2) maximum Q^*-values are presented for the exemplar in Table 3, where we also present a set of [0,1]-scaled diversity estimates, which we will elucidate below.

Table 2: Maximum distance matrix (D^C) for two species: (A, three diploid individuals) and (B, two diploid individuals, with ten different alleles (g-1) · · · (g-10); squared distances $(d_{jk}^2 = 0)$ if alleles are identical but $(d_{jk}^2 = 1)$ if different; analysis yields sample-frame dependent maximum *Q-values and their translations into maximum diversity estimates.**

	Species A						Species B				
Allele	g-1	g-2	g-3	g-4	g-5	g-6	g-7	g-8	g-9	g-10	Allele
g-1	0	1	1	1	1	1	1	1	1	1	g-1
g-2	1	0	1	1	1	1	1	1	1	1	g-2
g-3	1	1	0	1	1	1	1	1	1	1	g-3
g-4	1	1	1	0	1	1	1	1	1	1	g-4
g-5	1	1	1	1	0	1	1	1	1	1	g-5
g-6	1	1	1	1	1	0	1	1	1	1	g-6
g-7	1	1	1	1	1	1	0	1	1	1	g-7
g-8	1	1	1	1	1	1	1	0	1	1	g-8
g-9	1	1	1	1	1	1	1	1	0	1	g-9
g-10	1	1	1	1	1	1	1	1	1	0	g-10
Allele	g-1	g-2	g-3	g-4	g-5	g-6	g-7	g-8	g-9	g-10	Allele
	Species A						Species B				

Table 3: Observed *Q*-values for exemplar (Table 1) and maximum *Q*-values (Table 2), translated into observed, maximum, and [0,1]-scaled estimates for total diversity, a separate within-species estimate for each species, a weighted average within-species estimate, and an among-species estimate

Computed Criterion	Total Study	Among Species	AveWithin Species	Within Species A	Within Species B
Data Diversity Estimate	4.545	1.748	2.600	2.571	2.667
Data *Q*-Value	0.780	0.428	0.615	0.611	0.625
[0,1]- Scaled Diversity	0.867	0.892	0.762	0.733	0.833
Max Possible *Q*-Value	0.900	0.480	0.808	0.833	0.750
Max Diversity Possible	10.000	1.923	5.200	6.000	4.000

Extending the diversity partition downward

We typically sample multiple populations within each of our sample species, and for diploids (and also polyploids) we may also want to elaborate diversity within and among single individuals. For both purposes, we need to extend our diversity partition downward, so imagine six (6) populations, two (2) from species A and four (4) from species B. We construct a separate distance matrix within each population, extract their within-population (Q and Q^*) values, and convert those into separate diversity estimates for each population,

$$\alpha_{\mathrm{WP1}} = 1/(1 - Q_{\mathrm{WP1}}) \cdots \alpha_{\mathrm{WP6}} = 1/(1 - Q_{\mathrm{WP6}}) \quad [9]$$

Using the same strategy as for Eqq. [5] - [6], we compute a (within-population) average Q_{WP} value for two populations in Species A and the four populations in Species B, using quadratic sample-size weights of the form:

182 $$f_{\mathrm{P1}} = N_{\mathrm{P1}}^2/(N_{\mathrm{P1}}^2 + N_{\mathrm{P2}}^2) \qquad f_{\mathrm{P2}} = N_{\mathrm{P2}}^2/(N_{\mathrm{P1}}^2 + N_{\mathrm{P2}}^2)\ ,$$

183 and [10]

184 $$f_{\mathrm{P3}} = N_{\mathrm{P3}}^2/(N_{\mathrm{P3}}^2 + N_{\mathrm{P4}}^2 + N_{\mathrm{P5}}^2 + N_{\mathrm{P6}}^2) \cdots f_{\mathrm{P6}} = N_{\mathrm{P6}}^2/(N_{\mathrm{P3}}^2 + N_{\mathrm{P4}}^2 + N_{\mathrm{P5}}^2 + N_{P6}^2)\ .$$

185 We then translate those weighted-average (Q_{WP}) values into separate within-population diversity

186 estimates for Species A and B, respectively,

187 $$\alpha_{\mathrm{WPA}} = 1/(1 - Q_{\mathrm{WPA}}) \qquad \alpha_{\mathrm{WPB}} = 1/(1 - Q_{\mathrm{WPB}})\ .$$ [11]

188 If we average all six within-population *Q*-values (weighted by their respective quadratic sample

189 sizes), we obtain an average (Q_{WP}) for the whole study, and can translate that into a study-wide

190 average estimate of the within-population diversity,

191 $$\alpha_{\mathrm{WP}} = 1/(1 - Q_{\mathrm{WP}})\ .$$ [12]

192 We are not constrained to balanced sampling at any level, but we do need to account explicitly for

193 whatever sampling imbalance exists within the dataset.

194 For diploids, we can also extract estimates of within-individual (among-allele) diversity

195 for each locus, as each individual is represented by a (2 x 2) submatrix (Table 1 and Table 2); we

196 have *N* such sub-matrices for the study. We later illustrate this new analysis for a pair of strongly

197 outbred species [47], where subdivision of within-individual diversity is not very helpful. For

198 organisms showing non-random mating systems, however, extracting sub-individual diversity

199 components may prove valuable, and we describe their estimation in S1 Appendix.

200 Given that our populations are nested within species, we also need to compute traditional

201 among-population (Q_{AP}) and (Q_{AP}^*) values and to translate those into both estimated (β_{AP}) and

202 maximum possible (β_{AP}^*) diversity. By analogy with Eq. [7], we deploy

203 $$\beta_{\mathrm{AP}} = (\sigma_{\mathrm{WS}}/\alpha_{\mathrm{WP}}) = (1 - Q_{\mathrm{WP}})/(1 - Q_{\mathrm{WS}})\ ,$$ [13]

204 and noting that $\beta_{\mathrm{AP}} = 1/(1 - Q_{\mathrm{AP}})$, we back-translate Eq. [13] to extract

205 $$Q_{\mathrm{AP}} = (Q_{\mathrm{WS}} - Q_{\mathrm{WP}})/(1 - Q_{\mathrm{WP}})\ .$$ [14]

With similar definition and estimation of the within-individual diversity (ω_{WI}) and the among-individual diversity (ε_{AI}), both nested within single populations (S1 Appendix), we have now defined and elaborated (an RQE-derivative) diversity estimation cascade that elaborates the traditional three-level panoply into a multiplicative multi-level cascade,

$$\gamma = (\omega_{WI} \cdot \varepsilon_{AI} \cdot \beta_{AP} \cdot \delta_{AS}) \, . \qquad [15]$$

Diversity desiderata

Beyond sheer definition and estimation, well-behaved diversity components should exhibit a set of key features. The within-stratum components represent 'effective numbers' of (equi-frequent and equi-different) alleles within each level of the nested hierarchy, and these within-stratum estimates should satisfy the condition,

$$1 \leq \omega_{WI} \leq \alpha_{WP} \leq \sigma_{WS} \leq \gamma \leq 2N \, , \qquad [16]$$

which our estimates do. The among-stratum components represent 'effective numbers' of non-overlapping allelic collections for (equi-frequent and equi-different) sub-strata: among individuals of a single population, among populations within a single species, and among species of the total collection. These among-stratum components are explicitly defined so that

$$\alpha_{WP} = (\varepsilon_{AI}) \cdot (\omega_{WI}) \qquad \sigma_{WS} = (\beta_{AP}) \cdot (\alpha_{WP}) \qquad \gamma = (\delta_{AS}) \cdot (\sigma_{WS}) \, . \qquad [17]$$

Given strict nesting of the *Q*-values, and within the constraints of Eqq. [7-8, 13-14], all of the components are free to vary independently. Our within- and among stratum diversity estimates meet all of those conditions.

If we add genetic variety at any level, diversity must increase. Consider a single population (P_1), nested within a species (S_A). For $\mathbf{D}^C$ coding, if any existing allele (within that population) is replaced by a novel allele (for that population), the within-population diversity

228 (α_{WP1}) will increase. That also increases the within-species diversity (σ_{WSA}) of the species,
229 within which (P_1) is nested. If other populations within (S_A) show some genetic overlap but do
230 not have this novel variant, the among-populations diversity will also increase. For $\mathbf{D}^R$ coding,
231 'degree of difference' also matters, and if the novel variant in (P_1) is beyond the 'size range' of
232 previously represented alleles in (P_1), the internal diversity (α_{WP1}) of that population will
233 increase. If it is also beyond the size range of the species (S_A), within which it is nested, so will
234 be (σ_{WS}) and (β_{AP}), etc. Our estimation protocols ensure that all of our diversity estimates meet
235 the desiderata.

236 Maximum and [0,1] - scaled diversity

237 Without scaling, a minimum achievable diversity estimate at any level is '1' by
238 construction. If we were to compute the diversity cascade, achievable from a ($2N$ x $2N$) matrix
239 with every off-diagonal element being identically '1', our diversity components would attain
240 (sample-frame constrained) maximum values (Table 2). That would maximize all the *Q*-values
241 and their diversity translations. . It is usual to estimate and compare diversity metrics from
242 (modest and typically unbalanced) samples, and it is often useful to gauge those estimates,
243 relative to the minima and maxima achievable, given the sampling limitations. We will
244 henceforth denote the max-diversity distance matrix (and all of its derivate summations, *Q*-
245 values, and diversity transforms) with an ($*$). If all genets (at any level) are equally different and
246 represented once each, the maximum diversity values become the raw numbers of those
247 elements. The most diverse collection attainable has 2*N* equally different alleles. A pair of
248 equally sampled species, sharing no alleles in common, yields ($\delta^*_{AS} = 2$), while a trio of equally
249 sampled populations, sharing no alleles in common, yields ($\beta^*_{AP} = 3$), etc. With unbalanced

sampling, those maxima are reduced, but whether sampling is balanced or not, all diversity estimates are explicitly (sample frame) bounded, both above and below,

$$1 \le \gamma \le \gamma^* \qquad 1 \le \sigma_{WS} \le \sigma^*_{WS} \qquad 1 \le \alpha_{WP} \le \alpha^*_{WP} \qquad 1 \le \omega_{WI} \le \omega^*_{WI} ,$$

and [18]

$$1 \le \delta_{AS} \le \delta^*_{AS} \qquad 1 \le \beta_{AP} \le \beta^*_{AP} \qquad 1 \le \varepsilon_{AI} \le \varepsilon^*_{AI} .$$

We can scale an estimate of shared diversity among strata at any given level, ranging from 0 (no sharing) to 1 (complete sharing and identical frequencies of) all elements [20]. Starting from that criterion, we can define a complementary estimate of non-overlap, ranging from 0 (total sharing and identical frequencies of elements) to 1 (no sharing of elements). For the RQE-derivative diversity metrics above, that translation yields a remarkably convenient and easily computed set of [0,1] - scaled diversity estimates [48] (S2 Appendix),

$$\gamma^{\sim} = (Q/Q^*) \quad (\sigma^{\sim}_{WS}) = (Q_{WS}/Q^*_{WS}) \quad (\alpha^{\sim}_{WP}) = (Q_{WP}/Q^*_{WP}) \quad (\omega^{\sim}_{WI}) = (Q_{WI}/Q^*_{WI}) ,$$

and [19]

$$(\delta^{\sim}_{AS}) = (Q_{AS}/Q^*_{AS}) \qquad (\beta^{\sim}_{AP}) = (Q_{AP}/Q^*_{AP}) \qquad (\varepsilon^{\sim}_{AI}) = (Q_{AI}/Q^*_{AI}) .$$

Returnning to our example array (Table 1 and Table 2), the [0,1]-scaled diversity estimates (third line) in (Table 3), are obtained by computing the corresponding ratios of the data *Q*-estimates from the line just above and the (Q^*) maxima from the line just below. For Table 3, we compute $(\gamma^{\sim} = (0.780/0.900) = 0.867)$, and similarly, for the other estimates. Each element of Eq. [19] is thus explicitly [0,1]-scaled for the sampling frame itself. Such [0,1] - scaling provides a useful sense of 'how large or small the diversity is' at any given level, relative to 'how large or small it cold be', given the sampling frame. The translation is that '0' represents no genetic diversification of the elements under consideration, and that '1' represents maximum achievable diversification (no overlap), given the sampling frame. .

Statistical inference on diversity components

By recasting the RQE argument in distance matrix form, we can convert any (genetically sensible) Euclidean (positive semi-definite) inter-allelic distance matrix **D** for the *N* individuals (2*N* alleles per locus) into a nested cascade of estimated diversity components. Beyond estimation, we can (and should) assess the statistical credibility of whatever we estimate. The total diversity provides a system-wide baseline, but since we would not conduct the exercise in the absence of meaningful genetic diversity, a test of whether $(\gamma^{\sim}) > 0$ would be rather pointless. A more interesting set of questions would be whether there is credible diversity among species $(\delta_{\mathrm{AS}}^{\sim}) > 0$ or diversity among populations $(\beta_{\mathrm{AP}}^{\sim}) > 0$ within them, or even (in some cases) whether there is diversity among individuals $(\varepsilon_{\mathrm{AI}}^{\sim}) > 0$ within the same population. The traditional variance-derivative tests of inter-population divergence, such as (F_{ST}) and (G_{ST}), have been challenged as poorly bounded, but alternative criteria that are [0,1]-scaled and better-behaved have been offered. We show (S2 Appendix) that [0,1] - scaled among-population diversity $(\beta_{\mathrm{AP}}^{\sim})$ is an extension of Jost's (*D*) criterion [49-50] to the more general (unbalanced sampling) case. We extend that treatment upward to the among-species $(\delta_{\mathrm{AS}}^{\sim})$ and downward to the among-individual $(\varepsilon_{\mathrm{AI}}^{\sim})$ levels of scaled diversity estimates, both appropriately bounded and well behaved.

We might also find it useful to test whether: (a) the separate within-species diversity estimates $(\sigma_{\mathrm{WS}}^{\sim})$ are credibly homogeneous from species to species, or (b) whether the within-population diversity estimates $(\alpha_{\mathrm{WP}}^{\sim})$ are credibly homogeneous from population to population (within, or even among species), or even (c) whether the within-individual diversity estimates $(\omega_{\mathrm{WI}}^{\sim})$ are credibly homogeneous among individuals, within or among populations or species. We show (S3 Appendix) that a test of the hypothesis of homogeneous within species diversity

297 values is tantamount to a Bartlett's test [51] of homogeneity of the corresponding within-species
298 variances. The same equivalence applies at the within-population and within-individual levels.
299 Failure of any of our within-stratum homogeneity tests would provide signals of differential
300 demographic, ecological and/or evolutionary pressures that have shaped such diversity in
301 different fashions or to different degrees within different sampling strata.

302 Normal or multinomial statistical theory assumptions are too restrictive for the wide array
303 of data sets and contextual situations under real world consideration, so we deploy here a set of
304 non-parametric test criteria, with (locus by locus) permutation of alleles among the strata under
305 consideration, while holding the realized sampling frame constant (S2 and S3 Appendices).
306 These estimation and testing protocols are embedded within the QDIVER routine, now available
307 within GenAlEx 6.51 (http://biology.anu.edu.au/GenAlEx/; [52-53]).

308 Diversity analysis of paired *Antechinus* congeners

309 Here we illustrate these new tools with the Australasian marsupial genus *Antechinus*,
310 comprised of small ground-dwelling and climbing predators of forests, woodlands and
311 heathlands. Morphological and phylogenetic research in recent decades has identified several
312 previously unrecognized species-level splits within the genus. Both *A. stuartii* and *A. agilis* were
313 once viewed as a single species (*A. stuartii*), but recent research has indicated that they are
314 separate species, with a geographic break approximately 200 km south of Sydney, New South
315 Wales [54]. Based on nuclear (*IRBP*, *RAG*1, *bFib*7) and mitochondrial (*cyt-b*, 12sRNA,
316 16sRNA) sequence analysis [55], these congeners are thought to have diverged in the early
317 Pliocene. Much of the published life history research on *Antechinus* was conducted within the
318 range of *A. agilis*, predating recognition of two species, but *A. stuartii* is demographically quite
319 similar. Both are small, semelparous carnivores (approx. 15-50g); polyandrous females give

birth to (6 - 10) offspring (there is geographic variation in teat number) each spring. Most females die after weaning their first clutch, but a few survive to breed in a second year. The males die shortly after an intense breeding season in their second year.

Regional allopatry

We sampled *A. stuartii* from Booderee National Park (BNP) in New South Wales and *A. agilis* from the Victoria Central Highlands (VCH) in Victoria (Banks *et al*. 2011), separated by about 500 km (Fig 1). Within each species, sampling involved a trio of spatially separated trapping areas, each treated here as a separate 'population' for illustrative purposes (Fig 1). Each species is common within its own range, and there are no overt habitat discontinuities or overt barriers to gene flow, barring the effects of dispersal distance itself. BNP populations are spread out along a peninsula, with GRP1 most seaward (and most constrained), about twice as far from GRP3 (the most landward population) as it is to GRP2. For VCH populations, CAM6 is about four times as far from BLR5 as is the latter from MUR4. The population samples themselves are more widely separated for *A. agilis* (VCH) than for *A. stuartii* (BNP). The average pre-mating dispersal distance of males is over 1000 m, while that for females averages less than 100 m [56-58]. Genetic isolation (over 10s of km) may well impact our within-species decomposition ($\sigma_{WS} = \beta_{AP} \cdot \alpha_{WP}$) for these organisms.

Fig 1: Map of the study locations for *Antechinus stuartii* (Booderee National Park, New South Wales) and *A. agilis* (Victoria Central Highlands, Victoria), Australia. Photo credits go to Stephen Mahony and Esther Beaton, respectively.

Nuclear microsatellite markers

For this study, we have analyzed eight nuclear SSR loci (Aa7d, Aa2e, Aa2g, Aa4d, Aa7f, Aa7m, Aa4k, Aa2b) for each of 50 individuals for each population, a grand total of $(2N = 600)$ alleles for each locus (SSR lab protocols in S4 Appendix). The loci used here are a subset of those previously assayed for these species, filtered here for an absence of null-alleles, as well as conformance to regular allele step size criteria [57]. The regularity restriction was applied to remove a trio of loci with a high proportion of allelic step sizes that were less than the length of the microsatellite repeat motif itself. We illustrate with a pair of allele size distributions (Aa4d and Aa7d), each with a two-nucleotide repeat motif, illustrating non-trivial 'ladder offset' between the allelic batteries of the two species, in spite of some allelic overlap and sharing (Fig 2).

Fig 2: A pair of typical two-nucleotide step SSRs (Aa4d and Aa7d) for *Antechinus stuartii* (in Booderee National Park (BNP), New South Wales) and *A. agilis* (in Victoria Central highlands (VCH), Victoria) Australia

Different is different ($\mathbf{D}^{\mathbf{C}}$) and degree of difference ($\mathbf{D}^{\mathbf{R}}$) coding

We used traditional (0 vs 1) coding for our ($\mathbf{D}^{\mathbf{C}}$) treatment (Table 1), but for each of the eight loci separately, then added the eight matrices for all loci, element by element, as is standard for multi-locus distance analysis. We then divided each element in the multi-locus $\mathbf{D}^{\mathbf{C}}$ matrix by

$(\mathrm{L} = 8)$, reducing each matrix to (average) single-locus form and ensuring that $[0 \leq Q < 1]$, so that all $\mathbf{D}^{\mathrm{C}}$ - derived diversity elements are properly bounded and well behaved. The two species occupy somewhat different (though overlapping) sectors of the eight nSSR ladders (Fig 2), so for each locus of our $\mathbf{D}^{\mathrm{R}}$ matrix, we computed the squared distance between any pair of alleles as the square of the number of (two-nucleotide) steps between them, divided by the maximum squared distance(d^2_{max}) for that locus. Finally, we added the eight single-locus matrices together, and we divided each element of that summation by $(\mathrm{L} = 8)$, ensuring that all $\mathbf{D}^{\mathrm{R}}$ - derived diversity components are properly bounded and well-behaved. This (R_{ST}) 'degree of difference' metric is convenient, but not the only possible choice, a matter to which we will return in the **Discussion**.

Diversity within populations

For each of the six sampled populations, we present the [0,1] - scaled values $(\tilde{\alpha}_{\mathrm{WP}})$, as well as separate species averages and a pooled study-wide average for both $\mathbf{D}^{\mathrm{C}}$and $\mathbf{D}^{\mathrm{R}}$ coding (Table 4). The geographically most isolated populations have less internal diversity within both species, $(\tilde{\alpha}_{\mathrm{P1}} < \tilde{\alpha}_{\mathrm{P2}} < \tilde{\alpha}_{\mathrm{P3}})$ within *A. stuartii* (BNP) and $(\tilde{\alpha}_{\mathrm{P4}} > \tilde{\alpha}_{\mathrm{P5}} > \tilde{\alpha}_{\mathrm{P6}})$ within *A. agilis* (VCH). The first of these (involving peninsular GRP1) is marginally significant; the second (simply more isolated CAM6) is not. The within-population components for *A. agilis* (VCH) are about twice as large as those for *A. stuartii* (BNP) for either $\mathbf{D}^{\mathrm{C}}$ or $\mathbf{D}^{\mathrm{R}}$ coding. Having scaled $\mathbf{D}^{\mathrm{C}}$ elements by $(d^2_{max}) = 1$ and those for $\mathbf{D}^{\mathrm{R}}$ by $(d^2_{max}) = 275$, all of our within-population components are an order of magnitude smaller for $\mathbf{D}^{\mathrm{R}}$ than they are for $\mathbf{D}^{\mathrm{C}}$ coding (Table 4).

Table 4: Scaled within-population allelic diversity ($\alpha_{WP}^{\sim}$) for both $\mathbf{D}^{C}$ and $\mathbf{D}^{R}$ coding within *Antechinus stuartii* (Booderee National Park), within *A. agilis* (Victoria Central Highlands), and within the entire study, with Bartlett's tests of within-population homogeneity

Antechinus stuartii (BNP)			$\mathbf{D}^{C}$ Coding	*Antechinus agilis* (VCH)		
GRP1	GRP2	GRP3	[0,1] - Scaled Diversity	MUR4	BRL5	CAM6
0.435	0.453	0.476	($\alpha_{WP}^{\sim}$) Estimates	0.841	0.836	0.831
Average & Bartlett's Test			Average & Bartlett's Test	Average & Bartlett's Test		
$\alpha_{WP}^{\sim} = 0.455$ ($P < 0.041$)			$\alpha_{WP}^{\sim} = 0.645$ ($P < 0.001$)	$\alpha_{WP}^{\sim} = 0.836$ ($P > 0.64$)		
Antechinus stuartii (BNP)			$\mathbf{D}^{R}$ Coding	*Antechinus agilis* (VCH)		
GRP1	GRP2	GRP3	[0,1] - Scaled Diversity	MUR4	BRL5	CAM6
0.031	0.036	0.041	($\alpha_{WP}^{\sim}$) Estimates	0.072	0.073	0.069
Average & Bartlett's Test			Average & Bartlett's Test	Average & Bartlett's Test		
$\alpha_{WP}^{\sim} = 0.036$ ($P < 0.07$)			$\alpha_{WP}^{\sim} = 0.054$ ($P < 0.001$)	$\alpha_{WP}^{\sim} = 0.071$ ($P > 087$)		

Partitioning diversity along the taxonomic hierarchy

Study-wide (and within-species) diversity cascades are presented for both $\mathbf{D}^{C}$ and $\mathbf{D}^{R}$ coding in (Table 5). The total $\gamma^{\sim} = (Q/Q^{*})$ and among-species $(\delta_{AS}^{\sim}) = (Q_{AS}/Q_{AS}^{*})$ estimates are explicitly defined for the two species jointly, but we also define within-species $(\sigma_{WS}^{\sim}) = (Q_{WS}/Q_{WS}^{*})$ and among-population $(\beta_{AP}^{\sim}) = (Q_{AP}/Q_{AP}^{*})$ components for each of the species separately, to augment the average within-population $(\alpha_{WP}^{\sim}) = (Q_{WP}/Q_{WP}^{*})$ diversities within those same species, the latter drawn from (Table 4). Total diversity is the product of all

components in the multiply-nested cascade, averaged over 600 alleles for each of $(\mathrm{L} = 8)$ nSSR loci. It is large for both $\mathbf{D}^{\mathrm{C}}$ $(\gamma^{\sim} = 0.807)$ and for $\mathbf{D}^{\mathrm{R}}$ $(\gamma^{\sim} = 0.125)$ coding, but reflects the scaling difference between the two coding schemes.

These two species are substantially divergent, with $(\delta^{\sim}_{\mathrm{AS}} = 0.891, P < 0.001)$ for $\mathbf{D}^{\mathrm{C}}$ coding and $(\delta^{\sim}_{\mathrm{AS}} = 0.150, P < 0.001)$ for $\mathbf{D}^{\mathrm{R}}$ coding, signatures of phylogenetic diversification for congeners separated since the early Pliocene [55]. Both of the within-species diversity components are substantial, but that within *A. agilis* (VCH) is about twice as large as that within *A. stuartii* (BNP), both for $\mathbf{D}^{\mathrm{C}}$ $(\sigma^{\sim}_{\mathrm{WS-VCH}} = 0.841 > \sigma^{\sim}_{\mathrm{WS-BNP}} = 0.461, P < 0.001)$ and for $\mathbf{D}^{\mathrm{R}}$ $(\sigma^{\sim}_{\mathrm{WS-VCH}} = 0.071 > \sigma^{\sim}_{\mathrm{WS-BNP}} = 0.036, P \leq 0.001)$ coding.

Neither regional landscape shows any overt barriers to gene flow, but the dispersal challenge posed by sheer distance is greater for *A. agilis* (VCH) than for *A. stuartii* (BNP) populations (Fig 1). Given the greater isolation of CAM6 from (BLR5 and MUR4) than of GRP1 from (GRP2 and GRP3), we might anticipate greater among-population diversity within *A. agilis* (VCH) than within *A. stuartii* (BNP). As anticipated, $\mathbf{D}^{\mathrm{C}}$ coding yields the expected pattern $(\beta^{\sim}_{\mathrm{AP-VCH}} = 0.088, P \leq 0.001)$ and $(\beta^{\sim}_{\mathrm{AP-BNP}} = 0.025, P \leq 0.001)$, but population subdivision is virtually nil $(\beta^{\sim}_{\mathrm{AP-VCH}} = 0.001 = \beta^{\sim}_{\mathrm{AP-BNP}})$ within either species for $\mathbf{D}^{\mathrm{R}}$ coding. Populations diverge somewhat for allelic composition within either species, but with no net 'ladder offset' among those populations. Given the deep phylogenetic history of this genus, the two taxa should be substantially more diverse at the species level than at the population level (within either of them), and that is what we find. While the among-species $(\sigma^{\sim}_{\mathrm{AS}})$ component is an order of magnitude larger than the among-population $(\beta^{\sim}_{\mathrm{AP}})$ component for 'different is different' $(\mathbf{D}^{\mathrm{C}})$ coding, however, it is two orders of magnitude larger for 'degree of difference' $(\mathbf{D}^{\mathrm{R}})$ coding.

The nSSR 'ladder offset' between these congeners (Fig 2) constitutes a compelling diversity signature of long-term evolutionary separation.

Table 5: Scaled diversity values for both D^C and D^R coding, and for both *Antechinus stuartii* and *A. agilis*: study total ($\gamma^{\sim}$), among-species ($\delta^{\sim}_{AS}$), within-species ($\sigma^{\sim}_{WS}$), among-populations ($\beta^{\sim}_{AP}$), and within-populations ($\alpha^{\sim}_{WP}$), with Bartlett's homogeneity tests of the within stratum components

Antechinus stuartii (BNP) Booderee National Park Diversity Components & Statistical Tests	$\mathbf{D}^C$ Study-Wide Average Diversity & Statistical Tests	*Antechinus agilis* (VCH) Victoria Central Highlands Diversity Components & Statistical Tests
$\sigma^{\sim}_{WS} = 0.461$	$\gamma^{\sim} = 0.807$ $\delta^{\sim}_{AS} = 0.891\ (P < 0.001)$ $\sigma^{\sim}_{WS} = 0.651\ (P < 0.001)$	$\sigma^{\sim}_{WS} = 0.841$
$\beta^{\sim}_{AP} = 0.025\ (P < 0.001)$	$\beta^{\sim}_{AP} = 0.040\ (P < 0.001)$	$\beta^{\sim}_{AP} = 0.088\ (P < 0.001)$
$\alpha^{\sim}_{WP} = 0.455\ (P < 0.041)$	$\alpha^{\sim}_{WP} = 0.645\ (P < 0.001)$	$\alpha^{\sim}_{WP} = 0.836\ (P > 0.64)$
Antechinus stuartii (BNP) Booderee National Park Diversity Components & Statistical Tests	$\mathbf{D}^R$ Study-Wide Average Diversity & Statistical Tests	*Antechinus agilis* (VCH) Victoria Central Highlands Diversity Components & Statistical Tests
$\sigma^{\sim}_{WS} = 0.036$	$\gamma^{\sim} = 0.125$ $\delta^{\sim}_{AS} = 0.150\ (P < 0.001)$ $\sigma\ ^{\sim}_{WS} = 0.054\ (P < 0.001)$	$\sigma^{\sim}_{WS} = 0.071$
$\beta^{\sim}_{AP} = 0.001\ (P > 0.90)$	$\beta^{\sim}_{AP} = 0.001$	$\beta^{\sim}_{AP} = 0.001\ (P > 0.74)$
$\alpha^{\sim}_{WP} = 0.036\ (P > 0.07)$	$\alpha^{\sim}_{WP} = 0.053\ (P < 0.001)$	$\alpha^{\sim}_{WP} = 0.071\ (P > 0.87)$

Discussion

Overview of outcomes

We have elaborated a classic approach for estimating $(q = 2)$ genetic diversity metrics that meet the standard desiderata of diversity measures. We first defined pairwise genetic distances between all pairs of (2*N*) alleles for each genetic locus, and packed those into a square distance matrix, using both $\mathbf{D}^{\mathrm{C}}$ and $\mathbf{D}^{\mathrm{R}}$ coding schemes. We then divided each element by the largest in the matrix (d_{max}^2), extracted a bounded form of Rao's quadratic entropy $[0 \leq Q < 1]$, and converted that to a measure of diversity for the whole collection (γ). We extended the treatment to a multiply-nested partition of the total diversity for the general case of unbalanced sampling, top to bottom of the hierarchy. We scaled each of the diversity components [0,1], using sample frame restrictions. Finally, we deployed tests for the among-species, among-population, and among-individual components, as well as novel homogeneity tests for the within-stratum components. All of these innovations are now encoded within the QDiver routine of GenAlEx 6.51 (http://biology.anu.edu.au/GenAlEx/; [52-53]).

We illustrated this new analysis with two *Antechinus* congeners (*A. stuartii* and *A. agilis*), using eight nSSR loci, treated in both $(\mathbf{D}^{\mathrm{C}})$ and $(\mathbf{D}^{\mathrm{R}})$ fashion. There is large (phyletic) diversity between the two species, but about twice as much diversity within *A. agilis* as within *A. stuarti*. The 'population structure' within *A. agilis* was also greater than that within *A. stuartii*. There are two possible explanations: (a) greater frequency of disturbance (wildfire) for BNP (*A. stuartii*) than for VCH (*A. agilis*), which could induce local bottlenecks and slow recovery of local population panmixis [59-61]; and (b) greater spatial dispersion of sampling sites within VCH than within BNP. There is some confounding of regional fire history with regional spatial separation here, but the spatial dispersion differences seem the more likely explanation. Finally,

451 the ratio of among-species to among-populations diversity was an order of magnitude larger for
452 $\mathbf{D}^{R}$ than for $\mathbf{D}^{C}$ coding, with minor allele frequency divergence (but no ladder shifts) among
453 populations within either species, coupled with major ladder shifts between the two species.

454 Scaling considerations

455 We have here deployed standard 'degree of difference' (R_{ST}) coding for the $\mathbf{D}^{R}$
456 treatment. Beyond some level of phylogenetic separation, however, the use of $\mathbf{D}^{R}$ coding may
457 not be linear with phylogenetic time, given the inherent mutational homoplasy of microsatellite
458 substitution [62]. Particularly with small sample sizes, small numbers of SSR loci, and deep
459 time depth, it is possible to under-estimate divergence with classic (R_{ST}) scaling, and that
460 estimation error increases with evolutionary time. Various workers have suggested using
461 negative binomial coding [43, 45, 63-64], for which 'degree of difference' scaling is log-linear
462 (rather than linear) with increasing phylogenetic time. More generally, 'degree of difference'
463 scaling is a consequential choice for diversity estimation, testing and interpretation, and such
464 scaling will warrant careful attention as we move forward.

465 Large NGS panels are now available [65-66], containing both synonymous (presumably
466 neutral) and non-synonymous (possibly adaptive) substitutions [67-68]. Methods such as
467 sequence capture of ultra-conserved elements (UCE's) enable interspecific comparisons of
468 evolutionary processes using standardized sequence datasets [69], and efforts are increasing to
469 sort among myriad markers for smaller subsets that may represent important adaptive signals
470 within and/or among the taxa examined [70-71]. With newer types of genetic markers becoming
471 available, each with its own coding conventions, the choice of Euclidean metrics has obvious
472 implications for diversity exposition.

473

Translation between evolution and ecology

The use of multiple characters for quantitative taxonomic analysis dates to the 1960s [72], and has been a recurring theme in population genetics. More recently, there has been a suggestion to use taxonomic subdivision itself as a 'degree of difference' metric to quantify diversity, using simple code, say $(d_{jk} = 0)$ for individuals in the same species, $(d_{jk} = 1)$ for different species but same genus, and $(d_{jk} = 2)$ for different genera [32]. Others have used more elaborate phylogenetic time depth estimates as 'degree of difference' metrics [45, 73-77]. There have also been increasing attempts to translate ecological separation into derivative evolutionary diversity outcomes [74-75, 77-89].

Both the need and our ability to communicate across the boundary between Evolution and Ecology continue to develop [90], and there should be three larger payoffs from what we have done here. (a) We have improved our ability to deal with 'different is different' coding, have scaled it [0,1], and have configured diversity analysis for convenient statistical evaluation. (b) By extending treatment of diversity into 'degree of difference' coding, we can attack problems where the scale of divergence itself is a part of the story. (c) Translation between diversity-metric and variance-metric methods provides access to a large panoply of quadratic estimation and testing methodology. Cross-disciplinary analytical translation will be of increasing importance and value, as evolutionary ecology continues to develop.

Conclusion

We have here articulated a novel quadratic approach for partitioning genetic diversity within and among strata of a hierarchical sampling design that exhibits the desirable properties of diversity criteria. Importantly, this approach is unique among diversity treatments to date, in

496 providing a statistical comparison of within-stratum diversity components at any given level. It
497 also enables diversity analysis of a wide range of inter-individual genetic coding schemes that
498 emerge from modern genomic work, as well as being extendable to organisms of virtually any
499 ploidy level. This new approach promises to be informative and useful across a wide range of
500 ecological and evolutionary studies.

501 Acknowledgments

502 The thank David Lindenmayer and Chris MacGregor for the BNP regional samples,
503 Stephen Mahony and Esther Beaton for permission to use the *Antechinus* photo-images in
504 (Fig 1), Bill Sherwin, Oscar Gaggiotti, and a trio of anonymous reviewers for helpful critique on
505 earlier drafts of the manuscript. PES was supported by the USDA and the New Jersey
506 Agricultural Experiment Station, Hatch project NJ17160; SCB was supported by Australian
507 Research Council Future Fellowship FT130100043.

508 Statement on animal usage

509 The animal use protocols for *Antechinus* sampling and handling were covered by
510 A2015/60 and A2012/49 permits (Australian National University).

511 Conflict of interest statement

512 The authors affirm that they have no competing or conflicting interests.

513 Accessibility arrangements

514 The *Antechinus* data are archived in Excel workbook form, along with listings of the
515 QDIVER results extracted from GenAlEx6.51 (http://biology.anu.edu.au/GenAlEx/). $\mathbf{D}^{C}$ data and
516 analyses are presented in S5 Appendix, and $\mathbf{D}^{R}$ data and analyses are presented in S6 Appendix.
517

Literature cited

1. Smouse PE, Robledo-Arnuncio JJ, Measuring the genetic structure of the pollen pool as the probability of paternal identity. Heredity. 2005;94: 640-649.
2. Grivet D, Smouse PE, Sork VL, A new approach to the study of seed dispersal: a novel approach to an old problem. Molecular Ecol. 2005;14: 3585-3595.
3. Grivet D, Robledo-Arnuncio JJ, Smouse PE, Sork VL, Relative contribution of contemporary pollen and seed dispersal to the neighborhood size of a seedling population of California valley oak (*Quercus lobata*, Née). Molecular Ecol. 2009;16: 3967-3979.
4. Gonzales E, Hamrick JL, Smouse PE, A Comparison of clonal diversity in mountain and Piedmont populations of *Trillium cuneatum* (Melanthiaceae - Trilliaceae**)**, a forest understory species. Amer Journal Bot. 2008;95: 1-9.
5. Scofield DG, Sork VL, Smouse PE, Influence of acorn woodpecker social behavior on transport of Coastal Live Oak (*Quercus agrifolia* Née) acorns in a southern California oak savanna. Journal Ecol. 2010;98: 561-572.
6. Scofield DG, Smouse PE, Karubian J, Sork VL, Use of α, β, and γ diversity measures to characterize seed dispersal by animals. American Nat. 2012;180: 719-732.
7. Sork VL, Smouse PE, Scofield DG, Grivet D, Impact of asymmetric male and female gamete dispersal on allelic diversity and spatial genetic structure in valley oak (*Quercus lobata* Née). Evolution & Ecol. 2015;29: 927-945.
8. Excoffier L, Smouse PE, Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: Molecular variance parsimony. Genetics. 1994; 136: 343-359.

9. Bandelt H-J, Forster P, Sykes BC, Richards MB, Mitochondrial portraits of human populations using median networks. Genetics. 1995;141: 743-753.

10. Bandelt H-J, Forster P, Röhl P, Median-joining networks for inferring intraspecific phylogenies. Molecular Biol & Evol. 1999;16: 37-48.

11. Bandelt H-J, Macaulay V, Richards M, Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. Molecular Phylogenetics & Evol. 2000;16: 8-28.

12. Leigh JW, Bryant D, Popart: Full-feature software for haplotype network construction. Methods Ecol & Evol. 2015;6: 1110–1116

13. Kimura M, Crow JF, The number of alleles that can be maintained in a finite population. Genetics. 1964;49: 725-734.

14. Ohta T, Kimura M, A model of mutat8on appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genetics Res. 1973;22: 201-204.

15. Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigell JE, et al., Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. Annual Rev Ecol Syst. 1987;18: 489-522.

16. Richardson RH, Smouse PE, Richardson ME, Patterns of molecular variation. II. Associations of electrophoretic mobility and larval substrate within species of the *Drosophila mulleri* complex. Genetics. 1977;85: 141-154.

17. Felsenstein J, Phylogenies and quantitative characters. Annual Rev Ecol Syst. 1988;19: 445-471.

562 18. Goldstein DB, Ruíz-Linares A, Cavalli-Sforza LL, Feldman MW, An evaluation of genetic
563 distances for use with microsatellite loci. Genetics. 1995;139: 463-471.
564 19. Slatkin M, A measure of population subdivision based on microsatellite allele frequencies.
565 Genetics. 1995;139: 457-462.
566 20. Jost L, Partitioning diversity into independent alpha and beta components. Ecology.
567 2007;88: 2427–2439
568 21. Allen B, Kon M, Bar-Yam Y, A new phylogenetic diversity measure generalizing the
569 Shannon index and its application to Phyllostomid bats. American Nat. 2009;174: 236-243.
570 22. Sherwin WB, Entropy and information approaches to genetic diversity and its expression:
571 genomic geography. Entropy. 2010;12: 1765-1798.
572 23. Leinster T, Cobbold CA, Measuring diversity: the importance of species similarity.
573 Ecology. 2012;93: 477-488.
574 24. Rao CR, Analysis of diversity: A unified approach. In: Statistical Decision Theory and
575 Related Topics III, Vol. 2, SS Gupta, JO Berger, eds. Academic Press, NY, 1982a: pp. 235-
576 250.
577 25. Rao CR, Diversity and dissimilarity coefficients: a unified approach. Theoretical Pop Biol.
578 1982;21: 24-43.
579 26. Rao CR, Diversity, its measurement, decomposition, apportionment and analysis. Sankhya.
580 1982;44: 1-21.
581 27. Rao CR, Gini-Simpson index of diversity: A characterization, generalization and
582 applications. Util Math. 1982;21: 273-282.
583 28. Rao CR, Rao's axiomatization of diversity measures. Encyclopedia of Statistical Sciences,
584 S Kotz, NL Johnson (eds). Wiley, NewYork. 1986; pp. 614-617.

29. Rao CR, Quadratic entropy and analysis of diversity. Sankyhā: Indian J Stat. 2010;72A**:** 70-80.

30. Jost L, Entropy and diversity. Oikos. 2006:113: 363–375.

31. Ricotta C, Szeidl L, Towards a unifying approach to diversity measures: Bridging the gap between the Shannon entropy and Rao's quadratic index. Theoretical Pop Biol. 2006;70: 237-243.

32. Ricotta C, Szeidl L, Diversity partitioning of Rao's quadratic entropy. Theoretical Pop Biol. 2009;76: 299-302.

33. De Bello F, Lavergne S, Meynard CN, Lepš J, Thuiller W, The partitioning of diversity: showing Theseus a way out of the labyrinth. J Vegetation Science. 2010;21: 1992-1000.

34. Sherwin WB, Jabot F, Rush R, Rossetto M, Measurement of biological information with applications from genes to landscapes. Molecular Ecol. 2006;15: 2857-2869.

35. Chiu C-H, Chao A, Distance-based functional diversity measures and their decomposition: A framework based on Hill numbers. PLoS One. 2014;9(7): e100014.

36. Excoffier L, Smouse PE, Quattro JM, Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics. 1992; 31: 479-491.

37. Peakall R, Smouse PE, Huff DR, Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss (*Buchloë dactyloides* (Nutt.) Engelm.). Molecular Ecol. 1995;4: 135-147.

38. Excoffier L, Lischer HEL, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecol Resour. 2010;10: 564-567

608 39. Anderson MJ, A new method for non-parametric multivariate analysis of variance.
609 Australian Ecol. 2001:26**:** 32-46.
610 40. Anderson MJ, PERMANOVA: a FORTRAN computer program for permutational
611 multivariate analysis of variance. Dept Statistics, Univ Auckland, New Zealand. 2005.
612 41. McArdle BH, Anderson MJ, Fitting multivariate models to community data: A comment
613 on distance based redundancy analysis. Ecology. 2001; 82: 290-297.
614 42. Nievergelt CM, Libiger O, Schork NJ, Generalized analysis of molecular variance. PLoS
615 Genet. 2007;3(4): e51. doi:10.1371/journal.pgen.0030051.
616 43. Coulson TN, Pemberton JM, Albon SD, Beaumont M, Marshall TC, Slate J, et al.,
617 Microsatellite reveals heterosis in red deer. Proc Roy Soc Lond B. 1998;256: 489-495.
618 44. Otter KA, Stewart IRK, McGregor PK, Terry AMR, Dabelsteen T, Burke T, Extra-pair
619 paternity among great tits *Parus major* following manipulation of male signals. J Avian
620 Biol. 2001;32: 338-344.
621 45. Bruvo RU, Michiels NK, D'Souza TG, Schulenburg H, A simple method for the
622 calculation of microsatellite genotype distances irrespective of ploidy level. Molecular
623 Ecol. 2004;13: 2101-2106.
624 46. Gower JC, Legendre P, Metric and Euclidean properties of dissimilarity coefficients.
625 Journal Classif. 1986;3: 5-48.
626 47. Banks SC, McBurney L, Blair D, Davies ID, Lindenmayer DB, Where do animals come
627 from during post-fire population recovery? Implications for ecological and genetic patterns
628 in post-fire landscapes. Ecography. 2017; DOI: 10.1111/ecog.02251.

48. Smouse PE, Whitehead MR, Peakall R, An informational diversity analysis framework, illustrated with sexually deceptive orchids in early stages of speciation. Molecular Ecol Resour. 2015;15: 1375-1384.

49. Jost L, *G*ST and its relatives do not measure differentiation. Molecular Ecol. 2008;17: 4015–4026.

50. Jost L, *D vs. G*ST: response to Heller and Siegismund (2009) and Ryman and Leimar (2009). Molecular Ecol. 2009;18: 2088–2091.

51. Bartlett MS, Properties of sufficiency and statistical tests. Proc Roy Stat Soc. 1937;160: 268-282.

52. Peakall R, Smouse PE, GenAlEx 6: Genetic Analysis in Excel. Population genetic software for teaching and research. Molecular Ecol Notes. 2006;6: 288-295.

53. Peakall R, Smouse PE, GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research - an update. Bioinformatics. 2012;28: 2537-2539.

54. Dickman CR, Parnaby HE, Crowther MS, King DH, *Antechinus agilis* (Marsupialia: Dasyuridae), a new species from the *A. stuartii* complex in south-eastern Australia Australian J Zool. 1998;46: 1-26.

55. Westerman M, Krajewski C, Kear BP, Meehan L, Meredith RW, Emerling CA, et al., Phylogenetic relationships of dasyuromorphian marsupials revisited. Zoological J Linn Soc. 2016;176: 686-701

56. Cockburn A, Sccot MP, Scotts DJ, Inbreeding avoidance and male-biased natal dispersal in *Antechinus* spp (Marsupialia, Dasyuridae). Animal Behav. 1985;33: 908-915.

57. Banks SC, Finlayson GR, Lawson SJ, Lindenmayer DB, Paetkau D, Ward SJ, et al., The effects of habitat fragmentation due to forestry plantation establishment on the demography

652 and genetic variation of a marsupial carnivore, *Antechinus agilis*. Biological Conserv.
653 2005;122: 581-597.
654 58. Banks SC, Lindenmayer DB, Inbreeding avoidance, patch isolation and matrix
655 permeability influence dispersal and settlement choices by male agile antechinus in a
656 fragmented landscape. J Animal Ecol. 2014;83: 515-524.
657 59. Banks SC, Dujardin M, McBurney L, Blair D, Barker M, Starting points for small mammal
658 population recovery after wildfire: recolonisation or residual populations? Oikos.
659 2011;120: 26-37
660 60. Davies ID, Cary GC, Landguth EL, Lindenmayer DB, Banks SC, Implications of recurrent
661 disturbance for genetic diversity. Ecology & Evol. 2016;6: 1181-1196.
662 61. Lindenmayer D, Blanchard W, MacGregor C, Barton P, Banks SC, Crane M, et al.,
663 Temporal trends in mammal responses to fire reveals the complex effects of fire regime
664 attributes. Ecological Appl. 2016;26: 557-573.
665 62. Gaggiotti OE, Lange O, Rassmann K, Gliddon C, A comparison of two indirect methods
666 for estimating average levels of gene flow using microsatellite data. Molecular Ecol.
667 1999;8: 1513-1520.
668 63. Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB, Mutational
669 processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci (USA).
670 1994;91: 3166-3170.
671 64. Fu Y-X, Chakraborty R, Simultaneous estimation of all the parameters of a stepwise
672 mutation model. Genetics. 1998;150: 487–497.

65. Nicotra AB, Chong C, Brag JG, Ong CR, Aitken NC, Chuah A, et al., Population and phylogenomic decomposition via genotyping-by-sequencing in Australian *Pelargonium*. Molecular Ecol. 2016;25: 2000–2014.

66. Trumbo DR, Epstein B, Hohenlohe PA, Alford RA, Schwarzkopf L, Storfer A, Mixed population genomics support for the central marginal hypothesis across the invasive range of the cane toad (*Rhinella marina*) in Australia. Molecular Ecol 2016;25: 4161–4176.

67. Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L, Widespread signals of convergent adaptation to high altitude in Asia and America. American J Hum Genet. 2014;95: 394-407.

68. Enard D, Cai L, Gwennap C, Dmitri A Petrov DA, Viruses are a dominant driver of protein adaptation in mammals. eLIFE. 2016;5: 1-25 (DOI: 10.7554/eLife.12469).

69. Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT, Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Systematic Biol. 2014;63: 83-95.

70. Excoffier L, Hofer T, Foll M, Detecting loci under selection in a hierarchically structured population. Heredity. 2009;103: 285-298.

71. Coop G, Witonsky D, Di Rienzo A, Pritchard JK, Using environmental correlations to identify loci underlying local adaptation. Genetics. 2010;185: 1411-1423.

72. Sokal RR, Sneath PHA, Principles of Numerical Taxonomy. W.H. Freeman and Company, San Francisco. 1963: 359+ pp

73. Izsák J, Papp L, Application of the quadratic entropy index for diversity studies on drosophilid species assemblages. J Environmental Ecol Stat. 1995;2: 213-224.

74. Webb CO, Exploring the phylogenetic structure of ecological communities: An example for rain forest trees. American Nat. 2000;156: 145-155.
75. Hardy OJ, Senterre B, Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. Journal Ecol. 2007;95: 493-506.
76. Hardy OJ, Jost L, Interpreting and estimating measures of community phylogenetic structuring. Journal Ecol. 2008;96: 849-852.
77. Chao A, Chiu C-H, Jost L, Phylogenetic diversity measures based on Hill numbers. Phil Trans Roy Soc B. 2010;365: 3599-3609.
78. Faith DP, Conservation evaluation and phylogenic diversity. Biological Conserv. 1992;61: 1-10.
79. Warwick RM, Clarke KR, New `biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. Marine Ecol Prog Ser. 1995;129: 301-305.
80. Shimatani K, On the measurement of species diversity incorporating species differences. Oikos. 2001;93: 135-147.
81. Botta-Dukat Z, Rao's quadratic entropy as a measure of functional diversity based on multiple traits. J Vegetation Sci. 2005;: 533-540.
82. Ricotta C, A note on functional diversity measures. Basic Appl Ecol. 2005;6: 479-486.
83. Ricotta C, Additive partitioning of Rao's quadratic diversity: a hierarchical approach. Ecol Model. 2005;183: 365-371
84. Chao A, Jost L, Chiang S, Jiang Y, Chazdon R, A two-stage probabilistic approach to multiple-community similarity indices. Biometrics. 2008;64: 1178-1186.
85. Hardy OJ, Jost L, Interpreting and estimating measures of community phylogenetic structuring. Journal Ecol. 2008;96: 849-852

718 86. Ricotta C, Di Nepi M, Guglietta D, Celesti-Grapow L, Exploring taxonomic filtering in
719 urban environments. J Vegetation Sci. 2008;19: 229-238.
720 87. Villéger S, Mouillot D, Additive partitioning of diversity including species differences: a
721 comment on Hardy & Senterre (2007). J Ecology. 2008;96: 845-848.
722 88. Eastman JM, Paine CET, Hardy OJ, SpacodiR: structuring of phylogenetic diversity in
723 ecological communities. Bioinformatics. 2011;27: 2437-2438.
724 89. Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L, Widespread signals of convergent
725 adaptation to high altitude in Asia and America. Amer J Hum Genet. 2014;95: 394-407.
726 90. Chao A, Chiu C-H, Bridging the variance and diversity decomposition approaches to beta
727 diversity via similarity and differentiation measures. Methods Ecol Evolution. 2016;7: 919-
728 928.

729

730 Supplemental materials

731
732 **S1 Appendix: Partitioning within-population diversity into sub-components**

733
734 **S2 Appendix: Scaling diversity components [0,1]**

735
736 **S3 Appendix: Homogeneity testing of within-stratum diversity components**

737
738 **S4 Appendix: Laboratory microsatellite protocols**

739
740 **S5 Appendix: Analysis of Rao Diversity (QDiver) for D^C Matrix**

741
742 **S6 Appendix: Analysis of Rao Diversity (QDiver) for D^R Matrix**

743

S1 Appendix: Partitioning within-population diversity into sub-components

With sexual diploids, there will be situations where uniting gametes will be either more (negative assortative) or less (positive assortative) diverse than expected under random mating. For any particular individual, say the j^{th}, and scaling relative to (d^2_{max}), we compute a Q-value for the j^{th} (2 x 2) within individual matrix

$$Q_{WI\text{-}j} = (\Sigma_{WI\text{-}j} / 2^2) \quad \rightarrow \quad \omega_{WI\text{-}j} = (1 - Q_{WI\text{-}j})^{-1} \quad . \qquad [S1.1]$$

The sampling within any single individual is too limited to provide useful replication for statistical testing of homogeneity, of course, but we can certainly obtain a weighted average (within-individual) diversity for any particular population, within which, the relative weight for each individual is $(4 / 4N_{Pk}) = (N_{Pk})^{-1}$, translating [S1.1] into population average form,

$$Q_{WI,Pk} = \text{average } (Q_{WI\text{-}j}) \text{ for population } P_k \quad \rightarrow \quad \omega_{WI,Pk} = (1 - Q_{WI,Pk})^{-1} \, . \qquad [S1.2]$$

To extract the inter-individual diversity within the k^{th} population, we compute

$$\varepsilon_{AI,Pk} = (\alpha_{WPk} / \omega_{WI,Pk}) = [1 - Q_{WI,Pk}] / [\, 1 - Q_{WPk}] = [1 - Q_{AI,Pk}]^{-1} \quad . \qquad [S1.3]$$

Just as the among-species (δ_{AS}) and among population (β_{AP}) diversity components can be back-translated into functions of the appropriate Q-values (Text Eqq. [7, 13]), we also have

$$Q_{AI,Pk} = [Q_{WPk} - Q_{WI,Pk}] / [1 - Q_{WI,Pk}] = (\varepsilon_{AI,Pk} - 1) / \varepsilon_{AI,Pk} \quad . \qquad [S1.4]$$

We can compute average among-individual (ε_{AI}) and within-individual (ω_{WI}) components, combining all the individuals within a single species, or even across the whole study, as needed. Finally, our partition of the total diversity in the study, takes the expanded form

$$\gamma = (\alpha_{WP} \cdot \beta_{AP} \cdot \delta_{AS}) = (\omega_{WI} \cdot \varepsilon_{AI} \cdot \beta_{AP} \cdot \delta_{AS}) \, . \qquad [S1.5]$$

744

S2 Appendix: Scaling diversity components [0,1]

Jost [1] defined a [0,1]-scaled index of diversity overlap (labeled θ here) for any two elements of the study, which when applied (for example) to alleles within the total sample, yields [0,1] scaling, '0' for no sharing (total diversity) and '1' for total sharing (no diversity). Using (γ^*) here to denote the maximum possible value, we can write that scaling as:

$$0 \leq \theta = [\ (\gamma)^{-1} - \ (\gamma^*)^{-1}\ ] / [\ 1 - (\gamma^*)^{-1}\ ] \leq 1\ , \qquad \text{[S2.1]}$$

We have elsewhere [2] defined a complementary divergence index (γ'), with '0' = (total sharing) and '1' = (non-overlap) for the total sample:

$$0 \leq \gamma' = (1 - \theta) = \cdots = \gamma^* \cdot (\gamma - 1) / (\gamma^* - 1) \cdot \gamma \leq 1\ , \qquad \text{[S2.2a]}$$

which reduces the [0,1] - scaled diversity [S2.2a] to very convenient form:

$$[\ 0 \leq \gamma' = (Q / Q^*) \leq 1\ ]\ . \qquad \text{[S2.2b]}$$

Similar [0,1]-scaling for each of nested diversity components (δ'_{AS} = among species, σ'_{WS} = within species, β'_{AP} = among populations, α'_{WP} = within populations, ε'_{AI} = among individuals, and ω'_{WI} = within individuals) yields analogous outcomes, specifically

Strata	Among-Strata	Within-Strata	
Species	$[\ 0 \leq \delta'_{AS} = (Q_{AS} / Q^*_{AS}) \leq 1\ ]$	$[\ 0 \leq \sigma'_{WS} = (Q_{WS} / Q^*_{WS}) \leq 1\ ]$,	[S2.3a]
Populations	$[\ 0 \leq \beta'_{AP} = (Q_{AP} / Q^*_{AP}) \leq 1\ ]$	$[\ 0 \leq \alpha'_{WP} = (Q_{WP} / Q^*_{WP}) \leq 1\ ]$,	[S2.3b]
Individuals	$[\ 0 \leq \varepsilon'_{AI} = (Q_{AI} / Q^*_{AI}) \leq 1\ ]$	$[\ 0 \leq \omega'_{WI} = (Q_{WI} / Q^*_{WI}) \leq 1\ ]$.	[S2.3c]

The Q^*-value maxima are strictly functions of the replication, determined by the sampling frame, and because that sampling frame does not change with permutational testing procedures, a test of any of the raw diversities (Text Eq. [14]) is monotonic and 1:1 with a test of the scaled ratios in

745

S3 Appendix: Homogeneity testing of within-stratum diversity components

If we scale our distance matrix **D** by its maximum entry $(d_{max})^2$, then average the elements, both for the total collection and for any relevant subdivisions (e.g., within single species, within single populations, or even within single individuals), we can translate each of the resulting *Q*-values into a corresponding sum of squares. For the grand total collection of $(2N)$ alleles in the study, for the $(2N_{Sg})$ alleles within the g^{th} species ($g = 1, \ldots, G$), or for the $(2N_{Pk})$ alleles within the k^{th} population ($k = 1, \ldots, K$), the corresponding *Q* - values are simple multiples of the corresponding sums of squares and variances:

$$Q = (\text{sum}\mathbf{D} / 4N^2) = (SS / N) = V \cdot (2N - 1) / N \ , \qquad [S3.1a]$$

where SS and V are the grand total sum of squares and total variance, respectively. Similar extractions for the within-species, within-population, and within-individual *Q*-values yield

$$Q_{WSg} = \cdot\cdot\cdot = (SS_{WSg} / N_{Sg}) = V_{WSg} \cdot (2N_{Sg} - 1) / N_{Sg} \ , \qquad [S3.1b]$$

$$Q_{WPk} = \cdot\cdot\cdot = (SS_{WPk} / N_{Pk}) = V_{WPk} \cdot (2N_{Pk} - 1) / N_{Pk} \ , \qquad [S3.1c]$$

$$Q_{WIj} = \cdot\cdot\cdot = (SS_{WI\text{-}j} / N_{Ij}) = V_{WIj} \cdot (2N_{Ij} - 1) / N_{Ij} \ . \qquad [S3.1d]$$

Averages of the within-stratum components of [S3.1b - S3.1d] can be computed as needed.

We convert *Q*-values to their diversity analogues, and then use their maxima to scale each of those components from (0 = no diversity) to (1 = maximum diversity possible). Given the sampling realities of the study, those maxima take very simple forms,

$$Q^* = (2N - 1) / 2N \qquad Q^*_{WSg} = (2N_{Sg} - 1) / 2N_{Sg} \ ,$$

$$Q^*_{WPk} = (2N_{Pk} - 1) / 2N_{Pk} \qquad Q^*_{WI\text{-}j} = (2N_{Ij} - 1) / 2N_{Ij} = (\tfrac{1}{2}) \ , \qquad [S3.2]$$

recalling (for diploids) that, $N_{Ij} = 2$ alleles.

746

S4 Appendix: Laboratory microsatellite protocols

DNA extraction and genotyping of *Antechinus* samples

We extracted DNA from all samples using Qiagen 96-well plate DNeasy extraction kits. We genotyped all antechinus samples with the microsatellite loci Aa2B, Aa2E, Aa2G, Aa4A, Aa4K, Aa7D, Aa7F and Aa7M [1-2], as well as a Y chromosome SRY locus [3], not used for this paper. PCR amplification conditions are presented in Table S4.1 and protocols followed those in the papers describing the primer sequences themselves [1-2], except that each forward primer was tagged with the (-21) M13 sequence (TGTAAAACGACGGCCAGT) to enable dye-labelling of PCR products, using the approach of Schuelke [4]. Following amplification, PCR products were run on an ABI3100 sequencer and scored with GeneMapper software at AGRF laboratories. The microsatellite panels used here were the best-performing subset of loci from those previously used for these species in these regions [1, 5-6].

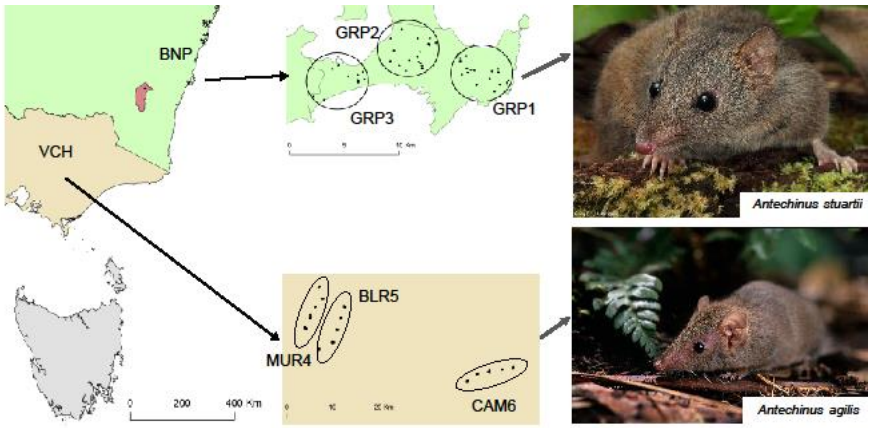
Table S4.1. PCR amplification conditions for the eight microsatellite loci used in this study, originally described by [1-2].

Locus	Primer	DNA Sequence	Annealing Temperature	[$MgCl_2$] (mM)
Aa2B	Forward	GTACCACAAGATGCACCTAC	54	1.5
	Reverse	TTCACAGCCTAACTAATGCTCCTA		
Aa2E	Forward	TCTCGGCTCCTGTCAGTT	54	2
	Reverse	TCACATAGGGCAGCTTTCCTCCTG		
Aa2G	Forward	TTACACACATGCCCATTCAC	50	2
	Reverse	AGTTCTAAAACAGAGGTTCTT		
Aa4A	Forward	TTTGATCCTCAGAGACTTGAT	50	2.25
	Reverse	CCAAATCTAAAATATCC		

748

749

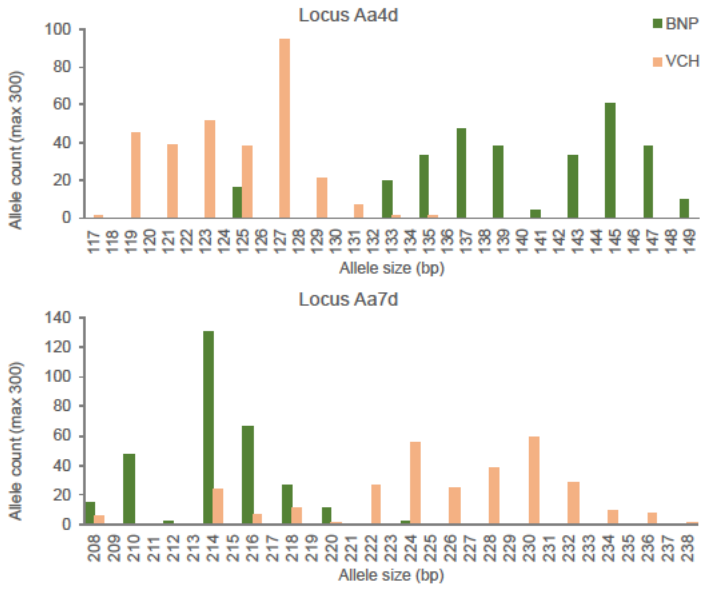
750 Figure 1



751

752

753 Figure 2



754