

RESEARCH ARTICLE

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

Peter E. Smouse¹, Sam C. Banks²*, Rod Peakall³

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

* These authors contributed equally to this work.

* Sam.Banks@anu.edu.au



OPEN ACCESS

Citation: Smouse PE, Banks SC, Peakall R (2017) Converting quadratic entropy to diversity: Both animals and alleles are diverse, but some are more diverse than others. PLoS ONE 12(10): e0185499. <https://doi.org/10.1371/journal.pone.0185499>

Editor: Wolfgang Arthofer, University of Innsbruck, AUSTRIA

Received: May 24, 2017

Accepted: September 13, 2017

Published: October 31, 2017

Copyright: © 2017 Smouse et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data is within the paper and its Supporting Information files. The *Antechinus* data are archived in Excel workbook form, along with listings of the QDIVER results extracted from GenAlEx6.51 (<http://biology.anu.edu.au/GenAlEx/>). DC data and analyses are presented in [S5 Appendix](#), and DR data and analyses are presented in [S6 Appendix](#).

Funding: PES was supported by the USDA National Institute of Food and Agriculture Hatch Project 1005333, and the New Jersey Agricultural Experiment Station, Hatch project NJ17160; SCB

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.

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,

was supported by Australian Research Council Future Fellowship FT130100043. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

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 $(j, k = 1, \dots, J)$ different alleles for that locus, we define Q for the total collection of N individuals ($2N$ alleles) as:

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

where the relative frequencies of the j^{th} and k^{th} alleles are p_j and p_k . Eq (1) can be rewritten in matrix form, using a matrix (\mathbf{D}) of squared distances between all $(4N^2)$ allelic-pairs, and as a

vector, $\mathbf{P}' = [(1/2N), \dots, (1/2N)]$, one entry per sampled allele

$$Q = \mathbf{P}'\mathbf{D}\mathbf{P} = (\text{sum}\mathbf{D}/4N^2), \tag{2}$$

where $\text{sum}\mathbf{D}$ is the total of $(4N^2)$ squared distances within matrix \mathbf{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, ($d_{jk}^2 = 0$, same) or ($d_{jk}^2 = 1$, different), to construct a multi-allelic distance matrix \mathbf{D}^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 $d_{jk}^2 = (r_j - r_k)^2$, we can pack those squared repeat-number distances into a \mathbf{D}^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 ($2N$ alleles), that translation takes the simple form

$$\gamma = 1/(1 - Q), \tag{3}$$

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}^C matrix, Q is properly bounded by virtue of its (0 vs 1) construction. For the ‘degree of difference’ \mathbf{D}^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}^R matrix should be divided by the maximum squared distance for that locus, $d_{max}^2 = (r_{max} - r_{min})^2$, thus ensuring that $[0 \leq Q < 1]$. To complete our scaling, we sum the squared distance values over loci, for either \mathbf{D}^C or \mathbf{D}^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 \mathbf{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 ($2N_A = 6$, $2N_B = 4$), and start with the example (five-allele array) illustrated in (Table 1). The average of all $(4N^2 = 100)$ elements in (\mathbf{D}^C) is $Q = (\text{sum}\mathbf{D}^C/4N^2) = (78/100) = 0.78$, 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, $Q_{WA} = (22/36)$ and $Q_{WB} = (10/16)$ and then convert the Q -values into separate diversity estimates within each

Table 1. 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.

| Allele | 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 | |
| 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 | | | | |

<https://doi.org/10.1371/journal.pone.0185499.t001>

of the species,

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

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

$$f_{WA} = 4N_{WA}^2 / (4N_{WA}^2 + 4N_{WB}^2) \quad f_{WB} = 4N_{WB}^2 / (4N_{WA}^2 + 4N_{WB}^2), \quad (5)$$

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

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

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

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

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

$$Q_{AS} = (Q - Q_{WS})/(1 - Q_{WS}) = (\delta_{AS} - 1)/\delta_{AS} = 0.428. \quad (8)$$

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

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 Eqs (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.

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

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.

| Allele | 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 | | |
| g-1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | g-1 |
| g-2 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | g-2 |
| g-3 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | g-3 |
| g-4 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | g-4 |
| g-5 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | g-5 |
| g-6 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | g-6 |
| g-7 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | g-7 |
| g-8 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | g-8 |
| g-9 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | g-9 |
| g-10 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | g-10 |
| Allele | g-1 | g-2 | g-3 | g-4 | g-5 | g-6 | g-7 | g-8 | g-9 | g-10 | Allele | |

<https://doi.org/10.1371/journal.pone.0185499.t002>

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,

$$\alpha_{WP1} = 1/(1 - Q_{WP1}) \cdots \alpha_{WP6} = 1/(1 - Q_{WP6}) \tag{9}$$

Using the same strategy as for Eqs (5) and (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:

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

and (10)

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

We then translate those weighted-average (Q_{WP}) values into separate within-population diversity estimates for Species A and B, respectively,

$$\alpha_{WPA} = 1/(1 - Q_{WPA}) \quad \alpha_{WPB} = 1/(1 - Q_{WPB}). \tag{11}$$

If we average all six within-population Q -values (weighted by their respective quadratic sample sizes), we obtain an average (Q_{WP}) for the whole study, and can translate that into a study-wide average estimate of the within-population diversity,

$$\alpha_{WP} = 1/(1 - Q_{WP}). \tag{12}$$

We are not constrained to balanced sampling at any level, but we do need to account explicitly for whatever sampling imbalance exists within the dataset.

For diploids, we can also extract estimates of within-individual (among-allele) diversity for each locus, as each individual is represented by a (2 x 2) submatrix (Table 1 and Table 2); we have N such sub-matrices for the study. We later illustrate this new analysis for a pair of strongly outbred species [47], where subdivision of within-individual diversity is not very

helpful. For organisms showing non-random mating systems, however, extracting sub-individual diversity components may prove valuable, and we describe their estimation in [S1 Appendix](#).

Given that our populations are nested within species, we also need to compute traditional among-population (Q_{AP}) and (Q_{AP}^*) values and to translate those into both estimated (β_{AP}) and maximum possible (β_{AP}^*) diversity. By analogy with Eq (7), we deploy

$$\beta_{AP} = (\sigma_{WS}/\alpha_{WP}) = (1 - Q_{WP})/(1 - Q_{WS}), \tag{13}$$

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

$$Q_{AP} = (Q_{WS} - Q_{WP})/(1 - Q_{WP}). \tag{14}$$

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,

$$\gamma = (\omega_{WI} \cdot \epsilon_{AI} \cdot \beta_{AP} \cdot \delta_{AS}) \tag{15}$$

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,

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

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

$$\alpha_{WP} = (\epsilon_{AI}) \cdot (\omega_{WI}) \quad \sigma_{WS} = (\beta_{AP}) \cdot (\alpha_{WP}) \quad \gamma = (\delta_{AS}) \cdot (\sigma_{WS}). \tag{17}$$

Given strict nesting of the Q -values, and within the constraints of Eqs (7, 8, 13 and 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 D^C coding, if any existing allele (within that population) is replaced by a novel allele (for that population), the within-population diversity (α_{WP1}) will increase. That also increases the within-species diversity (σ_{WSA}) of the species, within which (P_1) is nested. If other populations within (S_A) show some genetic overlap but do not have this novel variant, the among-populations diversity will also increase. For D^R coding, ‘degree of difference’ also matters, and if the novel variant in (P_1) is beyond the ‘size range’ of previously represented alleles in (P_1), the internal diversity (α_{WP1}) of that population will increase. If it is also beyond the size range of the species (S_A), within which it is nested, so will be (σ_{WS}) and (β_{AP}), etc. Our estimation protocols ensure that all of our diversity estimates meet the desiderata.

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 |

<https://doi.org/10.1371/journal.pone.0185499.t003>

Maximum and [0,1]—Scaled diversity

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

$$1 \leq \gamma \leq \gamma^* \quad 1 \leq \sigma_{WS} \leq \sigma_{WS}^* \quad 1 \leq \alpha_{WP} \leq \alpha_{WP}^* \quad 1 \leq \omega_{WI} \leq \omega_{WI}^*,$$

and

$$1 \leq \delta_{AS} \leq \delta_{AS}^* \quad 1 \leq \beta_{AP} \leq \beta_{AP}^* \quad 1 \leq \epsilon_{AI} \leq \epsilon_{AI}^*. \tag{18}$$

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),

$$\gamma^{\sim} = (Q/Q^*) \quad (\sigma_{WS}^{\sim}) = (Q_{WS}/Q_{WS}^*) \quad (\sigma_{WS}^{\sim}) = (Q_{WS}/Q_{WS}^*) \quad (\omega_{WI}^{\sim}) = (Q_{WI}/Q_{WI}^*),$$

and

$$(\delta_{AS}^{\sim}) = (Q_{AS}/Q_{AS}^*) \quad (\beta_{AP}^{\sim}) = (Q_{AP}/Q_{AP}^*) \quad (\epsilon_{AI}^{\sim}) = (Q_{AI}/Q_{AI}^*). \tag{19}$$

Returning 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 could 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 \mathbf{D} for the N individuals ($2N$ 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 $(\delta_{AS}^{\sim}) > 0$ or diversity among populations $(\beta_{AP}^{\sim}) > 0$ within them, or even (in some cases) whether there is diversity among individuals $(\epsilon_{AI}^{\sim}) > 0$ 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 (β_{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 (δ_{AS}^{\sim}) and downward to the among-individual (ϵ_{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 (σ_{WS}^{\sim}) are credibly homogeneous from species to species, or (b) whether the within-population diversity estimates (α_{WP}^{\sim}) are credibly homogeneous from population to population (within, or even among species), or even (c) whether the within-individual diversity estimates (ω_{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 values is tantamount to a Bartlett’s test [51] of homogeneity of the corresponding within-species variances. The same equivalence applies at the within-population and within-individual levels. Failure of any of our within-stratum homogeneity tests would provide signals of differential demographic, ecological and/or evolutionary pressures that have shaped such diversity in different fashions or to different degrees within different sampling strata.

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

Diversity analysis of paired *Antechinus* congeners

Here we illustrate these new tools with the Australasian marsupial genus *Antechinus*, comprised of small ground-dwelling and climbing predators of forests, woodlands and heathlands. Morphological and phylogenetic research in recent decades has identified several previously unrecognized species-level splits within the genus. Both *A. stuartii* and *A. agilis* were once viewed as a single species (*A. stuartii*), but recent research has indicated that they are separate

species, with a geographic break approximately 200 km south of Sydney, New South Wales [54]. Based on nuclear (*IRBP*, *RAG1*, *bFib7*) and mitochondrial (*cyt-b*, 12sRNA, 16sRNA) sequence analysis [55], these congeners are thought to have diverged in the early Pliocene. Much of the published life history research on *Antechinus* was conducted within the range of *A. agilis*, predating recognition of two species, but *A. stuartii* is demographically quite 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.

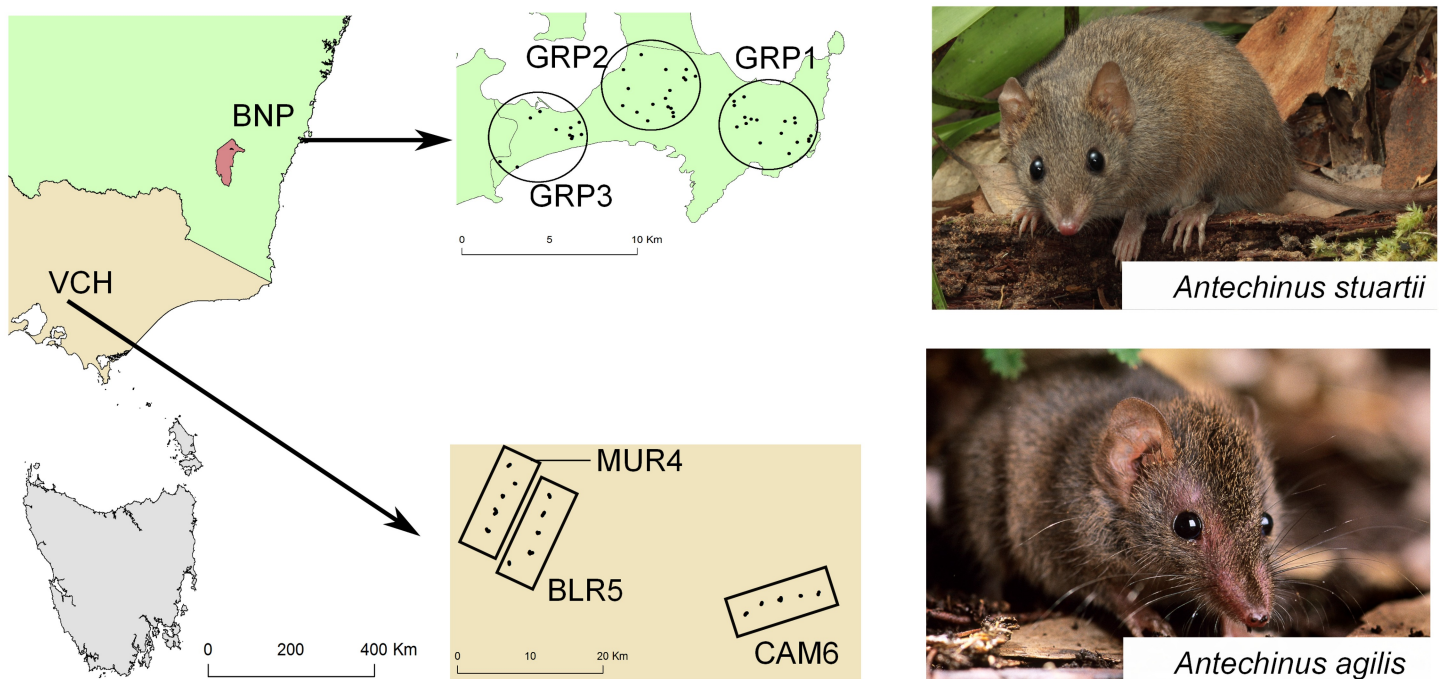


Fig 1. Map of the study locations for *Antechinus stuartii* (Booderee National Park, New South Wales) and *A. agilis* (Victoria Central Highlands, Victoria), Australia. Photo credits go to Stephen Mahony and Esther Beaton, respectively.

<https://doi.org/10.1371/journal.pone.0185499.g001>

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).

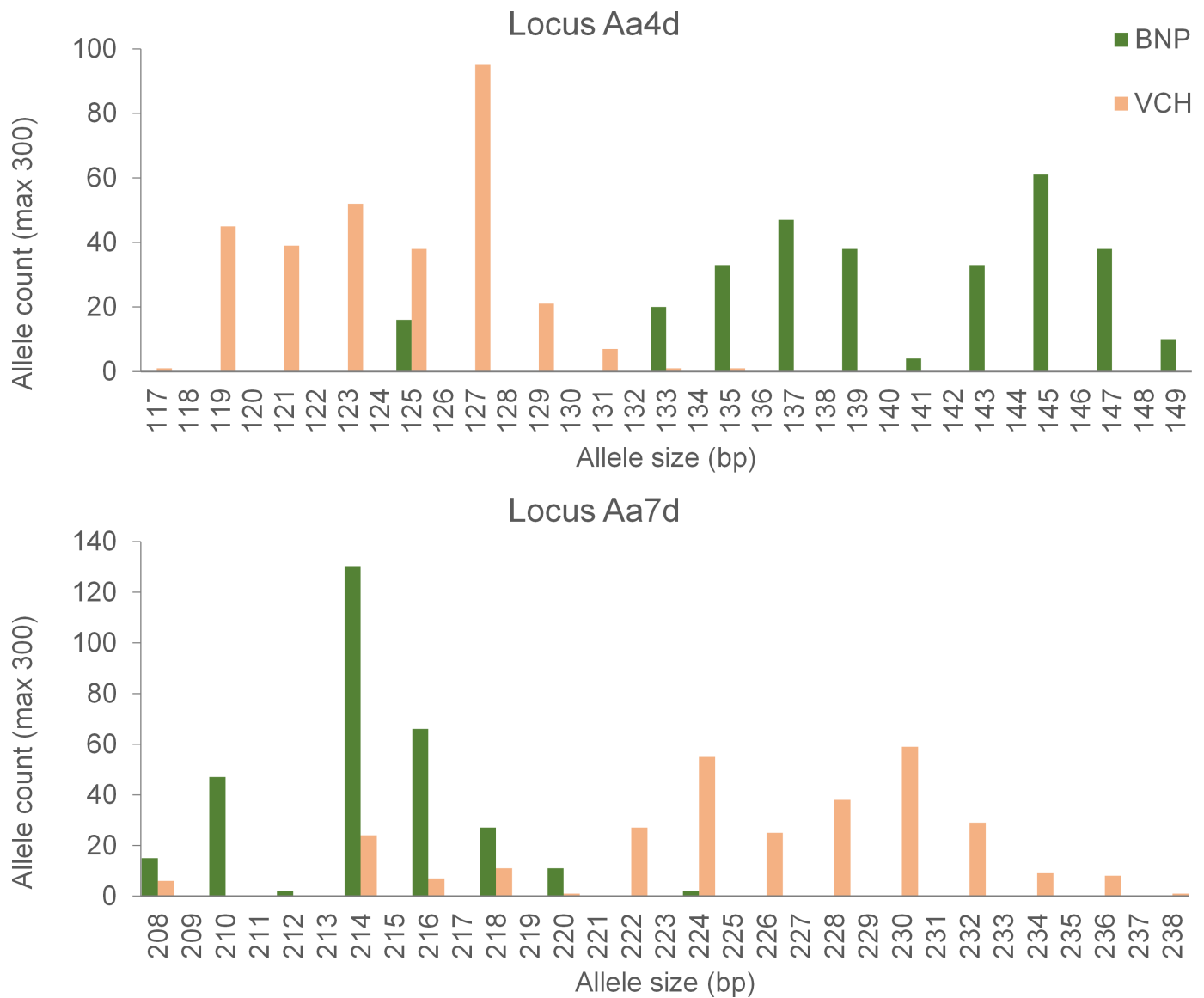


Fig 2. A pair of typical two-nucleotide step SSRs (Aa4d and Aa7d) for *Antechinus stuartii* (in Booderee National Park (BNP), New South Wales) and *A. agilis* (in Victoria Central highlands (VCH), Victoria) Australia.

<https://doi.org/10.1371/journal.pone.0185499.g002>

Different is different (D^C) and degree of difference (D^R) coding

We used traditional (0 vs 1) coding for our (D^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 D^C matrix by ($L = 8$), reducing each matrix to (average) single-locus form and ensuring that $[0 \leq Q < 1]$, so that all D^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 D^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_{max}^2) for that locus. Finally, we added the eight single-locus matrices together, and we divided each element of that summation by ($L = 8$), ensuring that all D^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 (α_{WP}), as well as separate species averages and a pooled study-wide average for both D^C and D^R coding (Table 4). The geographically most isolated populations have less internal diversity within both species, ($\alpha_{p1} < \alpha_{p2} < \alpha_{p3}$) within *A. stuartii* (BNP) and ($\alpha_{p4} > \alpha_{p5} > \alpha_{p6}$) 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 D^C or D^R coding. Having scaled D^C elements by ($d_{max}^2 = 1$) and those for D^R by ($d_{max}^2 = 275$), all of our within-population components are an order of magnitude smaller for D^R than they are for D^C coding (Table 4).

Partitioning diversity along the taxonomic hierarchy

Study-wide (and within-species) diversity cascades are presented for both D^C and 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

Table 4. Scaled within-population allelic diversity (α_{WP}) for both D^C and 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.

| <i>Antechinus stuartii</i> (BNP) | | | D^C Coding | <i>Antechinus agilis</i> (VCH) | | |
|--|-------|-------|--|---------------------------------------|-------|-------|
| GRP1 | GRP2 | GRP3 | [0,1]-Scaled Diversity | MUR4 | BRL5 | CAM6 |
| 0.435 | 0.453 | 0.476 | (α_{WP}) Estimates | 0.841 | 0.836 | 0.831 |
| Average & Bartlett’s Test | | | Average & Bartlett’s Test | Average & Bartlett’s Test | | |
| $\alpha_{WP} \sim = 0.455 (P < 0.041)$ | | | $\alpha_{WP} \sim = 0.645 (P < 0.001)$ | $\alpha_{WP} \sim = 0.836 (P > 0.64)$ | | |
| <i>Antechinus stuartii</i> (BNP) | | | D^R Coding | <i>Antechinus agilis</i> (VCH) | | |
| GRP1 | GRP2 | GRP3 | [0,1]-Scaled Diversity | MUR4 | BRL5 | CAM6 |
| 0.031 | 0.036 | 0.041 | (α_{WP}) Estimates | 0.072 | 0.073 | 0.069 |
| Average & Bartlett’s Test | | | Average & Bartlett’s Test | Average & Bartlett’s Test | | |
| $\alpha_{WP} \sim = 0.036 (P < 0.07)$ | | | $\alpha_{WP} \sim = 0.054 (P < 0.001)$ | $\alpha_{WP} \sim = 0.071 (P > 0.87)$ | | |

<https://doi.org/10.1371/journal.pone.0185499.t004>

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

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

<https://doi.org/10.1371/journal.pone.0185499.t005>

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

These two species are substantially divergent, with ($\delta_{AS}^{\sim} = 0.891, P < 0.001$) for D^C coding and ($\delta_{AS}^{\sim} = 0.150, P < 0.001$) for D^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 D^C ($\sigma_{WS-VCH}^{\sim} = 0.841 > \sigma_{WS-BNP}^{\sim} = 0.461, P < 0.001$) and for D^R ($\sigma_{WS-VCH}^{\sim} = 0.071 > \sigma_{WS-BNP}^{\sim} = 0.036, P \leq 0.001$) 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, D^C coding yields the expected pattern ($\beta_{AP-VCH}^{\sim} = 0.088, P \leq 0.001$) and ($\beta_{AP-BNP}^{\sim} = 0.025, P \leq 0.001$), but population subdivision is virtually nil ($\beta_{AP-VCH}^{\sim} = 0.001 = \beta_{AP-BNP}^{\sim}$) within either species for D^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 (σ_{AS}^{\sim}) component is an order of magnitude larger than the among-population (β_{AP}^{\sim}) component for 'different is different' (D^C) coding, however, it is two orders of magnitude larger for 'degree of difference' (D^R) coding. The nSSR 'ladder offset' between these congeners (Fig 2) constitutes a compelling diversity signature of long-term evolutionary separation.

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 ($2N$) alleles for each genetic locus, and packed those into a square distance

matrix, using both \mathbf{D}^C and \mathbf{D}^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}^C) and (\mathbf{D}^R) fashion. There is large (phyletic) diversity between the two species, but about twice as much diversity within *A. agilis* as within *A. stuartii*. 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, the ratio of among-species to among-populations diversity was an order of magnitude larger for \mathbf{D}^R than for \mathbf{D}^C coding, with minor allele frequency divergence (but no ladder shifts) among populations within either species, coupled with major ladder shifts between the two species.

Scaling considerations

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

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

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

Statement on animal usage

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

Accessibility arrangements

The *Antechinus* data are archived in Excel workbook form, along with listings of the QDiver results extracted from GenAlEx6.51 (<http://biology.anu.edu.au/GenAlEx/>). D^C data and analyses are presented in [S5 Appendix](#), and D^R data and analyses are presented in [S6 Appendix](#).

Supporting information

S1 Appendix. Partitioning within-population diversity into sub-components.
(PDF)

S2 Appendix. Scaling diversity components [0,1].
(PDF)

S3 Appendix. Homogeneity testing of within-stratum diversity components.
(PDF)

S4 Appendix. Laboratory microsatellite protocols.
(PDF)

S5 Appendix. Analysis of Rao Diversity (QDiver) for D^C Matrix.
(XLSX)

S6 Appendix. Analysis of Rao Diversity (QDiver) for D^R Matrix. (XLSX)

Acknowledgments

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

Author Contributions

Conceptualization: Peter E. Smouse, Rod Peakall.

Data curation: Sam C. Banks.

Formal analysis: Peter E. Smouse, Sam C. Banks, Rod Peakall.

Funding acquisition: Peter E. Smouse, Sam C. Banks, Rod Peakall.

Investigation: Peter E. Smouse, Rod Peakall.

Methodology: Peter E. Smouse, Rod Peakall.

Project administration: Peter E. Smouse.

Resources: Sam C. Banks.

Software: Rod Peakall.

Validation: Peter E. Smouse, Rod Peakall.

Visualization: Peter E. Smouse, Sam C. Banks.

Writing – original draft: Peter E. Smouse.

Writing – review & editing: Peter E. Smouse, Sam C. Banks, Rod Peakall.

References

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. <https://doi.org/10.1038/sj.hdy.6800674> PMID: 15940275
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. PMID: [8138170](#)
9. Bandelt H-J, Forster P, Sykes BC, Richards MB, Mitochondrial portraits of human populations using median networks. *Genetics*. 1995; 141: 743–753. PMID: [8647407](#)
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. PMID: [14156929](#)
14. Ohta T, Kimura M, A model of mutation 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. PMID: [838268](#)
17. Felsenstein J, Phylogenies and quantitative characters. *Annual Rev Ecol Syst*. 1988; 19: 445–471.
18. Goldstein DB, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW, An evaluation of genetic distances for use with microsatellite loci. *Genetics*. 1995; 139: 463–471. PMID: [7705647](#)
19. Slatkin M, A measure of population subdivision based on microsatellite allele frequencies. *Genetics*. 1995; 139: 457–462. PMID: [7705646](#)
20. Jost L, Partitioning diversity into independent alpha and beta components. *Ecology*. 2007; 88: 2427–2439 PMID: [18027744](#)
21. Allen B, Kon M, Bar-Yam Y, A new phylogenetic diversity measure generalizing the Shannon index and its application to Phyllostomid bats. *American Nat*. 2009; 174: 236–243.
22. Sherwin WB, Entropy and information approaches to genetic diversity and its expression: genomic geography. *Entropy*. 2010; 12: 1765–1798.
23. Leinster T, Cobbold CA, Measuring diversity: the importance of species similarity. *Ecology*. 2012; 93: 477–488. PMID: [22624203](#)
24. Rao CR, Analysis of diversity: A unified approach. In: *Statistical Decision Theory and Related Topics III*, Vol. 2, Gupta SS, Berger JO, eds. Academic Press, NY, 1982a: pp. 235–250.
25. Rao CR, Diversity and dissimilarity coefficients: a unified approach. *Theoretical Pop Biol*. 1982; 21: 24–43.
26. Rao CR, Diversity, its measurement, decomposition, apportionment and analysis. *Sankhya*. 1982; 44: 1–21.
27. Rao CR, Gini-Simpson index of diversity: A characterization, generalization and applications. *Util Math*. 1982; 21: 273–282.
28. Rao CR, Rao's axiomatization of diversity measures. *Encyclopedia of Statistical Sciences*, Kotz S, Johnson NL (eds). Wiley, New York. 1986; pp. 614–617.
29. Rao CR, Quadratic entropy and analysis of diversity. *Sankhyā: 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. <https://doi.org/10.1371/journal.pone.0100014> PMID: 25000299
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
39. Anderson MJ, A new method for non-parametric multivariate analysis of variance. *Australian Ecol*. 2001; 26: 32–46.
40. Anderson MJ, PERMANOVA: a FORTRAN computer program for permutational multivariate analysis of variance. Dept Statistics, Univ Auckland, New Zealand. 2005.
41. McArdle BH, Anderson MJ, Fitting multivariate models to community data: A comment on distance based redundancy analysis. *Ecology*. 2001; 82: 290–297.
42. Nievergelt CM, Libiger O, Schork NJ, Generalized analysis of molecular variance. *PLoS Genet*. 2007; 3(4): e51. <https://doi.org/10.1371/journal.pgen.0030051> PMID: 17411342
43. Coulson TN, Pemberton JM, Albon SD, Beaumont M, Marshall TC, Slate J, et al., Microsatellite reveals heterosis in red deer. *Proc Roy Soc Lond B*. 1998; 256: 489–495.
44. Otter KA, Stewart IRK, McGregor PK, Terry AMR, Dabelsteen T, Burke T, Extra-pair paternity among great tits *Parus major* following manipulation of male signals. *J Avian Biol*. 2001; 32: 338–344.
45. Bruvo RU, Michiels NK, D'Souza TG, Schulenburg H, A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecol*. 2004; 13: 2101–2106.
46. Gower JC, Legendre P, Metric and Euclidean properties of dissimilarity coefficients. *Journal Classif*. 1986; 3: 5–48.
47. Banks SC, McBurney L, Blair D, Davies ID, Lindenmayer DB, Where do animals come from during post-fire population recovery? Implications for ecological and genetic patterns in post-fire landscapes. *Ecography*. 2017; <https://doi.org/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, GST and its relatives do not measure differentiation. *Molecular Ecol*. 2008; 17: 4015–4026.
50. Jost L, *D* vs. GST: response to Heller and Siegmund (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, GenAIEx 6: Genetic Analysis in Excel. Population genetic software for teaching and research. *Molecular Ecol Notes*. 2006; 6: 288–295.
53. Peakall R, Smouse PE, GenAIEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*. 2012; 28: 2537–2539. <https://doi.org/10.1093/bioinformatics/bts460> PMID: 22820204
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, Scott 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 and genetic variation of a marsupial carnivore, *Antechinus agilis*. *Biological Conserv*. 2005; 122: 581–597.
58. Banks SC, Lindenmayer DB, Inbreeding avoidance, patch isolation and matrix permeability influence dispersal and settlement choices by male agile antechinus in a fragmented landscape. *J Animal Ecol*. 2014; 83: 515–524.
59. Banks SC, Dujardin M, McBurney L, Blair D, Barker M, Starting points for small mammal population recovery after wildfire: recolonisation or residual populations? *Oikos*. 2011; 120: 26–37
60. Davies ID, Cary GC, Landguth EL, Lindenmayer DB, Banks SC, Implications of recurrent disturbance for genetic diversity. *Ecology & Evol*. 2016; 6: 1181–1196.

61. Lindenmayer D, Blanchard W, MacGregor C, Barton P, Banks SC, Crane M, et al., Temporal trends in mammal responses to fire reveals the complex effects of fire regime attributes. *Ecological Appl.* 2016; 26: 557–573.
62. Gaggiotti OE, Lange O, Rassmann K, Gliddon C, A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecol.* 1999; 8: 1513–1520.
63. Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB, Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci (USA)*. 1994; 91: 3166–3170.
64. Fu Y-X, Chakraborty R, Simultaneous estimation of all the parameters of a stepwise mutation model. *Genetics*. 1998; 150: 487–497. PMID: [9725863](https://pubmed.ncbi.nlm.nih.gov/9725863/)
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 (<https://doi.org/10.7554/eLife.12469>). PMID: [27187613](https://pubmed.ncbi.nlm.nih.gov/27187613/)
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. <https://doi.org/10.1038/hdy.2009.74> PMID: [19623208](https://pubmed.ncbi.nlm.nih.gov/19623208/)
71. Coop G, Witonsky D, Di Rienzo A, Pritchard JK, Using environmental correlations to identify loci underlying local adaptation. *Genetics*. 2010; 185: 1411–1423. <https://doi.org/10.1534/genetics.110.114819> PMID: [20516501](https://pubmed.ncbi.nlm.nih.gov/20516501/)
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 phylogenetic 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; 16: 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. <https://doi.org/10.1111/j.1541-0420.2008.01010.x> PMID: [18355386](https://pubmed.ncbi.nlm.nih.gov/18355386/)
85. Hardy OJ, Jost L, Interpreting and estimating measures of community phylogenetic structuring. *Journal Ecol.* 2008; 96: 849–852
86. Ricotta C, Di Nepi M, Guglietta D, Celesti-Grapow L, Exploring taxonomic filtering in urban environments. *J Vegetation Sci.* 2008; 19: 229–238.

87. Villéger S, Mouillot D, Additive partitioning of diversity including species differences: a comment on Hardy & Senterre (2007). *J Ecology*. 2008; 96: 845–848.
88. Eastman JM, Paine CET, Hardy OJ, SpacodiR: structuring of phylogenetic diversity in ecological communities. *Bioinformatics*. 2011; 27: 2437–2438. <https://doi.org/10.1093/bioinformatics/btr403> PMID: 21737436
89. Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L, Widespread signals of convergent adaptation to high altitude in Asia and America. *Amer J Hum Genet*. 2014; 95: 394–407. <https://doi.org/10.1016/j.ajhg.2014.09.002> PMID: 25262650
90. Chao A, Chiu C-H, Bridging the variance and diversity decomposition approaches to beta diversity via similarity and differentiation measures. *Methods Ecol Evolution*. 2016; 7: 919–928.