BIODIVERSITY RESEARCH

Partitioning diversity for conservation analyses

Lou Jost[1]*, Philip DeVries[2], Thomas Walla[3], Harold Greeney[4], Anne Chao[5] and Carlo Ricotta[6]

[1]*Via a Runtun, Baños, Tungurahua Province, Ecuador,* [2]*Department of Biological Sciences, University of New Orleans, New Orleans, LA 70148, USA,* [3]*Department of Biology, Mesa State College, 1100 North Avenue, Grand Junction, CO 81501, USA,* [4]*Yanayacu Biological Station & Center for Creative Studies, Cosanga, Napo Province, Ecuador,* [5]*Institute of Statistics, National Tsing Hua University, Hsin-Chu 30043, Taiwan,* [6]*Department of Plant Biology, University of Rome "La Sapienza", Piazzale Aldo Moro, 5, 00185 Rome, Italy*

ABSTRACT

Aim Differentiation of sites or communities is often measured by partitioning regional or gamma diversity into additive or multiplicative alpha and beta components. The beta component and the ratio of within-group to total diversity (alpha/gamma) are then used to infer the compositional differentiation or similarity of the sites. There is debate about the appropriate measures and partitioning formulas for this purpose. We test the main partitioning methods, using empirical and simulated data, to see if some of these methods lead to false conclusions, and we show how to resolve the problems that we uncover.

Location South America, Ecuador, Orellana province, Rio Shiripuno.

Methods We construct sets of real and simulated tropical butterfly communities that can be unambiguously ranked according to their degree of differentiation. We then test whether beta and similarity measures from the different partitioning approaches rank these datasets correctly.

Results The ratio of within-group diversity to total diversity does not reflect compositional similarity, when the Gini–Simpson index or Shannon entropy are used to measure diversity. Additive beta diversity based on the Gini–Simpson index does not reflect the degree of differentiation between *N* sites or communities.

Main conclusions The ratio of within-group to total diversity (alpha/gamma) should not be used to measure the compositional similarity of groups, if diversity is equated with Shannon entropy or the Gini–Simpson index. Conversion of these measures to effective number of species solves these problems. Additive Gini–Simpson beta diversity does not directly reflect the differentiation of *N* samples or communities. However, when properly transformed onto the unit interval so as to remove the dependence on alpha and *N*, additive and multiplicative beta measures yield identical normalized measures of relative similarity and differentiation.

Keywords

Additive partitioning, alpha diversity, beta diversity, differentiation, multiplicative partitioning, Neotropical butterflies.

*Correspondence: Lou Jost, South American Explorers Club, Apdo 17-21-431, Eloy Alfaro, Quito, Ecuador.
E-mail: loujost@yahoo.com

INTRODUCTION

To effectively conserve regional biodiversity, conservationists need to know how diversity is distributed geographically within the region. Does the region consist of many distinct communities, or is it homogeneous? How much does each community contribute to the regional diversity? How different are the communities? The answers to these questions

determine how conservation resources should be allocated among sites in the region.

Compositional differentiation and similarity of groups is often analysed by partitioning a regional or 'gamma' diversity measure into within- and between-group components, 'alpha' and 'beta'. For a given number of groups, biologists often use the between-group or beta component of diversity to measure the degree of compositional differentiation between the groups

(Whittaker, 1972; Veech *et al.*, 2002; Magurran, 2004). Biologists also generally use the ratio of within-group diversity to total diversity to judge the compositional similarity of the groups (Lande, 1996; Veech *et al.*, 2002). These measures are used to guide regional conservation strategies (e.g. Gering *et al.*, 2003; Summerville *et al.*, 2003; Chandy *et al.*, 2006). These same forms of reasoning are also important in conservation genetics, where the diversity of interest is allele diversity rather than species diversity.

If our regional conservation strategies are to have a sound basis in science, it is important that these inferences about similarity and differentiation be correct. Are these inferences generally valid, or do they depend on the diversity measure and partitioning scheme used? If these inferences are not valid, our conservation recommendations will not withstand the scrutiny of fellow scientists, development agencies, and policy makers, and may lead to the extinction of species or unique ecosystems.

Additive partitioning framework

A popular method of partitioning diversity in ecology and genetics is additive partitioning of concave 'diversity' measures (Nei, 1973; Lande, 1996; Veech *et al.*, 2002; Couteron & Pelissier, 2004). In its standard form, this method is applied to species richness,

$$H_{SR} = \sum_{i=1}^{S} p_i^0, \tag{1}$$

Shannon entropy,

$$H_{Sh} = -\sum_{i=1}^{S} p_i \log p_i, \tag{2}$$

and the Gini–Simpson index:

$$H_{GS} = 1 - \sum_{i=1}^{S} p_i^2. \tag{3}$$

The Gini–Simpson index is also the standard measure of 'diversity' in genetics, where it is called heterozygosity. In the above expressions, p_i is the population frequency of the ith species or allele, and S is the total number of species or alleles in the whole population.

These complexity measures are additively partitioned into additive within- and between-group components:

Gamma = alpha + beta.

Alpha is the mean index value for the groups. Alpha and gamma are calculated directly from the data, and beta is derived from them by subtraction. In this framework, regional homogeneity (or community similarity) is given by the ratio of the within-group component to the total diversity measure, alpha/gamma (Lewontin, 1972; Lande, 1996; Veech *et al.*, 2002). This has a maximum value of unity, which is supposed to indicate complete homogeneity (all groups are identical).

Geneticists use this same framework. They subtract the homogeneity measure alpha/gamma from unity, obtaining the standard measure of genetic differentiation, G_{ST} (Nei, 1973):

$$G_{ST} = 1 - (\text{within-group 'diversity'})/(\text{total 'diversity'})$$
$$= (\text{additive between-group 'diversity'})/(\text{total 'diversity'}).$$

These measures are often used to form recommendations about allocation of conservation resources in ecology and genetics (e.g. Veech *et al.*, 2002 and papers cited therein; Summerville *et al.*, 2003; Ribeiro *et al.*, 2008; Torres *et al.*, 2003; Petit *et al.*, 1998) and to test ecological hypotheses (Summerville *et al.*, 2006; Patzkowsky & Holland, 2007).

Problems with the additive partitioning framework

Recent theoretical analyses of the mathematics of diversity in ecology and population genetics (Jost, 2006, 2007, 2008, 2009; Hardy & Jost, 2008) have shown that the additive partitioning of Shannon entropy and especially the Gini–Simpson index is inadequate in some respects. First, its fundamental premise, that diversity can be equated with complexity measures such as Shannon entropy and the Gini–Simpson index (heterozygosity), leads to logical contradictions in conservation analyses (Jost, 2009). Second, the additive between-group or beta component of some measures of complexity necessarily approaches zero when diversity is high, regardless of the degree of differentiation of the groups (Jost, 2006, 2007). Third, for these complexity measures, the 'similarity' alpha/gamma necessarily approaches unity when diversity is high, regardless of the compositional similarity of the groups (Jost, 2006, 2007).

Another problem with some of these measures arises at the other end of the scale, when diversity is very low. When there are many communities, each dominated by a single species, the additive similarity measure alpha/gamma based on the Gini–Simpson index will be approximately or exactly zero, supposedly indicating maximal differentiation. This would make sense if the dominant species were different in each community, but the measure can still be exactly zero, supposedly indicating no similarity between groups, even if all but one of the communities are identical. (If every community is dominated by the same species, the measure is undefined.) This is the ecological analogue of a problem with G_{ST} first pointed out by Wright (1978) and Gregorius (1987); it occurs because G_{ST} and its ecological analogue is really a fixation index, not a measure of relative differentiation as often thought.

Multiplicative framework

The problems caused by using Shannon entropy or the Gini–Simpson index to measure diversity were recognized by Whittaker (1972) and MacArthur (1965, 1972). They arise because these measures are nonlinear with respect to species addition, even when all species are equally common. All else being equal, each added species leads to a smaller increment in

'diversity' than the species added before it. In the case of the Gini–Simpson index, the effect is so extreme that the diversity measure approaches an asymptote of unity, no matter how many species are added. MacArthur (1972) and Hill (1973) resolved these problems by converting Shannon entropy and the Gini–Simpson index to 'effective number of species', which has the same linear metric as species richness. The effective number of species of a community is the species richness of a perfectly even community (all species equally common) with the same diversity as the original community.

Shannon entropy is converted to effective number of species by taking its exponential, $\exp(H_{Sh})$, and the Gini–Simpson index is converted by the transformation $1/(1-H_{GS})$, which is the inverse Simpson concentration $1/\left(\sum_{i=1}^{S} p_i^2\right)$. Hill (1973) gave a one-parameter family of measures with the same linear metric as species richness,

$$^{q}D = \left(\sum_{i=1}^{S} p_i^q\right)^{1/(1-q)} \tag{4}$$

where q is the parameter that determines the measure's sensitivity to species frequencies. Jost (2007) called ^{q}D a 'true diversity' because it has the mathematical properties needed to make common forms of reasoning about diversity valid. Jost (ibid.) derived the partitioning of these 'true diversities', and any measures derived from them (such as Shannon entropy and the Gini–Simpson index), into independent within- and between-group components. This partitioning is unique for any given measure. For effective number of species (equation 4), it is

gamma = alpha × beta,

which is Whittaker's (1972) multiplicative partitioning. The alpha diversity is calculated by converting the alpha of the additive approach to effective number of species. In general, if we are interested in a beta that is monotonically related to the degree of differentiation between N groups, alpha and gamma should be calculated with all group weights set to $1/N$, unless $q = 0$ (species richness) or $q = 1$ (exponential of Shannon entropy). [See Appendix S1 in Supporting Information, and Jost, 2007 for formulas and further explanation. Note that there is a misplaced bracket in the expression for gamma above (Equation 16a of Jost, 2007)].

Comparing differentiation and similarity measures of the additive and multiplicative frameworks

In this study, we extend our earlier studies by examining the core meanings of 'differentiation' and 'similarity' in conservation biology. We identify a basic criterion which must be satisfied by all admissible measures of compositional differentiation and similarity in this field. Any measure that fails this test cannot be interpreted as a measure of compositional differentiation or similarity. We test additive and multiplicative beta to see whether, for a fixed number of communities or samples, they meet our criterion for a measure of differentiation. We also test additive and multiplicative alpha/gamma to

see whether these meet our criterion for a measure of compositional similarity. We illustrate the problems revealed by our tests, and give their solutions, by presenting the first comparative analysis of a real dataset using the two partitioning approaches. Finally, we show that normalized measures of relative differentiation provide a partial bridge between traditional additive and multiplicative approaches, and we derive an alternative form of additive partitioning from the multiplicative approach.

METHODS

Criterion for rejecting measures of compositional differentiation or similarity

Differentiation measures may legitimately disagree in the way they rank datasets, as they may emphasize dominant species, or they may ignore species frequencies altogether. Furthermore, different measures may weigh each community of a dataset differently depending on its size, causing more ambiguity. Because of these liberties, in general there can be no single 'correct' ranking of datasets in terms of their differentiation. However, for datasets that control for these variables, such an absolute ordering is possible, and legitimate measures of differentiation for conservation purposes must preserve this absolute ranking.

To eliminate ambiguity in ranking caused by varying emphasis on dominant species or communities, we designed test datasets in which all communities are equally large and equally diverse, and all species within a community are equally common. We imposed complete symmetry by assuming that there is a set of k species shared by all communities, and that no other species are shared by any communities. The point of using these artificial and unrealistic datasets is to test metrics under conditions so clear that there would be no doubt as to how they should behave. If measures do not work under these clean conditions, they will not be interpretable under the messy and complex conditions of the real world.

Suppose we have such a completely even and symmetric multi-community dataset. What can we do to this dataset to unambiguously increase (or at least not decrease) its differentiation? We could add a new species (one not previously present in any community) to Community 1, but doing just this would imbalance the communities, and someone might argue that for some purposes, communities should be weighed differently depending on their diversity. To eliminate the possibility of such counter-explanations for unexpected behaviours, we also add a different new species to Community 2, another to Community 3, etc., until each community has gained an additional 'endemic' species not shared by any other community in the dataset. Suppose that we add these new species with the same abundances as the pre-existing species in each community, so that the augmented communities are still completely even and symmetric. As we have added a unique endemic species to each community, avoiding dominance effects and unequal community sizes and asymmetric overlaps,

the differentiation of the new dataset cannot possibly be less than that of the original dataset.

A sequence of communities with non-decreasing differentiation

To apply this criterion, we created a pair of communities that were initially identical (no differentiation). Each community initially had the same two species, 10 individuals of each species. Thus it satisfies our condition for a perfectly symmetric dataset. We then transformed the communities using the technique just described. At each step, we increased the differentiation of the communities by adding 10 individuals of a unique species to one of them, and 10 individuals of a different unique species to the other. We repeated this species-adding step 20 times, building a sequence of datasets.

Any differentiation measure that *decreased*, and any similarity measure that *increased*, as unique species are added in this way, would lead to obviously wrong conservation decisions. This is a *minimal* requirement for differentiation and similarity measures. As discussed later, there are other properties we might reasonably require of such measures. This is just the most basic and uncontroversial requirement that we can find. If a measure does not satisfy this criterion, it must be rejected, but if a measure does satisfy this criterion, it may still have other flaws that render it unsuitable.

A sequence of communities based on a real dataset

Real communities are not perfectly even, as in our artificial datasets. To test the measures in a real situation with uneven species-abundance distributions, we apply them to the Neotropical fruit-feeding butterfly guild, which has been intensively sampled and studied in the last two decades using a stratified sampling method first developed by DeVries (1988). The dataset we examine here, 'RS', consists of a 1-year sample of 1247 individuals from an intact Amazonian rain forest near the Rio Shiripuno, Parque Nacional Yasuni, Ecuador.

Table 1 The 10 most abundant species in the RS dataset. These 10 species contain 51% of the individuals in the RS dataset. With only one exception (*Catonephele orites*) the dominant species of each community are very rare in the other community. Less abundant species show similar patterns.

Species	Canopy	Understorey
Historis acheronta	100	1
Historis odius	58	3
Catonephele orites	45	31
Panacea divalis	30	4
Bia actorion	0	98
Morpho achilles	1	89
Taygetis sp-1	1	54
Nessea obrina	1	47
Euptychia sp.	0	36
Nessea hewitsonii	1	32

This butterfly dataset can be divided into two communities: canopy and understorey. As in previous studies (DeVries, 1988; DeVries *et al.*, 1997, 1999; DeVries & Walla, 2001; Walla *et al.*, 2004), this dataset demonstrates high differentiation between the canopy and understorey communities at the species, genus and subfamily levels. The abundances of the 10 most common species clearly show the high degree of differentiation in the dataset (Table 1); the less common species are also highly differentiated. A biologically meaningful beta should reflect this high degree of differentiation between canopy and understorey communities, and a biologically meaningful homogeneity measure should show that these two communities are dissimilar in composition (the tropical rain forest fruit-feeding butterfly guild is not homogeneous across canopy and understorey).

To create a sequence of communities with a clear ranking in terms of differentiation, relative to the actual dataset, we apply some simple transformations of the real data:

Dataset 1: Zero differentiation. We created a zero-differentiation dataset by replacing the understorey data of the RS dataset with a copy of the canopy data.

Dataset 2 (RS): Moderate to high differentiation. The actual Rio Shiripuno (RS) dataset, with high species diversity and high relative and absolute differentiation.

Dataset 3: Complete differentiation. We give new names to all shared species in the canopy community, so that canopy and understorey communities shared no species.

Dataset 4: Complete differentiation, with higher diversity. We quadrupled each of the two communities in Dataset 3, by adding to each community three copies of itself. We renamed each species in each copy. The result was a pair of completely differentiated communities (no shared species), each with four times the number of species as the corresponding community of Dataset 3.

All calculations used equal statistical weights for each community, so that beta reflects actual differences in community composition, without confounding this with community size differences (Jost, 2007; see Table S1 for the formulae used). We treated the species frequencies as if they were known exactly; we were not testing sampling effects but rather the fundamental mathematical and logical consistency of partitioning approaches.

Test 1: For a fixed number of communities, does beta measure differentiation?

A valid measure of differentiation must be non-decreasing for our sequence of perfectly symmetric artificial communities, as we increase the number of endemic species in each community. At each step, we calculated additive beta based on species richness, Shannon entropy and the Gini–Simpson index, and we calculated multiplicative beta based on species richness, the exponential of Shannon entropy and the inverse Simpson concentration. We graphed these additive and multiplicative beta values as a function of the number of endemic species added to each community (Fig. 1b,d). The same calculations
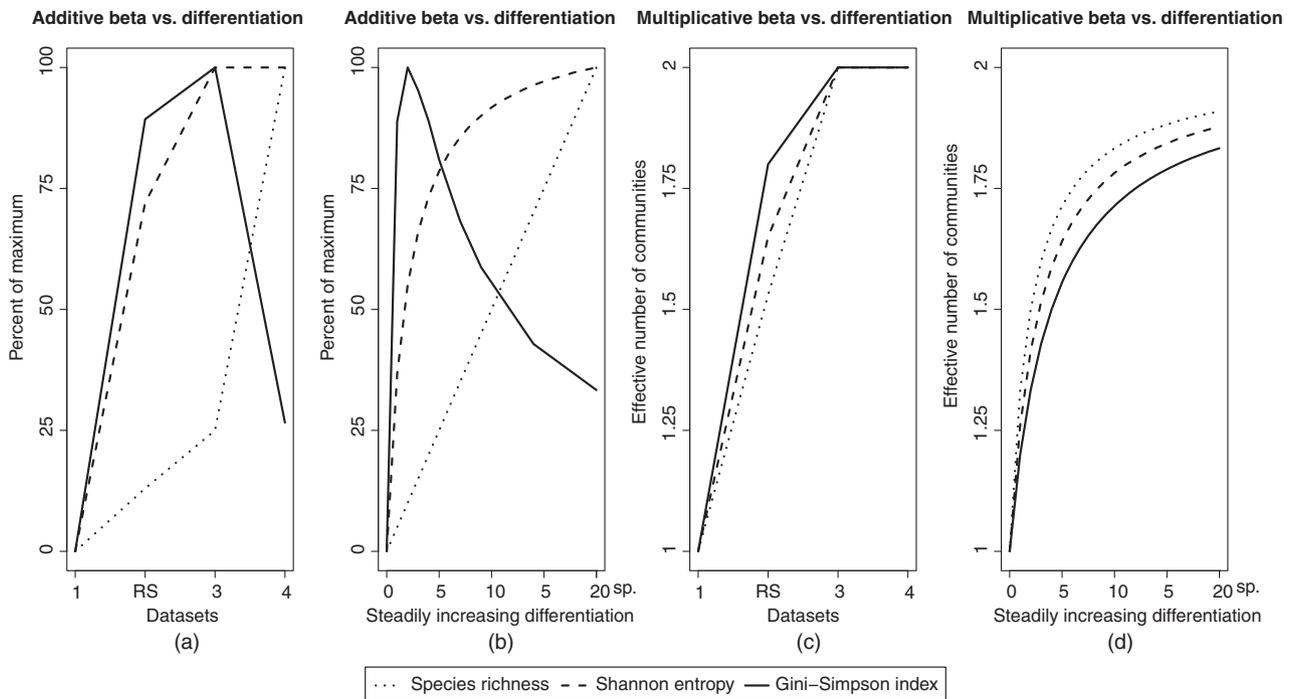
Figure 1 Additive and multiplicative beta versus differentiation. Additive beta (a) and multiplicative beta (c) for RS dataset and artificial Datasets 1, 3 and 4 (see Methods). Datasets increase in differentiation from left to right. Additive beta units differ for each index, so all were normalized by dividing by their maximum value on the interval. (b,d) Two communities are identical initially (at left in each graph), and become successively more differentiated as more unique species are added to each community in equal numbers (see Methods). *X*-axis is number of unique species added to each community, *Y*-axis is additive beta (b) and multiplicative beta (d).
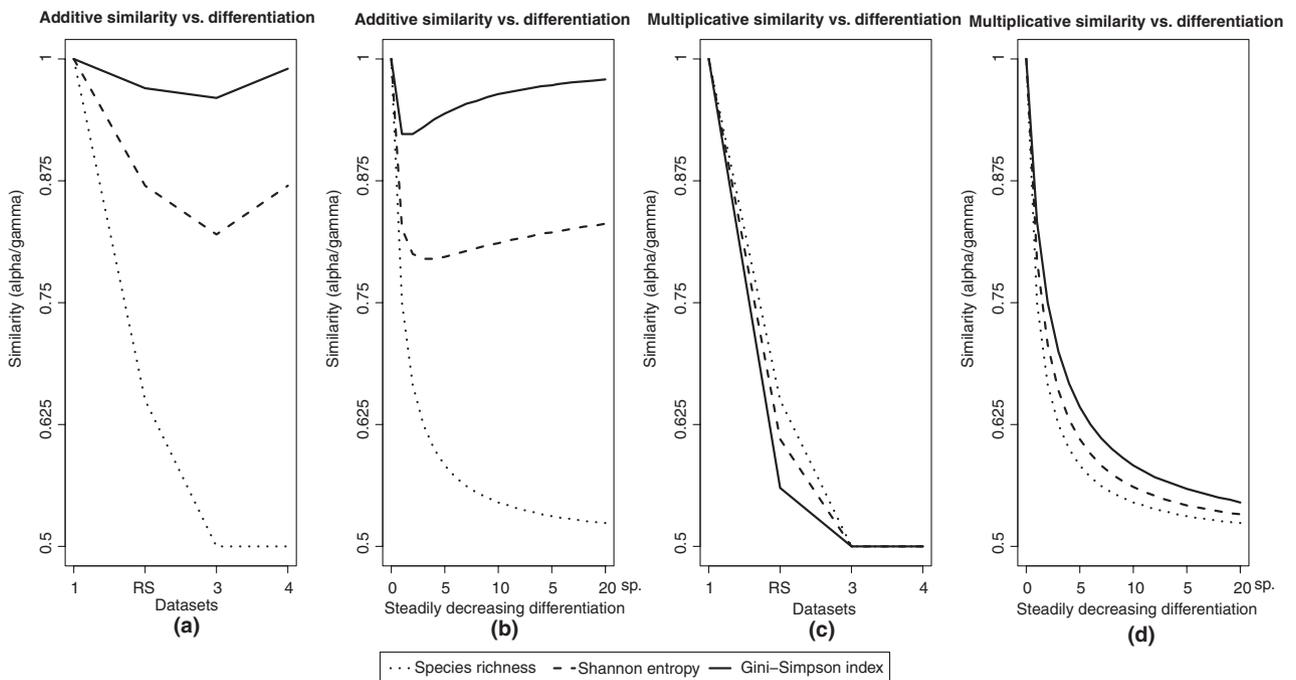

Figure 2 Additive and multiplicative similarity versus differentiation. Additive similarity (a) and multiplicative similarity (c) for RS dataset and artificial Datasets 1, 3 and 4 (see Methods). Datasets decrease in similarity from left to right. (b,d) Two communities are identical initially (at left in each graph), and become successively less similar as more unique species are added to each community in equal numbers (see Methods). *X*-axis is number of unique species added to each community; *Y*-axis is additive similarity (b) and multiplicative similarity (d).

were done using Datasets 1–4 based on the real butterfly data, and their additive and multiplicative beta values were graphed in Fig. 1(a,c).

Test 2: Does the homogeneity or similarity measure alpha/gamma decrease with increasing differentiation?

A valid measure of similarity must be non-increasing for our sequence of perfectly symmetric communities, as we increase the number of endemic species in each community. At each step, we calculated additive similarity alpha/gamma based on species richness, Shannon entropy and the Gini–Simpson index, and we calculated multiplicative similarity alpha/gamma based on species richness, the exponential of Shannon entropy and the inverse Simpson concentration. We graphed these additive and multiplicative similarity values as a function of the number of endemic species added to each community (Fig. 2b,d). The same calculations were done using Datasets 1–4 based on the real butterfly data, and their additive and multiplicative similarity values were graphed in Fig. 2(a,c).

Test 3: Are comparisons of similarity measures across indices meaningful?

It is well known that indices differ in their sensitivity to common and rare species (Keylock, 2005). Species richness is equally sensitive to rare and common species, the Gini–Simpson index and inverse Simpson concentration are most sensitive to the dominant species, and Shannon measures are intermediate (Hill, 1973; Jost, 2006). The similarity measure alpha/gamma will therefore have different sensitivities to rare and common species depending on which complexity measure is used.

We compared the ratio alpha/gamma of the real butterfly dataset RS across indices to see if the differences between them were informative (Table 2). In this test, which involves only real data, we calculated 95% confidence intervals (calculated as ± 2 SE centred at the estimator value) using a bootstrap approach, so that ecologists can compare these values with

those of other communities. The bootstrap resampling was performed with replacement. Extensive simulations have shown this to be the most accurate way to estimate the standard error of these estimators (Chao *et al.*, 2008).

RESULTS

Test 1: For a fixed number of communities, does beta increase with increasing differentiation?

Figure 1(b,d) shows the values of additive and multiplicative beta diversity for the artificial dataset, as endemic species are repeatedly added to each community in equal numbers (as described above). According to our criterion for a valid differentiation measure, these values cannot decrease as endemic species are added to each community. The additive Gini–Simpson beta fails this basic test, so it cannot be considered a measure of compositional differentiation. The same pattern emerges when beta is graphed for Datasets 1–4 based on the real butterfly data (Fig. 1a,c). Additive Gini–Simpson beta is smaller for the completely differentiated and highly diverse Dataset 4 than it is for the less diverse and less differentiated real butterfly dataset.

Test 2: Does similarity decrease with increasing differentiation?

Figure 2(b,d) shows the additive and multiplicative versions of the similarity measure alpha/gamma applied to the artificial dataset, as endemic species are repeatedly added to each community in equal numbers. According to our criterion, a valid measure of compositional similarity can never increase when endemic species are added to each community in this way. The additive Gini–Simpson and Shannon similarity measures fail this basic test; they can increase as real similarity decreases. Figure 2(b) shows the same pattern when the additive similarity measures are applied to Datasets 1–4 derived from the butterfly communities. For additive similarity alpha/gamma based on Shannon entropy or the Gini–Simpson index, the completely differentiated communities of Dataset 4

Table 2 Beta and similarity ± 2 SE for the RS dataset. Numbers in parentheses represent theoretical range of the value for two equally weighted communities. Uncertainties are obtained from a bootstrap approach and represent 2 SE.

Measure	Additive		Multiplicative	
	Beta	Theoretical range	Beta	Theoretical range
Species richness	30.5 ± 3.1	0–infinity	1.53 ± 0.06	1–2
Shannon	0.502 ± 0.03	0–0.6931	1.65 ± 0.05	1–2
Simpson	0.030 ± 0.004	0–0.5	1.80 ± 0.05	1–2
	Similarity	Theoretical range	Similarity	Theoretical range
Species richness	0.65 ± 0.03	0.5–1	0.65 ± 0.03	0.5–1
Shannon	0.87 ± 0.01	0–1	0.61 ± 0.02	0.5–1
Simpson	0.97 ± 0.004	0–1	0.56 ± 0.02	0.5–1

are incorrectly classified as more homogeneous (more similar to each other) than the moderately differentiated communities of the RS dataset (Fig. 2a).

Test 3: Are comparisons of similarity measures across indices meaningful?

The additive similarity measures strongly increase with increasing emphasis on dominant species. In contrast, the multiplicative measures strongly decrease with increasing emphasis on the dominant species. (See Fig. 2. It should be noted that the uncertainties were all < 0.03 so are not visible in the figure.)

DISCUSSION

Why do some additive partitioning measures fail Tests 1 and 2?

The anomalous additive partitioning test results have clear mathematical explanations (Jost, 2006, 2007). The first and most fundamental problem of the traditional additive approach is its premise that Shannon entropy or the Gini–Simpson index quantify conservation biology's concept of diversity. Conservation biologists treat diversity as something that can be conserved or lost, but Shannon entropy and the Gini–Simpson index are logically inconsistent when used in this way (Jost, 2009). Consider an idealized 'thought experiment' whose symmetry eliminates other possible explanations for contradictory behaviour (Jost, 2009). Suppose we are concerned with tree conservation on an archipelago of 20 islands of equal size and diversity. Suppose that no island shares species with any other. Suppose that the species on each island have the same abundance distribution as the tree abundances on Barro Colorado Island in Panama (Hubbell *et al.*, 2005). Conservation biologists might be asked how many of these islands, each with a completely distinct set of species, should be saved to conserve most of the diversity of the archipelago. If we equate diversity with the Gini–Simpson index, the 'diversity' of any single island is calculated to be 0.95, whereas the 'diversity' of the whole archipelago (pooling all islands) is calculated to be 0.998. This means we can preserve most of the regional 'diversity' by saving just one of the twenty unique islands; saving one fragment supposedly conserves 95% (0.95/0.998) of the regional 'diversity'. Yet at the same time, the percentage of regional 'diversity' *lost* in this scenario, using the same 'diversity' measure, is 99.9% (the pooled 'diversity' of the 19 sacrificed fragments is 0.997, which is 99.9% of the total 'diversity', 0.998). The same conservation plan both saves and sacrifices almost all of the archipelago's 'diversity'. By focusing on one or the other of these numbers, a conservation biologist could prove whatever he or she wanted about this or any other conservation plan. Similar self-contradictory results are found using Shannon entropy: saving one island is sufficient to save most of the archipelago's 'diversity', as each island contains 57% of the archipelago's

'diversity' (3.96/6.95). Yet the 19 sacrificed islands contain 99% of the archipelago's 'diversity' (6.90/6.95). More than half the 'diversity' is saved, and simultaneously almost all of it is lost [see Jost (2009), for further discussion of this example.] Analogous paradoxes can be constructed using alleles instead of species.

These measures are self-contradictory because they do not possess Hill's (1973) doubling property, known in economics as the replication principle (Hannah & Kay, 1977; Ricotta, 2008). The property can be stated as follows: If N equally large, completely distinct communities (no shared species) each have diversity X, then the pooled diversity of the communities is $N \cdot X$.

Any measure satisfying the replication principle (such as species richness, the exponential of Shannon entropy and the inverse Simpson concentration) gives the intuitively reasonable, unambiguous, self-consistent answer: saving one of the twenty islands conserves exactly 5% (1/20) of the regional diversity, and sacrificing the other 19 islands causes the loss of 95% (19/20) of the regional diversity. Using any such measure, biologists would arrive at the intuitive and logically consistent conclusion that we must protect more than half of the islands to protect most of the archipelago's diversity. For such measures, when there are N equally diverse, equally large, completely distinct communities, the diversity of the portion saved plus the diversity of the portion destroyed add up to the regional diversity. If conservation plans are to stand up to close scrutiny, they should be based on measures of diversity that obey the replication principle. This property is also required if the ratio alpha/gamma is to reflect compositional similarity between groups.

Another problem of the traditional additive approach is that the Gini–Simpson index, also called heterozygosity or Tsallis entropy of order 2, is not additive (Aczel & Daroczy, 1975; Tsallis, 1988). When an additive decomposition is imposed on it, the resulting alpha and beta components are not independent of each other. This is especially clear in high-diversity, high-differentiation ecosystems such as our butterfly test community. The alpha Gini–Simpson index can be close to unity for high-diversity ecosystems (for our butterfly data, it was 0.93). The gamma Gini–Simpson index is always greater than alpha, but it cannot exceed unity. This means that when local diversity is high, both alpha and gamma are very close to unity for this index, regardless of the degree of differentiation among communities. If beta equals gamma minus alpha, and both gamma and alpha are close to unity, then beta will be approximately zero, regardless of the degree of differentiation among the communities being measured. Beta can only be high when alpha is low. This strong negative relationship between alpha and additive Gini–Simpson beta explains why the additive Gini–Simpson beta of fully differentiated butterfly Dataset 4 is less than the additive Gini–Simpson beta of the less diverse and incompletely differentiated RS dataset (Fig. 1a). The greater alpha diversity of Dataset 4 drives beta to a lower value despite the complete differentiation of this dataset. A beta value of 0.001 may therefore mean that two communities

are completely differentiated, nearly identical, or anything in between.

Some researchers choose to avoid these partitioning problems by falling back on species richness, even when abundance data are available. This may be legitimate in some conservation applications, where the mere presence of a rare species may be sufficient to ensure its conservation. However, ecologically and statistically more meaningful characterizations depend on differences in species frequencies between communities, not their mere presence/absence.

Comparing diversity measures with different sensitivities to rare and common species (Test 3)

On the RS butterfly dataset, the additive similarity measure alpha/gamma sharply increases with increasing sensitivity to dominant species (Table 2). Similarity of canopy and understorey according to species richness is 0.65 ± 0.03, close to the minimum possible value of 0.5, but for Shannon entropy it increased to 0.87 ± 0.01, and for the Gini–Simpson index it increased to 0.97 ± 0.004, close to the maximum possible value of 1.00 (which would indicate that the canopy and understorey communities are identical). This same pattern was observed in another large neotropical rainforest canopy/ understorey dataset, Garzacocha (Table 3, 'GC'; $n = 11861$; DeVries & Walla, 2001). An unpublished dataset (Table 3, 'CR'; $n = 3576$) from Costa Rica also showed this pattern (DeVries, unpublished). In fact, a very high similarity value using the Gini–Simpson index is universal in additive partitioning studies of high-diversity ecosystems (e.g. DeVries *et al.*, 1997; Gering *et al.*, 2003; Summerville *et al.*, 2003, 2006; Stendera & Johnson, 2005; Chandy *et al.*, 2006; Ribeiro *et al.*, 2008). The high observed values of additive Gini–Simpson similarity compared to the other similarities suggests that canopy and understorey butterfly communities share most of their dominant species and differ primarily in their less common species. This interpretation is commonly made in additive partitioning studies. The universality of the pattern suggests that some general ecological principle is responsible.

However, precisely the opposite pattern is observed when the multiplicative similarity measures are applied to the same canopy–understorey systems (Fig. 3 and Table 3). According
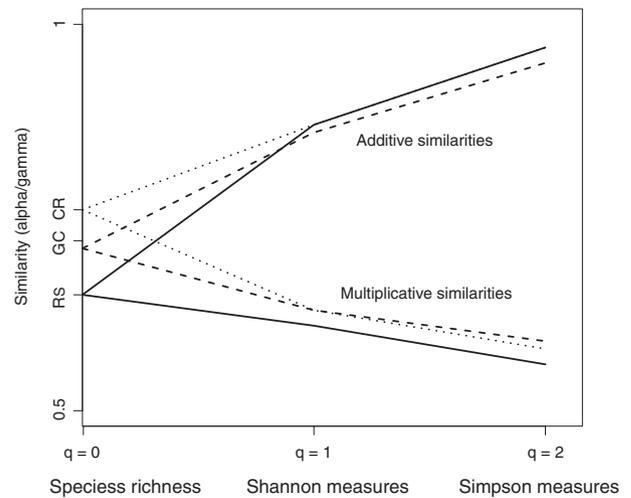

Figure 3 Additive and multiplicative similarities for CR, GC, RS datasets as the diversity measure goes from order 0 (species richness) to order 1 (Shannon) to order 2 (Gini–Simpson). Uncertainties (not indicated) are all less than ± 0.03 (2 SE).

to these similarity measures, we would infer that the dominant species in the canopy are very different from the dominant species in the understorey. Again, the universality of the pattern suggested that a general ecological principle is at work, but a principle exactly contrary to the one derived from additive partitioning studies.

Which is the correct ecological principle? Our tests above and Jost's (2006, 2007) mathematical analysis demonstrate that additive Shannon and especially Gini–Simpson similarity measures necessarily approach unity for high-diversity ecosystems, even if the communities are completely differentiated. The pattern observed in additive partitioning similarity measures can be explained by these purely mathematical effects. A glance at the raw data and the Morisita overlap measure (Table 3) confirms that the dominant canopy species are very different from the dominant understorey species, the opposite of the conclusion drawn from comparison of the additive partitioning similarity measures across indices. Additive alpha/gamma cannot be meaningfully compared across indices, and doing so may lead to incorrect conclusions.

Table 3 Additive and multiplicative canopy versus understorey similarity ± 2 SE for three butterfly datasets. RS = Rio Shiripuno, Ecuador; GC = Garzacocha, Ecuador (DeVries & Walla, 2001); and CR = Costa Rica (DeVries, unpublished). The well-known Morisita overlap measure (rightmost column) can range from 0.00 when both samples are completely different to *c.* 1.00 when both samples are identical. Uncertainties are obtained from a bootstrap approach and represent 2 SE.

Dataset	Additive species richness similarity	Additive Shannon similarity	Additive Gini–Simpson similarity	Multiplicative species richness similarity	Multiplicative Shannon similarity	Multiplicative Gini–Simpson similarity	Overlap (Morisita)
RS	0.65 ± 0.03	0.87 ± 0.01	0.97 ± 0.004	0.65 ± 0.03	0.61 ± 0.02	0.56 ± 0.02	0.11 ± 0.03
GC	0.71 ± 0.02	0.86 ± 0.003	0.95 ± 0.002	0.71 ± 0.02	0.63 ± 0.01	0.59 ± 0.01	0.19 ± 0.01
CR	0.76 ± 0.03	0.87 ± 0.01	0.97 ± 0.002	0.76 ± 0.03	0.63 ± 0.01	0.58 ± 0.01	0.16 ± 0.02

Multiplicative partitioning in practice

The anomalous behaviour of the similarity measure alpha/gamma when used with Shannon entropy can be easily corrected by converting the alpha and gamma indices to the exponential of Shannon entropy (MacArthur, 1965; Whittaker, 1972; Hill, 1973; Jost, 2006, 2007, 2009). Existing additive partitioning software can be used with Shannon entropy, as long as the components are converted by taking their exponentials before interpretation. Alternatively, multiplicative partitioning of Shannon entropy can be used throughout; the results are equivalent (Ricotta, 2005; Jost, 2007).

Partitioning non-Shannon abundance-based measures requires some caution. Given a set of communities of fixed sizes, biologists usually expect beta diversity to increase monotonically with increasing compositional differentiation. In order for beta to have this property and still be independent of alpha, it is important to calculate both alpha and gamma indices giving equal weights to each community (Jost, 2007). When this condition is satisfied, standard complexity measures such as the Gini–Simpson index can be converted to effective number of species (see Table S1) and then partitioned. The ratio alpha/gamma is then a meaningful measure of homogeneity or similarity.

Bridges between additive and multiplicative approaches

For N equally weighted samples, multiplicative beta is a measure of relative differentiation ranging from 1 to N, and it passes all of our tests. Additive species richness beta (but not Shannon beta or Gini–Simpson beta) is a measure of absolute differentiation, the average number of regional species absent from a sample. This has no upper limit, but it too passes all our tests. Therefore, for species richness, we could use either concept, depending on our purpose. Ricotta (2005) noticed that these two beta concepts for species richness have a simple relationship:

Additive beta $\equiv$ gamma $-$ alpha

$= $ (multiplicative beta $-$ 1)alpha.

The right-hand side of this equation gives the gamma effective number of species minus the alpha effective number of species, hence it is an additive partitioning of the effective number of species:

$$^qD_\alpha(^qD_\beta - 1) = ^qD_\alpha{}^qD_\beta - ^qD_\alpha = ^qD_\gamma - ^qD_\alpha, \qquad (5)$$

where $^qD_\alpha$, $^qD_\beta$ and $^qD_\gamma$ are defined in Table S1 for a general order of q. It is interpreted as the effective number of species absent from the average community. Unlike traditional additive partitioning, this additive beta necessarily passes our Test 1 (which requires that beta be non-decreasing with unambiguously increasing differentiation), because both multiplicative beta and alpha are non-decreasing as communities become more differentiated in that test. It is important that

community weights be set equal to $1/N$, unless $q = 0$ or 1 [see Jost (2007), for a discussion of this restriction]. Economo & Keitt (2008) used this additive partitioning of effective number of species to analyse Hubbell's neutral theory of biodiversity.

It is often useful to transform multiplicative beta onto the unit interval, to obtain a normalized measure of similarity or differentiation that can be compared across datasets with different numbers of samples or communities. These transformations depend only on beta and N, the number of communities. Perhaps, the most useful normalized relative similarity measure is the overlap measure C_{qN} (Chao et al., 2008):

$$C_{qN} = [(1/^qD_\beta)^{q-1} - (1/N)^{q-1}]/[1 - (1/N)^{q-1}], \qquad (6)$$

where $^qD_\beta$ denotes the multiplicative beta diversity for any order q (see Table S1). This family includes such well-known similarity measures as the Sorensen index ($q = 0$), the Horn index of similarity ($q = 1$) and the Morisita–Horn index ($q = 2$). The complement of C_{qN} is a useful measure of relative differentiation. Taking $q = 2$ (as in Simpson measures, or heterozygosity), this gives the new measure D that should replace G_{ST} as a measure of differentiation in genetics (Jost, 2008):

Genetic differentiation $D = 1 - C_{2N}$

$$= [(H_{GS\gamma} - H_{GS\alpha})]/[(1 - H_{GS\alpha})(1 - 1/N)]. \qquad (7)$$

For N equally weighted samples, additive beta may also be transformed onto the unit interval, in much the same way as multiplicative beta. These transformations will, in general, be functions not only of additive beta and N but also of additive alpha, so they will not always be monotonic with respect to additive beta. To find a general formula for such a transformation, we first generalize the traditional additive approach (Lande, 1996) to 'Tsallis entropies' of any order of q, which include Shannon entropy and the Gini–Simpson indices as special cases (Keylock, 2005). The within-group 'diversity' (Tsallis entropy) is

$$^qH_\alpha = \frac{1}{(q-1)} \sum_{j=1}^{N} \frac{1}{N} \left(1 - \sum_{i=1}^{S} p_{ij}^q\right)$$

and the total 'diversity' is

$$^qH_\gamma = \frac{1}{(q-1)} \left(1 - \sum_{i=1}^{S} \bar{p}_i^q\right),$$

where $\bar{p}_i = (1/N) \sum_{j=1}^{N} p_{ij}$. We can prove that additive beta has the following range:

$$0 \leq {}^qH_\gamma - {}^qH_\alpha \leq \frac{(1 - N^{1-q})[1 - (q-1)(^qH_\alpha)]}{q-1}. \qquad (8)$$

Dividing additive beta (which is not monotonic with respect to differentiation, as shown in Fig. 1b) by its maximum possible value, consistent with the observed value of alpha and N, produces a monotonic measure of differentiation D_{qN}:

$$D_{qN} = \frac{(q-1)({}^qH_\gamma - {}^qH_\alpha)}{(1 - N^{1-q})[1 - (q-1)({}^qH_\alpha)]}. \tag{9}$$

With some algebra, we can show that this generalized differentiation measure is exactly $1 - C_{qN}$, where C_{qN} is given in equation (6) and derived from multiplicative partitioning. For $q = 2$, this is the same genetic differentiation measure D given in the preceding paragraph. As q tends to 1, equation (9) leads to the complement of C_{1N}, the multiple-community generalization of the Horn index, which can also be derived within the multiplicative framework. The case $q = 0$ gives differentiation measures based on species richness. Let $\bar{S}$ denote the average number of species per group. $D_{0N} = (S/\bar{S} - 1)/(N - 1)$, which is the complement of C_{0N}, the multiple-community generalization of the Sørensen index, also derivable within the multiplicative framework (Chao *et al.*, 2008). These normalized measures of relative similarity and differentiation are identical in the additive and multiplicative approaches for all values of q, providing a bridge between the two approaches.

Hedrick (2005) used a similar approach to normalize the standard differentiation measure of genetics, G_{ST}, the complement of the additive similarity measure alpha/gamma. In this case, Hedrick's normalization did not produce a relative differentiation measure that is independent of within-group diversity, except at the extremes of complete differentiation and complete similarity. Normalization is not a general cure for problems of independence, but it does work for additive beta based on Tsallis entropies using the upper bound in equation (8).

A more restrictive differentiation criterion

We used a mild criterion to filter out measures that do not quantify biologically useful concepts of differentiation and similarity. Although some additive measures failed, most measures in the literature pass this filter. For some conservation purposes, a tougher filter may be useful. We might require that a differentiation measure increases any time a new endemic species is added to any community, with any abundance. This seems to be the meaning of differentiation in many conservation contexts.

If we apply this filter to the additive and multiplicative measures discussed here, we find that few measures pass. The only differentiation measures that pass are additive and multiplicative beta based on species richness and Shannon entropy. The only similarity measures that pass are alpha/gamma based on species richness (which is the same in both the multiplicative and additive approaches), and multiplicative alpha/gamma based on the exponential of Shannon entropy. This is just one of many special mathematical properties obeyed by Shannon measures and species richness. Other measures fail this test because of their differential treatment of rare and common species.

CONCLUSIONS

Neither Shannon entropy nor the Gini–Simpson index quantifies a logically consistent concept of diversity for conservation analyses. Furthermore, when these indices are applied to a set of N samples or communities, the 'similarity' measure alpha/gamma is not monotonically related to compositional similarity, and the additive beta based on the Gini–Simpson index is not monotonically related to compositional differentiation. Likewise, heterozygosity does not quantify a logically consistent concept of diversity for conservation genetics, and the standard 'differentiation' measures derived from it, such as F_{ST} and G_{ST}, do not measure genetic differentiation. Many of these problems can be fixed by converting the measures to effective number of species, and partitioning these either additively or multiplicatively.

To conserve as much of Earth's biodiversity as possible, sites with high diversity and high differentiation from others should be given conservation priority. These are exactly the systems which produce the strongest mathematical artefacts when Shannon entropy and the Gini–Simpson index are equated with diversity and additively partitioned. When diversity and differentiation are high, as in the tropical butterfly datasets presented here, the additive partitioning similarity measure used with these two indices severely underestimates the degree of differentiation between sites, potentially skewing conservation decisions towards irreversible loss of biodiversity. As our examples show, ecological conclusions will also often be wrong when these measures are used. The effects are not small; for typical tropical ecosystems, relative compositional similarity can appear to be nearly 100% even when there is no overlap in species composition between sites. The same magnitude of error occurs in conservation genetics when high-diversity microsatellites are analysed (e.g. Balloux *et al.*, 2000). Misrankings are apparently common in the genetic literature; a meta-analysis (Heller & Siegismund, 2009) revealed that 25% of recent studies would have ranked populations differently according to their differentiation if Jost's D (equation 7), developed within the multiplicative framework, had been used instead of the misleading traditional measures F_{ST} and G_{ST}.

To avoid potentially irreversible conservation mistakes, we recommend using the exponential of Shannon entropy or the inverse Simpson concentration as abundance-based measures of diversity. We recommend the compositional overlap C_{qN} (Chao *et al.*, 2008) as a meaningful normalized measure of compositional similarity, and its complement as a meaningful measure of differentiation (Jost, 2008). These normalized measures follow from both the additive and multiplicative approaches to partitioning diversity, so they transcend debates about this subject.

ACKNOWLEDGEMENTS

REFERENCES

Aczel, J. & Daroczy, Z. (1975) *On measures of information and their characterization*. Academic Press, New York.

Balloux, F., Brunner, H., Lugon-Moulin, N., Hausser, J. & Goudet, J. (2000) Microsatellites can be misleading: an empirical and simulation study. *Evolution*, **54**, 1414–1422.

Chandy, S., Gibson, D.A. & Robertson, P.J. (2006) Additive partitioning of diversity across hierarchical spatial scales in a forested landscape. *Journal of Applied Ecology*, **43**, 792–801.

Chao, A., Jost, L., Chiang, S.C., Jiang, Y.-H. & Chazdon, R.L. (2008) A two-stage probabilistic approach to multiple-community similarity indices. *Biometrics*, **64**, 1178–1186.

Couteron, P.E. & Pelissier, R. (2004) Additive apportioning of species diversity: towards more sophisticated models and analyses. *Oikos*, **107**, 215–221.

DeVries, P.J. (1988) Stratification of fruit-feeding nymphalid butterflies in the Costa Rican rainforest. *Journal of Research on Lepidoptera*, **26**, 98–108.

DeVries, P.J. & Walla, T.R. (2001) Species diversity and community structure in neotropical butterflies. *Biological Journal of the Linnean Society*, **74**, 1–15.

DeVries, P.J., Murray, D. & Lande, R. (1997) Species diversity in vertical, horizontal, and temporal dimensions of a fruit-feeding butterfly community in an Ecuadorian rainforest. *Biological Journal of the Linnean Society*, **62**, 343–364.

DeVries, P.J., Walla, T.R. & Greeney, H.F. (1999) Species diversity in spatial and temporal dimensions of fruit-feeding butterflies from two Ecuadorian rainforests. *Biological Journal of the Linnean Society*, **68**, 333–353.

Economo, E.P. & Keitt, T.H. (2008) Species diversity in neutral metacommunities: a network approach. *Ecology Letters*, **11**, 52–62.

Gering, J.C., Crist, T.O. & Veech, J.A. (2003) Additive partitioning of species diversity across multiple spatial scales: implications for regional conservation of biodiversity. *Conservation Biology*, **17**, 488–499.

Gregorius, H.-R. (1987) The relationship between the concepts of genetic diversity and differentiation. *TAG Theoretical and Applied Genetics*, **74**, 397–401.

Hannah, L. & Kay, J.A. (1977) *Concentration in the modern industry: theory, measurement, and the U.K. experience*. MacMillan, London.

Hardy, O.J. & Jost, L. (2008) Interpreting and estimating measures of community phylogenetic structuring. *Journal of Ecology*, **96**, 849–852.

Hedrick, P.W. (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Heller, R. & Siegismund, H.R. (2009) Relationship between three measures of genetic differentiation G_{ST}, D_{EST} and G'_{ST}: how wrong have we been? *Molecular Ecology*, **18**, 2080–2083.

Hill, M. (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology*, **54**, 427–432.

Hubbell, S.P., Condit, R. & Foster, R.B. (2005) *Barro Colorado Forest Census plot data*. Available at: http://www.ctfs.si.edu/datasets/bci (accessed January 2007).

Jost, L. (2006) Entropy and diversity. *Oikos*, **113**, 363–375.

Jost, L. (2007) Partitioning diversity into independent alpha and beta components. *Ecology*, **88**, 2427–2439.

Jost, L. (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Jost, L. (2009) Mismeasuring biological diversity: response to Hoffmann and Hoffmann (2008). *Ecological Economics*, **68**, 925–928.

Keylock, C.J. (2005) Simpson diversity and the Shannon-Wiener index as special cases of a generalized entropy. *Oikos*, **109**, 203–207.

Lande, R. (1996) Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos*, **76**, 5–13.

Lewontin, R. (1972) The apportionment of human diversity. *Evolutionary Biology*, **6**, 381–398.

MacArthur, R.H. (1965) Patterns of species diversity. *Biological Reviews*, **40**, 510–533.

MacArthur, R.H. (1972) *Geographical ecology*. Harper & Row, New York.

Magurran, A. (2004) *Measuring biological diversity*. Blackwell Publishing, Oxford.

Nei, M. (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences USA*, **70**, 3321–3323.

Patzkowsky, M.E. & Holland, S.M. (2007) Diversity partitioning of a Late Ordovician marine biotic invasion: controls on diversity in regional ecosystems. *Paleobiology*, **33**, 295–309.

Petit, R., El Mousadik, A. & Pons, O. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844–855.

Ribeiro, D.B., Prado, P.I., Brown, K.S. & Freitas, A.V.L. (2008) Additive partitioning of butterfly diversity in a fragmented landscape: importance of scale and implications for conservation. *Diversity and Distributions*, **14**, 961–968.

Ricotta, C. (2005) On hierarchical diversity decomposition. *Journal of Vegetation Science*, **16**, 223–226.

Ricotta, C. (2008) Computing additive β-diversity from presence and absence scores: a critique and alternative parameters. *Theoretical Population Biology*, **73**, 244–249.

Stendera, S.E.S. & Johnson, R.K. (2005) Additive partitioning of aquatic invertebrate species diversity across multiple spatial scales. *Freshwater Biology*, **50**, 1360–1375.

Summerville, K.S., Boulware, M.J., Veech, J.A. & Crist, T.O. (2003) Spatial variation in species diversity and composition of forest Lepidoptera in eastern deciduous forests of North America. *Conservation Biology*, **17**, 1045–1057.

Summerville, K.S., Wilson, T.D., Veech, J.A. & Crist, T.O. (2006) Do body size and diet breadth affect partitioning of species diversity? A test with forest Lepidoptera *Diversity and Distributions*, **12**, 91–99.

Torres, E., Iriondo, J. & Perez, C. (2003) Genetic structure of an endangered plant, *Antirrhinum microphyllum* (Scrophulariaceae): allozyme and RAPD analysis. *American Journal of Botany*, **90**, 85–92.

Tsallis, C. (1988) Possible generalization of Boltzman-Gibbs statistics. *Journal of Statistical Physics*, **52**, 479–487.

Veech, J.A., Summerville, K.S., Crist, T.O. & Gering, J.C. (2002) The additive partitioning of species diversity: recent revival of an old idea. *Oikos*, **99**, 3–9.

Walla, T.R., Engen, S., DeVries, P.J. & Lande, R. (2004) Modelling vertical beta-diversity in tropical butterfly communities. *Oikos*, **107**, 610–618.

Whittaker, R.H. (1972) Evolution and measurement of species diversity. *Taxon*, **21**, 213–251.

Wright, S. (1978) *Evolution and the genetics of populations*, Vol. 4. University of Chicago Press, Chicago.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1 Formulas for additive and multiplicative alpha, beta and gamma diversity.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

BIOSKETCHES

Lou Jost studies the mathematics of diversity in ecology and genetics, and has set up a system of biodiversity reserves in Ecuador (http://www.ecominga.com).

Philip DeVries studies the diversity, ecology and evolution of tropical butterflies.

Thomas Walla studies interactions of lepidoptera, plants, parasitoids, and predators and teaches field courses in Ecuador.

Harold Greeney untangles the natural history of the web of life in his cloud forest reserve in Ecuador (http://www.yanayacu.org).

Anne Chao investigates the mathematics of diversity, and develops novel statistical tools for data analysis (SPADE, at http://chao.stat.nthu.edu.tw/download.html).

Carlo Ricotta works on the conceptual foundations of diversity, with current focus on quantifying functional or phylogenetic diversity.

Author contributions: L.J. made theoretical contributions and led the writing; P.D., T.W. and H.G. did the field work and added ecological perspectives; A.C. did theoretical derivations, statistical analyses and figures; C.R. made theoretical contributions.

Editor: Risto Heikkinen