Molecular Ecology (2008) **17**, 4015–4026 doi: 10.1111/j.1365-294X.2008.03887.x

G_{ST} and its relatives do not measure differentiation

LOU JOST
Via Runtun, Baños, Tungurahua, Ecuador

Abstract

G_{ST} and its relatives are often interpreted as measures of differentiation between subpopulations, with values near zero supposedly indicating low differentiation. However, G_{ST} necessarily approaches zero when gene diversity is high, even if subpopulations are completely differentiated, and it is not monotonic with increasing differentiation. Likewise, when diversity is equated with heterozygosity, standard similarity measures formed by taking the ratio of mean within-subpopulation diversity to total diversity necessarily approach unity when diversity is high, even if the subpopulations are completely dissimilar (no shared alleles). None of these measures can be interpreted as measures of differentiation or similarity. The derivations of these measures contain two subtle misconceptions which cause their paradoxical behaviours. Conclusions about population differentiation, gene flow, relatedness, and conservation priority will often be wrong when based on these fixation indices or similarity measures. These are not statistical issues; the problems persist even when true population frequencies are used in the calculations. Recent advances in the mathematics of diversity identify the misconceptions, and yield mathematically consistent descriptive measures of population structure which eliminate the paradoxes produced by standard measures. These measures can be directly related to the migration and mutation rates of the finite-island model.

Keywords: differentiation, diversity, G_{ST}, heterozygosity, mutation rate, partition

Received 23 December 2007; revision received 1 May 2008; accepted 6 July 2008

Introduction

Gene diversity and genetic differentiation between subpopulations are the fundamental descriptive variables of population genetics, with important applications in molecular ecology, evolutionary and conservation biology, anthropology, immunology, and more. In population genetics, gene diversity is usually equated with heterozygosity, and genetic differentiation between subpopulations is usually measured by G_{ST}, F_{ST}, or their relatives. Most geneticists interpret these as a stand-alone measure of relative differentiation, with values close to zero supposedly indicating little differentiation, and values close to unity supposedly indicating nearly complete differentiation. Similarity between subpopulations is often measured by taking the ratio of within-subpopulation diversity (heterozygosity) to total diversity (heterozygosity). Values near unity supposedly indicate high similarity. If mean within-subpopulation diversity is much greater than between-subpopulation diversity, geneticists believe that the subpopulations show low differentiation.

When this approach was developed by Wright (1951), Lewontin (1972), Nei (1973), and others, low-diversity systems were the focus of attention, and for such systems these measures gave qualitatively reasonable descriptions of subdivided populations. Standard measures such as G_{ST} and its relatives unquestionably provide important insights into population structure (e.g. Crow 2004) and selectable variance. However, the interpretation of G_{ST} and additive between-group heterozygosity as measures of differentiation rest on multiple mathematical misconceptions, and when diversity and differentiation are high, they give nonsensical results. These are not sampling issues; they are undiminished even if the true population allele frequencies are used. Now that technology permits detection of many alleles per locus, conclusions about diversity, population differentiation, gene flow, relatedness, and conservation priority will often be wrong if based on the standard approach.

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

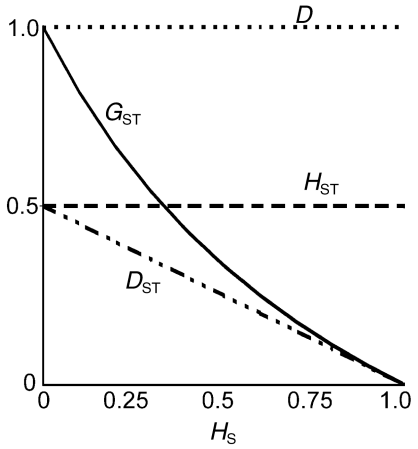

Fig. 1 Differentiation measures applied to two completely differentiated subpopulations (no shared alleles) for various values of mean within-subpopulation heterozygosity. G_{ST} and D_{ST} approach zero when mean heterozygosity is high, even though differentiation is 100% for these subpopulations. New measures D and H_{ST} correctly reflect differentiation regardless of the value of mean heterozygosity. All values are calculated using actual population frequencies, not sample frequencies; this is not a sampling issue.

Warning signs

Cracks in the logic of the standard approach to diversity analysis have been appearing for some time in both the empirical and theoretical literature. In a study of population structure in two races of the common shrew *Sorex araneus* (Balloux *et al*. 2000), an analysis of Y-chromosome microsatellites gave an F_{ST} of 0.19, indicating little differentiation and moderate gene flow. Yet the authors showed that both these conclusions were false; their analysis of microsatellite allele lengths in the two races revealed that none of the alleles were shared between races, and gene flow was essentially zero. The authors conclude that 'population structure estimates based solely on autosomal microsatellites would have led to a strong overestimate of the extent of gene flow between the two races'. The correct mathematical treatment of differentiation, given below, shows that the differentiation between the two races is not 19% but 100%, implying that migration rate is near zero.

Carreras-Carbonell *et al*. (2006) studied microsatellites in two subspecies of the fish *Tripterygion delaisi* that were distinguishable behaviourally, geographically, and via mtDNA. The authors were puzzled that the differentiation measure F_{ST} based on highly polymorphic microsatellites was very low between subspecies, even at loci with no shared alleles between subspecies. Like many other authors, they observed that 'estimates of F_{ST} seem to decline with increasing polymorphism'. In fact, F_{ST} and G_{ST} necessarily decline with increasing polymorphism (Fig. 1). The correct treatment shows that the between-subspecies differentiation at their most diverse locus (Td06) is not 0.05 (the value of F_{ST}) but 1.00.

Another puzzle in the literature involves the amount of differentiation shown by microsatellites vs. allozyme loci. The mutation rates of microsatellites are known to be higher than the mutation rates of allozyme loci. Microsatellites should therefore show more differentiation between subpopulations than allozyme loci in any given organism. However, the opposite pattern is widely reported in the literature; G_{ST} (or its relatives) based on microsatellites is usually lower than G_{ST} based on allozyme loci (e.g. Sanetra & Crozier 2003). This is a predictable effect of the mathematical problems in the definition of G_{ST}. When these are corrected, microsatellites usually show the expected high differentiation. For example, in the study of Carreras-Carbonell *et al*. (2006) discussed in the preceding paragraph, the real average differentiation exhibited by the three highest-diversity microsatellite loci is actually 0.74, not the 0.05 given by F_{ST}.

Theoreticians have also been aware of these problems but have not identified their origin, and perhaps have underestimated their importance. Nei (1973) himself recognized that under some circumstances, his G_{ST} 'is not a good measure of differentiation'. Nagylaki (1998) showed that G_{ST} is an appropriate measure of differentiation if and only if the gene diversity is low. Even when diversity is low, Charlesworth (1998) recognized that the dependence of G_{ST} on within-subpopulation heterozygosity makes it unsuitable as a measure of differentiation in comparisons between populations with different mean heterozygosities. The growing use of high-diversity microsatellite loci has made the limitations of G_{ST} more obvious. Writing about microsatellites, Balloux & Lugon-Moulin (2002) concluded that interpreting G_{ST} and related measures 'can be a dangerous task' and 'it can be difficult or misleading to give a biological meaning to these values'.

Indeed, for highly polymorphic loci, G_{ST} by itself provides virtually no information about the actual degree of differentiation of subpopulations, and is not monotonic with respect to increasing differentiation. A G_{ST} of 0.001 can indicate low, moderate, or maximal differentiation, depending on the value of the mean within-subpopulation heterozygosity. When within-subpopulation heterozygosity is very high and differentiation is 100%, G_{ST} will always incorrectly show that differentiation is near 0% (Fig. 1). G_{ST} can even decrease as differentiation increases (Fig. 2), and so G_{ST} will not rank data sets correctly according to their differentiation: completely differentiated populations will often have lower G_{ST} values than poorly differentiated ones.

For example, Table 1 gives population allele frequencies for three species: species A consists of two identical subpopulations, species B consists of two somewhat differentiated

	Species A		Species B		Species C	
Allele	Subpop. 1	Subpop. 2	Subpop. 1	Subpop. 2	Subpop. 1	Subpop. 2
1	0.5	0.5	0.2	0.8	0.095	0
2	0.5	0.5	0.8	0.2	0.08	0
3	0	0	0	0	0.11	0
4	0	0	0	0	0.08	0
5	0	0	0	0	0.095	0
6	0	0	0	0	0.06	0
7	0	0	0	0	0.07	0
8	0	0	0	0	0.096	0
9	0	0	0	0	0.094	0
10	0	0	0	0	0.08	0
11	0	0	0	0	0.03	0
12	0	0	0	0	0.06	0
13	0	0	0	0	0.05	0
14	0	0	0	0	0	0.15
15	0	0	0	0	0	0.16
16	0	0	0	0	0	0.12
17	0	0	0	0	0	0.13
18	0	0	0	0	0	0.17
19	0	0	0	0	0	0.14
20	0	0	0	0	0	0.13
Measures of differentiation; should increase with increasing differentiation:						
	Species A		Species B		Species C	
D_{ST}	0		0.18		0.06(!)	
G_{ST}	0		0.36		0.06(!)	
H_{ST}	0		0.26		0.5	
Δ_{ST}	1.00		1.36		2.00	
D	0		0.53		1.00	
Measures of similarity; should decrease with increasing differentiation:						
	Species A		Species B		Species C	
H_S/H_T	1.00		0.64		0.94(!)	
Δ_S/Δ_T	1.00		0.74		0.50	

Table 1 Subpopulation differentiation and similarity according to traditional and new measures, for three species. For each species, the true subpopulation frequencies of each allele are given. The subpopulations of species A are undifferentiated, those of species B are moderately differentiated, and those of species C are maximally differentiated. D_{ST}, G_{ST}, H_S, and H_T are defined in eqns. 1–3 following notation of Nei (1973); Δ_{ST}, H_{ST}, and D are defined in eqns. 7–11. All measures are calculated using the exact subpopulation frequencies. Standard measures marked (!) fail to rank these species correctly in terms of differentiation or similarity. New measures H_{ST}, and D rank them correctly

subpopulations, and species C consists of two completely differentiated subpopulations. All geneticists agree that the subpopulations of species B are less differentiated than the subpopulations of species C (which are completely differentiated). Yet both G_{ST} and the standard similarity measure indicate that the subpopulations of species C are *less* differentiated than those of species B.

It does not help to use P values; the difference in G_{ST} between two populations may be highly statistically significant but this tells nothing about the relative differentiation of the two populations. For example, in Table 1 the G_{ST} for species B is greater than G_{ST} for species C, and the P value of the difference will be highly significant if sample size is large, but species C is clearly more differentiated than species B.

There has been speculation that the problems of standard measures like G_{ST} are due to homoplasy or other poorly understood biological effects (Sanetra & Crozier 2003; O'Reilly *et al*. 2004), but these same problems also arise in nongenetic applications of the standard approach (Jost 2006, 2007). The real causes of these anomalies are several fundamental misunderstandings about the mathematics of diversity and differentiation measures. These mathematical misconceptions can be pinpointed and corrected, resulting in a new logically consistent mathematical framework for describing diversity and differentiation in population genetics and other sciences (Jost 2007).

Pinpointing the misconceptions underlying the standard approach

Nei (1973) defined gene diversity as

$$H \equiv 1 - \sum_{i=1}^{k} p_i^2. \qquad \text{(eqn 1)}$$

where p_i is the population frequency of the i-th allele, and k is the total number of alleles. For randomly mating species, this is just the heterozygosity, and we use that term rather than 'gene diversity' to avoid confusion with

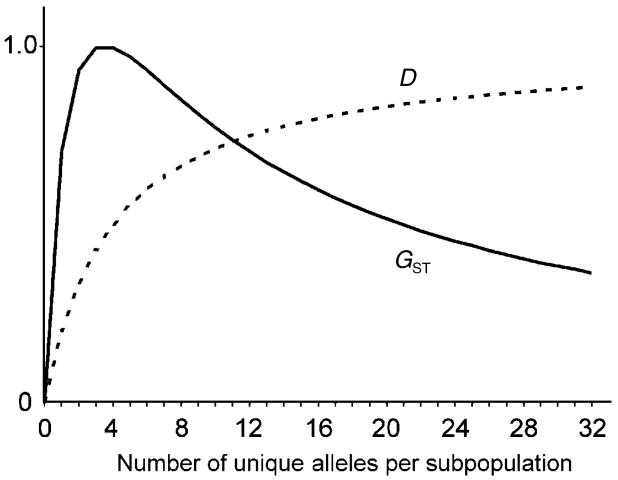

Fig. 2 Behaviour of G_{ST} and D as differentiation increases. We start with two identical subpopulations (four equally common alleles, 1000 individuals per allele per subpopulation). We then successively add unique alleles to each subpopulation (1000 individuals per allele) and graph G_{ST} and D (the measure of differentiation defined in the text). G_{ST} is normalized by dividing by its maximum value (0.0345). Even though differentiation increases steadily from left to right, G_{ST} reaches its maximum and then falls back to zero. G_{ST} is calculated from exact population allele frequencies, so this is not a sampling issue.

other diversity measures discussed later. It should be understood that heterozygosity here refers to the quantity defined in eqn 1.

Partitioning issues

In Nei's derivation of G_{ST}, he additively partitioned total heterozygosity (H_T, the heterozygosity of the pooled subpopulations) into mean within-subpopulation heterozygosity H_S (the mean heterozygosity of the individual subpopulations) and between-subpopulation diversity D_{ST}:

$$H_T = H_S + D_{ST} \text{ or } H_T - H_S = D_{ST} \quad \text{(eqn 2)}$$

D_{ST} is calculated by subtracting H_S from H_T. Nei called D_{ST} a measure of absolute differentiation, and he divided this by total diversity to obtain what he termed the relative differentiation, G_{ST}:

$$G_{ST} \equiv D_{ST}/H_T = (H_T - H_S)/H_T \quad \text{(eqn 3)}$$

The first misconception of the standard approach is this belief that additive partitioning of heterozygosity yields a between-subpopulation component (D_{ST}) that measures differentiation. Nei's additive decomposition of total heterozygosity is analogous to the additive decomposition of total variance into two independent variances, but H_S (the within-subpopulation component of heterozygosity) and D_{ST} are not independent. Heterozygosity is always less than unity, so H_S and D_{ST} are related through the constraint $H_S + D_{ST} < 1$. When H_S is large, D_{ST} has to be small regardless of the differentiation of the subpopulations. D_{ST} (and therefore G_{ST}) is strongly confounded with mean within-subpopulation heterozygosity. Additive partitioning of heterozygosity does not produce pure within- and between-subpopulation components; it is an incomplete partitioning.

This is demonstrated in Fig. 1, a graph of G_{ST} vs. H_S for two equally diverse, equally large, completely differentiated subpopulations, whose heterozygosities are allowed to vary from zero to unity. (G_{ST} is here calculated from the actual population frequencies of the alleles, so its anomalous behaviour has nothing to do with sampling effects.) A real measure of relative differentiation would have a flat graph showing 100% differentiation regardless of the value of H_S. Yet this graph of G_{ST} sharply decreases from 100% to 0% as H_S becomes large. G_{ST} can be 100% wrong when used with highly polymorphic loci. This shows that G_{ST} is not a stand-alone measure of relative genetic differentiation between subpopulations (and D_{ST} is not a measure of absolute differentiation).

Another way of seeing the gravity of the problem with G_{ST} is to take a pair of genetically identical subpopulations, and increase their differentiation by successively adding different unique alleles to each subpopulation (Fig. 2). G_{ST} starts out at zero (since the subpopulations start out with identical alleles and allele frequencies). After adding one unique allele to a subpopulation, and a different unique allele to the other subpopulation, G_{ST} begins to rise, as a measure of differentiation should. However, as we repeat the process of adding different novel alleles to each subpopulation, G_{ST} soon reaches a maximum and then begins to fall, even though differentiation is increasing. As more novel alleles are added, G_{ST} falls all the way back to zero. This happens because of the negative dependence of G_{ST} on diversity. (Again, this is not a statistical issue; the graph in Fig. 2 is produced using the exact population frequencies of the alleles.)

Equating diversity with heterozygosity or entropy

Geneticists have an alternative method of measuring the differentiation or similarity of a set of subpopulations, involving the ratio of average within-subpopulation diversity to total diversity. Geneticists reason that if this ratio is close to unity (that is, the average within-group diversity is close to the total diversity), all the groups must be similar to each other (else the diversity of the pooled groups would be significantly larger than the average diversity of the individual groups). This reasoning was first used in genetics by Lewontin (1972) in his classic study on the apportionment of gene diversity within and between

human races. In that study, he found that the mean ratio of within-group Shannon entropy to total Shannon entropy was 85.4%. He concluded that less than 15% of human genetic diversity is accounted for by differences between races, and so 'human races and populations are remarkably similar to each other [genetically].'

While this reasoning is intuitively plausible, it is not generally valid. When diversity is equated with Shannon entropy or heterozygosity, the ratio of average within-subpopulation diversity to total diversity necessarily approaches unity when diversity is high, regardless of the similarity or dissimilarity among subpopulations. Species C in Table 1 is an example of this; the two subpopulations are completely distinct, sharing no alleles, yet their 'similarity' is 94% when heterozygosity is equated with diversity. When diversity is very high, subpopulations with no alleles in common can have 'similarities' of 99.99% or higher. Following Lewontin's reasoning, such subpopulations which do not share any alleles would wrongly be considered very similar to each other genetically.

This same problem affects applications of genetics to conservation biology. Suppose geneticists study an endangered species whose population is subdivided into 20 equally large subpopulations, each with 100 equally common alleles at the locus. Let each subpopulation be completely distinct genetically (no shared alleles between subpopulations), so that each subpopulation contributes equally to the total genetic diversity of the species at the studied locus. Suppose conservation managers ask geneticists how many subpopulations would need to be saved in order to conserve half the genetic diversity of this species. The intuitive answer is that since each subpopulation contributes equally to the total pool of alleles, 10 of the 20 subpopulations must be conserved to achieve this goal. However, a geneticist who equates genetic diversity with heterozygosity would find that the 'diversity' of any single subpopulation is 0.99, while the 'diversity'of the total population is 0.9995. This geneticist would therefore tell the managers of the species that conservation of just one subpopulation would conserve 99% (0.99/0.9995) of the species' total genetic diversity at the studied locus. The geneticist would reach the paradoxical conclusion that the loss of almost all the common alleles of the species would have virtually no effect on the species' genetic diversity. From almost any practical or theoretical standpoint, this is an incorrect conclusion which would have negative consequences for the species' long-term survival.

This issue is not connected to the partitioning problem discussed earlier. Its roots lie deeper, in the widespread belief that any increasing measure of compositional complexity (such as heterozygosity or Shannon entropy) can be equated with diversity. The last two examples showed that common forms of reasoning about diversity are false when diversity is equated with heterozygosity or entropy. Therefore, one could conclude that neither heterozygosity nor Shannon entropy precisely match the intuitive concept of diversity as actually used by geneticists, ecologists, and other scientists.

It is possible to pinpoint the mathematical reason why heterozygosity and Shannon entropy produce nonsensical results when used in ratio similarity measures and other kinds of ratio comparisons. They fail because they are nonlinear with respect to pooling of subpopulations. When there are many common alleles in each subpopulation, the 'diversity' of the pooled subpopulations will not be much greater than the 'diversities' of the subpopulations themselves, so the ratio of within-subpopulation diversity to total diversity will approach unity regardless of the similarity between subpopulations. Therefore a 'similarity' value of 0.999 may mean that the subpopulations are nearly identical, completely distinct, or anything in between. This 'similarity' measure, by itself, tells virtually nothing about the similarity between subpopulations when diversity is high. Yet this defective measure of similarity or other equally invalid ratio comparisons are common in the literature. Even such simple statements as 'the post-treatment diversity dropped by 5%' use an invalid ratio comparison and can be highly misleading when diversity is equated with heterozygosity.

The mathematics of diversity

Why not derive new mathematically self-consistent descriptive measures of diversity and differentiation that really behave the way that geneticists thought their traditional measures behaved? The first step in such a program is to find a measure of genetic diversity that behaves correctly in common ratio comparisons and conservation genetics problems such as those just mentioned. We can then derive a formula to partition this diversity into truly independent within- and between-subpopulation components. The resulting pure between-subpopulation component can then be transformed into a meaningful, logically and mathematically consistent measure of relative differentiation to replace G_{ST}.

Diversity

As shown above, ratio comparisons are invalid when diversity is equated with heterozygosity or entropy. What mathematical properties should diversity measures possess if these common forms of reasoning about diversity and differentiation in genetics are to give correct answers? The key property implicit in the biological concept of diversity was first identified by Hill (1973) in ecology: suppose we have two equally large, equally diverse, completely distinct subpopulations. *If these two subpopulations with no*

shared alleles are pooled and treated as one population, the diversity of that population should be double the diversity of either of the original subpopulations. Economists have independently arrived at an analogous criterion (Hannah & Kay 1977). If a diversity measure has this kind of linear metric with respect to subpopulation pooling, then geneticists' ratio comparisons will accurately reflect compositional similarity, and the conservation genetics problem posed in the preceding paragraph will give the intuitively correct answer that 10 subpopulations should be preserved. The general formula for diversity measures with linear metrics is (Hill 1973; Jost 2007):

$$\text{Diversity } \Delta \equiv \left(\sum_{i=1}^{k} p_i^q\right)^{1/(1-q)} \qquad \text{(eqn 4)}$$

where p_i is the population frequency of the i-th allele and the exponent q determines the measure's sensitivity to allele frequencies. When $q = 0$, eqn 4 gives the allele number. When q approaches unity, eqn 4 gives (via calculus) the exponential of Shannon entropy

$$\Delta_{(q\to 1)} = \exp\left(-\sum_{i=1}^{k} p_i \ln p_i\right)$$

and when $q = 2$ it gives Kimura & Crow's (1964) effective number of alleles

$$\Delta_{(q=2)} = \left[\sum_{i=1}^{k} (p_i^2)\right]^{-1} \qquad \text{(eqn 5)}$$

which is the reciprocal of Nei's (1973) gene identity

$$J \equiv \sum_{i=1}^{k} p_i^2$$

These diversities have the linear behaviour implicit in the geneticists' concept of diversity, while heterozygosity and Shannon entropy do not. For this reason, allele number, the exponential of Shannon entropy and effective number of alleles deserve to be called true diversities, and they should be used in place of traditional measures when diversity is the quantity of interest. For an example of this use, see Nagylaki (1998).

A general treatment of the mathematics of true diversities is given in Jost (2007). For genetics, the treatment may be greatly simplified by assuming that the statistical weights of all subpopulations are equal, as in standard genetic models such as the finite-island model (Crow 1986; Rousset 2004). This is also the weighting scheme that must be used when calculating the actual relative differentiation between a set of n subpopulations. For this task, the sizes of the subpopulations are irrelevant (two subpopulations that are completely differentiated must have a differentiation of 100% even when one subpopulation is much smaller than the other). Under this restriction, eqn 5 is the most useful diversity measure for genetics, since it is directly connected to heterozygosity H and gene identity J:

$$\Delta_{(q=2)} = 1/(1 - H) = 1/J, \qquad \text{(eqn 6)}$$

and so it can be easily connected to evolutionary parameters of simple genetic models like the finite-island model. While J is often interpreted as homozygosity, in this context it is sometimes useful to note that J is also the expected value of allele frequency in the population (more precisely, the expected value of the population allele frequency of a randomly chosen gene in the population; see Hill 1973).

True diversity has an easily interpretable metric, unlike heterozygosity. A drop of, say, 20% in diversity is equivalent to a loss of 20% of the alleles of a perfectly even population (a population whose alleles are all equally common). In contrast, a drop of 20% in heterozygosity has no easy interpretation. When heterozygosity is high, a drop of 20% represents an enormous loss of diversity, but when heterozygosity is low, a drop of 20% may be inconsequential. If we take species C of Table 1 as an example, a loss of subpopulation 1 would clearly result in a serious loss of the species' genetic diversity. True diversity accurately reflects the magnitude of the loss, since it drops by 56% [(15.8 – 6.95)/15.8], but heterozygosity only drops by 9% [(0.936 – 0.856)/0.936].

Partitioning true diversity

It can be proven (Jost 2007) that the decomposition of any true diversity Δ_T into pure within- and between-subpopulation components (Δ_S and Δ_{ST}, respectively) must be multiplicative: the effective number of alleles in the pooled subpopulations (Δ_T) must equal the effective number of alleles per subpopulation times the effective number of distinct subpopulations:

$$\Delta_T = \Delta_S \cdot \Delta_{ST}. \qquad \text{(eqn 7)}$$

If Δ_S and Δ_{ST} are to be independent of each other (that is, if they are to be pure within- and between-subpopulation components), then Δ_S for n equally weighted subpopulations must be the reciprocal of the average of the gene identities of the subpopulations:

$$\Delta_S = (1 - H_S)^{-1} = [(1/n)(J_1 + J_2 + \ldots + J_n)]^{-1} \equiv (J_S)^{-1}$$

(Jost 2007). The between-subpopulation component of diversity, the effective number of distinct subpopulations, is therefore:

$$\Delta_{ST} = \Delta_T/\Delta_S = (J_S/J_T) \qquad \text{(eqn 8)}$$

where J_T is the gene identity of the pooled subpopulations and p_{ij} is the population frequency of the i-th allele in the j-th population ($\sum_{i=1}^{k} p_{ij} = 1$ for each j):

$$J_T \equiv \sum_{i=1}^{k}\left[\sum_{j=1}^{n}(p_{ij}/n)\right]^2$$

For n equally weighted subpopulations, Δ_{ST} (the between-subpopulation component of diversity, or the effective number of distinct subpopulations) has a minimum of unity when all subpopulations are identical, and a maximum of n when all n subpopulations are completely distinct. It is independent of the within-subpopulation diversity. The reciprocal of eqn 8 gives the ratio of within-subpopulation diversity to total diversity. Unlike the ratio of within-subpopulation heterozygosity to total heterozygosity, this measure correctly ranks the species in Table 1 according to the degree of differentiation or similarity of their subpopulations.

The partitioning formula for true diversities, eqn 7, can also yield the proper partitioning formulas for other quantities of interest to geneticists. For example, it follows from eqn 7 that the total gene identity of a population can be partitioned into independent within- and between-subpopulation components according to:

$$J_T = J_S \cdot J_{ST}$$

Similarly, the partitioning formula for heterozygosity can be derived by using eqn 6 to write each term of eqn 7 in terms of heterozygosities:

$$1/(1 - H_T) = [1/(1 - H_S)] \cdot [1/(1 - H_{ST})].$$

This leads to the unique equation that partitions heterozygosity into pure, independent within- and between-subpopulation components H_S and H_{ST}:

$$H_T = H_S + H_{ST} - H_S H_{ST} \text{ or } H_{ST} = (H_T - H_S)/(1 - H_S). \quad \text{(eqn 9)}$$

This formula has been derived by other means in information theory (Aczel & Daroczy 1975) and physics (Tsallis & Brigatti 2004). For n equally large subpopulations, H_{ST} ranges from zero (no differentiation) to $1–1/n$ (complete differentiation). The value of H_{ST} for completely differentiated subpopulations is independent of H_S, in contrast to the corresponding value of D_{ST} (Fig. 1). Comparison of eqs 2 and 9 reveals that Nei's D_{ST} is actually $H_{ST} - H_S \cdot H_{ST}$, which explains the strong negative dependence of D_{ST} (and G_{ST}) on H_S. This new between-subpopulation component of heterozygosity correctly ranks the species in Table 1 according to their degree of subpopulation differentiation. It should replace D_{ST} when the goal is to describe subpopulation differentiation. It should be converted to effective number of distinct subpopulations Δ_{ST}, using eqn 6, before interpretation. The reciprocal of this gives the proportion of total diversity contained in the average subpopulation.

Relative differentiation

Δ_{ST} is an absolute measure of subpopulation differentiation, since it gives the effective number of distinct subpopulations. To replace G_{ST} we need a relative measure of differentiation, one that ranges from zero (no differentiation) to unity (complete differentiation). Relative differentiation is especially useful when it is necessary to compare data sets with different numbers of subpopulations. A wide range of transformations of Δ_{ST} onto the unit interval is available, depending on the desired purpose. Each transformation produces a differentiation or similarity measure with specific properties. All inherit Δ_{ST}'s independence from within-subpopulation diversity, a desirable property.

Perhaps the most useful measure of differentiation would be linear in the proportion of total diversity that is contained in the average subpopulation. This proportion is Δ_S/Δ_T which equals $1/\Delta_{ST}$. For n subpopulations, it has a maximum of unity when all n subpopulations are identical, and a minimum of $1/n$ when all n subpopulations are completely distinct (sharing no alleles). The linear transformation of this onto the unit interval which increases with increasing differentiation is:

$$D \equiv [(\Delta_S/\Delta_T) - 1]/[(1/n) - 1] \quad \text{(eqn 10)}$$

It can be written in terms of standard genetic quantities:

$$D = (J_T/J_S - 1)/[(1/n) - 1] \text{ or } [(H_T - H_S)/(1 - H_S)]\,[n/(n - 1)]. \quad \text{(eqn 11)}$$

If two subpopulations consist of k equally common alleles, this measure gives the proportion of each subpopulation's alleles that are unique to that subpopulation. This is a measure of pure differentiation, independent of average within-subpopulation heterozygosity, and it should replace G_{ST} when differentiation is the quantity of interest. Other transformations of Δ_S/Δ_T onto the unit interval may also be useful. Monotonic transformations of Δ_S/Δ_T are the only measures of differentiation that are mathematically consistent with (will never contradict) conclusions about subpopulation similarity based on the effective number of alleles (Jost 2007).

Compositional differentiation and similarity have nothing to do with the relative sizes of the subpopulations, so formulas such as eqn 11 involve only allele frequencies, not their raw abundances. The statistical weights of the subpopulations do not enter into the formulae. However, for some purposes, as when studying the contributions of hierarchical levels to total gene diversity, the weights are important. For applications in which weights play an essential role, diversity measures based on heterozygosity have fundamental mathematical limitations (Jost 2007). The only measures whose between-subpopulation components

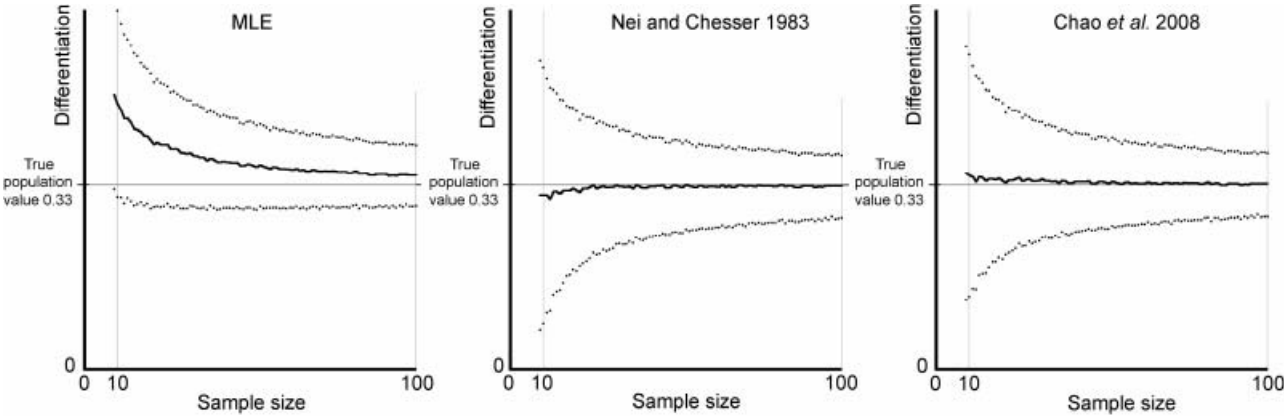

Fig. 3 Comparison of three estimators of differentiation D vs. sample size. Means of 1000 runs. Envelopes are +/− 1 standard deviation centred on the means. These estimators are applied to a large population consisting of two moderately differentiated subpopulations. The true differentiation is 0.33. Left to right: no estimator (eqn 11 using sample frequencies in place of population frequencies); eqn 12, estimator based on Nei & Chesser (1983); eqn 13, estimator based on Chao *et al*. (2008).

are independent of within-subpopulation diversity, and at the same time monotonic with respect to differentiation, are measures based on Shannon entropy or its exponential (Jost 2007).

Estimating differentiation from population samples

True differentiation D, like G_{ST}, must generally be estimated from small samples taken from a much larger population. Estimates of D and G_{ST} are biased when calculated directly using the sample frequencies in place of population frequencies in H_S and H_T. This bias can be reduced by using the nearly unbiased estimators of H_T and H_S derived by Nei & Chesser (1983). Define

n, number of subpopulations

$\tilde{N}$, harmonic mean of the subpopulation sample sizes

p'_{ij}, frequency of allele i in the sample from the jth subpopulation

$$H'_j = 1 - \sum_{i=1}^{k} (p'_{ij})^2$$

$$H'_S = (1/n) \sum_{j=1}^{n} H'_j$$

$$H'_T = 1 - \left\{ \sum_{i=1}^{k} \left[(1/n) \sum_{j=1}^{n} p'_{ij} \right]^2 \right\}$$

Nei & Chesser's (1983) estimators for a diploid organism are:

$$H_{S\_est} = \left[2\tilde{N}/(2\tilde{N} - 1) \right] H'_S$$

$$H_{T\_est} = H'_T + (H_{S\_est})/(2\tilde{N}n)$$

These were developed to estimate G_{ST} (by substituting H_{S_est} and H_{T_est} into eqn 3, yielding G_{ST_est}) but they can also be used in eqn 11 to estimate D:

$$D_{est} = [(H_{T\_est} - H_{S\_est})/(1 - H_{S\_est})]\,[n/(n-1)]. \qquad \text{(eqn 12)}$$

Simulations (Fig. 3) show that the D_{est} is nearly unbiased, and that its variance is less than or equal to the relative variance of G_{ST_est}.

Differentiation D turns out to be the complement of the n-community generalization of the popular and robust Morisita–Horn similarity measure used by ecologists (Chao *et al*. 2008). For two communities or subpopulations, this has a well-known unbiased estimator due to Morisita (Krebs 1999). Chao *et al*. (2008) generalized this unbiased estimator to apply to their multiple-community similarity measure C_{2n}. When that estimator is subtracted from unity, an unbiased estimator of D is obtained:

$$D_{est\_Chao} = 1 - (a/b)$$

$$a = \sum_{i=1}^{k} \left\{ \left[\left(\sum_{j=1}^{n} N_{ij}/N_j \right)^2 - \sum_{j=1}^{n} (N_{ij}/N_j)^2 \right] \Big/ (n-1) \right\}$$

$$b = \sum_{i=1}^{k} \sum_{j=1}^{n} \left\{ N_{ij}[N_{ij} - 1]/[N_j(N_j - 1)] \right\}$$

(eqn 13)

where N_{ij} is the number of examples of allele i sampled from subpopulation j, and N_j is the total number of alleles sampled. (Note that for a diploid organism, N_j is twice the number of individuals sampled from subpopulation j.) The

variance of this estimator can be estimated using the bootstrap estimate described in Chao *et al*. (2008).

Figure 3 compares three estimators of D: eqn 11 (with sample frequencies used as estimates of population frequencies), D_{est} using Nei & Chesser's (1983) estimators of H_T and H_S, and D_{est_Chao}. The latter two estimators give very similar nearly unbiased results.

Application to the finite-island model

The actual differentiation D (eqs 10, 11) can be easily connected to the parameters of common genetic models. Chao *et al*. (2008) show that the multiple-community generalization of the Morisita–Horn similarity index in ecology (which is $1 - D$) is the ratio of G_D (the probability that two alleles randomly drawn from different populations are identical) to G_S (the probability that two alleles drawn from the same population are identical). D can thus be written:

$$D = 1 - G_D/G_S \qquad \text{(eqn 14)}$$

Most genetic models yield expressions for G_D and G_S, and these can be inserted directly into eqn 14 to give the equilibrium differentiation for those models.

For example, the finite-island model with infinite alleles is often used to investigate the effects of migration and mutation on diversity and population structure. In this model as developed by Crow (1986) and Rousset (2004), there are n subpopulations, each of effective size N; the proportion of migrants in each subpopulation in a given generation is m, and the mutation rate per generation is μ. Migration is assumed to be equally likely between any two different subpopulations, and mutations are assumed to always produce new alleles. For this model, the expected value of the ratio G_D/G_S at equilibrium can be calculated by dividing Rousset's (2004) eqn 3.11 by his eqn 3.10. Inserting this into eqn 14 shows that differentiation at equilibrium is:

$$D = \{1 + (1 - g)/[ng(2\mu - \mu^2)]\}^{-1} \qquad \text{(eqn 15)}$$

where $g \equiv [1 - mn/(n-1)]^2$. This can be greatly simplified when $m < 1$ and $\mu << m$, using Crow's (1986) approximate expressions for G_S and G_D:

$$G_S = [m + \mu(n-1)]/[m + 4Nmn\mu + (n-1)\mu], \qquad \text{(eqn 16)}$$

$$G_D = [m]/[m + 4Nmn\mu + (n-1)\mu]$$

(Crow 1986). Under this approximation, the differentiation D at equilibrium is

$$D = 1 - G_D/G_S = [\mu(n-1)]/[\mu(n-1) + m] \qquad \text{(eqn 17)}$$

which has a very simple form when $m << 1$ and $\mu n << m$:

Table 2 G_{ST} vs. actual differentiation D. Expected values under the finite-island model with infinite alleles

n	N	m	μ	G_{ST}	D
5	100	0.01	0.001	**0.127**	**0.282**
5	1 000			**0.014**	**0.282**
5	10 000			**0.001**	**0.282**
10	10 000			**0.002**	**0.469**
20	10 000			**0.002**	**0.651**
40	10 000			**0.002**	**0.793**
80	10 000			**0.002**	**0.886**
160	10 000			**0.002**	**0.940**
320	10 000			**0.002**	**0.969**
640	10 000			**0.002**	**0.984**
1280	10 000			**0.002**	**0.992**
2560	10 000			**0.002**	**0.996**
2560	1 000			**0.022**	**0.996**
2560	100			**0.183**	**0.996**
2560	10			**0.691**	**0.996**
2560	10		0.0001	**0.709**	**0.962**
2560	10		0.00001	**0.711**	**0.716**
2560	10		0.000001	**0.711**	**0.201**

Number of demes, n; size of each deme, N; migration rate, m; mutation rate, μ. G_{ST} defined in eqn 3, D defined in eqn 11.

$$D \approx \mu(n-1)/m$$

$$\approx \mu/m \text{ for } n = 2;$$

$$\approx \mu n/m \text{ for moderate } n$$

When n is large and $\mu n << m$, the actual differentiation at equilibrium is therefore approximately proportional to mutation rate and number of subpopulations, and inversely proportional to migration rate. It is independent of deme size. This is a dramatically different picture from that painted by G_{ST}, which is not independent of deme size but is independent of mutation rate and number of subpopulations when n is large and $\mu n << m$. The strong dependence of real differentiation on mutation rate implies that expected differentiation may vary from locus to locus, if mutation rates are significantly different between loci.

Table 2 gives expected values of G_{ST} and actual differentiation D at equilibrium for several combinations of n, N, m, and μ. The actual differentiation behaves very differently from G_{ST} when n is large and mutation rate is high. This may have consequences for theories of speciation.

G_{ST} is often used to estimate gene flow between subpopulations. In the finite-island model with infinite alleles, the exact expected value of G_{ST} at equilibrium is given by Takahata & Nei's (1984) eqn 5. Their m must be transformed to $mn/(n-1)$ to correspond with Crow's and Rousset's usage of m. After making this substitution, Takahata and Nei's expression can be solved for m:

Table 3 Calculating migration rates and mutation rates from structural parameters of the finite-island model

n	N	G_{ST}	D	Actual m	Estimated m from eqn 19	Estimated m from classical eqn 20	Actual μ	Estimated μ (eqn 23) from D, estimated m via eqn 19	Estimated μ (eqn 24) from D, G_{ST}, n, N
2	**10 000**	0.000	0.007	**0.1**	0.100	0.141	**0.001**	0.000714	0.001005
2	**100**	0.004	0.007		0.100	0.141		0.000714	0.001005
20		0.019	0.138		0.101	0.120		0.000852	0.001010
200		0.020	0.629		0.101	0.119		0.000859	0.001010
2		0.004	0.000		0.100	0.141	**0.00001**	0.000007	0.000010
20		0.019	0.002		0.100	0.118		0.000008	0.000010
200		0.021	0.017		0.100	0.117		0.000009	0.000010
2		0.010	0.017	**0.05**	0.050	0.059	**0.001**	0.000860	0.001010
20		0.039	0.259		0.051	0.055		0.000938	0.001019
200		0.043	0.786		0.051	0.055		0.000942	0.001020
2		0.011	0.000		0.050	0.059	**0.00001**	0.000009	0.000010
20		0.040	0.003		0.050	0.054		0.000009	0.000010
200		0.044	0.036		0.050	0.054		0.000009	0.000010
2		0.055	0.088	**0.01**	0.010	0.011	**0.001**	0.001017	0.001050
20		0.169	0.651		0.011	0.011		0.001076	0.001095
200		0.181	0.951		0.011	0.011		0.001081	0.001100
2		0.057	0.001		0.010	0.010	**0.00001**	0.000010	0.000010
20		0.182	0.018		0.010	0.010		0.000010	0.000010
200		0.196	0.164		0.010	0.010		0.000010	0.000010

Number of demes, n; size of each deme, N; migration rate, m; mutation rate, μ. Estimated m and μ are calculated from n, N, and expected values of G_{ST} and D.

$$m = \{1 - [(1-\mu)(w+1)^{1/2}]^{-1}\}\{[n-1]/n\} \quad \text{(eqn 18)}$$

where $w \equiv [(1-G_{ST})/G_{ST}][(n-1)/n][1/(2N)]$.

Since μ is always very small compared to unity, the factor $1-\mu$ has virtually no effect, so eqn 18 is approximately:

$$m \approx [1 - (w+1)^{-1/2}][(n-1)/n] \quad \text{(eqn 19)}$$

This provides a nearly exact estimate for m, which can be simplified when w is small, using a power series expansion:

$$m \approx [(1-G_{ST})/(4N \cdot G_{ST})][(n-1)/n]^2. \quad \text{(eqn 20)}$$

This is the classical equation for m (Crow & Aoki 1984). Values of m based on eqns. 19 and 20 are compared in Table 3.

The actual differentiation D provides another way to estimate migration rate, if mutation rate is known or can be estimated and if the population is in drift–migration equilibrium. The exact eqn 15 can be solved for m:

$$m = \{1 - [(n((1/D)-1)(2\mu-\mu^2)) + 1]^{-1/2}\}\{[n-1]/n\}. \quad \text{(eqn 21)}$$

If μ^2 is small compared to μ, this is approximately

$$m \approx \mu(n-1)(1-D)/D. \quad \text{(eqn 22)}$$

This correctly yields a migration rate of zero when differentiation is complete.

When geneticists apply the finite-island model to a real population, both m and μ are usually unknown. By using G_{ST} and D, both m and μ can be calculated from empirical population data if the structural parameters N and n are known or can be estimated, and if one is confident that the population is at or near equilibrium. The most precise way to do this would be to set the two exact equations for migration rate, eqns 18 and 21, equal to each other, and solve numerically for μ. However, the approximate equations given above are robust, and they lead to a simple algebraic solution for μ. Equations 19 or 20 provide estimates for m in terms of G_{ST}. This estimate, m_{est}, can be substituted for m in eqn 22, and the resulting equation can be solved for μ:

$$\mu = (D \cdot m_{est})/[(1-D)(n-1)] \quad \text{(eqn 23)}$$

Equations 20 and 23 can be combined to give a simple approximate expression for μ that does not require knowledge of migration rate; it depends only on the structural parameters n and N and the descriptive parameters D and G_{ST}:

$$\mu \approx [D/(1-D)]\,[(1-G_{ST})/G_{ST}]\,[(n-1)/(n^2 \cdot 4N)]. \quad \text{(eqn 24)}$$

This simplifies considerably when written in terms of standard genetic quantities:

$$\mu \approx (1-G_S)/(4NnG_D) = (H_S)/[4Nn(1-H_D)] \quad \text{(eqn 25)}$$

where $H_D = 1 - G_D$. Equation 25 can easily be verified by substituting the expressions for G_D and G_S given in eqn 16. Table 3 compares estimates of m (via Eqns. 19 and 20) and μ (via eqns 23 and 24) for a variety of values of n, N, m, and μ. The table shows that N plays no role in these estimates, and that the approximate formulas for m and μ are reasonably accurate. As with traditional estimates, the real accuracy of these formulae depend on how well the population in question fits the finite-island model and whether it really is in drift–mutation equilibrium.

Conclusions

Some standard forms of reasoning in population genetics are invalid:

1 In conservation genetics, it is often said that if mean within-subpopulation 'diversity' H_S is relatively high with respect to between-subpopulation 'diversity' D_{ST}, or if G_{ST} is low, then the subpopulations are not highly differentiated and it is not critical to maintain many different subpopulations of the species (e.g. Hunter 1995). This is a dangerous myth. When there are many common alleles, the within-subpopulation heterozygosity H_S is *always* much higher than D_{ST}, and G_{ST} is *always* close to zero, *even when each subpopulation consists entirely of private alleles* (Fig. 1 and Table 1). The same invalid forms of reasoning are common in molecular taxonomy and other applications of population genetics. The new partitioning and differentiation measures derived here must replace the traditional measures in these applications.

2 If G_{ST} of one species is higher than that of another species, geneticists regularly infer that the species with the higher G_{ST} has more highly differentiated subpopulations. The dependence of G_{ST} on mean subpopulation heterozygosity H_S (Fig. 1) is virtually never taken into account. G_{ST} does not necessarily rank species correctly in terms of differentiation (e.g. Table 1, species B vs. C). The true differentiation D does rank them correctly.

3 It is often said that under the finite-island model, when the number of absolute migrants per generation is greater than unity, G_{ST} will be low, so subpopulations will not evolve significant differentiation. This rule is based on interpretation of G_{ST} as a measure of differentiation. Since G_{ST} is not a measure of differentiation, the rule is irrelevant. In fact, subpopulation differentiation in the finite-island model depends on the ratio of mutation rate to migration rate, not on the absolute number of migrants, and subpopulation differentiation can occur even when there is more than one migrant per generation (Table 2).

G_{ST} has some legitimate uses but it is not a measure of differentiation. G_{ST} can decrease as differentiation increases (Fig. 2); it can be arbitrarily close to zero even when subpopulations are fully differentiated (Fig. 1), and it can misrank populations in terms of their differentiation (Table 1). When an accurate description of diversity and differentiation is required, diversity should be equated with Kimura & Crow's (1964) effective number of alleles rather than heterozygosity, and G_{ST} should be replaced by the actual differentiation D (eqn 11). These changes in common practice would resolve widespread paradoxes in the literature, and would alter many conclusions based on G_{ST}. Conclusions based on Hedrick's standardized G_{ST} (2005) are more robust, because standardized G_{ST} approaches the true differentiation D when diversity is high.

While the measures of differentiation discussed here connect nicely to standard genetic models, they are accurate model-independent descriptive measures, so they will be especially useful when such models are unavailable.

Acknowledgements

This work was supported by grants from John Moore to the Population Biology Foundation, and from Steven K. Beckendorf and Cynthia L. Hill to the EcoMinga Foundation. I acknowledge constructive criticism from Drs Olivier Hardy, James Crow, Anne Chao, Montgomery Slatkin, Steven K. Beckendorf, Brad Jost, Tom Walla, and Don Waller.

References

Aczel J, Daroczy Z (1975) *On Measures of Information and Their Characterization*. Academic Press, New York.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**, 155–165.

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.

Carreras-Carbonell J, Macpherson E, Pascual M (2006) Population structure within and between subspecies of the Mediterranean triplefin fish *Tripterygion delaisi* revealed by highly polymorphic microsatellite loci. *Molecular Ecology*, **15**, 3527–3539.

Chao A, Jost L, Chiang SC, Jiang YH, Chazdon R (2008) A two-stage probabilistic approach to multiple-community similarity indices. *Biometrics*, in press.

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular and Biological Evolution*, **15**, 538–543.

Crow J (1986) *Basic Concepts in Population, Quantitative, and Evolutionary Genetics*. Freeman, San Francisco, California.

Crow J (2004) *Evolutionary Theory and Processes: Modern Horizons, Papers in Honour of Eviatar Nevo*, Wasser SP ed. Kluwer, Dordrecht, The Netherlands.

Crow J, Aoki K (1984) Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. *Proceedings of the National Academy of Sciences, USA*, **81**, 6073–6077.

Hannah L, Kay J (1977) *Concentration in the Modern Industry: Theory, Measurement, and the UK Experience*. Macmillan, London.

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

Hill M (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology*, **54**, 427–432.
Hunter M (1995) *Fundamentals of Conservation Biology*. Blackwell Science, Cambridge, Massachusetts.
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.
Kimura M, Crow J (1964) The number of alleles that can be maintained in a finite population. *Genetics*, **49**, 725–738.
Krebs C (1999) *Ecological Methodology*. Harper & Row, New York.
Lewontin R (1972) The apportionment of human diversity. *Evolutionary Biology*, **6**, 381–398.
Nagylaki T (1998) Fixation indices in subdivided populations. *Genetics*, **148**, 1325–1332.
Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences, USA*, **70**, 3321–3323.
Nei M, Chesser R (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, **47**, 253–259.
O'Reilly P, Canino M, Bailey K, Bentzen P (2004) Inverse relationship between F_{ST} and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Molecular Ecology*, **13**, 1799–1814.
Rousset F (2004) *Genetic Structure and Selection in Subdivided Populations*. Princeton Press, Princeton, New Jersey.
Sanetra M, Crozier R (2003) Patterns of population subdivision and gene flow in the ant *Nothomyrmecia macrops* reflected in microsatellite and mitochondrial DNA markers. *Molecular Ecology*, **12**, 2281–2295.
Takahata N, Nei M (1984) F_{ST} and G_{ST} statistics in the finite island model. *Genetics*, **107**, 501–504.
Tsallis C, Brigatti E (2004) Nonextensive statistical mechanics: a brief introduction. *Continuum Mechanics and Thermodynamics*, **16**, 223–235.
Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Lou Jost (www.loujost.com) was trained as a mathematician and physicist but now studies the biogeography of orchids in the Andean cloud forests of Ecuador where he lives. His goal is to understand the processes of speciation in tropical forests. He cofounded the EcoMinga Foundation, www.ecominga.net, which identifies Andean forests with high levels of endemism and buys them for conservation.
