MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) **11**, 5–18 doi: 10.1111/j.1755-0998.2010.02927.x

INVITED TECHNICAL REVIEW

Assessing population structure: F_{ST} and related measures

PATRICK G. MEIRMANS* and PHILIP W. HEDRICK†
*Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, PO Box 94248, 1090GE, Amsterdam, The Netherlands, †School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA

Abstract

Although F_{ST} is widely used as a measure of population structure, it has been criticized recently because of its dependency on within-population diversity. This dependency can lead to difficulties in interpretation and in the comparison of estimates among species or among loci and has led to the development of two replacement statistics, F'_{ST} and D. F'_{ST} is the normal F_{ST} standardized by the maximum value it can obtain, given the observed within-population diversity. D uses a multiplicative partitioning of diversity, based on the effective number of alleles rather than on the expected heterozygosity. In this study, we review the relationships between the three classes of statistics (F_{ST}, F'_{ST} and D), their estimation and their properties. We illustrate the relationships between the statistics using a data set of estimates from 84 species taken from the last 4 years of *Molecular Ecology*. As with F_{ST}, unbiased estimators are available for the two new statistics D and F'_{ST}. Here, we develop a new unbiased F'_{ST} estimator based on G_{ST}, which we call G''_{ST}. However, F'_{ST} can be calculated using any F_{ST} estimator for which the maximum value can be obtained. As all three statistics have their advantages and their drawbacks, we recommend continued use of F_{ST} in combination with either F'_{ST} or D. In most cases, F'_{ST} would be the best choice among the latter two as it is most suited for inferences of the influence of demographic processes such as genetic drift and migration on genetic population structure.

Keywords: D, fixation, F-statistics, G'_{ST}, heterozygosity, population differentiation

Received 30 May 2010; revision received 17 August 2010; accepted 2 September 2010

The trouble with F_{ST}

Quantifying population structure using F_{ST}

Determining the genetic structure of natural populations forms an important part of population genetics and has many applications in evolutionary biology, conservation, forensics, and plant and animal breeding. The method most frequently used to assess population structure is the calculation of F_{ST}, a summary statistic first introduced by Sewall Wright (1943a, 1965). Wright originally developed his F-statistics as inbreeding coefficients, defined as a correlation between uniting gametes. This was long before the advent of allozymes and other molecular genetic markers, and Wright therefore assumed loci to be biallelic. Like most geneticists of his time, he focused on morphological characters with simple Mendelian inheritance. In fact, his landmark paper on isolation by distance and its effects on the distribution of genetic variation (Wright 1943a) was directly followed by another paper in which he illustrated his findings with a data set on the distribution of white and blue flowers in *Linanthus parryae* in the Mojave desert (Wright 1943b). Later, when allozymes were introduced as a convenient marker to assess the genetic diversity of a population, Wright's F_{ST} was adapted for use with multiallelic loci, redefined as a ratio of genetic variances (Cockerham 1973). This led to the development of several statistical frameworks to estimate F_{ST} statistics from small samples from a limited number of populations (e.g. Weir & Cockerham 1984; Nei 1987). Nei's (1987) G_{ST} is a direct expansion of Wright's work and is based on a comparison of the expected heterozygosity (gene diversity) within and among populations. The method of Weir & Cockerham (1984) uses an ANOVA approach to estimate within- and among-population variance components, which are then used to estimate their F_{ST} analogue θ. These F_{ST} analogues became widely used for analysing allozyme data, primarily because of their

Correspondence: Patrick Meirmans, Fax: +31 20 525 7832;
E-mail: p.g.meirmans@uva.nl

ability to describe the genetic population structure in a single summary statistic and the direct link between F_{ST} and the rate of gene flow (which we will refer to as the 'migration rate' below).

With the discovery of new, more variable, genetic markers, new F_{ST} analogues were developed to take the special properties of these markers into account. Excoffier *et al.* (1992) used the ANOVA approach of Weir & Cockerham (1984), but performed this on a matrix of squared Euclidean distances between DNA haplotypes, from which the F_{ST} analogue φ_{ST} is calculated. Slatkin (1995) developed R_{ST}, which is especially suited for markers with a stepwise mutation model such as some microsatellites. In addition to these new methods, the older G_{ST} and θ statistics are still widely used for the analysis of highly variable microsatellite markers. Nowadays, a large number of different marker types are available for population genetic studies, with a large range of allelic diversities, from SNPs that are essentially biallelic to microsatellites that can have over 50 alleles at a single locus (e.g. Peijnenburg *et al.* 2006). Despite their different diversities, all these markers are analysed with what is basically still the same F_{ST} statistic that was originally developed for biallelic data. Only recently have biologists started to become aware of the limitations of *F*-statistics for analysing data from highly variable loci (Charlesworth 1998; Hedrick 1999, 2005; Balloux *et al.* 2000; Balloux & Lugon-Moulin 2002; Long & Kittles 2003; Gregorius *et al.* 2007; Jost 2008; Gregorius 2010).

Dependency on H_S

When defined as a ratio of genetic variances (Cockerham 1973), F_{ST} and its analogues work by relating the amount of genetic variation among populations to the total genetic variation over all populations. For biallelic markers, this makes sure that F_{ST} is bounded between zero and one, with zero representing no differentiation and one representing fixation of different alleles within populations. For multiallelic markers, however, the maximum possible value is not necessarily equal to one, but is instead determined by the amount of within-population diversity (Charlesworth 1998; Hedrick 1999). The reason for this can be best understood by looking at G_{ST}, which is defined as (Nei 1987)

$$G_{ST} = \frac{(H_T - H_S)}{H_T},$$

where H_T is the total gene diversity, and H_S is the within-population gene diversity (equal to the expected heterozygosity for diploids). Because the true population value of H_T is necessarily bigger than or equal than that of H_S and the maximum possible value for H_T is one, it follows that the maximum value of G_{ST} equals $1-H_S$ (Charlesworth 1998; Hedrick 1999; Jost 2008). For highly variable loci, this can lead to a very small possible range of G_{ST} values. To illustrate this relationship, Fig. 1 gives the joint values of F_{ST} and H_S found in the past 4 years in *Molecular Ecology* (expanded from Heller & Siegismund 2009; see also Table S1, Supporting information). Notice that the observed range of F_{ST} is always less than H_S and that the range of F_{ST} becomes very small when H_S is large. For example when $H_S = 0.9$, a value that is commonly encountered for microsatellite markers, the maximum possible value of F_{ST} is 0.1. Such a value of F_{ST} is generally interpreted as representing a rather weak population structure. However, here it represents the case with maximum differentiation among the populations, meaning that the populations do not share any alleles at all. It is important to realize that this is not a statistical issue, deriving from the sampling of individuals from populations, but that the problem also occurs when the actual population allele frequencies are used (Jost 2008).

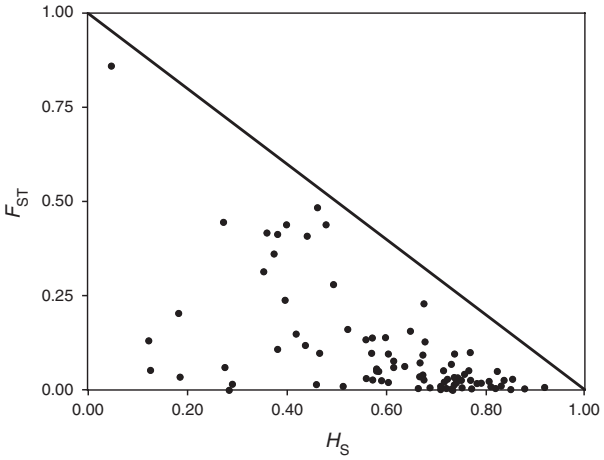

Fig. 1 The maximum possible value of F_{ST} as a function of the expected heterozygosity within-population H_S (solid line). The closed circles represent values from 84 species published in *Molecular Ecology* over the last 4 years (expanded from Siegismund and Heller, 2009).

Although the dependency on the amount of within-population variation has mostly been discussed for G_{ST} (Hedrick 2005; Jost 2008; Ryman & Leimar 2008), it is also present for other F_{ST} estimators such as θ and φ_{ST} (Balloux *et al.* 2000; Meirmans 2006). For these statistics, the calculation of the maximum possible value is less straightforward than for G_{ST} and requires calculating the maximum possible among-population variance component, given the within-population variance in the sample (Meirmans 2006). However, in most cases, G_{ST}, θ, and φ_{ST} give highly similar values, so that their maximum values will also generally be close or equal to $1 - H_S$.

One statistic that is not affected by the amount of within-population variation is R_{ST}, which was especially

developed for markers with a stepwise mutation model, such as some microsatellites (Slatkin 1995). Slatkin showed that estimates of the number of migrants, calculated from R_{ST}, were essentially unbiased over a range of mutation rates and were much better than estimates calculated from F_{ST}. Similar results were obtained by Balloux *et al.* (2000) who found that estimates of R_{ST} were mostly unbiased for highly variable microsatellite loci with up to 30 alleles, while F_{ST} was severely underestimated. However, estimates for R_{ST} were only satisfactory when the mutations strictly followed the stepwise mutation model (Balloux *et al.* 2000). When a small proportion of random mutations were added, most of the ''memory'' in the mutation process was lost, and estimates of R_{ST} were not reliable. As in practice microsatellites hardly ever follow a strict stepwise mutation model, the use of R_{ST} is best avoided (but see Excoffier & Hamilton 2003). In fact, even when mutation is strictly stepwise, R_{ST} is not necessarily always a better estimator than F_{ST} (Balloux & Goudet 2002). For example, this is the case when the timescale of interest is short and the influence of mutation is relatively small (Slatkin 1995).

Difficulties in interpretation

Obviously, the dependency of many F_{ST} estimators on the level of diversity will cause difficulties in their interpretation. This will be especially the case when markers are compared that have different mutation rates or when species are compared with different effective population sizes. Comparisons of *F*-statistics within species can also be difficult when different parts of the distribution are compared that differ in diversity. For example, an invasive species may have a lower diversity in the invaded area than in the original distribution area, which may give problems when F_{ST} is used to compare the population structure within the two areas. Indeed, several authors have remarked that their estimates of F_{ST} did not conform to the expectations based on what was known about their study organism or that the estimates varied over loci with different mutation rates (e.g. Balloux *et al.* 2000; O'Reilly *et al.* 2004; Carreras-Carbonell *et al.* 2006).

In a comparison of two chromosomal races of the common shrew, Balloux *et al.* (2000) found that the estimates of genetic differentiation were much lower than expected based on earlier studies. For example, for one highly variable Y-chromosomal microsatellite, the value of θ was only 0.19, even though no alleles were shared between the two races. Ten autosomal microsatellite loci showed an even lower overall θ value of 0.10. In contrast, a biallelic mtDNA marker where the two alleles were shared by the two races had a θ value of 0.56. Balloux *et al.* (2000) then conducted simulations to show that indeed the estimates of θ were strongly affected by the mutation rates of the loci. Therefore, they concluded that, despite the low θ values, the two races were genetically strongly differentiated as a result of almost complete reproductive isolation.

O'Reilly *et al.* (2004) used 14 microsatellite loci to study the population structure of a marine fish, Walleye pollock. The loci varied dramatically in the number of alleles (6–43), resulting in expected heterozygosities ranging from 0.68 to 0.96. They found that the estimates of θ declined significantly with increasing heterozygosity, leading them to conclude that 'mutation rates of some microsatellite loci are sufficiently high to limit resolution of weak genetic structure' (O'Reilly *et al.* 2004). They attributed the observed correlation to size homoplasy, downplaying the effect of heterozygosity itself as they regarded the population structure too weak to be affected by this. However, also when the population structure is weak, F_{ST} will be affected by the level of heterozygosity and this can therefore readily explain the observed correlation.

The proposed solutions

Hedrick's G'_{ST}

Having noted earlier (Hedrick 1999) that the diversity restricts the possible range of G_{ST}, Hedrick (2005) suggested standardizing G_{ST} by the maximum value it can obtain given the observed within-population diversity. This method of standardization was inspired by Lewontin's (1964) measure of linkage disequilibrium *D'*, which is the standard measure *D*, divided by the maximum possible value given the observed allele frequencies.

Hedrick used the original (Nei 1973) definition of G_{ST} as $(H_T\text{-}H_S)/H_T$ and found that its maximum value ($G_{ST(max)}$) is a function of the expected heterozygosity, H_S, and the number of sampled populations *k*

$$G_{ST(max)} = \frac{(k-1)(1-H_S)}{k-1+H_S}$$

Hedrick then defined the standardized G_{ST}, which he called G'_{ST}, as (equation 4b in Hedrick 2005)

$$G'_{ST} = \frac{G_{ST}}{G_{ST(max)}} = \frac{G_{ST}(k-1+H_S)}{(k-1)(1-H_S)} \qquad (1)$$

When *k* is large, $G_{ST(max)}$ becomes equal to $1 - H_S$, the same value that was obtained above (Charlesworth 1998; Hedrick 1999; Jost 2008). The standardization ensures that G'_{ST} has an upper limit of 1, which is reached when the populations have nonoverlapping sets of alleles or when all populations are fixed for a single allele ($H_S = 0$) and there are two or more different alleles over all populations.

Hedrick defined his standardized measure only for G_{ST}, but the rationale is also applicable to other F_{ST} analogues. Meirmans (2006) developed a method to estimate the standardized measure φ'_{ST} based on an analysis of molecular variance (AMOVA, Excoffier *et al.* 1992). In a normal AMOVA, the summary statistic φ_{ST} is defined as a function of the between-population variance component σ_2^a and the within-population variance component σ_2^b:

$$\varphi_{ST} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2}$$

The maximum value of φ_{ST} given the amount of within-population variation can then be found by maximizing the among-population variance σ_2^a. In an AMOVA, the variance components are calculated from a matrix of pairwise squared Euclidean distances between individuals. For a single locus, the maximum among-group variance can then be found by setting all distances between pairs of individuals from different populations to a value of one.

Jost's D

Jost (2008) argued that there are in fact two separate problems with G_{ST} and developed a new framework for analysing population differentiation based on ecological diversity theory. The first problem recognized by Jost is the same one that we saw above, where H_S puts a limit on the maximum possible differentiation. Jost argued that the additive partitioning that is used for G_{ST}, where the total diversity is the sum of the within-population and among-population diversity, is inadequate to describe the among-population diversity. The second problem recognized by Jost is that the expected heterozygosity is an unsuitable metric for describing the diversity, leading to unintuitive results. For example, the heterozygosity does not scale linearly with an increase in diversity. Going from two equally frequent alleles to 20 equally frequent alleles does not give a tenfold change in heterozygosity, but a more moderate change from 0.5 to 0.95. Changing from 20 to 200 equally frequent alleles gives even less change in heterozygosity from 0.95 to 0.995. For some uses, this is a desirable property, for example because the heterozygosity very well fits our human interpretation of diversity where changes in small numbers (from 2 to 20) are often considered more important than changes in large numbers (from 20 to 200) (Hubalek 2000). However, this quality makes the heterozygosity less suitable for a statistical breakdown of diversity. One additional advantage of using heterozygosity is its easy interpretation, because it presents the probability that a pair of randomly drawn genes are different.

Jost (2008) then developed a new framework for estimating genetic differentiation that avoids these two problems. Instead of using heterozygosity, Jost based his statistic D on the effective number of alleles, which Jost (2006, 2008) simply calls the 'true diversity'. The effective number of alleles scales linearly with an increase in equally frequent alleles, which, according to Jost, gives a more intuitive diversity estimate. A disadvantage of this diversity index is that it depends on the sample size, so rarefaction to a standard sample size is needed before estimates can be compared (note that this does not affect the estimation of D). The effective number of alleles is directly related to the heterozygosity and can be defined as $1/(1 - H_S)$. However, unlike the heterozygosity, the effective number of alleles does scale linearly with increases in diversity. Jost (2007, 2008) then developed a multiplicative approach to partition the diversity, where the total diversity is the product of the within-population and among-population diversity and shows that this approach is mathematically more robust than the additive partitioning used by Nei. He then transformed the among-population diversity into a summary statistic, D, which ranges from zero to one. Although this statistic is not directly based on the heterozygosity as an index of diversity, it can nevertheless be expressed as a function of the total and within-population heterozygosities (equation 11 in Jost 2008):

$$D = \left(\frac{k}{k-1}\right)\left(\frac{H_T - H_S}{1 - H_S}\right) \quad (2)$$

If $H_S = 0$, then $D = H_T\, k/(k-1)$. This means that when $k = 2$, D equals $2H_T$ and when k becomes large, D approaches H_T.

Relationships between G_{ST}, G'_{ST} and D

To visualize the relationships between the three summary statistics, G_{ST}, G'_{ST} and D, Heller & Siegismund (2009) collected data on 43 species from 34 studies published in *Molecular Ecology* between 2006 and 2008. They included all species for which estimates were given for both H_S and an F_{ST} analogue and used these estimates to calculate (by approximation) G_{ST}, H_T, D and G'_{ST}. We extended their data set and added another 41 species from 36 studies published in *Molecular Ecology* between January 2009 and March 2010. The strong positive correlation between G'_{ST} and D that was reported by Heller & Siegismund (2009) was also present in the extended data set, though with a slightly lower value for the correlation coefficient ($r = 0.85$ for the extended data set, vs. $r = 0.99$ for the smaller set).

As G_{ST}, G'_{ST} and D can all be expressed in terms of H_S, H_T and k, it is possible to directly analyse the relationships among these three statistics (Heller & Siegismund 2009). The relationship between G'_{ST} and G_{ST} is simple

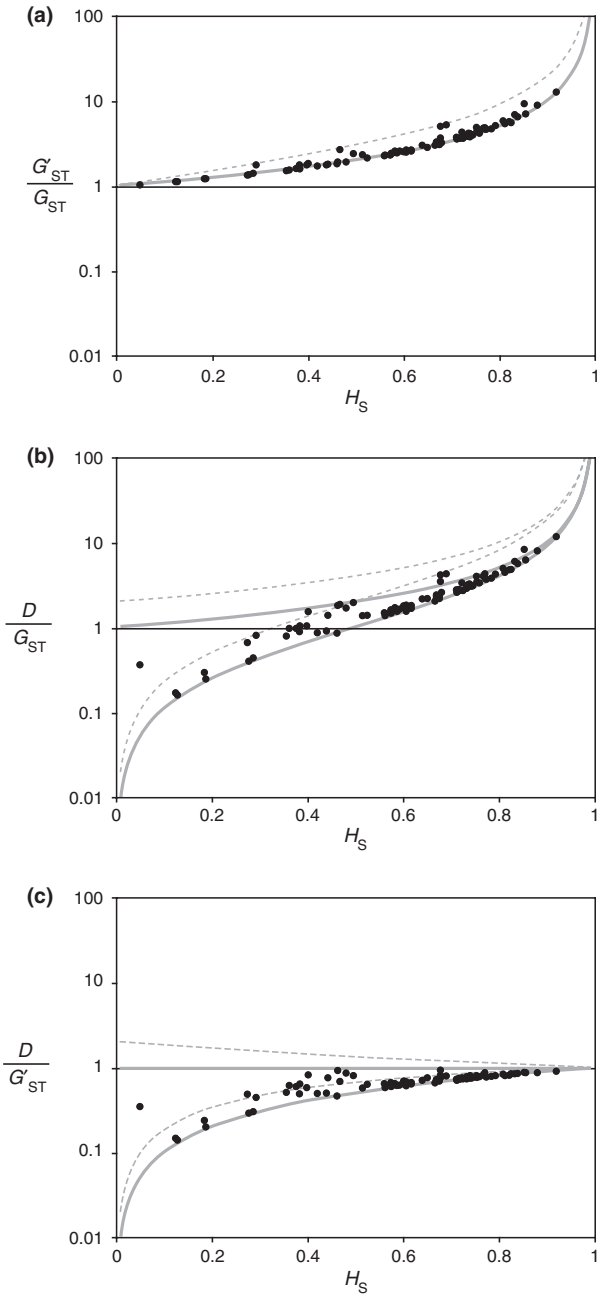

Fig. 2 The relationships between G_{ST}, G'_{ST} and D, as a function of the amount of within-population diversity H_S. (a) The ratio between G'_{ST} and G_{ST}, (b) the ratio between D and G_{ST}, (c) the ratio between D and G'_{ST}. The thick grey lines outline the possible range of the expected relationship for a large number of populations, the thin dotted lines outline the possible range for $k = 2$. These upper and lower limits, respectively, assume that $H_T = 1$ and $H_T = H_S$. The black symbols represent values from the literature (expanded from Siegismund and Heller, 2009).

(see Fig. 2a), because G'_{ST} is by definition always larger than G_{ST}. For large values of k, the ratio between the two statistics is (Hedrick 2005):

$$\lim_{k\to\infty} \frac{G'_{ST}}{G_{ST}} = \frac{1}{1 - H_S} \quad (3a)$$

The relationship between D and G_{ST} is less straightforward, but for large k, this simplifies to:

$$\lim_{k\to\infty} \frac{D}{G_{ST}} = \frac{H_T}{1 - H_S} \quad (3b)$$

This means that for $H_S > 0.5$, D is larger than G_{ST} (see Fig. 2b), while for small values of H_S, D is almost always smaller than G_{ST} (see Fig. 2b). Low values of H_S generally indicate a strong effect of genetic drift and/or a low mutation rate. For D, this means that low values of H_S often lead to low values of D relative to G_{ST}. The literature data show that in practice, D is mostly larger than G_{ST}, because in most of the analysed studies the within-population diversity is high. This indicates that G_{ST} indeed underestimates the amount of differentiation among populations. However, there are also some notable cases where G_{ST} shows stronger differentiation than D. The relationship between D and G'_{ST} can also most suitably be expressed as their ratio (see Fig. 2c), which for large k simply becomes (Heller & Siegismund 2009):

$$\lim_{k\to\infty} \frac{D}{G'_{ST}} = H_T \quad (3c)$$

From the above equations, we see that when k is large, the relationships between the three summary statistics become rather simple. However, when k is small, the relationships between the statistics can diverge from those presented above (dotted lines in Fig. 2). This can especially be the case for pairwise comparisons between populations, which are often used for detecting isolation by distance. Another interesting relationship between these statistics is that D is in fact equivalent to another method of standardization. Nei (1973) defined the between-subpopulation diversity as $D_{ST} = H_T - H_S$ (Nei 1973), which he used to calculate G_{ST} as D_{ST}/H_T. It can be proven that Jost's D is equivalent to a standardization of Nei's D_{ST} relative to its maximum possible value: $D = D'_{ST} = D_{ST}/D_{ST(max)}$ (Anne Chao unpublished res).

Heller & Siegismund (2009) also found that there was a significant negative correlation between H_S and the values of G_{ST}, G'_{ST} and D. However, they argued that the correlations of G'_{ST} and D with H_S were mainly caused by three outliers with 'somewhat extreme' demographic histories. In our extended data set, there was a significant relationship between H_S and G_{ST} ($r = -0.58$, $P = 0.001$, two-sided Monte Carlo test with 999 permutations), and between H_S and G'_{ST} ($r = -0.37$, $P = 0.001$), but not between H_S and D ($r = -0.18$, $P = 0.085$). Unlike for the data from Heller & Siegismund, the relationship between H_S and G'_{ST} was not caused by a few outliers.

Estimation

Correction for sampling bias

Genetic summary statistics are generally estimated from small samples taken from a small number of populations from a much larger metapopulation. When only a small number of individuals are sampled from a population, using the sample allele frequencies to directly calculate H_S and H_T will lead to a bias. Therefore, Nei & Chesser (1983) developed nearly unbiased estimators of H_S and H_T, intended for the calculation of G_{ST}. Later, Nei (1987, p. 164–165) introduced additional estimators to also correct for the amount of inbreeding, although the differences with the estimators of Nei & Chesser are generally rather small. When calculating G_{ST}, G'_{ST} and D, it is important to always use such unbiased estimates (Jost (2008) distinguishes between the theoretical D statistic and its estimator D_{EST} based on bias-corrected H_S and H_T). For calculating the overall values, H_S and H_T should be calculated giving equal weights to all populations, independently of the real population sizes or the sample sizes. In addition, Jost (2008) showed another method for estimating D without bias, derived from the unbiased estimation of the Morisita–Horn similarity index, which is frequently used in ecology (Chao *et al*. 2008).

For multiple loci, the estimates of H_S and H_T should first be averaged over loci before calculating G_{ST} or G'_{ST} (Nei 1973; Weir & Cockerham 1984). First calculating G_{ST} or G'_{ST} separately and then averaging these estimates is also frequently carried out (Culley *et al*. 2002), though this should be discouraged because it can lead to a downward bias, especially when few populations are sampled. Jost's D was developed as a single-locus measure, but a multilocus value can be obtained by taking the harmonic mean across loci (Crawford 2010). However, the above technique of first averaging H_S and H_T over loci also gives good results for multilocus estimates of D and has been used for all calculations below.

Nei (1987, p. 164–165) noted that G_{ST} implicitly includes a comparison of every population with itself, which leads to an underestimation when the number of sampled populations (k) is small. He therefore derived an unbiased version, which he called G'_{ST} (we will refer to this as $G'_{ST(Nei)}$ to avoid confusion with Hedrick's G'_{ST}):

$$G'_{ST(Nei)} = \frac{k(H_T - H_S)}{kH_T - H_S}$$

Hedrick's (2005) G'_{ST} suffers from the same underestimation when only a small number of populations is sampled and should therefore also be corrected. It can be shown (Appendix 1) that the maximum possible value of $G'_{ST(Nei)}$ is equal to $(1 - H_S)$. Therefore, the corrected version of Hedrick's G'_{ST} becomes

$$G''_{ST} = \frac{G'_{ST(Nei)}}{1 - H_S} = \frac{k(H_T - H_S)}{(kH_T - H_S)(1 - H_S)} \quad (4)$$

This statistic should be used for estimating Hedrick's standardized measure whenever the number of sampled populations is small, especially for pairwise comparisons. D does not suffer from this bias, as its definition already includes the correction term $k/(k - 1)$ (eqn 2).

To illustrate the effects of the number of sampled populations, we used the program Easypop (Balloux 2001) to simulate a metapopulation consisting of 100 populations, each with 20 individuals. Migration between populations followed an island model with a migration rate of 0.05 per generation. We simulated 20 neutral genetic markers with a maximum of 20 possible allelic states and a mutation rate of 0.0001. Simulations were run for 10 000 generations, with 100 replicates. From every data set of the 100 simulated data sets, we then randomly sampled from 2 to 50 populations and calculated G_{ST}, $G'_{ST(Nei)}$, G'_{ST}, G''_{ST} and D. Figure 3 shows the averages of the estimates of these statistics. Clearly, both G_{ST} and G'_{ST} are underestimated when the number of sampled populations is small, while $G'_{ST(Nei)}$, G''_{ST} and D are essentially unbiased. When k gets very large, the asymptotic values of G_{ST} and G'_{ST} become equal to those of $G'_{ST(Nei)}$ and G''_{ST}. Although the estimates of G_{ST} and G'_{ST} are always biased for small k, the actual extent of the bias depends for a large part on population parameters such as the migration rate and the mutation rate.

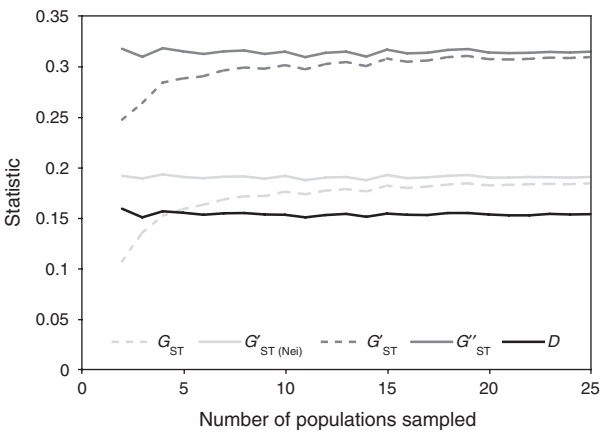

Fig. 3 Values of five different summary statistics as a function of an increasing number of sampled populations. Subsamples were taken of 2–50 subpopulations from simulated data sets of a metapopulation of 100 populations. Estimates of the summary statistics were calculated for every subsample taken and averaged over 100 replications.

Estimation of other F'_{ST} analogues

In general, the rationale behind Hedrick's standardization of G_{ST} can be applied to any F_{ST} analogue when its maximum value can be obtained. Unbiased estimates of φ'_{ST} can be obtained through an AMOVA using the method described by Meirmans (2006). When analysing population structure at multiple hierarchical levels, one has to keep in mind that the maximization has to be performed at the correct level. So for estimating φ'_{SC}, the degree of structure among populations within groups, the between populations variation should be maximized. For calculating φ'_{CT}, the degree of structure among groups of populations, the between populations variation should be maximized.

Weir & Cockerham's (1984) θ can also be standardized by dividing it by its maximum value. As with an AMOVA, the maximum value of θ can be obtained by maximizing the among-population variance component through the corresponding sum of squares. Unfortunately, the derivation of direct equations for θ_{max} is difficult. However, when sample sizes are large and equal and Hardy–Weinberg equilibrium can be assumed within populations, it can be shown that the value of θ_{max} equals the homozygosity or 1-H_S (Steven Kalinowski, in prepration), the same value as the maximum of $G'_{ST(Nei)}$ that we found above.

In addition, the maximum value of θ and most other F_{ST} analogues can be calculated using a simple modification of the allelic data (Meirmans 2006). Maximization of the among-population diversity, while maintaining the within-population diversity, can be obtained by recoding the alleles in such a way that every population only contains alleles unique to that population. For example, when two populations both contain alleles A_1 and A_2 (possibly in different frequencies), the value of $F_{ST(max)}$ can be obtained by coding all alleles in the second population to A_3 and A_4 and then recalculating the statistic. Following the method of Hedrick, the standardized measure F'_{ST} can then be found by dividing the original F_{ST} value by the maximum value.

SNP data

The method of calculating the maximum possible value of G_{ST} that is used for calculating G'_{ST} makes implicit use of a dummy data set in which all populations only have uniquely different alleles. This dummy data set is a mathematical abstraction, comparable to Lewontin's (1964) assumption of a maximally linked data set that he used to standardize the measure of linkage disequilibrium, which generally does not pose any problems. However, it may lead to counterintuitive situations when used with markers that have only a few possible allelic states. For example, SNP markers have an absolute maximum of four allelic states corresponding to the four nucleotides (and generally have only two allelic states). Depending on the number of sampled populations, calculating G'_{ST} can have the implicit assumption of much more than four allelic states. However, for biallelic markers (like most SNPs), the classic F_{ST} is appropriate as it is and no standardization is necessary. The same is generally the case for markers with few allelic states.

Sequence data

Sequence data are of a different nature than allelic data as they contain information on the evolutionary relationships between haplotypes. G_{ST}, G'_{ST} and D do not take this information into account so their calculation is difficult for sequence data (however, an appropriate version of D is being developed: Chao *et al.* in preparation). However, standardization is actually not necessary for sequence data when the relationships between haplotypes are taken into account. This can, for example, be obtained by performing a standard AMOVA on the sequence data where a matrix of pairwise differences between haplotypes (Excoffier *et al.* 1992) is used to calculate the φ_{ST} statistic. Simulations have shown that unlike F_{ST}, φ_{ST} is independent of the mutation rate when calculated for sequence data (Kronholm *et al.* 2010).

Available software

As far as we know, there are currently four programs that can directly calculate estimates of the new differentiation statistics (see Table 1). The most complete is the program GENODIVE (Meirmans & Van Tienderen 2004), which can calculate per-locus and multilocus G'_{ST}, G''_{ST}, D, as well as φ'-statistics from a hierarchical AMOVA. The web-based program SMOGD (Crawford 2010) can estimate G'_{ST} and D per locus, plus a multilocus version of D by taking the harmonic mean across loci. The program THETAMAX (Kalinowski in preparation) can calculate the maximum value of θ, which it uses to obtain θ'. Lastly, there is the program SPADE (Chao & Shen 2009), which can calculate D on genetic data, despite being originally designed for ecological studies.

In addition to the above three programs, it is possible to calculate G'_{ST}, D and G''_{ST} indirectly using any program that can calculate the appropriate bias-corrected estimates of H_S and H_T [e.g. ARLEQUIN (Excoffier & Lischer 2010), FSTAT (Goudet 1995) or GENALEX (Peakall & Smouse 2005)], with the help of eqns 1, 2 and 4 above. Other F_{ST} analogues can be standardized using the data-recoding trick above, which can be used with any general population genetics program. A small utility called RECODEDATA (Meirmans 2006) is available to

Table 1 Software that can be used to calculate the new measures of genetic differentiation

Program	Statistics	Reference	Platform	Website
GenoDive	G'_{ST}, G''_{ST}, D, φ'_{ST}	Meirmans & Van Tienderen 2004	Mac OS X	http://www.patrickmeirmans.com
SMOGD	G'_{ST}, G''_{ST}, D	Crawford 2010	Website	http://people.bu.edungcrawfo/smogd
ThetaMax	θ'	Kalinowski, in preparation	Windows	http://www.montana.edu/kalinowski
SPADE	D	Chao & Shen 2009	Windows	http://chao.stat.nthu.edu.tw/softwareCE.html
RecodeData	G'_{ST}, G''_{ST}, φ'_{ST}, θ' (all indirectly)	Meirmans 2006	Mac OS X, Windows	http://www.patrickmeirmans.com

perform such recoding for data files in FSTAT format (Goudet 1995).

Properties of the new statistics

Variance and standard error

The variance in the estimates among loci can be much higher for G''_{ST} and D than for G_{ST}. This is especially the case when the heterozygosity is high. However, this higher variance is expected because of the much larger range of G''_{ST} and D for high H_S. The high among-locus variance, which translates into a higher standard error for estimates of G''_{ST} and D therefore generally does not constitute a large problem, especially when sufficient loci and/or populations are available. Nevertheless, it is advisable to perform an analysis of the standard error or confidence interval of estimates, e.g. using a bootstrapping (SPADE, SMOGD) or jackknifing (GENODIVE) approach.

The high standard error can give problems when multiple samples are unknowingly taken from a single population. To illustrate this, we used Easypop (Balloux 2001) to simulate a single large population (2000 individuals) and varied the mutation rate to obtain a series of 20 data sets with H_S values ranging from 0.05 to 0.99. From every data set, we took two samples of 20 individuals each and calculated $G'_{ST(Nei)}$, G''_{ST} and D, averaged over the 20 simulated loci. We repeated this 1000 times for every data set and analysed the range of observed values. The use of the unbiased H_S and H_T estimator correction (Nei & Chesser 1983; Nei 1987) can lead to negative estimates of the differentiation statistics. This is a desired effect as it keeps the average centred around zero, which is the expected value here because the samples are from the same population. Figure 4 shows that for all three statistics, the overall averages indeed always fall very close to zero, as they should. However, Fig. 4 also shows that for very high diversities, the observed range of values gets very large for G''_{ST} and D, and rather extreme negative and positive values can be observed. So for example, for $H_S = 0.99$, values of G''_{ST} and D as low as −0.25 or as high as 0.25 are no exception. Normally, negative values are interpreted as zero, and even when large, these should not present a problem in the interpretation. However, the large positive values can be more of a problem as such values are generally interpreted as representing moderately strong differentiation. This again shows that it is important to perform an analysis of the standard error of estimates of these differentiation statistics.

An interesting case occurs when every individual only has unique alleles. This can be the case when analysing sequence data as allelic data. Although rare, such cases have been observed in certain species. For example, in the planktonic chaetognath *Sagitta setosa*, Peijnenburg *et al.* (2004) found that all 85 individuals they sampled from 12 locations had unique sequences for the mitochondrial cytochrome oxidase II gene. In such cases, the value of H_S can equal one and as a result, both D and G''_{ST} are undefined owing to a division by zero. This makes sense, as in such a case, it is not possible to distinguish whether they represent two samples from a single very diverse population or from multiple very diverse populations.

Approach to equilibrium

Jost's (2008) D statistic has been criticized for taking much longer to reach equilibrium than G_{ST} (Ryman & Leimar 2009). As an example, Ryman & Leimar (2009) used a model where ten populations, which were initially in mutation–drift equilibrium, started diverging at $t = 0$ and then tracked the change in D and G_{ST} over time (based on recurrence equations from Nei 1975 and Li 1976; see also Ryman & Leimar 2008). Their results show that at the highest of the two mutation rates they used, both D and G_{ST} reached equilibrium relatively rapidly, though the equilibrium value of G_{ST} was much lower because of its dependence on the mutation rate. At a low mutation rate, G_{ST} again started to increase almost immediately, while D maintained a value close to zero for thousands of generations. We used their model to further assess the relationship between the mutation rate and the

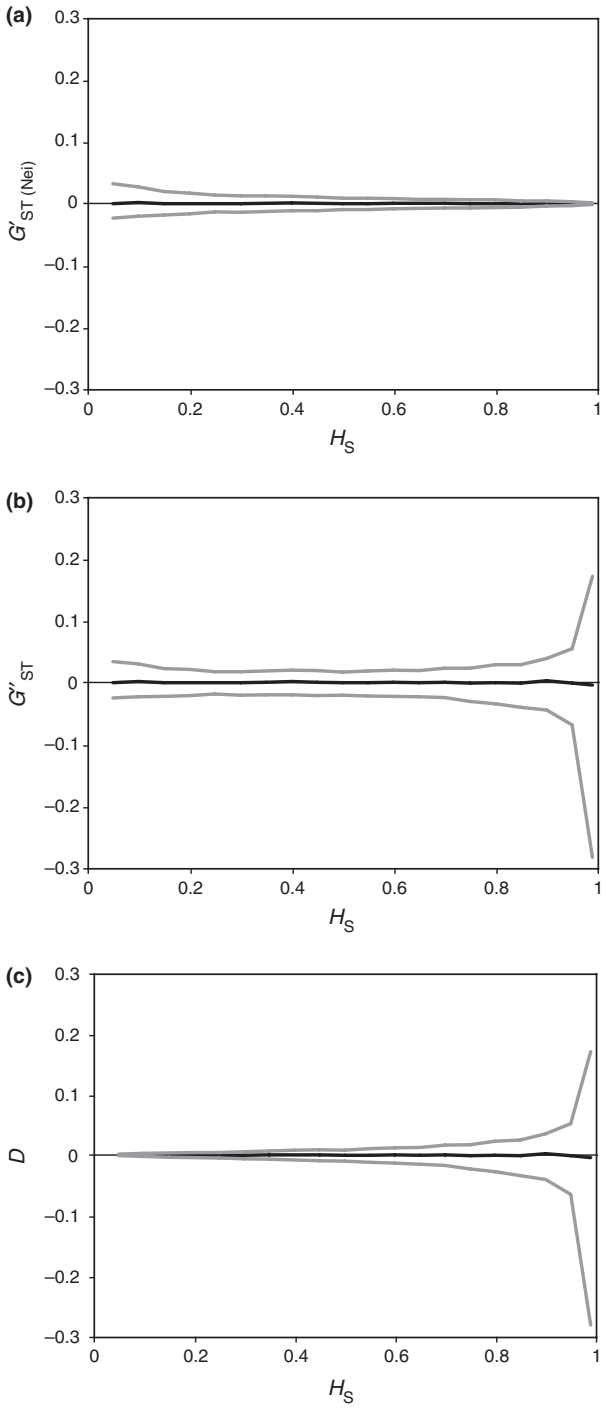

Fig. 4 The effect of taking two samples from a single population as a function of H_S for three different summary statistics. (a) $G'_{ST(Nei)}$, (b) G''_{ST}, (c) D. Two samples of 20 individuals were drawn from a simulated population of 2000 individuals. The black line shows the average value of the statistics over 1000 replications, with the grey lines indicating the upper and lower 2.5% percentiles.

time to equilibrium, not only for D and G_{ST} (for which we used the estimator $G'_{ST(Nei)}$) but also for G'_{ST} (using the estimator G''_{ST}). Figure 5 shows that for high mutation rates, the time taken to reach 95% of the equilibrium value is practically identical for the three statistics. However, when the mutation rate is low, the time to equilibrium is still the same for $G'_{ST(Nei)}$ and G''_{ST}, but much higher for D.

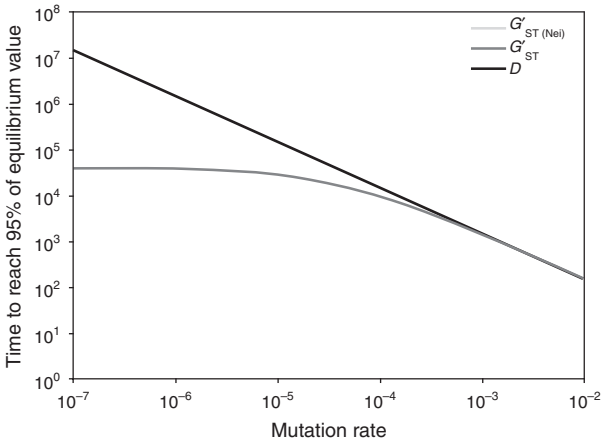

Fig. 5 The time required to reach 95% of the equilibrium value as a function of the mutation rate for three summary statistics. The lines of G''_{ST} and $G'_{ST\ (Nei)}$ are completely overlapping, so only the former is visible. Note the logarithmic scale on both axes.

Ryman & Leimar (2009) concluded that D is in fact more strongly affected by the mutation rate than G_{ST}, unless equilibrium conditions can be assumed. This makes D less suitable for practical applications such as the estimation of migration rates from the observed value of D. Jost (2009) responded to these criticisms by stating that this is in fact the expected behaviour of D. In the absence of migration, the value of D is only determined by the mutation rate. For low mutation rates, it simply takes a long time before enough mutations have accumulated to allow for any allelic differentiation among populations.

Migration

One reason why *F*-statistics became so widely used is that under the island model of population structure, there is a direct relationship between the migration rate and F_{ST}. This relationship is expressed in Sewall Wright's (1943) famous equation

$$F_{ST} \approx \frac{1}{1 + 4Nm} \quad (5)$$

Here, N is the size of the populations and m is the rate of migration among populations. However, this equation is actually a simplified version of another, less famous, equation that also includes the mutation rate u (Wright 1943, Cockerham & Weir 1993):

$$F_{ST} \approx \frac{1}{1 + 4Nu + 4Nm}$$

Equation 5 makes the assumption that the mutation rate is much lower than the migration rate. This assumption is unlikely to be met when hypervariable markers such as microsatellites are used.

Hedrick (2005) did not present a direct relationship between G'_{ST} and migration, though he did analyse under which conditions mutation influences estimates of the migration rate, using his equation 9b.

$$Nm = \frac{1 - F_{ST}[1 + H_S/(1 - H_S)]}{4F_{ST}} \quad (6)$$

Hedrick then used this equation to compare the estimates of the number of migrants obtained using Wright's simplified equation (eqn 4) to the actual number of migrants. He concluded that for a given F_{ST} value, the estimate of the number of migrants obtained using eqn 5 only gets reduced significantly when the heterozygosity is high.

Above, we saw that $F_{ST}/(1 - H_S)$ can be taken as a general definition of F'_{ST}, which is valid for both G_{ST} and θ. This means that F'_{ST} can be substituted for the term F_{ST} $(1 + H_S/(1 - H_S))$ in eqn 6, which simplifies the equation to

$$Nm = \frac{1 - F'_{ST}}{4F_{ST}}$$

So we see that when an equilibrium Island model can be assumed, an estimate of the number of migrants that is unaffected by H_S can be obtained through a combination of F_{ST} and F'_{ST}.

Jost (2008) did provide an extensive discussion of the relationship between D and the migration rate, showing that the equilibrium value of D is a complex combination of the mutation rate, migration rate, and the number of subpopulations (Jost's equation 17). This relationship simplifies drastically when it can be assumed that $m \ll 1$ and $uk \ll m$:

$$D \approx \frac{u(k-1)}{m}$$

Remarkably, under an Island model, the value of D is not influenced by the population size N. This result seems counterintuitive as it is generally thought that the effective population size determines the strength of genetic drift. In classical population genetics, the population size is therefore seen as one of the main drivers of population differentiation. This independence of the population size is also the reason why D takes such a long time to reach equilibrium: its value is only determined by population divergence because of mutations and (lack of) migration. Jost (2009) acknowledged that for these reasons, D is less suitable for estimating migration than G_{ST}, but also pointed out that this is not the purpose of the statistic.

The relationships between these summary statistics and the migration rate are only valid under an island model of population structure that is assumed to be under equilibrium. As discussed earlier, the long time required to reach its equilibrium value makes this assumption unrealistic for the D statistic. Besides equilibrium, the island model makes a large number of other assumptions, such as nonspatial migration, equal population sizes and no selection. Real populations are very likely to violate these assumptions, so any estimates of the migration rate or the number of migrants obtained from these summary statistics will be highly unreliable (Whitlock & McCauley 1999). However, this does not mean that F-statistics cannot be used as a rough indicator of the degree of population connectivity (Lowe & Allendorf 2010). For example, a strong relationship between the F_{ST} values and dispersal mode and other life history traits has repeatedly been found across many plant species (Hamrick & Godt 1996; Nybom 2004, Meirmans *et al.* in preparation).

Applications

The elusive ideal summary statistic

The most common use of F-statistics is to make inferences on demographic processes taking place within and among populations, such as migration, genetic drift, extinction and colonization. Therefore, the ideal summary statistic would provide information only on such demographic processes, and not about the purely genetic process of mutation (Ryman & Leimar 2008). The process of mutation is mostly a quality of the markers that are used, and in the great majority of cases, we are not interested in the markers per se, which are generally chosen to be selectively neutral. None of the discussed summary statistics are ideal in this respect, because they all are dependent on mutational processes in one way or another. The only statistics that are truly independent of mutational processes are those that make use of the 'memory' of the mutation process, such as φ_{ST} for sequence data or R_{ST} for microsatellites (Slatkin 1995). However, the latter statistic is only independent of the mutation rate when mutation takes place in a strictly stepwise fashion. Even a small percentage of random mutations can severely affect the value of R_{ST} estimates (Balloux *et al.* 2000).

Under an island model of population structure, the value of F_{ST} is determined by the migration rate, population size and mutation rate. The value of F'_{ST} is less dependent on the mutation rate and mostly determined by the migration rate and population size (but see Ryman

& Leimar 2008). D is determined by the migration rate, the number of populations, and the mutation rate, but not by the population size. Although its range is not restricted by the value of H_S as is the case for F_{ST}, the long time required for D to reach its equilibrium value means that during this long transition period, it is actually more affected by H_S than F_{ST} (Ryman & Leimar 2009). Jost's (2009) response that this correctly reflects the allelic differentiation at these loci is beside the point: in the great majority of cases, we are not interested in the mutational differentiation of the markers, but we are interested in describing the populations' demography.

Using a divergence model, Ryman & Leimar (2008) noted that the effect of mutation on G_{ST} is relatively limited and may take time to develop. This is because under such a scenario, the effect of genetic drift works much faster than mutation, which is the same reason why G_{ST} reaches equilibrium more quickly than D. Ryman & Leimar (2008) also noted that during the early phase of population divergence, G_{ST} is a better measure of divergence than G'_{ST}, because the standardized measure gives much larger differences in value for loci with different mutation rates, suggesting that different loci had experienced different amounts of drift.

One scenario where the difference between D and F_{ST}-like statistics is very clear is when different alleles have gone to fixation in different populations. Being fixation indexes, both F_{ST} and F'_{ST} have a value of one in such a case. However, this value reflects the fixation and not the differentiation in allele frequencies among the populations. Imagine a metapopulation where 99 populations are fixed for the same allele and only one population for another allele. In this case, the values of F_{ST} and F'_{ST} equal one, even though most populations are exactly the same. D, on the other hand, will have a value close to zero, which better reflects the similarity among the populations. However, this seemingly incorrect behaviour of F_{ST} and F'_{ST} makes sense when viewed in a different light. Imagine a population with three equally frequent alleles. When this population splits into three populations and there is no migration or mutation, these alleles will go to fixation in the three subpopulations. There are then three different possibilities: the same allele goes to fixation in all three subpopulations (11% chance), in every subpopulation a different allele goes to fixation (22% chance), or one allele goes to fixation in one of the subpopulations and another allele in the other two (66%). These are three outcomes of the same demographic process that can easily occur simultaneously at multiple loci within a single species. Nevertheless, D gives three widely different values: 0, 0.67 and 1 for the three possible outcomes, respectively. In contrast, the values of F_{ST} and F'_{ST} are one (though they are undefined when there is no diversity at all). So from these examples, we see that D performs better at measuring the actual differentiation in allele frequencies among populations, while F_{ST} and F'_{ST} are better at describing the influence of demographic events on the distribution of genetic variation.

When to use which statistic?

As there are now three classes of differentiation statistics (F_{ST}, F'_{ST} and D), it is useful to provide guidelines on the use of these statistics. However, all three have their advantages and drawbacks, so it is difficult to point out a single best all-purpose summary statistic. For one thing, we think it is important that the original F_{ST} is always presented. This statistic has been used for several decades and, despite its shortcomings, continued use will allow a better comparison with those past studies, especially because reanalysis of the old data is mostly not possible (Neigel 2002). If it is suspected that the value of F_{ST} has been influenced by the heterozygosity, for instance when highly variable markers are used, one of the alternative statistics (F'_{ST} and in some cases D) should be used in addition to F_{ST}.

Above, we saw that D is best suited for describing the allelic differentiation among populations, while F_{ST} and F'_{ST} are better suited for demographic inferences. Jost (2009) made a similar distinction between using D for measuring differentiation and F_{ST} for estimating migration. Although useful, this distinction is not as clear-cut as it seems. For example, there are many applications where it is not clear whether the analysis involves differentiation or migration. Is isolation by distance a case of differentiation or of limited migration between remote demes? Such a question is difficult to answer. Nevertheless, when the objective of a study is clear, then so is the choice of summary statistic to use for the task.

F'_{ST}, estimated using G''_{ST}, θ' or φ'_{ST}, is most suited for inferences on demographic history and migration. Above, we showed how F'_{ST} can be combined with F_{ST} to yield an estimate of the number of migrants that is not affected by the mutation rate. However, the theoretical links between F'_{ST} and various population processes need to be explored more fully. In the last decade, the indirect estimation of migration rates using F-statistics has fallen into disuse once biologists started to realize that this involves making a large number of assumptions that are unlikely to be met in practice (Whitlock & McCauley 1999). Nevertheless, F_{ST} is still regarded as a useful tool for comparative analyses of gene flow (Neigel 2002). In addition, F_{ST} is still useful as a fixation index, measuring the level of inbreeding at different hierarchical levels. One advantage of Hedrick's (2005) standardization approach is that it is very flexible and can be applied

to every F_{ST} analogue for which the maximum value can be obtained. This makes it possible to take into account the evolutionary distances and a hierarchical population structure (e.g. using Meirmans' (2006) AMOVA method). However, one should be cautious with applying the standardization to cases where the calculation of the maximum value is ambiguous.

When one is interested in allelic differentiation, D can be used. D may e.g. be useful in conservation genetics of rare species, when decisions have to be made on the selection of populations based on their contribution to the overall genetic differentiation (see Caballero *et al.* 2010 for a comparison of using heterozygosity-based and allelic diversity based F-statistics for this purpose). We feel that D is less suited for inferences of populations' demography as it is insensitive to the population size and can take a very long time to reach equilibrium. D is currently only formally described as a single-locus statistic for use with allelic data (Jost 2008). It can therefore not yet take evolutionary distances into account, nor can it be calculated for a hierarchical population structure. However, the statistical framework around D is still under development and such extensions can be expected in the near future (Lou Jost personal communication).

Other uses of F_{ST}

F_{ST} is used for many different purposes (Holsinger & Weir 2009) and the question arises whether the newly developed alternatives suit all these purposes. One increasingly popular use of F-statistics is for the detection of loci under selection (Beaumont & Balding 2004; Excoffier *et al.* 2009). Loci under divergent selection are expected to have higher F_{ST} values than neutral loci, while loci under balancing selection are expected to have lower F_{ST} values (Lewontin & Krakauer 1973). However, when loci differ in mutation rates, the range of F_{ST} values becomes influenced by the heterozygosity, which will lower the F_{ST} values for the more variable loci. Therefore, one of the most frequently used methods for detecting selection does not only compare F_{ST} values to detect outliers but performs a joint analysis of F_{ST} and H_S (Beaumont & Balding 2004). Possibly, the comparison with H_S may not be necessary when F'_{ST} (possibly estimated using G''_{ST}) or D is used. Recently, Neff & Fraser (2010) developed a computer program that uses resampling approaches to compare values of either θ or G'_{ST} among loci to detect loci under selection.

Another use of F_{ST} is to compare its value calculated from neutral marker loci with that of an analogous statistic, Q_{ST}, calculated from quantitative traits (Spitze 1993). This allows one to detect traits under selection by looking whether Q_{ST} is significantly lower or higher than F_{ST}. However, for highly variable markers, the interpretation can be difficult when the F_{ST} estimates are lowered by the influence of H_S, possibly leading to an excess of false positives where $Q_{ST} > F_{ST}$. However, in a review of 27 studies that used this approach, the difference between Q_{ST} and F_{ST} was found to be more pronounced for allozymes than for microsatellites (Merilä & Crnokrak 2001), suggesting that correcting for the within-population diversity may not be necessary. In fact, in a study on pied flycatchers, Lehtonen *et al.* (2009) explicitly stated that they did not standardize their F_{ST} estimates as they considered this to be not applicable to $Q_{ST} - F_{ST}$ comparisons. On the other hand, Q_{ST} itself is based on the classical definition of F_{ST} and its value may therefore also be dependent on the amount of variation within populations. In that case, highly variable characters are expected to show smaller values of Q_{ST} than less variable characters, and some standardization of Q_{ST} may also be necessary. Indeed, simulations have shown that the value of Q_{ST} strongly depends on the mutation rate and mutation model of the underlying quantitative trait loci (Kronholm *et al.* 2010).

Conclusions

Where for decades only F_{ST} has been used to assess population structure, we now have three main classes of summary statistics: the classical F_{ST} (Wright 1943), the standardized F'_{ST} (Hedrick 2005) and Jost's (2008) D. Although these statistics have different theoretical backgrounds, when the number of samples is large, they are connected to each other through relatively simple relationships, because all three can be expressed as ratios of H_S and H_T. For F_{ST}, it has long been common practice to use nearly unbiased estimators for calculating its value, and we advise that the same is done for the two new statistics. F'_{ST} can be estimated using the G''_{ST} statistic introduced here, or alternatively using θ' (Kalinowski in preparation) or φ'_{ST} (Meirmans 2006). For heterozygosity-based estimators such as D and G''_{ST}, it is important to also use unbiased estimators for H_S and H_T (Nei & Chesser 1983; Nei 1987). Several software packages are available for these purposes (Table 1).

Even though the two new statistics F'_{ST} and D indeed correct the dependency of F_{ST} on the amount of within-population variation (Heller & Siegismund 2009), they are not universally applicable. F'_{ST} has the implicit assumption of a 'dummy' data set where every population only has unique alleles. As a result, some may find it problematic to apply F'_{ST} to markers where there are only a few allelic states possible, such as SNPs or sequence data. D is independent of the population size and therefore not well suited for inferences of the effect of

demographic factors on the population structure. In addition, D can take a very long time to reach equilibrium (Ryman & Leimar 2009). The ideal summary statistic only provides information on demographic processes and not on genetic processes such as mutation that are specific to the used marker system (Ryman & Leimar 2009). Unfortunately, none of the three classes of statistics has these qualities, but their combined use will enable more robust analyses of population structure than what is possible with only F_{ST}.

Acknowledgements

We thank Noah Rosenberg, Lukas Keller, Pirmin Nietlisbach, Lou Jost, Alex Buerkle and an anonymous reviewer for their helpful comments on an earlier version of the manuscript. Furthermore, we thank Steven Kalinowski for sharing his results on the standardization of θ, Nick Crawford for information on the workings of his program SPADE and Rasmus Heller for sharing the data from his meta-analysis of genetic summary statistics.

References

Balloux F (2001) EASYPOP (Version 1.7): a Computer Program for Population Genetics Simulations. *Journal of Heredity*, **92**, 301–302.

Balloux F, Goudet J (2002) Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Molecular Ecology*, **11**, 771–783.

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.

Beaumont MA, Balding D (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Caballero A, Rodríguez-Ramilo S, Ávila V, Fernández J (2010). Management of genetic diversity of subdivided populations in conservation programmes. *Conservation Genetics*, DOI 10.1007/s10592-009-0020-0.

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 S, Jiang Y, Chazdon R (2008) A two-stage probabilistic approach to multiple-community similarity indices. *Biometrics*, **64**, 1178–1186.

Chao A, Shen T-J (2009) *Program SPADE (Species Prediction and Diversity Estimation)*. Program and user's guide at http://chao.stat.nthu.edu.tw.

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

Cockerham CC (1973) Analysis of gene frequencies. *Genetics*, **74**, 679–700.

Cockerham CC, Weir BS (1993) Estimation of gene flow from *F*-statistics. *Evolution*, **47**, 855–863.

Crawford N (2010) SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources*, **10**, 556–557.

Culley TM, Wallace LE, Gengler-Nowak KM, Crawford DJ (2002) A comparison of two methods of calculating G_{ST}, a genetic measure of population differentiation. *American Journal of Botany*, **89**, 460–465.

Excoffier L, Hamilton G (2003) Comment on ''Genetic structure of human populations.''. *Science*, **300**, 1877.

Excoffier L, Lischer H (2010) ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial-DNA restriction data. *Genetics*, **131**, 479–491.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate *F*-statistics. *Journal of Heredity*, **86**, 485–486.

Gregorius HR (2010) Linking diversity and differentiation. *Diversity*, **2**, 370–394.

Gregorius HR, Degen B, Koenig A (2007) Problems in the analysis of genetic differentiation among populations – a case study in *Quercus robur*. *Silvae Genetica*, **56**, 190–199.

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **351**, 1291–1298.

Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313–318.

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

Heller R, Siegismund H (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.

Holsinger KE, Weir BS (2009) Fundamental concepts in genetics: genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nature Reviews Genetics*, **10**, 639–650.

Hubalek Z (2000) Measures of species diversity in ecology: An evaluation. *Folia Zoologica*, **49**, 241–260.

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). D vs. G_{ST}: response to Heller and Siegismund (2009) and Ryman and Leimar (2009). *Molecular Ecology*, **18**: 2088–2091.

Kronholm I, Loudet O, De Mieux J (2010) Influence of mutation rate on estimators of genetic differentiation-lessons from *Arabidopsis thaliana*. *BMC Genetics*, **11**, 33.

Lehtonen P, Laaksonen T, Artemyev A *et al*. (2009) Geographic patterns of genetic differentiation and plumage colour variation are different in the pied flycatcher (*Ficedula hypoleuca*). *Molecular Ecology*, **18**, 4463–4476.

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, **49**, 49–67.

Lewontin R, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175.

Li W (1976) Effect of migration on genetic distance. *American Naturalist*, **110**, 841–847.

Long JC, Kittles RA (2003) Human genetic diversity and the nonexistence of biological races. *Human Biology*, **75**, 449–471.

Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? *Molecular Ecology*, **19**, 3038–3051.

Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution*, **60**, 2399–2402.

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Merilä J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, **14**, 892–903.

Neff B, Fraser B (2010) A program to compare genetic differentiation statistics across loci using resampling of individuals and loci. *Molecular Ecology Resources*, **10**, 546–550.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321–3323.
Nei M (1975). *Molecular Population Genetics and Evolution*. North-Holland Publishing Co., Amsterdam, The Netherlands.
Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
Nei M, Chesser R (1983) Estimation of fixation indexes and gene diversities. *Annals of Human Genetics*, **47**, 253–259.
Neigel JE (2002) Is F_{ST} obsolete? *Conservation Genetics*, **3**, 167–173.
Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology*, **13**, 1143–1155.
O'Reilly PT, Canino MF, Bailey KM, Bentzen P (2004) Inverse relationship between FST and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Molecular Ecology*, **13**, 1799–1814.
Peakall R, Smouse PE (2005) GenAlEx 6: Genetic Analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.
Peijnenburg KTCA, Breeuwer JAJ, Pierrot-Bults AC, Menken SBJ (2004) Phylogeography of the planktonic chaetognath *Sagitta setosa* reveals isolation in European seas. *Evolution*, **58**, 1472–1487.
Peijnenburg KTCA, Fauvelot C, Breeuwer AJ, Menken SBJ (2006) Spatial and temporal genetic structure of the planktonic *Sagitta setosa* (*Chaetognatha*) in European seas as revealed by mitochondrial and nuclear DNA markers. *Molecular Ecology*, **15**, 3319–3338.
Ryman N, Leimar O (2008) Effect of mutation on genetic differentiation among nonequilibrium populations. *Evolution*, **62**, 2250–2259.
Ryman N, Leimar O (2009) G_{ST} is still a useful measure of genetic differentiation – a comment on Jost's D. *Molecular Ecology*, **18**, 2084–2087.
Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.
Spitze K (1993) Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics*, **135**, 367–374.
Weir BS, Cockerham C (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: F_{ST}!= 1/(4*Nm*+1). *Heredity*, **82**, 117–125.
Wright S (1943a) Isolation by distance. *Genetics*, **28**, 114.
Wright S (1943b) An analysis of local variability of flower color in *Linanthus parryae*. *Genetics*, **28**, 139.
Wright S (1965) The interpretation of population structure by *F*-statistics with special regard to systems of mating. *Evolution*, **19**, 395–420.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 Studies included in the meta-analysis, expanded from Heller & Siegismund (2009).

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

Appendix

Nei (1987, p. 164–165) defined:

$$D'_{ST} = (H_T - H_S) \cdot k/(k-1)$$

and

$$H'_T = H_S + D'_{ST} = H_S + (H_T - H_S) \cdot k/(k-1)$$

that he used to calculate

$$G'_{ST} = \frac{H'_T - H_S}{H'_T} = \frac{H_S + (H_T - H_S) \cdot k/(k-1) - H_S}{H_S + (H_T - H_S) \cdot k/(k-1)}$$

this can be rewritten as:

$$G'_{ST} = \frac{k \cdot (H_T - H_S)}{k \cdot H_T - H_S}$$

thus,

$$G'_{ST(max)} = \frac{k \cdot (H_{T(max)} - H_S)}{k \cdot H_{T(max)} - H_S}$$

Using Hedrick's (2005)

$$H_{T(max)} = (k - 1 + H_S)/k$$

this becomes:

$$G'_{ST(max)} = \frac{k \cdot ((k-1+H_S)/k - H_S)}{k \cdot (k-1+H_S)/k - H_S}$$

which can be rewritten to be much simpler:

$$G'_{ST(max)} = 1 - H_S$$

This makes sense as Hedrick found that $G_{ST(max)}$ approaches $1 - H_S$ when the number of populations goes to infinity, so this value is reasonable for a statistic that is supposed to be independent of the number of populations sampled.

Hedrick defined another G'_{ST}

$$G'_{ST} = \frac{G_{ST}}{G_{ST(max)}}$$

combining this with $G'_{ST(Nei)}$ gives:

$$G''_{ST} = \frac{G'_{ST}}{G'_{ST(max)}} = \frac{k \cdot (H_T - H_S)}{(k \cdot H_T - H_S) \cdot (1 - H_S)}$$