Current Zoology 61 (5): 886–897, 2015

Statistical measures of genetic differentiation of populations: Rationales, history and current states

Liang MA[1,2,3], Ya-Jie JI[1], De-Xing ZHANG[1,3,4*]

[1] Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

[2] Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing 100190, China

[3] Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

[4] National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, Beijing 100190, China

Abstract Population differentiation is a fundamental process of evolution, and many evolutionary studies, such as population genetics, phylogeography and conservation biology, all require the inference of population differentiation. Recently, there has been a lot of debate over the validity of F_{ST} (and its analogue G_{ST}) as a measure for population genetic differentiation, notably since the proposal of the new index D in 2008. Although several papers reviewed or explored specific features of these statistical measures, a succinct account of this bewildering issue with an overall update appears to be desirable. This is the purpose of the present review. The available statistics generally fall into two categories, represented by F_{ST} and D, respectively. None of them is perfect in measuring population genetic differentiation. Nevertheless, they each have advantages and are valuable for current research. In practice, both indices should be calculated and a comparison of them can generate useful insights into the evolutionary processes that influence population differentiation. F_{ST} (G_{ST}) has some unique irreplaceable characteristics assuring its standing as the default measure for the foreseeable near future. Also, it will continue to serve as the standard for any alternative measures to contrast with. Instead of being anxious about making choice between these indices, one should pay due attention to the equilibrium status and the level of diversity (especially H_S) of the populations, since they largely sway the power of a given statistic to address a specific question. We provide a multi-faceted comparative summary of the various statistics, which can serve as a basic reference for readers to guide their applications [*Current Zoology* 61 (5): 886–897, 2015].

Keywords Population structure and subdivision, Coefficient of inbreeding, Fixation index, Gene identity, Gene flow, Non-equilibrium conditions, Heterozygosity, F_{ST} and selection

1 Introduction

Population differentiation (subdivision) is a fundamental process of evolution. It is recognized as "fundamental" because it is a process which every species unavoidably undergoes during evolution, and it may lead to speciation or extinction under certain conditions. Actually, if one adopts Wright's (1932) perspective on evolution, the problem of speciation can be reduced to the problem of how a single population splits into two populations on different "adaptive peaks" (Barton and Charlesworth, 1984). Consequently, many studies in population genetics, phylogeography and conservation biology benefit from the inference of population differentiation. Therefore, determining and measuring population differentiation is of central importance. Essentially, genetic differentiation of populations is the result of uneven (nonrandom) spatial distribution of genetic variation in a species (Hartl and Clark, 1997), reflecting a departure from panmixia (Box 1). Accordingly, Sewall Wright, one of the trio key figures known as the founders of theoretical population genetics, developed the so-called *F*-statistics (also known as fixation indices) in the 1950s to measure population differentiation (Wright, 1951; and see below for a historical account). These statistics, in particular the index F_{ST}, have since been widely employed in a great diversity of research fields.

Recently, there has been some heated debate in molecular ecology over the validity of F_{ST} (and its analogue G_{ST}) as a measure for population genetic differentiation, notably since the proposal of the new index D in 2008 (e.g. Jost, 2008, 2009; Ryman and Leimar, 2008, 2009; Whitlock, 2011; Leng and Zhang, 2011). This has stimulated field-wide careful rethinking on current perceptions and methods for measuring population differentiation. As a consequence, several papers have been published, either reviewing the general relationships between the popular statistics such as D, F_{ST} (G_{ST}) and

Received June 30, 2015; accepted July 20, 2015.

∗ Corresponding author. E-mail: dxzhang@ioz.ac.cn, dxzhang@big.ac.cn

Box 1 Glossary

Equilibrium: A state of population in which the actions of various evolutionary forces such as mutation, migration, genetic drift and natrual selection, are in balance such that the gene (allele) frequencies remain unchanged in the population. In practice, it is difficult to know whether a population is already under equilibrium. Given the significant impact of Pleistocene glaciations on the distribution of plants and animals, and considering the time scale required for populations to reach equilibrium (often in the order of the reciprocal of the mutation rate; Takahata & Nei, 1984), it is likely that many species are far from reaching equilibrium at many genetic loci (i.e. they are still under non-equilibrium).

Haplotype, homozygote, heterozygote: For a diploid individual, there exists a pair of homologous alleles at any genomic locus, with the two sequence copies being either identical (in this case, the individual is homozygous or is a homozygote) or nonidentical (in this case, the individual is heterozygous or is a heterozygote). The molecular description of the particular DNA segments at a locus in an individual is called its 'genotype', with each distinct copy being termed a 'haplotype'. From a genomic viewpoint, the term haplotype refers to a distinct set of nucleotide sites linked on the same chromosome and inherited together in meiosis; from a population genetic viewpoint, it refers collectively to a set of identical alleles in populations (Huang et al., 2008).

Infinite allele model (IAM): A mutation model proposed by Kimura and Crow in 1964. It assumes that every mutation occurred in the population will produce an unique state of allele. It implies that mutation is independent and there exist an infinite number of nucleotide sites to mutate, that is known as **infinite site mutation model (ISM)**. IAM is mathematically very convinent and has been popularly employed for describing mutations in nucleotide sequences in molecular evolution and population genetics. A special case of IAM is known as the ***K*-alleles model (KAM)** under which there exist a finite number of allelic states (e.g. $K = 999$), and mutation will generate a new allele of any possible allelic state at random.

Island model: A model for decribing population structure proposed by Sewall Wright in which the total population is divided into subgroups, each breeding at random within itself, except for a certain proportion of migrants drawn at random from the whole (Wright, 1943). Often, it is further assumed that each subpopulation (subgroup) exchanges genes at the same rate with every other subpopulation.

Panmixia, population structure, population differentiation, population subdivision: Panmixia refers to a condition in which the population is a single entity with complete random mating. Such a condition can be broken by various factors, for example, selection makes certain individuals to have mating advantages over the others, or a reduction in gene flow allows subpopulations in two geographical localities to be influenced independently by genetic drift. This will ultimately result in differentiation of allele frequencies between these subpopulations, leading to population structure - a process known as population differentiation or population subdivision.

Private alleles: Alleles found in only one local population (or deme).

Stepwise mutation model (SMM): A mutation model proposed by Ohta and Kimura in 1973 which assumes that the allelic states can be expressed by integers and that, mutation will change the allelic state by moving either one (or finite number) step in the positive direction or in the negative direction in the allele space. SMM is commonly used in modeling mutaions of microsatellite loci. The simplest SMM is known as the **single-step mutation model (SSM)** under which the new allele size is either increased or decreased by one step.

their standardized forms (Meirmans and Hedrick, 2011), or inspecting the behaviours of these statistics on short timescales (Raeymaekers et al., 2012), or examining specifically non-equilibrium conditions (e.g. Leng and Zhang, 2013; Box 1), or discussing applications of these statistics using a certain type of population genetic data (e.g. microsatellite; Putman and Carbone, 2014). Nevertheless, a succinct account of this rather puzzling issue with an overall update appears to be lacking. This is the purpose of the present review. We will focus on some important issues which have not been covered or not extensively discussed, while minimizing the topics already discussed in the aforementioned publications. We do not deal with the methods for estimating the various statistical measures. Mathematic formulae are also kept to a minimum except where they are associated with the logic flow of writing. Readers who are familiar with the various F_{ST}-like and alternative statistics can skip the third and fourth sections.

2 The Rationales

The key issue in determining and measuring popula-

tion differentiation is how to quantify the nonrandom distribution of genetic variation in populations. Considering a metapopulation (the total population), it consists of several local populations (the subpopulations). A basic rationale is to compute the difference in the genetic composition of the subpopulations from the expectation of random mating in the total population. There are various ways to realize this, depending on one's viewpoints; here are a few examples: (1) to compute the variance of allele frequencies between subpopulations (e.g. Wright, 1943); (2) to calculate differences in the (effective) number of alleles between subpopulations (e.g. Jost, 2008); (3) to estimate the difference in genetic distance between genes (or haplotypes, Box 1) in subpopulations (e.g. the analysis of molecular variance, AMOVA, approach; Excoffier et al., 1992); (4) to determine the difference in coalescence times of alleles within and between subpopulations (e.g. Slatkin, 1991); (5) to count the difference in allele size between subpopulations in some particular situations [e.g. Slatkin, 1995; note that this is for loci following stepwise mutation model (SMM, see Box 1) such as microsatellites, and hence in a sense is analogous to the previous methods dealing with coalescence times and genetic distance]; (6) to compare the discrepancy in heterozygosity (or homozygosity) between subpopulations and the total population (e.g. Nei, 1973); (7) to assess the deviation in inbreeding between subpopulations and the total population (e.g. Wright, 1921,1922); (8) to determine the co-ancestry coefficient (correlation or relatedness) for alleles within a subpopulation relative to the total population (Weir and Cockerham, 1984); and (9) to quantify the difference in private alleles (Box 1) between subpopulations (e.g. Slatkin, 1985; note that private alleles can be the results of drift, mutation, or other forces; a subset of the private alleles may have reached fixation, thus the proportion of differentially fixed alleles can also be a kind of measure).

Following these different approaches, different statistics have been or can be developed. Obviously, these routes are not independent and nor exclusive to each other. For example, although by definition some of the statistics cited above are not based on heterozygosity, they can be expressed, after some mathematical transformation, as a function of heterozygosities [e.g. as noticed by Meirmans and Hedrick (2011) for Jost's *D*].

However, in practice, all these approaches should be considered from both statistical and genomic perspectives, since what we obtained empirically are estimates of the statistics from population and genomic data. There exist at least four levels of stochastic variation that complicate all analyses: first, the effect of sample sizes; second, the effect of the (number of) subpopulations sampled; third, the effect of the number of loci used to monitor genetic variation in populations; and fourth, the genomic location of genetic loci. The first two can be referred to as population level sampling variance, and the last two genomic level sampling variance. Although well-defined statistical approaches exist to minimize bias for estimating parameters from data (Lehmann and Casella, 1998), e.g. maximum likelihood and Bayesian methods, thus dealing with the population level sampling variance, the genomic level sampling variance cannot be coped with just statistically. For example, great variation between individual loci (that is the genomic level sampling variance) is not unexpected, even if these loci are under identical evolutionary forces. This means that measuring population genetic differentiation in any applications must be based on a large number of unlinked loci besides reasonable population level sampling. In addition, different types of loci (*e.g.* microsatellite DNA *versus* SNPs) usually yield different estimates, a reflection of different modes of molecular and genomic evolution of these loci.

3 The Origin of F_{ST} as a Standard Measure of Population Subdivision

Technically, genetic differentiation can be formulated as differences of allele frequencies between subpopulations (Wright, 1943), reflecting a significant departure from random mating in the total population. Such departure can be assessed in terms of changes in heterozygosity and homozygosity of populations ("the relative amount of heterozygosis" or homozygosis in Wright's words; see below), for example, the relative reduction in heterozygosity due to non-random mating - this is what the standard measure of differentiation, Wright's F_{ST}, was founded on (Wright, 1951, 1965).

Although F_{ST} has been widely-employed as the *classic* and *standard* measure of population genetic differentiation, its original form was developed by Wright (1921,1922) as coefficient of inbreeding for depicting localized departure of allele frequency because of nonrandom mating from expectation under panmixia (assuming there is no change in allele frequency in the population - note that this is a rather strong assumption that is in turn based on the presumption that population size is large and mutation and selection are unimportant.

In one sense, it is here the deficiencies of F_{ST} and G_{ST} stem from). As population subdivision also leads to non-random mating in the total population, the coefficient of inbreeding F was later adopted for measuring the departure from random mating in structured population (Wright, 1951). But, in this situation, there exist several hierarchic levels to define the analysis unit, more than one inbreeding coefficients are thus needed. Hence, Wright referred to the set of inbreeding coefficients as F-statistics, and also designated them as the fixation indices (Wright, 1951). Among them, nevertheless, non-random mating as indicated by F in the broadest context, namely F_{ST}, has two different sources of origin: inbreeding within each subpopulation and population structure among subpopulations. From this perspective, F_{ST} clearly carries information about population subdivision.

For a population with two levels of structures, F_{IT} is the inbreeding coefficient of individual relative to gametes of the total population, F_{IS} is the inbreeding coefficient of individual relative to gametes of the subpopulation (average over all subpopulations), and F_{ST} which represents the correlation between gametes randomly drawn from within subpopulations relative to gametes of the total population is

$$F_{ST} = (F_{IT} - F_{IS})/(1 - F_{IS}). \quad (1)$$

It can also be written as,

$$F_{ST} = \frac{\sigma_p^2}{\bar{p}(1-\bar{p})} \quad (2)$$

σ_p^2 is the variance of the frequency of allele A of subpopulations and $\bar{p}$ is the average frequency of allele A in the total population for biallelic systems (Wright 1965). Wright thus summarized that there exist several interpretations of the F-statistics: "as correlations, as functions of the relative amount of heterozygosis, ... in some cases, as probabilities of identity by origin", and as "the ratio of the actual variance of gene frequencies of subdivisions to its limiting value, irrespective of their own structures" (Wright 1965). He also emphasized: "F_{ST} in the broad sense can always be obtained, at least empirically, for the variance of distribution of gene frequencies even in cases involving selection", from the formula (2) above. "The results, of course, apply only to the particular loci in question…".

A detail worthy of pointing out that will promote our understanding of Wright's F-statistics is that Wright's usage of the word "fixation" in that classic context is quite different from the current usage in population genetics (where fixation refers to the frequency of an allele reaching 100% in the population). Fixation, in the F-statistics context, refers to the fixation of a single allele at a locus to form a homozygote, i.e. homozygosis (by definition, the formation of a zygote by the union of two gametes that have one pairs of identical alleles for diploid). Thus, for a population, "the percentage of homozygosis measures the degree of fixation of heredity…"(*p*.129, Wright, 1921). The inbreeding coefficient "gives the departure from the amount of homozygosis under random mating toward complete homozygosis" (*p*.334, Wright, 1922). Initially, it was defined to assess the degree of inbreeding in a given population. Thus, it represents "a scale which runs from" "zero under random mating" to "1 under complete fixation" "while the percentage of homozygosis is running from 50 per cent. to 100 per cent.", "and F as the weighted average in the intermediate population" (Wright, 1922; 1951). On that account, for F = 0.59, it means that 59% of the heterozygotes expected under Hardy-Weinberg assumptions was replaced by homozygotes in the population (or the population contains 59% fewer heterozygotes than expected under Hardy-Weinberg assumptions) due to non-random mating. Therefore, the population is in a much greater degree towards complete fixation (F = 1) compared to random mating (F = 0). Hence, the logic of Wright to name his F-statistics generally as "fixation index" (Wright, 1951) is that they reflect, compared to the situation of random mating, the degree of approach toward the status of complete fixation in a population where every individual is homozygote at the locus of interest (i.e. complete homozygosis in Wright's words). As such, even if a population hosts two different alleles at a locus, if there exists no heterozygote, the population has reached complete fixation (i.e. under complete homozygosis), and thus F = 1.

4 A Brief Introduction of the F_{ST}-like Statistics

A number of related indices of genetic differentiation have been subsequently derived in link with the natures of the diagnostic genetic markers, such as G_{ST} (Nei 1973),Φ_{ST} (Excoffier et al., 1992), Q_{ST} (Prout and Barker, 1993; Spitze, 1993), R_{ST} (Slatkin, 1995). These are referred to as F_{ST}-like statistics for convenience.

G_{ST}. In practice, the most widely applied statistic for measuring population genetic differentiation is Nei's

G_{ST} (1973), an extension of F_{ST} for loci with multiple states of alleles. It analyzes allele frequency variation among subpopulations in terms of heterozygosity or gene diversity as defined by Nei (1973). Given a diploid population with K subpopulations and I allelic states at a locus. Denote the frequency of the i_{th} allele in the population as p_i, and the corresponding frequency in subpopulation k as p_{ki}. Let $H_T = 1–J_T$ be the total heterozygosity, i.e. the probability of genotypes with the union of two different states of alleles, of the total population, where $J_T = \sum_i p_i^2$ is the homozygosity (i.e. the probability of genotypes with the union of two identical states of alleles). Nei (1973) referred to H_T and J_T as gene diversity and gene identity of the total population, respectively. Extended from the definition of pairwise diversity of two populations, he defined D_{ST} as the average gene diversity between subpopulations. The total gene diversity is then linearly decomposed as $H_T = H_S + D_{ST}$, where H_S is defined as the (average) gene diversity within subpopulations, which can also be written in form of average gene identity within subpopulation as $H_S = 1 – J_S$. Nei regarded D_{ST} as a measure of absolute magnitude of gene differentiation. The differentiation relative to the total population, named by Nei the coefficient of gene differentiation is given by

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

For a neutral locus with only two types of allele, it can be shown that G_{ST} is identical to Wright's F_{ST}. For multiple allelic situations, G_{ST} is equal to the median of F_{ST} for all alleles, especially by definition expressed in equation (2). Note that for definition of F_{ST} in equation (1), F_{IS} and F_{IT} can be negative as they are similar to correlation coefficient; however, quantities used for defining G_{ST} are all nonnegative.

θ. By analogue of Wright's *F*-statistics, Weir and Cockerham (1984) derived a set of parameters *f, θ* and *F* to describe correlations of gene frequencies, by the variance of the allele frequencies between populations σ_w, the variance of the allele frequencies between individuals within populations σ_b, and the variance of the allele frequencies between gametes within individuals σ_a.

They defined

$$1 - \hat{F} = \frac{\sigma_a}{\sigma_a + \sigma_b + \sigma_w},$$

$$1 - \hat{f} = \frac{\sigma_a}{\sigma_a + \sigma_b},$$

$$\hat{\theta} = \frac{\sigma_w}{\sigma_a + \sigma_b + \sigma_w}.$$

and used $\hat{\theta}$ as an estimator of θ (the equivalent of Wright's F_{ST}). θ can be regarded as co-ancestry coefficient (or relatedness) for alleles within a subpopulation relative to the total population. A unique point is that their estimator also accounted for sampling variance of population and samples which are drawn from the population. $\hat{\theta}$ can be approximated by the sample mean and variance of allele frequency as,

$$\hat{\theta} = \frac{s^2}{\bar{p}(1-\bar{p})}.$$

where s^2 is the sample variance of allele A frequency over subpopulations and $\bar{p}$ is the average sample frequency of allele A and 1 - $\bar{p}$ the average sample frequency of allele a in the total population for biallelic situations (Weir and Cockerham, 1984). They further provided a jackknife procedure for estimating the variance of their estimator. For multiple loci cases, they omits one locus at a time and calculate the jackknife variance of $\hat{\theta}$, while for single locus case, they suggest to jackknife over (sub-) populations.

Coalescent F_{ST}. Also by analogue of Wright's *F*-statistics, Slatkin (1991) related F_{ST} with the time to most recent common ancestor (i.e. the coalescence time) for a pair of alleles chosen within the same subpopulation and drawn randomly from the total population, that is

$$F_{ST} = \frac{\bar{t} - \bar{t}_0}{\bar{t}},$$

where, $\bar{t}_0$ is the average coalescence time of two alleles drawn from the same subpopulation and $\bar{t}$ is the average coalescence time of two alleles drawn from the total population (the whole metapopulation). This definition of F_{ST} is independent on the assumptions of demography and expected to be roughly similar for all neutral loci (Whitlock, 2011).

Φ_{ST}. Another F_{ST} analogous statistic, Φ_{ST}, was developed by Excoffier (1992). It is based on the idea of analysis of variance (ANOVA) and was termed "analysis of molecular variance (AMOVA)". They extended the work of Cockerham (1973) and Weir and Cockerham (1984), which partitioned the overall variance into within and among populations components, to a comparable analysis of haplotypic diversity. A matrix of squared distances of each pair of haplotypes was constructed and used to calculate sum of squared deviations of different subdivisions. The distance metric can be customarily specified to any meaningful evolutionary or

genetic distance according to the research question. If a binary distance between haplotypes is used, one for identical haplotype and zero for different haplotypes, then Φ_{ST} is the same as θ or F_{ST}. To test the significance of each component of variances, a permutation procedure was conducted. The null distribution of component of variance was calculated from a large number of replicate data sets by reallocating each individual to a randomly chosen population.

R_{ST}. As microsatellite data become more popular in the analysis of natural populations, R_{ST}, which is also a F_{ST}-like statistic, specifically accounting for the mutational process of microsatellite loci, was introduced by Slatkin (1995). R_{ST} is the fraction of total variance of allele size from between populations (Slatkin, 1995). Allele size is measured as the number of repeat units in the short microsatellite DNA sequences. Slatkin showed that, for microsatellite loci following generalized stepwise mutation model, R_{ST} has very similar property as that of F_{ST} under a K-alleles mutation model (Box 1).

The F_{ST} analogues such as G_{ST}, θ and Φ_{ST}, in particular G_{ST}, have been criticized to be constrained by within subpopulation heterozygosity H_S (Hedrick, 2005; Jost, 2008; Edelaar and Bjorklund, 2011).

5 Alternative Statistics of Differentiation

There exist two additional measures of differentiation that are quite different from the F_{ST}-family statistics discussed above:

G'_{ST}. As seen from formula (3), the maximum value of G_{ST} is constrained by that of H_S; cases exist where G_{ST} is very small even when there are no shared alleles between any two subpopulations (Hedrick, 2005; Jost, 2008). As a rectification and inspired by Lewontin's standardization on the index of linkage (Lewontin, 1964), Hedrick (2005) introduced a standardized measure of differentiation based on the original G_{ST},

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

where, for K equally weighted subpopulations,

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

This ensures that the value 1 can be reached when there are no shared alleles among subpopulations. This standardization was later recommended to extend to other F_{ST} analogous such as Φ_{ST} or θ (but not R_{ST}) (Meirmans, 2006). However, G'_{ST} is largely a superficial transformation of G_{ST}, at a cost: it loses the inherent theoretical properties of G_{ST} and thus lacks a proper evolutionary interpretation (Ryman and Leimar, 2009; Whitlock, 2011; Leng and Zhang, 2011).

D. Jost (2008) further noticed that Nei's additive decomposition of total heterozygosity, $H_T = H_S + D_{ST}$, is an incomplete partitioning, because the two components on the right are not independent. To overcome this, he adopted an ecological concept of the so-called true diversity Δ_T (Jost 2007), which corresponds to, under the assumption of equal sizes of all subpopulations, the effective number of alleles in the term of Kimura and Crow (1964).

$$\Delta_T = 1/J_T,$$

which is actually equal to the reciprocal of Nei's (1973) gene identity. Jost then partitioned Δ_T into the within- and between-subpopulation components (Δ_S and Δ_{ST}, respectively) (Jost, 2008).

$$\Delta_T = \Delta_S \cdot \Delta_{ST}.$$

To make Δ_S and Δ_{ST} being independent of each other (i.e. being the *pure* within- and between-subpopulation components), "Δ_S for n equally weighted subpopulations must be the reciprocal of the average of the gene identities of the subpopulations" (*p.*4020, Jost, 2008). That is:

$$\Delta_S = 1/J_S.$$

The between subpopulation component Δ_{ST} is then defined as an absolute measure of subpopulation differentiation (because it represents the effective number of distinct subpopulations). It has a range between n (when all n subpopulations have no shared alleles) and 1 (when all subpopulations are identical in composition). From this perspective, Jost (2008) proposed a completely new relative measure of differentiation, named as D:

$$\begin{aligned} D &= [(\Delta_S/\Delta_T) - 1]/[(1/K) - 1] \\ &= (J_T/J_S - 1)/[(1/K) - 1] \\ &= [(J_S - J_T)/J_S]\,[K/(K-1)] \\ &= [(H_T - H_S)/(1 - H_S)]\,[K/(K-1)]. \end{aligned}$$

Jost (2008) pointed out that D is independent of within-subpopulation heterozygosity (H_S). However, one may argue that D is not really independent of H_S but is not constrained by it (recall that G_{ST} is constrained by H_S).

6 Critical Issues in the Application of the Statistics

We would like to emphasize that the distribution pattern of genetic variation within a species is an *inte-*

grated consequence of both historical and contemporary processes, the status of which depends on a number of factors: mutation, natural selection, random genetic drift, gene flow, the initial state of ancestral populations, and population split time (Leng and Zhang, 2013). Statistical measures discussed in this review do not presume the causal factors of population differentiation, nor do they discriminate between adaptive differentiation and non-adaptive differentiation. For example, there is no rigid assumption to constrain the application of the above statistics for estimating genetic differentiation of populations except for the assumption of equal subpopulation sizes. It is the investigators who seek the interpretation(s) and implication(s) of the estimates of these statistics. However, the situation will be radically different when one intends to estimate any parameters of population from these statistics. The aforementioned factors are acting interwindingly in populations. As a consequence, it is wishful thinking to attempt to infer the effect of just one particular factor while ignoring the others for natural populations, except under some simplified theoretical models. One must pay great attention to such (generally unrealistic) assumptions when applying the relevant shortcut methods.

6.1 Inference of the magnitude of gene flow

The level of gene flow is a very important population parameter. For practical reasons, the concern is usually on the absolute magnitude of gene flow, i.e. the number of individuals immigrated to and interbred in a population per generation, that is, N_em. One unique advantage of the statistic F_{ST} (G_{ST}) is that it allows the inference of the magnitude of gene flow N_em among populations under certain conditions (a marvelous characteristic of F_{ST} and G_{ST} that is in great favour of this application which is not aware of by the early population geneticists is just revealed recently: the equilibrium value of G_{ST} is largely not sensitive to violation of the assumption of mutation model, such as the infinite alleles assumption. See Leng and Zhang, 2011). Dobzhansky and Wright (1941) first explicitly showed that under the island model of population structure (Box 1) with the assumptions of no mutation and migration rate (m) being small, the following approximation exists among the fixation index F_{ST}, the effective population size N_e and m under equilibrium (Box 1),

$$F_{ST} \approx \frac{1}{4N_em+1}. \quad (4)$$

Actually, this formula was obtained in a different way earlier by Wright (1931). Therefore, if the equilibrium value of F_{ST} can be estimated from empirical genetic data, N_em can then be calculated.

About two years after his publication of G_{ST}, Nei derived a rather complex formula for G_{ST} under the island model at equilibrium (Nei, 1975). However, this can be reduced to

$$G_{ST} \approx \frac{1}{1+4N_e\frac{K}{K-1}(m+\mu)},$$

where μ is the mutation rate per generation and K is the number of subpopulations (Takahata and Nei, 1984). When the number of subpopulation K is sufficiently large, $m << 1$ and $\mu << m$, the right side of the above formula is approximately equivalent to that of equation (4). If the mutation rate is not negligible, the formula becomes

$$G_{ST} = \frac{1}{4N_em+4N_e\mu+1}$$

(Cockerham and Tachida, 1987; Cockerham and Weir, 1993).

Although there is no direct relationship between F'_{ST} (G'_{ST}) and m, Meirmans and Hedrick (2011) suggest that 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} (equally applicable to G'_{ST}),

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

Simulation studies revealed that the equilibrium value of D is highly sensitive to the assumed mutation model (Leng and Zhang, 2011). Because of this property, the utility of D in population parameter estimation is rather limited.

Here we would like to add a cautionary note to the story of estimation of N_em from F_{ST} (G_{ST}): Its reliability depends critically on whether G_{ST} has approached equilibrium (see below for more discussion). In contrast to the popular belief, $\mu << m$ is not a sufficient condition for using G_{ST} to estimate N_em. For example, If population size is large and gene flow relatively low, it can lead to seriously overestimated N_em even if the condition $\mu << m$ is fulfilled, because it is not safe to assume that G_{ST} has reached equilibrium (Leng and Zhang, 2013; Box 1). Additionally, one should carefully consider the biology of their study organisms before they employ equation (4) to estimate gene flow, and should not simply presume that their species follows the island model of population structure.

6.2 Properties of the differentiation measures under non-equilibrium conditions

As briefly discussed above, the inference of demographic parameters such as N_em under island model has to meet the premise that the statistic indices are at equilibrium, but many species are probably far from reaching equilibrium at many genetic loci (Box 1). Therefore, it is important to learn the properties of the differentiation measures under non-equilibrium conditions. Research focus thus far is mainly on the two most popular indices, G_{ST} and D, under island model. The following outcomes have emerged. First, it is observed that both G_{ST} and D take a fairly long time to reach equilibrium (and hence serve as a suitable measure of genetic differentiation) if gene flow is weak (e.g., $m < 10^{-4}$) and mutation rate is not very large (Leng and Zhang, 2013). Simulation study reveals that when gene flow is absent or very weak, G_{ST} and D can only reach equilibrium when H_S and H_T are both in equilibrium; but when migration rate is moderate or high (e.g., 10^{-3}) G_{ST} approaches equilibrium much quicker than H_T, H_S, and D. In general, D usually converges to its equilibrium value much slower than G_{ST} (Ryman and Leimar, 2009; Leng and Zhang, 2011, 2013; Whitlock, 2011). It is worthy of emphasis that it is the migration rate (m), not the absolute number of migrants (N_em), that determines the speed at which G_{ST} or D approaches equilibrium (Leng and Zhang, 2013).

Second, it appears that in non-equilibrium populations, drift plays a dominant role on G_{ST} whatever the level of gene flow, whereas after the initial stage of population differentiation, drift seems to only play a secondary role on D when subpopulations exchange individuals, even if the exchange is infrequent (Leng and Zhang, 2013). This property of D can be seen more clearly from its equilibrium value. When $m << 1$ and $K\mu << m$, at equilibrium (Jost, 2008),

$$D \approx \mu(K-1)/m$$
$$\approx \mu/m \text{ for } K = 2;$$
$$\approx \mu K/m \text{ for moderate } K.$$

Note that this formula does not contain the effective population size N_e.

Third, in non-equilibrium populations, G_{ST} builds up more quickly in time than D. Also, when very low level of genetic variation exists in populations, D is often unable to detect differentiation. Hence, G_{ST} should have a larger power to detect recent population genetic events than D (Leng and Zhang, 2013).

Fourth, although D is usually sensitive to mutation rate, in certain situations it can be much less sensitive to mutation rate heterogeneity than G_{ST}. For example, under SMM (but not IAM, i.e. the infinite allele model, see Box 1) and complete isolation, when population size is large, mutation rate shows a great impact on G_{ST} but only a mild influence on D. This is markedly different from what has been recognized in the equilibrium perspective, and provides a potential option for using D to inspect mutation rate heterogeneity across loci in large isolated populations (Leng and Zhang, 2013).

Finally, an important but often overlooked aspect is that the speed at which populations diverged, as measured by G_{ST} and D, are strongly dependent on the genetic diversity of these populations at split time. The greater the initial diversity was, the faster the populations diverged, and the earlier the equilibrium state was reached (Ryman and Leimar, 2008; Leng and Zhang, 2011). Therefore, the demographic history of the ancestral population may have some long-lasting impacts on population differentiation and can hold sway over our application of the statistical measures.

6.3 Detecting differentiation

There exist fundamental differences between the two most popular statistics G_{ST} and D, neither of them operates satisfactorily in all situations for quantifying differentiation. In addition to those properties of the two indices under non-equilibrium conditions just discussed above, the following messages are also worthy of consideration. First, as a general rule, G_{ST} can quantify differentiation fairly well when heterozygosity is low whatever the causes (e.g. low mutation rate, low initial heterozygosity of the ancestral population or short split time); however, when heterozygosity is high (whatever the causes, e.g. high mutation rate or high initial heterozygosity) and gene flow is moderate to strong, G_{ST} often fails to measure differentiation (Leng and Zhang, 2011). A good practice is to pay close attention to the relative levels of heterozygosities (especially H_S) in the data when G_{ST} is the index of interest.

Second, the accuracy of G_{ST} and D in signaling genetic differentiation varies depending on mutation regimes, and the two indices bear different insights for markers with high mutation rates. But the issue is complicated by other factors. For example, Alcala et al. (2014) claimed that D apparently has a better reflection to genetic diversity at weak mutation strength, while G_{ST} has better performance in detection of differentia-

tion when mutation rate is intermediate. At high mutation regime, both measures are unsatisfactory, but D is slightly better than G_{ST}, particularly if the number of subpopulations is small (Alcala et al., 2014). However, this probably only reflects one facet of the dice, since the behaviours of G_{ST} and D will be affected by population split time and the evolutionary process as well, as demonstrated in Raeymaekers et al. (2012) (see below). Quite interestingly, it is also demonstrated that when population size is not very small (e.g. $N \geq 1000$), G_{ST} can quantify differentiation quite linearly with time over a long duration when gene flow is absent or very weak even if mutation rate is not low (e.g. $\mu = 0.001$) (Leng and Zhang, 2011).

Third, D and G_{ST} have different sensitivities to processes shaping short-term population structure. An empirical study by Raeymaekers et al. (2012) suggests that for markers with moderate to high mutation rates, on short timescales and across strong clines in population size and connectivity, D is useful to infer colonization history, whereas G_{ST} is sensitive to more recent demographic events. The theoretical basis of this observation can be found in the simulation study of Leng and Zhang (2013) and is discussed above.

Fourth, because both G_{ST} and D are affected by a number of variables, including population size, heterozygosity, migration rate, mutation rate, mutation model, etc., a large value of G_{ST} or D that we observed may not necessarily means a greater degree of differentiation, and it may simply indicate a small population size under non-equilibrium conditions. In summary, if population is still in non-equilibrium state, a large value of D may mean several things: short population split time and small size, and/or low initial homozygosity (high initial heterozygosity), or low migration rate, or high mutation rate. Whereas, a large value of G_{ST} may mean small population size, or large number of subpopulations, or low mutation rate, or low migration rate. This general nature may bear some important implications. For example, it is known that the effective population size (Ne) vary across the genome and much of this variation still remains unexplained (Gossmann et al., 2011). Thus, outlier loci with high G_{ST} (F_{ST}) may not all be accounted for by selection. Similarly, these facts also strengthen the argument that comparisons of the values of these statistics across studies, not mentioning across different types of genetic markers, should be practiced judiciously. In particular, although the effect on G_{ST} and D of mutation rate heterogeneity across loci employed in a study remains to be explored, one should be prudent if H_S also manifests great heterogeneity across loci.

6.4 Estimation and test of significance

Under the assumption of equal effective population size for all subpopulations, statistics such as G_{ST}, G'_{ST}, and D can be all calculated from the total heterozygosity H_T and the average within subpopulations heterozygosity H_S. Although one can directly apply sampling frequencies of alleles as plugin for population frequencies and calculate the estimates of heterozygosities, and then compute the estimates of the above statistics, it is highly recommended to use the nearly unbiased estimators of H_T and H_S, such as Nei and Chesser (1983) or Nei (1987). Note that θ defined by Weir and Cockerham (1984) can be regarded as an estimator of F_{ST} corrected for sampling bias. In addition, Jost (2008) also derived the estimator of D according to Morisita-Horn similarity measure used by ecologists (Chao et al., 2008). For more exhaustive discussion on this issue, readers may consult a recent comprehensive review by Excoffier (2007).

Researchers are often concerned about the statistical confidence of these estimates. Sadly, no parametrical or statistical distributions of these measures have been derived, and the best solution for test of significance is to apply some non-parametric procedures. One way is to use a permutation test as Excoffier et al. (1992) for Φ_{ST}, which shuffle the genotypes of individuals among subpopulations a great many times, and calculate the empirical distribution of the statistic. Another way is to do a resampling procedure such as bootstrap or jackknife; the null distribution can be then obtained from the calculation of estimators from replicate samples (Meirmans and Hedrick, 2011). However, in practical applications, these seemingly simple resampling techniques are not as simple as they sound. The nature of the scientific questions (e.g. the hierarchical levels), the feature of genetic markers (e.g. linked or independent), the assumptions on population structuring (e.g. with structure or no structure) can all affect whether a given technique can be employed or how it should be applied (e.g. bootstrapping or permuting loci, individuals, or subpopulations?). Excoffier (2007) has provided a comprehensive analysis on these issues. Additionally, when only a small number of loci are used and resampling results are capricious, one should be vigilant against some unusually behaved alleles, loci, or population samples.

7 Challenges and Future Development

Although population differentiation is a fundamental process of evolution, and many evolutionary studies demand the inference of population differentiation, the field of theoretical research on statistic measures has been largely in stasis for decades. The popping up of D as a novel measure alternative to G_{ST} has unprecedentedly advanced our understanding of both the over half century old G_{ST} and the a few years old D, creating both challenges and opportunities for future development. The first remark to make is that both G_{ST} and D are useful indices but in the same time are not satisfactory. After all, F_{ST} (and thus G_{ST}) was tinkered up from a coefficient initially developed for assessing inbreeding within a population, and D was developed in a sense to overcome the incomplete partitioning of Nei's additive decomposition of the total heterozygosity (this leads it to be criticized of lacking any evolutionary interpretation as its theoretical foundation). Can an alternative statistic be set up specifically for quantifying the degree of population differentiation that complies with both the established population genetic theory as well as probability theory, for example? Under the shadow of thought of F_{ST}, various paths have already been explored and realized (see the section "The rationales"). Nevertheless, one track that has not been seriously examined is the inflation of gene identity in subpopulations caused by nonrandom mating since differentiation. This is the common consequence of population subdivision, thus a fundamental nature of the fundamental process. We are currently exploring this possibility and developing a novel statistic named the inflation index which will be publish elsewhere.

The second remark to make is that all the formulas of the statistics discussed in this review represent unrealistic approximation to natural populations. For example, they all assume equal subpopulation size and simplified formulas for estimating population parameters under the island model. Therefore, even if they can satisfactorily quantify population differentiation, their estimates are unlikely to be reliable in natural populations. Hence, a challenge for any novel statistic is that it should have the potential to be extended to more generalized situations, for example, allowing subpopulations to have different sizes and accommodating different models of population structure. Attempt has already been made to define a generalized form of G_{ST} to reconcile more demographic and evolutionary scenarios (Hössjer et al., 2014).

The third remark is that measuring population differentiation so far has been largely limited to summary statistics without much inference power. This seems to be increasingly in disfavour. While model-based inference is clearly the promoted promising direction, delving deeper into the existing statistics may also be a worthwhile effort. Our earlier simulation study indicates that, although it only explored a very limited parameter space, by contrasting the differential behaviors of D and G_{ST} under non-equilibrium conditions, it is potentially possible to make some inference of the evolutionary processes shaping population differentiation (Leng and Zhang, 2011).

8 Concluding Remarks

The available statistics fall into two categories, represented by F_{ST} and D, respectively. None of them is perfect in measuring population genetic differentiation (Leng and Zhang, 2011; Meirmans and Hedrick, 2011; Whitlock, 2011; Putman and Carbone, 2014). Nevertheless, they each have advantages and are all valuable for current research. In practice, both indices should be calculated and a comparison of them can generate useful insights into the evolutionary processes that influence population differentiation. F_{ST} (G_{ST}) has some unique irreplaceable characteristics assuring its standing as the default measure at present and in the near future. Also, it will continue to serve as the standard (a null model) for any alternative measures to contrast with. Instead of being anxious about making choice between these indices, one should pay due attention to the equilibrium status and the level of diversity (especially H_S) of the populations, since they largely sway the power of a given statistic to address a specific question. The multi-faceted comparative summary discussed in the section "Critical issues in the application of the statistics" provides a basic reference for readers to be familiar with the pros and cons of various statistics and then guide their applications.

Acknowledgements We are grateful to Hua Chen (Beijing Institute of Genomics) for valuable suggestions and Ziheng Yang for critical reading of an earlier version of the manuscript. This study is supported by a grant from the Ministry of Science and Technology of China (Grant No. 2011CB808001) and the Breakthrough Project of Strategic Priority Program of the Chinese Academy of Sciences, Grant No.XDB13030200.

References

Alcala N, Goudet J, Vuilleumier S, 2014. On the transition of genetic differentiation from isolation to panmixia: What we can learn from G_{ST} and D. Theoretical Population Biology 93: 75–84.

Barton NH, Charlesworth B, 1984. Genetic revolutions, founder effects, and speciation. Annual Review of Ecology and Systematics 15: 133–164.

Chao A, Jost L, Chiang SC, Jiang YH, Chazdon RL, 2008. A two-stage probabilistic approach to multiple-community similarity indices. Biometrics 64: 1178–1186.

Cockerham CC, 1973. Analyses of gene frequencies. Genetics 74: 679–700.

Cockerham CC, Weir BS, 1987. Correlations, descent measures: Drift with migration and mutation. Proceedings of the National Academy of Sciences of the United States of America 84: 8512–8514.

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

Dobzhansky T, Wright S, 1941. Genetics of natural populations. V. Relations between mutation rate and accumulation of lethals in populations of *Drosophila pseudoobscura*. Genetics 26: 23–51.

Edelaar P, Bjorklund M, 2011. If F_{ST} does not measure neutral genetic differentiation, then comparing it with Q_{ST} is misleading. Or is it? Molecular Ecology 20: 1805–1812.

Excoffier L, 2007. Analysis of population subdivision. In: Balding DJ, Bishop M, Cannings C eds. Handbook of Statistical Genetics. 3rd edn. West Sussex: John Wiley & Sons, 980–1020.

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.

Gossmann TI, Woolfit M, Eyre-Walker A, 2011. Quantifying the variation in the effective population size within a genome. Genetics 189: 1389–1402.

Hartl DL, Clark AG, 1997. Principles of Population Genetics. 3rd edn. Massachusetts: Sinauer Associates, Inc.

Hedrick PW, 2005. A standardized genetic differentiation measure. Evolution 59: 1633–1638.

Hössjer O, Olsson F, Laikre L, Ryman N, 2014. A new general analytical approach for modeling patterns of genetic differentiation and effective size of subdivided populations over time. Mathematical Biosciences 258: 113–133.

Huang ZS, Ji YJ, Zhang DX, 2008. Haplotype reconstruction for scnp DNA: A consensus vote approach with extensive sequence data from populations of the migratory locust *Locusta migratoria*. Molecular Ecology 17: 1930–1947.

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.

Kimura M, Crow JF, 1964. The number of alleles that can be maintained in a finite population. Genetics 49: 725–738.

Lehmann EL, Casella G, 1998. Theory of Point Estimation. Springer Science & Business Media.

Leng L, Zhang DX, 2011. Measuring population differentiation using G_{ST} or D? A simulation study with microsatellite DNA markers under a finite island model and nonequilibrium conditions. Molecular Ecology 20: 2494–2509.

Leng L, Zhang DX, 2013. Time matters: Some interesting properties of the population differentiation measures G_{ST} and D overlooked in the equilibrium perspective. Journal of Systematics and Evolution 51: 44–60.

Lewontin RC, 1964. The interaction of selection and linkage. I. General consideration: Heterotic models. Genetics 49: 49–67.

Meirmans PG, 2006. Using the AMOVA famework to estimate a standardized genetic differentiation measure. Evolution 60: 2399–2402.

Meirmans PG, Hedrick PW, 2011. Assessing population structure: F_{ST} and related measures. Molecular Ecology Resources 11: 5–18.

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. Amsterdam, Netherlands: North-Holland Publishing Co.

Nei M, 1987. Molecular Evolutionary Genetics. New York: Columbia University Press.

Nei M, Chesser RK, 1983. Estimation of fixation indices and gene diversities. Annals of Human Genetics 47: 253–259.

Ohta T, Kimura M, 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genetics Research 22: 201–204.

Prout T, Barker JS, 1993. F statistics in *Drosophila buzzatii*: Selection, population size and inbreeding. Genetics 134: 369–375.

Putman AI, Carbone I, 2014. Challenges in analysis and interpretation of microsatellite data for population genetic studies. Ecology and Evolution 4: 4399–4428.

Raeymaekers JAM, Lens LUC, Van den Broeck F, Van Dongen S, Volckaert FAM, 2012. Quantifying population structure on short timescales. Molecular Ecology 21: 3458–3473.

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, 1985. Rare alleles as indicators of gene flow. Evolution 39: 53–65.

Slatkin M, 1991. Inbreeding coefficients and coalescence times. Genetics Research 89: 479–487.

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.

Takahata N, Nei M, 1984. F_{ST} and G_{ST} statistics in the finite island model. Genetics 107: 501–504.

Weir BS, Cockerham CC, 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358–1370.

Whitlock MC, 2011. G'_{ST} and D do not replace F_{ST}. Molecular Ecology 20: 1083–1091.

Wright S, 1921. Systems of mating. II. The effects of inbreeding

on the genetic composition of a population. Genetics 6: 124–143.
Wright S, 1922. Coefficients of inbreeding and relationship. The American Naturalist 56: 330–338.
Wright S, 1931. Evolution in mendelian populations. Genetics 16: 290–290.
Wright S, 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proceedings of the VI International Congress of Genetics 1: 356–366.
Wright S, 1943. Isolation by distance. Genetics 28: 114–138.
Wright S, 1951. The genetical structure of populations. Annals of Eugenics 15: 323–354.
Wright S, 1965. The interpretation of population structure by *F*-statistics with special regard to systems of mating. Evolution 19: 395–420.