Theoretical Population Biology 93 (2014) 75–84

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

On the transition of genetic differentiation from isolation to panmixia: What we can learn from G_{ST} and D

Nicolas Alcala [a,*], Jérôme Goudet [a], Séverine Vuilleumier [a,b]

[a] Department of Ecology and Evolution, University of Lausanne, Biophore CH-1015 Lausanne, Switzerland

[b] Institute of Microbiology, University Hospital Center and University of Lausanne (CHUV–UNIL), CH-1011 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 28 November 2013
Available online 19 February 2014

Keywords:
Genetic differentiation
Genetic diversity
F_{ST}
D
G_{ST}

ABSTRACT

Population genetic differentiation characterizes the repartition of alleles among populations. It is commonly thought that genetic differentiation measures, such as G_{ST} and D, should be near zero when allele frequencies are close to their expected value in panmictic populations, and close to one when they are close to their expected value in isolated populations. To analyse those properties, we first derive analytically a reference function f of known parameters that describes how important features of genetic differentiation (e.g. gene diversity, proportion of private alleles, frequency of the most common allele) are close to their expected panmictic and isolation value. We find that the behaviour of function f differs according to three distinct mutation regimes defined by the scaled mutation rate and the number of populations. Then, we compare G_{ST} and D to f, and demonstrate that their signal of differentiation strongly depends on the mutation regime. In particular, we show that D captures well the variations of genetic diversity when mutation is weak, otherwise it overestimates it when panmixia is not met. G_{ST} detects population differentiation when mutation is intermediate but has a low sensitivity to the variations of genetic diversity when mutation is weak. When mutation is strong the domain of sensitivity of both measures are altered. Finally, we also point out the importance of the number of populations on genetic differentiation measures, and provide recommendations for the use of G_{ST} and D.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Quantification of population genetic differentiation (i.e. how different are allele frequencies between populations) is a long standing issue in population genetics. Wright (1951) first proposed the fixation index F_{ST}, which measures both the proportion of alleles that reach fixation and population allelic differentiation for a bi-allelic locus. Wright's fixation index was then extended to multi-allelic loci, G_{ST}, by Nei (1973) and Weir and Cockerham (1984)—also often referred as F_{ST} (Holsinger and Weir, 2009; Jakobsson et al., 2012). G_{ST} characterizes the ratio of within-population, H_s, to total, H_t, gene diversities (notations are summarized in Table 1):

$$G_{ST} = 1 - \frac{H_s}{H_t} \tag{1a}$$

where

$$H_s = 1 - \frac{1}{n}\sum_j \sum_i p_{ij}^2 \tag{1b}$$

$$H_t = 1 - \sum_i \left(\frac{1}{n}\sum_j p_{ij}\right)^2 \tag{1c}$$

p_{ij} is the frequency of allele i in population j and n is the number of populations. Thus H_s and H_t correspond to the probability that two genes randomly chosen, respectively, from the same population and from different populations, at a given locus, are different (Nei, 1973).

Nei (1973) considered G_{ST} as a statistics which characterizes the properties of sampled populations, while Weir and Cockerham (1984) considered that G_{ST} is independent of the sampling scheme and represents a parameter of the populations that can be estimated (Holsinger and Weir, 2009). In this study, we analyse the actual level of genetic differentiation of the populations, and not the properties of a sample, thus we consider genetic differentiation measures as parameters and not as statistics.

* Correspondence to: Department of Ecology and Evolution, University of Lausanne, Biophore, Room 4325. Lausanne, Switzerland.
E-mail address: Nicolas.Alcala@unil.ch (N. Alcala).

http://dx.doi.org/10.1016/j.tpb.2014.02.003
0040-5809/© 2014 Elsevier Inc. All rights reserved.

Table 1
Summary of notations.

Genetic differentiation measures	
G_{ST}	Differentiation measure based on allele fixation (Nei, 1973)
G'_{ST}	Normalized genetic differentiation measure based on allele fixation (Hedrick, 2005)
D	Genetic differentiation measure based on genetic composition (Jost, 2008)
D_{ST}	Absolute genetic differentiation measure (Nei, 1973)
Genetic diversity measures	
H_s	Within-population gene diversity (Nei, 1973)
H_t	Total gene diversity (Nei, 1973)
Δ_S	Within-population effective number of alleles (Jost, 2008)
Δ_T	Total effective number of alleles (Jost, 2008)
Summary statistics of allele frequencies	
p_{max}	Frequency of most frequent allele in the total population (Jakobsson et al., 2012)
$\overline{p^i_{max}}$	Frequency of most frequent allele in each population (Jakobsson et al., 2012)
p_{priv}	Proportion of private alleles (Slatkin, 1985)
σ	Mean singular value of the allele frequency table (Golub and Kahan, 1965)
Functions	
$f(M)$	Function describing the transition from isolation to panmixia
$f_G(M)$	Function describing the behaviour of G_{ST} as a function of the scaled migration rate
$f_D(M)$	Function describing the behaviour of D as a function of the scaled migration rate
M_T	Threshold migration value of function $f(M)$
M_G	Threshold migration value of function $f_G(M)$
M_D	Threshold migration value of function $f_D(M)$

Several issues related to G_{ST} have been raised (Charlesworth, 1998; Nagylaki, 1998; Hedrick, 1999; Jost, 2008). Authors showed that values of G_{ST} are constrained by the value of within-population gene diversity H_s; G_{ST} remains inferior to $1-H_s$ (see Fig. 1 from Jost, 2008). Therefore, when H_s is large, the range of the genetic differentiation signal is truncated and G_{ST} can be constrained by the frequency of the most frequent allele (Jakobsson et al., 2012). This has two main consequences; first, when H_s is high, G_{ST} cannot detect differentiation (as shown for high mutation rate loci in Balloux et al., 2000); second, when loci have different H_s values, corresponding G_{ST} values cannot be compared as G_{ST} does not rank populations by their degree of differentiation (Jost, 2008; Heller and Siegismund, 2009). For example, a set of populations with high H_s that does not share alleles has a lower G_{ST} than a set of population with low H_s but that shares alleles (Jost, 2008). In addition, G_{ST} can take a value of 1 when in each population the same allele is fixed (Gregorius, 2010).

To overcome the limitations of G_{ST}, new measures of differentiation were proposed. First, Hedrick (2005) derived a normalized value of G_{ST}, G'_{ST}, which ranges from 0 to 1 whatever the level of within-population gene diversity, H_s. Nevertheless, like G_{ST}, G'_{ST} cannot detect differentiation when the same allele is fixed in each population, as G'_{ST} is very close to G_{ST} when H_s is low (Hedrick, 2005). Second, Jost (2008) proposed a new measure of genetic differentiation, D, based on the ratio of the within-population Δ_S to the total Δ_T effective number of alleles (Kimura and Crow, 1964; also called "true diversity" in Jost, 2008).

$$D = \frac{n}{n-1}\left(1 - \frac{\Delta_S}{\Delta_T}\right) \tag{2}$$

with $\Delta_S = \frac{1}{1-H_s}$ and $\Delta_T = \frac{1}{1-H_t}$.

G'_{ST} and D provide similar estimations of population differentiation (Heller and Siegismund, 2009) when H_s is high. However, both have slow rate of convergence to their equilibrium values after a perturbation when mutation is low (Ryman and Leimar, 2008, 2009). Thus, following a demographic change (i.e. change in population size or migration rate), G'_{ST} and D values can depend on the prior population size and gene diversity. Impacts can be strong in situation where the within-population gene diversity H_s before the demographic change is far from the expected H_s after the change. On the contrary, when the mutation rate is high, D can converge faster than G_{ST}. When loci are under different mutation regimes their rate of convergence differs accordingly.

The discrepancies between G_{ST} and D have been analyzed and discussed in numerous recent studies (Jost, 2008; Ryman and Leimar, 2008, 2009; Jost, 2009; Gerlach et al., 2010; Gregorius, 2010; Kronholm et al., 2010; Meirmans and Hedrick, 2011; Whitlock, 2011; Wang, 2012). Those studies conclude that D and G_{ST} do not detect the same type of genetic differentiation. G_{ST} values could be interpreted as a measure of the level of allele fixation in populations (Whitlock, 2003), and also reflects the population's demographic properties (e.g. the number of migrants per generation under assumed mutation rate) independently of the analyzed loci. D values could be interpreted as measures of the difference in the genetic composition of populations, reflecting the properties of the analyzed loci and are related to genetic distance (Jost, 2009).

Many studies that compared G_{ST} and D values used one of them as a reference value (Ryman and Leimar, 2009). To compare G_{ST} and D, Heller and Siegismund (2009) used D as a reference value, which prevented the detection of any issue related to D. Similarly, Whitlock (2011) used the coalescent F_{ST} as a reference value, which is closer to G_{ST} than D when the mutation rate is low. On the contrary, Jost (2008) used a genetic differentiation definition to compare D and G_{ST} values and avoid circular arguments (Table 1 and Fig. 2 in Jost, 2008), however it involved only two populations and considered a restrictive number of illustrative examples. To have a deep understanding of the behaviour and the properties of measures of genetic differentiation such as G_{ST} and D, an independent reference is necessary. This approach must provide a description of the behaviour of the measures of genetic differentiation given any number of populations considered and across a large range of migration and mutation rates.

Here, we do not propose a new measure of genetic differentiation, instead, we propose an innovative approach to characterize the transition of genetic differentiation from isolation to panmixia and to understand the properties of measures of genetic differentiation, G_{ST} and D. We first introduce a function, $f(M)$, that is implicitly defined and describes the degree to which populations are closer to isolation (i.e. no mating between individuals from different populations) or panmixia (i.e. random mating between individuals from all populations) and how transition between these two states occurs as a function of the number of migrants per generation. The derivation of $f(M)$ requires a complete understanding of the processes shaping genetic diversity and genetic differentiation and thus requires to specify a model with known parameters. We use the most common population structure model, the finite island model and the infinite allele model of mutation, under equilibrium conditions (Maruyama, 1970; Maynard Smith, 1970). Thus, $f(M)$ is not a model-independent statistics measurable from population data, but is a theoretical reference to which the values of G_{ST} and D can be compared under a known model.

Second, we demonstrate that the function, $f(M)$, captures the behaviour of the within-population gene diversity H_s and the total gene diversity, H_t, derived by Maruyama (1970) under the finite island model. The function $f(M)$ has a threshold migration value M_T that characterizes the number of migrants that leads to a transition between a behaviour of gene diversity close to isolation or close to panmixia. Interestingly, the behaviour of gene diversities (H_s and H_t) as a function of the number of migrants differs according to three distinct mutation regimes that are defined by the mutation

Download English Version:

https://daneshyari.com/en/article/4502344

Download Persian Version:

https://daneshyari.com/article/4502344

Daneshyari.com