Molecular Ecology (2009) **18**, 2088–2091

NEWS AND VIEWS

REPLY

D vs. G_{ST}: Response to Heller and Siegismund (2009) and Ryman and Leimar (2009)

LOU JOST
Via a Runtun, Baños, Tungurahua, Ecuador

Received 16 December 2008; revised 26 January 2009; accepted 3 February 2009

When to use *D* and when to use G_{ST}

The meta-analysis of Heller & Siegismund (2009) and the simulations of Ryman & Leimar (2009) confirm that *D* behaves very differently from G_{ST} not only in theory but in practice. This raises the question of which measure should be used in any given application. The answer is not a matter of finding parameter ranges where one is superior to the other. Rather, the choice of measure depends entirely on the kind of question being asked. G_{ST} and *D* measure different aspects of population structure. *D* measures the actual relative degree of differentiation of allele frequencies among the demes of a population, while G_{ST} is a useful tool for estimating one of the causes of that structure, the amount of migration between demes. These two tasks put different demands on genetic measures.

The task of measuring genetic differentiation answers a concrete question, 'How different are the allele frequencies of the subpopulations?' This is a purely descriptive task which does not depend on the validity of a particular genetic model, or on the achievement of any kind of equilibrium. Differentiation is completely determined by the actual allele frequencies in each deme at the moment of measurement. Allelic differentiation is controlled by multiple factors, including not only migration but also mutation rate, founder effects, genetic bottlenecks, etc., and a real differentiation measure must be sensitive to all of these. Allelic differentiation provides the genetic component of phenotypic differentiation, so it is the relevant quantity for understanding the evolutionary coherence or divergence of subpopulations. Allelic differentiation is also the appropriate guide for establishing conservation priorities, when the goal is to preserve the genetic diversity of a subdivided population of an endangered species. I derived *D* to describe this important population parameter in a logically and mathematically consistent manner.

The task of estimating migration rate, on the other hand, is theory laden. It depends on the approximate validity of the assumptions of the finite island model, which connects a population statistic to migration rate. If these assumptions (including the assumption that equilibrium has been reached) are met, both *D* and G_{ST} give exactly the same estimate of migration rate when the other parameters of the model (deme size *N*, mutation rate μ, and number of demes *n*) are known. However, because *D* is a measure of differentiation, its equilibrium value depends strongly on all the factors that control differentiation: *m*, μ and *n* (equations 15–17 in Jost 2008). G_{ST} on the other hand depends only weakly on these parameters (because it was not derived as a measure of differentiation). Since μ and *n* are usually unknown, G_{ST} is preferable to *D* for estimating migration rate (or absolute number of migrants) when the finite island model applies. G_{ST} does depend strongly on *N* (while *D* does not), but geneticists have learned to be content with an estimate of *Nm* instead of *m* itself, and this term can be extracted from the exact expression for G_{ST} (Takahata & Nei 1984) by making some approximations. As long as these approximations are valid and the island model holds, the estimate of *Nm* based on G_{ST} will be accurate, and it would be consistent with the estimate derived from *D* if the parameters *N*, μ, and *n* had been known. (By using *D* and G_{ST} together, we can calculate both *Nm* and one other parameter value.)

Ryman and Leimar claim that G_{ST} is at least as appropriate as *D* for the first task, the measurement of differentiation. However, G_{ST} and its relatives fail even the most basic tests of a logically consistent measure of allelic differentiation, and *D* passes all tests. G_{ST}-like measures are not monotonic with respect to unambiguously increasing allelic differentiation (Fig. 2 in Jost 2008), and they approach zero when H_S is high, even if the demes are completely differentiated (Fig. 1 in Jost 2008). A value of $G_{ST} = 0.001$ may mean either complete allelic differentiation between demes, almost no allelic differentiation between demes, or anything in between. Similar problems afflict the ecological equivalents of this measure (Jost 2006, 2007, 2009; Hardy & Jost 2008). Heller & Siegismund (2009) show that there is often a difference in ranking populations based on G_{ST} and *D*; Fig. 1 and Table 1 in Jost (2008) demonstrate that the correct ranking according to allelic differentiation is the one based on D. Ryman and Leimar defend G_{ST} as a measure of differentiation without addressing these seemingly decisive issues.

Ryman and Leimar's simulations nicely demonstrate why G_{ST} cannot be considered a measure of allelic differentiation. For example, their Fig. 1 presents graphs of G_{ST} and *D* vs. time for populations consisting of 10 completely isolated demes (which are initially identical in allele frequencies). One simulation uses a very high mutation rate and the other uses a low mutation rate. Beginning immediately after isolation, the high-mutation-rate population will rapidly accumulate more novel private alleles per deme than the low-mutation-rate population. Therefore, the allele compositions of the demes of the high-mutation-rate population will rapidly become more different from each other than those of the low-mutation-rate population. A legitimate measure of allelic differentiation must reflect these differences at all times. Ryman and Leimar's graph shows that *D* ranks the populations correctly at all times: after isolation, *D* is always greater for the high-mutation-rate population than for the low-mutation-rate one. However, this relation is reversed in

the graph of G_{ST}; G_{ST} is *lower* for the high-mutation-rate population than for the low-mutation-rate one. Ryman and Leimar's Fig. 2 shows the same thing: G_{ST} misranks those populations in terms of allelic differentiation while D ranks them correctly. This conclusively disproves their thesis that G_{ST} is a measure of allelic differentiation. The same property that makes G_{ST} a good measure of migration rate (independence from mutation rate) causes it to fail as a measure of allelic differentiation. No measure can do both jobs well.

Relation of D to μ and H_S

The pioneers of population genetics were surprised by the counterintuitive result that 'differentiation' G_{ST} at equilibrium did not depend on mutation rate. Crow (1986) wrote 'Something remarkable has happened here.' In fact, it can be proven mathematically (L. Jost, in preparation) that at equilibrium under the finite island model, any measure of relative allelic differentiation consistent with H_S and H_T must depend strongly on mutation rate.

The derivation of equation 17 in Jost (2008) makes explicit the strong connection between allelic differentiation and mutation rate at equilibrium in the finite island model:

$$D \approx [(n-1)\mu]/[(n-1)\mu + m] \qquad \text{(eqn 1)}$$

This equation contradicts Ryman and Leimar's statement that D 'cannot be interpreted exclusively in terms of basic population genetics quantities ... ' When $\mu << m$ and $n >> 1$, differentiation becomes

$$D \approx n\mu/m$$

D at equilibrium is approximately proportional to μ. The dependence of D at equilibrium on mutation rate is not a flaw (as Ryman and Leimar claim) but a reflection of reality, a mathematical consequence of the finite island model and an intuitively satisfying result. Doubling the mutation rate has approximately the same effect on allelic differentiation as halving the migration rate.

Ryman and Leimar use their Fig. 3 to clarify some points about the nature of D's independence from H_S. It is critically important to understand the difference between mathematical independence (orthogonality), on the one hand, and causal, statistical, or empirical independence on the other. My derivation of D depended on making a complete partition of total genetic diversity into independent within- and between-deme components. The within-deme component (and hence H_S) imposes no mathematical constraints on the between-deme component (which gives D), but this says nothing about the possibility of causal connections between H_S and D in the real world. The mathematical independence of H_S and D is like the mathematical independence of two orthogonal components of a vector. The value of one component puts no mathematical constraint on the value of the other component. However, in any given application, there can be empirical or causal relations between the two components, just as the laws of physics often relate the x- and y-components of vectors. Mathematical independence is important because then D is neutral with respect to these causal connections. It will accurately reflect the underlying genetic processes influencing allelic differentiation, without hiding these under purely mathematical effects. If the value of H_S had put mathematical constraints on the value of D (as it does for G_{ST}), then D could not function as a stand-alone measure of allelic differentiation; we would need to also know H_S in order to disentangle the purely mathematical effects from the real effects (as in Hedrick 2005). This can be demonstrated by modifying the initial conditions of the simulation presented in Ryman and Leimar's Fig. 1. As in their simulation, there are ten completely isolated demes, each with 1000 individuals. Mutation rate is 0.001 and migration rate is zero. Instead of starting with a uniform population in equilibrium, as they did, I let each deme be fixed for a different private allele, perhaps as a result of a founder effect or genetic bottleneck. The demes are therefore completely differentiated initially, and both D and G_{ST} are unity. Under the infinite allele model, the demes will remain completely differentiated at later times, because each new mutation is novel, and because migration is zero. However, within-deme heterozygosity H_S increases rapidly in the first few generations, because of the high mutation rate. The new measure D is not affected by the increasing H_S; it correctly shows that differentiation is complete at all times. In contrast, G_{ST} is mathematically constrained by the value of H_S. It is forced to take low values as H_S increases with time, wrongly indicating that differentiation is sharply decreasing. See my Fig. 1.

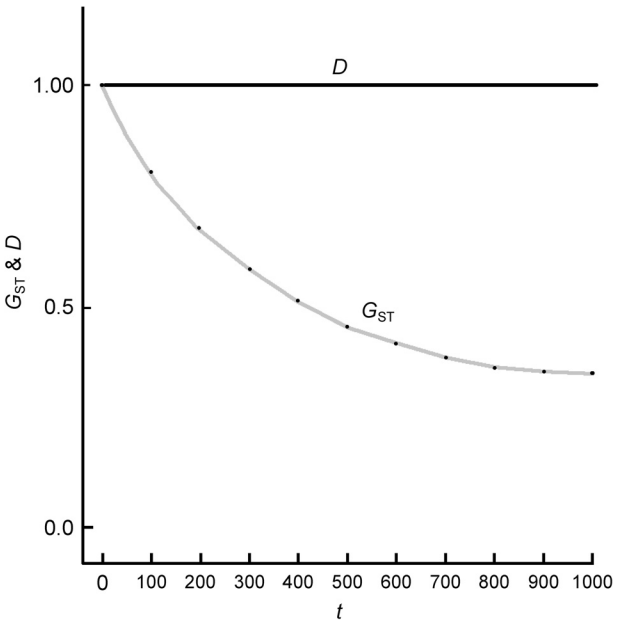

Fig. 1 Differentiation D and G_{ST} for ten demes under the finite island model with infinite alleles, using migration rate $m = 0$, mutation rate $\mu = 10^{-3}$, and number of individuals per deme $N = 1000$. Initially, at $t = 0$, a different (unique) allele is fixed in each deme. Mutation causes H_S to increase over time. Demes always remain completely differentiated, since there is no migration and since each mutation generates a novel allele. D correctly indicates complete differentiation at all times, while G_{ST} does not. Average of 40 runs.

Temporal evolution of *D*

Ryman and Leimar's Figs 1 and 2 show that *D* reaches its equilibrium value more slowly than G_{ST} under some circumstances (although both *D* and G_{ST} reach equilibrium much faster than their components H_S and H_T; see Crow 2004). This is a valuable observation, but again this is not a fault of *D*. Rather it is an accurate reflection of the time course of allelic differentiation in these scenarios. *D* always exactly tracks allelic differentiation, and its interpretation as differentiation is true at each moment in time. This interpretation does not depend upon any assumption about equilibrium. Of course *D*, like G_{ST}, cannot be used to estimate any parameters of the finite island model unless the population is in equilibrium, and for this kind of application the observation of Ryman and Leimar is important.

Ryman and Leimar's Fig. 3 shows the temporal evolution of *D* and G_{ST} for a system of 10 demes that are identical at $t = 0$, and then completely isolated. They graph two scenarios. The first scenario assumes the demes were in mutation–drift equilibrium at the time of isolation, and the other assumes the system had some bottleneck event that reduced the heterozygosity to a very low value at the time of isolation. The graph for the first scenario shows that *D*, unlike G_{ST}, is strikingly linear with respect to time following isolation. This shows that, under these conditions, *D* can be used to estimate divergence times over short timescales, just like Nei's (1972) genetic distance D_{Nei}. In fact, it can be shown that

$$D = 1 - e^{-D_{Nei}} \qquad \text{(eqn 2)}$$

Since $\exp(-x) \approx (1 - x)$ for small values of x, this implies that $D \approx D_{Nei}$ when D_{Nei} is small. Equation 2 reveals that the closely related concepts of allelic differentiation and genetic distance have a simple connection, providing a new interpretation of *D*. In contrast, G_{ST} is not monotonically related to genetic distance, another reason to reject it as a meaningful measure of differentiation.

Statistical issues

Some studies use G_{ST} primarily as a tool for generating a *P* value to test the null hypothesis of no differentiation. This approach focuses on the *P* value itself rather than the actual value of G_{ST}, whose actual magnitude is not given an interpretation. In this view, the often large differences observed by Heller & Siegismund (2008) between the magnitudes of G_{ST} and *D* (see their Fig. 1) may appear to be irrelevant, as long as either measure can be used to generate a *P* value. However, *P* values are not a substitute for real measures of effect size, and despite its popularity with researchers and journal editors, testing a null hypothesis is rarely the appropriate model in science (Morrison & Henkel 1969; Carver 1978; Gardner & Altman 1986; Yoccoz 1991; Johnson 1999; Anderson *et al*. 2000; Armstrong 2007). In natural populations, the null hypothesis of zero differentiation is virtually always false, and if sample size is large enough, this can be demonstrated with any desired degree of statistical significance. The important scientific question is the real magnitude of the differentiation, not the smallness of the *P* value (which confounds the magnitude of the effect with the sample size). Answering this question is a matter of parameter estimation, not hypothesis testing. In this approach, the final result should be an estimate of a meaningful measure of the magnitude of differentiation, accompanied by a confidence interval that describes the statistical uncertainty in this estimate. If the confidence interval includes zero, then the null hypothesis cannot be rejected. If the confidence interval does not include zero, then not only can we reject the null hypothesis, but we can have an idea of whether the real magnitude of the differentiation is large or small.

Obviously, this approach requires an interpretable measure of differentiation, such as *D*.

Perhaps the difficulty of interpreting the values of G_{ST} (and lack of familiarity with G'_{ST}) has forced geneticists to settle for mere testing of null hypotheses. The reverse may also be true: the widespread but misplaced emphasis on testing null hypotheses may have seemed to excuse researchers from the need to find meaningful, interpretable measure of differentiation. The introduction of *D* should make the parameter-estimation model more popular among geneticists. My collaborator Anne Chao and her students have written a free program, SPADE (Chao & Shen 2009), which calculates pairwise and global *D* and their confidence intervals. This program is aimed at ecologists estimating species diversity and community similarity and differentiation, but the same math applies to allele differentiation between subpopulations. Additional free software relating to *D* will be available at www.loujost.com. My collaborators and I are currently extending *D* to include multiple loci and molecular distance between alleles.

References

Anderson D, Burnham K, Thompson W (2000) Null hypothesis testing: problems, prevalence, and an alternative. *Journal of Wildlife Management*, **64**, 912–923.

Armstrong J (2007) Statistical significance tests are unnecessary even when properly done and properly interpreted: reply to commentaries. *International Journal of Forecasting*, **23**, 335–336.

Carver R (1978) The case against statistical significance testing. *Harvard Educational Review*, **48**, 378–399.

Chao A, Shen T-J (2009) *SPADE (Species Prediction and Diversity Estimation)*. http://chao.stat.nthu.edu.tw/softwareCE.html.

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

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

Gardner M, Altman D (1986) Confidence intervals rather than *P* values: estimation rather than hypothesis testing. *British Medical Journal*, **292**, 746–750.

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

Hedrick P (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.

Johnson D (1999) The insignificance of statistical significance testing. *Journal of Wildlife Management*, **63**, 763–772.

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

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

Jost L (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.
Jost L (2009) Mismeasuring biological diversity: response to Hoffmann and Hoffmann (2008). *Ecological Economics*, **68**, 925–928.
Morrison D, Henkel R (1969) Statistical tests reconsidered. *The American Sociologist*, **4**, 131–140.
Nei M (1972) Genetic distance between populations. *American Naturalist*, **106**, 283–292.
Ryman N, Leimar O (2009) G_{ST} is still a useful measure of differentiation-a comment on Jost's *D*. *Molecular Ecology*, **18**, 2084–2087.
Takahata N, Nei M (1984) F_{ST} and G_{ST} statistics in the finite island model. *Genetics*, **107**, 501–504.
Yoccoz N (1991) Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bulletin of the Ecological Society of America*, **72**, 106–111.

L.J. is interested in documenting, conserving, and explaining the biodiversity of the Andes. I am developing the mathematics of diversity in order to provide appropriate quantitative tools for these tasks. I put these tools to work by identifying areas of high biodiversity and endemism in Ecuador, and protecting the most important and distinctive of these through the foundation I co-founded, the EcoMinga Foundation (www.ecominga.net).

doi: 10.1111/j.1365-294X.2009.04186.x