

NEWS AND COMMENTARY

Gems from the *Heredity* archive

Splendor and misery of indirect measures of migration and gene flow

G Guillot

Heredity (2011) 106, 11–12; doi:10.1038/hdy.2010.55; published online 12 May 2010

When F_{ST} ceases being approximately equal to $1/(1 + 4Nm)$

Genetic variation among populations created by drift, mutations and selection is counteracted by migration. Wright (1931) in his seminal work derived a famous formula that relates approximately the amount of genetic variation among populations (as measured by the F statistic or F_{ST}) to the migration rate m : $F_{ST} \approx 1/(1 + 4N_e m)$, where N_e is the common effective size of populations. Together with the advent of molecular techniques, this formula opened the door to a simple and inexpensive way to quantify gene flow among populations through estimated F_{ST} . In a timely review, Whitlock and McCauley (1999) reminded us that an estimate is worth just as much as the model that has been used to obtain it, and that the assumptions underlying Wright's model are likely to be violated by most natural populations. Indeed, Wright's model assumed the existence of an infinite number of populations having a common and constant effective population size, that migrations occur at random at a common and constant rate among populations and that there is no mutation and no selection. Whitlock and McCauley (1999) noted that mutation might not have a crucial role in shaping population structure and focused their comments on the effect of selection, of asymmetries between populations and other departures from homogeneity in space and time. These aspects have been the object of active research in the last 10 years.

Islands with selection

Directional selection acting differentially among populations tends to increase F_{ST} values compared to the prediction of Wright's formula whereas balancing selection tends to decrease them (Nielsen, 2005; Holsinger and

Weir, 2009). In this context, all quantities involved in Wright's formula still make sense and in principle, the formula could be fixed to encompass the effect of selection. Actually, this fix would be of little value for estimating migration rate, as allele fitness is typically unknown. However, the detection of loci under selection has lately become an active area of research and another strategy can be undertaken. Indeed, increasingly sophisticated methods allow one to test the hypothesis of genetic neutrality of each locus (Vitalis *et al.*, 2001; Beaumont and Balding, 2004; Foll and Gaggiotti, 2008; Riebler *et al.*, 2008; Excoffier *et al.*, 2009). These methods have all the same flavor: they compare estimated F_{ST} 's to quantiles of the F_{ST} distribution expected under the null hypothesis of neutrality. See also Nielsen (2001) for alternative approaches. The recent work by Excoffier *et al.* (2009) stresses that an incorrect model (for example, ignoring hierarchical structure) may lead to incorrect attribution of outlier loci to the effects of selection. However, providing that the model assumed is correct, a valid approach to indirectly estimating the rate of migration consists in detecting genetically neutral markers by one of the above-mentioned methods and then resorting to Wright's formula.

More complex scenarios

Whitlock and McCauley (1999) pointed out that Wright's formula relies also on the assumption that populations have equal and constant size and exchange migrants at the same rate, in particular, the migration rate does not depend on the geographical distance separating populations. In this situation, using Wright's formula requires one to restrict estimation of F_{ST} and m to pairs of populations. This is carried out at the price of an increased estimation variance. Narrowing the data set to make it fit a model seems to be a brute

force strategy. A more efficient option consists in enriching the model to make it suitable for the whole data set. This has been made possible by the advent of computational statistical methods that allow one to estimate parameters under evolutionary scenarios that are more general than Wright's island model (Rousset, 1997; Beerli and Felsenstein, 2001; Lopes *et al.*, 2009). The latter is perhaps the most advanced method in this context. It is aimed at jointly estimating branching histories and population genetic parameters (effective population sizes, migration rates, mutation and recombination rates) for known panmictic populations in the isolation with migration model (Nielsen and Wakeley, 2001). The key idea is to assess the plausibility of a given scenario (that is, a set of parameters) by looking at how often simulation according to this scenario produces patterns similar to those displayed in data at hand. This task was considered prohibitive only 15 years ago and is now made possible by exploiting two key ideas: simulating data by coalescent and summarizing information contained in data by summary statistics (Beaumont *et al.*, 2002; Cornuet *et al.*, 2008).

... and $\hat{\theta} \neq \text{Argmax}_{\theta} L(\theta; \mathbf{x})$

The report that $F_{ST} \neq 1/(1 + 4N_e m)$ sounded like bad news, in particular because it did not provide straightforward solutions to the issues pointed out. However it became widely cited, presumably because of its cathartic power, but also because it corresponded to a change in the way people look at evolutionary biology. Around the corner were new, more realistic models than Wright's island model. Besides, detecting outlier F_{ST} values has been the key to the first methods of detecting loci under selection, showing that the model did have its uses. Ironically enough, the statistical methods by which the issues reported by Whitlock and McCauley have been solved have been mostly Bayesian (Beaumont and Balding, 2004), striking a blow to the likelihood dogma promoted almost 100 years ago by Fisher (1922), the other giant of evolutionary biology.

Conflict of interest

The author declares no conflict of interest.

Acknowledgements

Dr G Guillot is currently supported by Agence Nationale de la Recherche (ANR-09-BLAN-0145-01).

G Guillot is at the Department of Informatics and Mathematical Modelling, Technical University of Denmark, Richard Petersens Plads, Bygning 305, 2800 Lyngby, Copenhagen, Denmark.

e-mail: gigu@imm.dtu.dk

- Beaumont M, Balding D 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol* **13**: 969–980.
- Beaumont M, Zhang W, Balding D 2002. Approximate Bayesian computation in population genetics. *Genetics* **162**: 2025–2035.
- Beerli P, Felsenstein J 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc Natl Acad Sci USA* **98**: 4563–4568.
- Cornuet J, Santos F, Beaumont M, Robert C, Marin J, Balding D *et al.* 2008. Inferring population history with DIYABC: a user-friendly approach to Approximate Bayesian Computation. *Bioinformatics* **24**: 2713–2719.
- Excoffier L, Hofer T, Foll M 2009. Detecting loci under selection in a hierarchically structured population. *Heredity* **103**: 285–298.
- Fisher R 1922. On the mathematical foundations of theoretical statistics. *Philos Trans R Soc Lond A* **222**: 309–368.
- Foll M, Gaggiotti O 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**: 977–993.
- Holsinger K, Weir B 2009. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST} . *Nat Rev Genet* **10**: 639–650.
- Lopes J, Balding D, Beaumont MA 2009. PopABC: a program to infer historical demographic parameters. *Bioinformatics* **25**: 2747–2749.
- Nielsen R 2001. Statistical tests of neutrality at the age of genomics. *Heredity* **86**: 641–647.
- Nielsen R 2005. Molecular signatures of natural selection. *Ann Rev Genet* **39**: 197–218.
- Nielsen R, Wakeley J 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**: 885–896.
- Riebler A, Held L, Stephan W 2008. Bayesian variable selection for detecting adaptive genomic differences among populations. *Genetics* **178**: 1817–1829.
- Rousset F 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219–1228.
- Vitalis R, Dawson K, Boursot P 2001. Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**: 1811–1823.
- Whitlock M, McCauley D 1999. Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. *Heredity* **82**: 117–125.
- Wright S 1931. Evolution in Mendelian populations. *Genetics* **16**: 97–159.

Editor's suggested reading

- Larmuseau MHD, Raeymaekers JAM, Hellemans B, Van Houdt JKJ, Volckaert FAM (2010). Mitonuclear discordance in the degree of population differentiation in a marine goby. *Heredity* **105**: 532–542.
- Lachish S, Miller KJ, Storer A, Goldizen AW, Jones ME (2011). Evidence that disease-induced population decline changes genetic structure and alters dispersal patterns in the Tasmanian devil. *Heredity* **106**: 172–182.
- Guillemaud T, Beaumont MA, Ciosi M, Cornuet J-M, Estoup A. (2010). Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity* **104**: 88–99.
- Hermisson J (2009). Who believes in whole-genome scans for selection? *Heredity* **103**: 283–284.