

Supplementary text: The use and meaning of F -statistics

F -statistics are commonly used to estimate and summarize to what degree alleles are able to unite at random (Hedrick, 2000; Templeton, 2006). One can also think of it as summarizing how identity by descent or variance is affected by population structure and mating behavior (Templeton, 2006). Phenomena that influences the association of alleles are in/outbreeding and divisions into subpopulations (infrapopulations in this study) and populations. Four F -statistics describe the direction and/or magnitude of these divisions. The value of F_{IS} indicates the correlation between alleles in an individual, relative to allele frequencies in the infrapopulation. A value of zero suggests random union within the infrapopulation whereas positive and negative values respectively indicate that alleles in one individual are more or less likely to be identical, compared to a random union expectation for the infrapopulation. F_{SC} measures the degree to which alleles within an infrapopulation are more similar to each other than would be expected if the infrapopulations were composed randomly from alleles from the population. F_{CT} measures the degree to which alleles within a population are more similar to each other than would be expected if the population was composed randomly from alleles from all the populations together. F_{IT} measures the similarity of alleles in an individual if mating was random over the entire range, i.e. ignoring all lower levels of structuring.

The estimation and subsequent interpretation of F -statistics are however mired with complications (Whitlock and McCauley, 1999; Templeton, 2006; Meirmans and Hedrick 2011). The combined effects of F_{SC} and F_{CT} are summarized by F_{ST} , with $(1 - F_{SC})(1 - F_{CT}) = 1 - F_{ST}$. First, for two alleles F_{ST} has a maximum of one and results and interpretations can be compared between studies. However, as the number of alleles increases, the maximum value decreases and studies cannot be compared (Hedrick 1999). This is the case with markers such as microsatellites and Meirmans and Hedrick (2011) suggested corrections for these F -statistics to address this problem. Second, Wang (2015) illustrated that when demographic

processes that drives the structuring of alleles, such as drift and migration, outweighs the rate of mutation, then F -statistics can be trusted. On the other hand, when mutations outweigh demographic processes F -statistics can be misleading and reflect mutational processes rather than demographics. Wang (2015) illustrated that a significant negative correlation between the F -statistic of a level and the heterozygosity observed at that level of population structuring indicates that mutations have eroded the signal in F -statistics.

References

- Hedrick, P.W., 1999. Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* 53, 313-318.
- Hedrick, P.W., 2000. *Genetics of populations* 2nd ed. Jones and Bartlett Publishers, Sudbury.
- Meirmans, P.G., Hedrick, P.W. 2011. Assessing population structure: F_{ST} and related measures. *Molec. Ecol. Res.* 11, 5-18.
- Templeton, AR., 2006, *Population genetics and microevolutionary theory*. John Wiley and Sons, Inc., Hoboken.
- Wang, J., 2015. Does G_{ST} underestimate genetic differentiation from marker data? *Molec. Ecol.* 24, 3546-3558.
- Whitlock, M.C., McCauley, D.E., 1999. Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. *Heredity* 82, 117-125.