

## gIMble info documentation

gIMble info reports standard population genetic measures of divergence ( $d_{xy}$ ) and diversity ( $\pi$ ,  $H$ ) both for heterospecific pair blocks (dataset X) and for blocks defined for intraspecific sample pairs (datasets A and B). Since all gIMble parameter estimates are based on heterospecific pair blocks (dataset X) under the assumption that populations are panmictic, it is helpful to check for the effect of population structure and inbreeding by comparing average heterozygosity across individuals ( $H$ ) to genetic diversity ( $\pi$ ) estimated across all pairwise comparisons within species (Tajima, 1983, , eq. A3) ( $\pi$  for datasets A and B) which is computed as:

$$\pi_A = \sum_{i=1}^{n_A} (n_{hetA1,i} + n_{hetA2,i} + 4(n_{hetA12,i} + n_{fix,i})) / (6 * l * n_A)$$

, where  $n_{hetA1,i}$   $n_{hetA2,i}$  are the number of heterozygous sites unique to the two individuals 1 and 2 contributing to pair block  $i$  and  $n_{hetA12,i}$  are shared heterozygous sites.

For heterospecific pair blocks (dataset X), genetic diversity can only be measured using the mean heterozygosity (per base)  $H$ , which is computed as the average sum of  $n_{hetA}$  and  $n_{hetAB}$  counts across all  $n_X$  heterospecific blocks. E.g. for taxon A:

$$H_A = \sum_{i=1}^{n_X} (n_{hetA,i} + n_{hetAB,i}) / (l * n_X)$$

, where  $l$  is the block length (the calculation is analogous for B). Mean pairwise divergence  $d_{XY}$  (Nei, 1987, eq. 10.20) is computed as the average number of pairwise difference between A and B. Since we are averaging over the two haplotypes each individual contributes to a pair block  $i$ , the three types of heterozygous sites contribute 1/2:

$$d_{XY} = \sum_{i=1}^{n_X} (n_{fix,i} + (n_{hetA,i} + n_{hetB,i} + n_{hetAB,i})/2) / (l * n_X)$$

Given the above, we can compute  $F_{ST}$  between A and B (Slatkin, 1991, eq. 26) as:

$$F_{ST} = (d_{XY} - \overline{H}) / (d_{XY} + \overline{H})$$

, where  $\overline{H} = (H_A + H_B)/2$

## References

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

- M. Slatkin. Inbreeding coefficients and coalescence times. *Genetics Research*, 58(2):167–175, 1991.  
doi: 10.1017/S0016672300029827.
- F. Tajima. Evolutionary relationship of DNA sequences in finite populations. *Genetics*, 105(2):437–460, 1983. ISSN 0016-6731.