

Tags : #Redaction In RIDGE, there is two set of summary statistics that are used : the basic set, which is used to describe a single locus and an extended set which is use to describe an ensemble of loci. The basic set first :

- **bialsite** : number of bi allele sites in the locus
- **Fst** : Fst estimation using the method of Hudson (1992) elaborated by Bhatia et al. (2013).
- **divAB** : Estimate nucleotide divergence between two populations (dxy) within a given region, which is the average proportion of sites (including monomorphic sites not present in the data) that differ between randomly chosen pairs of chromosomes, one from each population.
- **netDivAB** : nucleotide divergence without ancestral polymorphism (Da). The ancestral polymorphism is asume to mean the mean of population polymorphism.  $Da = Dxy - (A+B)/2$
- **piA** , **piB** : Estimate nucleotide diversity within a given region, which is the average proportion of sites (including monomorphic sites not present in the data) that differ between randomly chosen pairs of chromosomes.
- **thetaA** , **thetaB** : Estimate nucleotide diversity with watterson theta
- **DtajA**, **DtajB** : Tajima's D
- **ss** : proportion of polymorphism shared between population
- **sf** : proportion of polymorphism with fixed difference between population
- **sxA**, **sxB** : proportion of polymorphism specific to a population

To describe the ensemble of loci, the mean (avg) , the standard error (std) and the median (median) is calculated for each previously described summary statistics. In addition new variable describe : - **pearson\_r\_pi** : pearson correlation between the pi of each population. - **pearson\_r\_theta** : pearson correlation between the Watterson's theta of each population. - **pearson\_r\_divAB\_netdivAB** : pearson correlation between the dxy and Da - **pearson\_r\_divAB\_FST**: pearson correlation between dxy and FST - **pearson\_r\_netdivAB\_FST**: pearson correlation between Da and FST - **ss\_sf** : number of loci which contain shared polymorphism and fixed difference - **ss\_noSf** : number of loci which contain shared polymorphism and ==no== fixed difference - **noSs\_sf** : number of loci which contain ==no== shared polymorphism and fixed difference - **noSs\_noSf** : number of loci which contain ==no== shared polymorphism and ==no== fixed difference

All the computation part is done by **scikit-allel** and **numpy** packages.

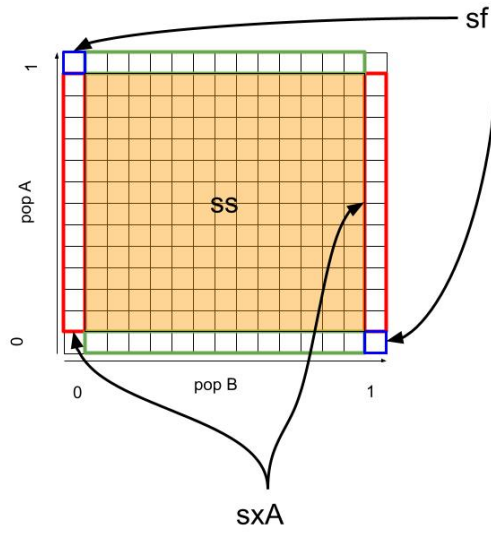


Figure 1: **Fig 1** :Joint sfs spectrum representation with explanation on where to find *ss*(central orange zone), *sf* (blue zone), *sxA*(red zones) and *sxB* (green zones). To calculate them, it is the sum of the number of SNP inside the zone and divide by the total number of SNP.