**Sliding window for genetic diversity and differentiation**

**Tutorials**

<https://speciationgenomics.github.io/sliding_windows/>

<https://www-users.york.ac.uk/~dj757/popgenomics/workshop5.html>

<https://eacooper400.github.io/gen8900/exercises/slide-window.html>

**About input data**

*“****Dxy and pi require monomorphic sites to be present*** *in the dataset, whereas* ***FST and fd are only computed on bi-allelic sites****.*

*Thus, it is important to filter out indels and multi-allelic sites and to keep monomorphic sites (no maf filter).” --from* [*https://speciationgenomics.github.io/*](https://speciationgenomics.github.io/)

**Your decision**

1. Perform BCFtools call for all sites (true SNPs + invariants), all congeneric species, and at per main chromosome.

Command for all site call (array job):

##-2.1- Set up array job input

echo $SLURM\_ARRAY\_TASK\_ID

**CHROM=**$(awk -v lineid=$SLURM\_ARRAY\_TASK\_ID 'NR==lineid{print;exit}' $OUTDIR/Geum\_CHRname.txt) ### This is a list containing all main chromosome names

##-2.2- Run BCFtools in arrary

time bcftools mpileup -a AD,DP,SP -Q 30 -q 30 -Ou -f $REF -b $bamList -r **"$CHROM"** | bcftools call **-m** -f GQ,GP -Oz -o $OUTDIR/2401118\_Geum**\_"$CHROM"**.vcf.gz

1. Filtration

* **Don’t** filter for QUAL (variant quality score) because you will lose all invariant sites
* **Don’t** filter for maf or mac (it is mentioned in one tutorial, see above), but I don’t know why…)
* **Do filter** quite **stringently on DP**: ideal score is > 30, but for yours, use > 8 or > 10. Let gimbleprep do this.

P.S. gimbleprep default:

minDP > 8

maxDP < 2\*Genome-wide\_mean\_depth\_per\_sample

* **Do filter** strictly to **keep bi-allelic sites only**. If you are using pixy program, you don’t need to do this.
* **Do filter** quite stringently on missingness per site. Have seen people use keeping completeness > 0.8. You can try 0.9 or 1.