

Pipeline for “Recombination suppression and selection affect local ancestries in genomes of a migratory songbird”

Local PCA

Here I show how local PCA was run with three example chromosomes, 12, 20, and 21, which are listed in `local_PCA/list/chromosomes/list`

```
dirbase=$PWD/local_PCA
dirin=$dirbase/input
dirvcf=$dirin/vcf
dirtab=$dirin/table
dirlist=$dirbase/list
dirout=$dirbase/output
dirlocalpca=$dirout/local_PCA
dirlocalpcamds=$dirout/local_PCA_MDS
dirscripts=$dirbase/scripts
```

Download filtered VCF files for chromosomes 12, 20, 21.

Make list of sites from VCF.

```
while read chr
do
    bcftools query -f '%POS\n' $dirvcf/$chr.vcf.gz > $dirtab/$chr.sites.list
done <$dirlist/chromosomes.list
```

Make genotype table from VCF.

```
while read chr
do
    sbatch $dirscripts/vcf2table.sh $dirvcf/$chr.vcf.gz $dirtab/$chr.table
done <$dirlist/chromosomes.list
```

Run local PCA. `local_PCA/scripts/local_PCA.sh` submits `local_PCA/scripts/local_PCA.R` via slurm. Check `local_PCA/scripts/local_PCA.sh` and `local_PCA/scripts/local_PCA.R` for detail.

```
module load R/3.5.3
```

```
while read chr
do
    sbatch $dirscripts/local_PCA.sh $dirtab $dirtab $dirlocalpca $dirlocalpcamds $dirlist
```

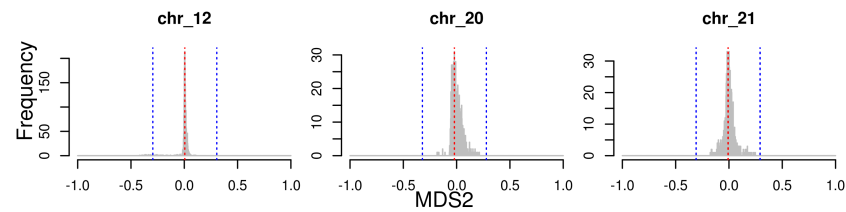
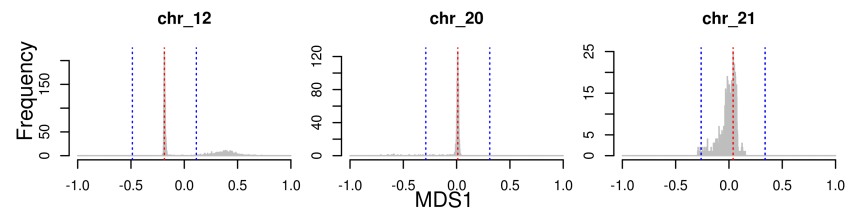
```
done <$dirlist/chromosomes.list
```

Concatenate the output for the three chromosomes.

```
while read chr
do
    cat $dirlocalpcams/local_PCA_MDS_$chr.txt
done <$dirlist/chromosomes.list | awk 'NR==1{print $0}NR>1{if($1!="chr")print $0}' > $dirlo
```

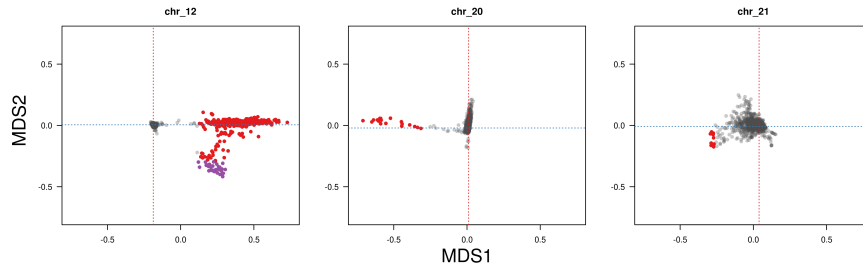
Get threshold of MDS values and plot MDS distribution.

```
Rscript $dirsripts/plot_local_PCA_MDS_distribution.R --input $dirlocalpcams/local_PCA_MDS_
```



Plot MDS1 vs MDS2.

```
Rscript $dirsripts/plot_local_PCA_MDS_MDS1-vs-MDS2.R --input $dirlocalpcams/local_PCA_MDS_
```

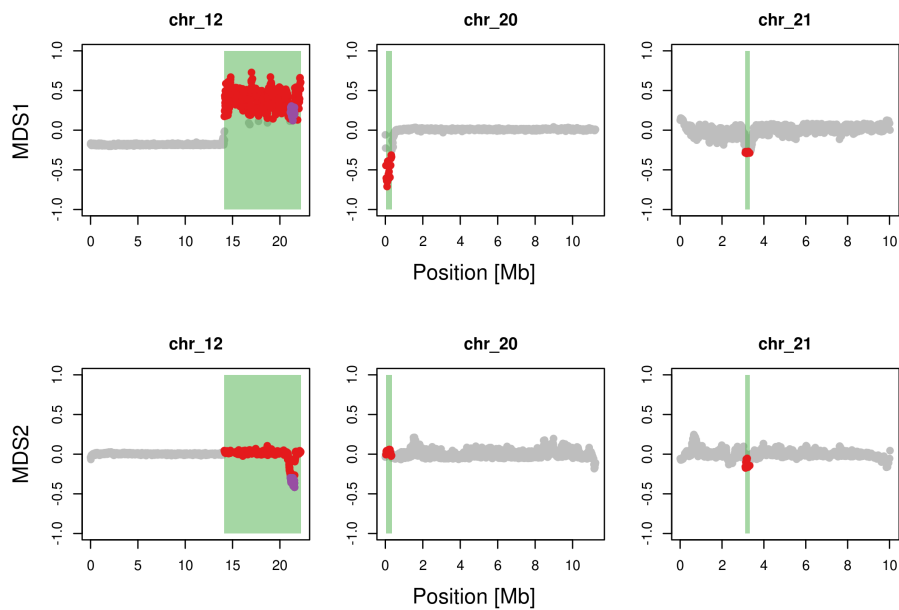


Get coordinates of outlier windows.

```
Rscript $dirscripts/getOutliers.R --input $dirlocalpcamds/local_PCA_MDS_3_chromosomes.txt --
```

Make Manhattan plots.

```
Rscript $dirscripts/plot_local_PCA_MDS_manhattan.R --input $dirlocalpcamds/local_PCA_MDS_3_c
```



PCA in local PCA outliers

Here I show how PCA was run for three local PCA outlier regions of chromosomes 12, 20, and 21, which represent class-1, 2, and 3 outliers.

```
dirbase=$PWD/PCA
dirin=$dirbase/input
dirvcf=$dirin/vcf
```

```

dirlist=$dirbase/list
dirout=$dirbase/output
dirscripts=$dirbase/scripts

```

Link VCF files from local_PCA/input/vcf and index them.

```
ln $dirbase/./local_PCA/input/vcf/*vcf.gz $dirvcf
```

```

while read chr
do
    echo $chr
    bcftools index $dirvcf/$chr.vcf.gz
done<$dirlist/chromosomes.list

```

Copy BED file for coordinates of local PCA outliers of chromosomes 12, 20, and 21.

```
ln $dirbase/./local_PCA/output/local_PCA_MDS/local_PCA_MDS_outlier.bed $dirlist
```

Extract SNPs within the local PCA outlier regions.

```

while read chr from to
do
    bcftools view -O z \n
        -r $chr:$from-$to \n
        -S $dirlist/blackcap_id.list \n
        $dirvcf/$chr.vcf.gz \n
        | vcftools --gzvcf - \n
        --max-missing 0.9 \n
        --recode \n
        --recode-INFO-all -c \n
        |bgzip > $dirvcf/${chr}_${from}_$to.vcf.gz
done < $dirlist/local_PCA_MDS_outlier.bed

```

Run PLINK for PCA.

```

while read chr from to
do
    sbatch $dirscripts/plink_pca.sh \n
        $dirvcf/${chr}_${from}_$to \n
        $dirout/${chr}_${from}_$to
done < $dirlist/local_PCA_MDS_outlier.bed

```

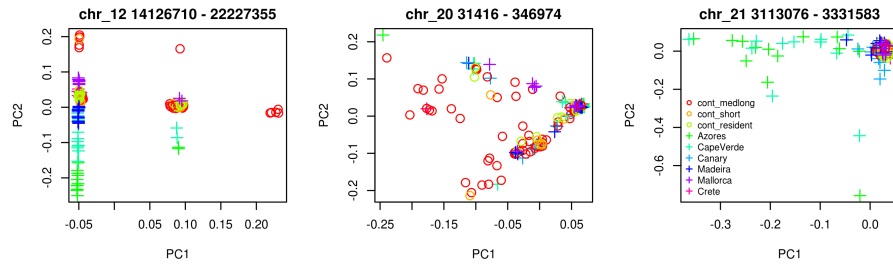
Plot PCA results.

```

Rscript $dirscripts/plot_pca_per_outlier.R --dirpca $dirout \n
    --poplist $dirlist/id_spp_pop_site_pheno.tsv \n

```

```
--outlierlist $dirlist/local_PCA_MDS_outlier.bed \n
--chrlist $dirlist/chromosomes.list
```

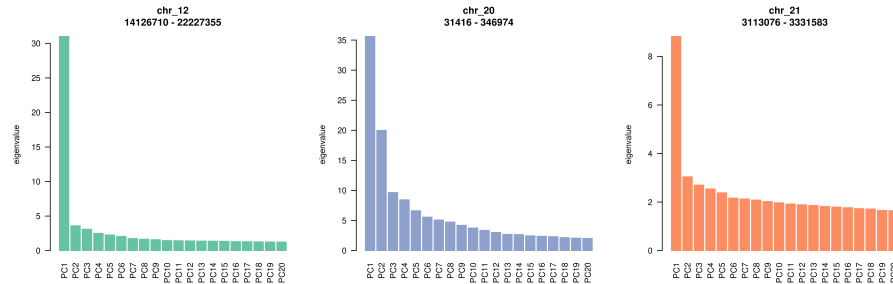


Get Eigenvalues.

```
while read chr pos1 pos2
do
    awk -v chr=$chr \n
        -v pos1=$pos1 \n
        -v pos2=$pos2 \n
        'BEGIN{printf "%s %s %s ", chr,pos1,pos2}{printf "%s ", $1}END{print ""}' \n
        $dirout/${chr}_${pos1}_${pos2}.eigenval
done < $dirlist/local_PCA_MDS_outlier.bed \n
| awk 'NR==1{printf "%s %s %s ", "chr", "pos1", "pos2";for(i=4;i<=NF;i++){printf "%s ", "PC"i-3}}
> $dirout/eigenvalues.txt
```

Plot Eigenvalues.

```
Rscript $dirsripts/plot_eigenval.R --dirpca $dirout
```



Class-1 genomic islands

```
dirbase=$PWD/class-1
dirin=$dirbase/input
dirvcf=$dirin/vcf
dirlist=$dirbase/list
dirout=$dirbase/output
dirscripts=$dirbase/scripts
```

Heterozygosity

Make link to VCF of chromosome 12, as an example chromosome harbouring a class-1 genomic island.

```
ln $dirbase/../../PCA/input/vcf/chr_12.vcf.gz $dirvcf
ln $dirbase/../../PCA/input/vcf/chr_12.vcf.gz.csi $dirvcf
```

Get coordinates of class-1 genomic island of chromosome 12.

```
awk -v OFS="\t" ' $1=="chr_12"{ $1=$1;print $0}' $dirbase/../../local_PCA/output/local_PCA_MDS/1
> $dirlist/class-1.chr_12.bed
```

Make list of individuals with AA, AB, and BB.

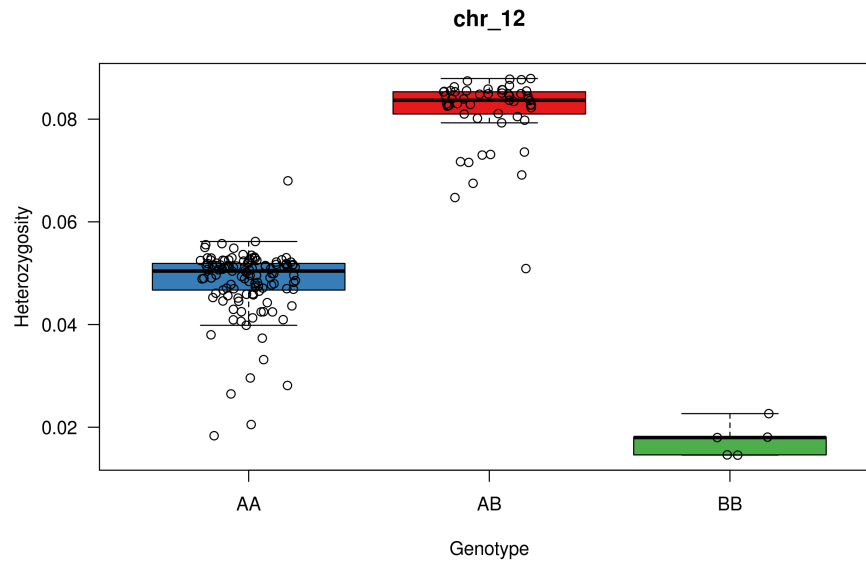
```
awk '$3<0{print $1}' $dirbase/../../PCA/output/chr_12_14126710_22227355.eigenvec > $dirlist/chr
awk '$3>0.15{print $1}' $dirbase/../../PCA/output/chr_12_14126710_22227355.eigenvec > $dirlist/
awk '$3>0&&$3<0.15{print $1}' $dirbase/../../PCA/output/chr_12_14126710_22227355.eigenvec > $d
```

Compute heterozygosity.

```
while read chr pos1 pos2
do
    for geno in AA AB BB
    do
        while read id
        do
            # echo $id $chr $geno
            bcftools query -f '[%GT ]\n' -r $chr:$pos1-$pos2 -s $id $dirvcf/$chr
            done<$dirlist/${chr}.$geno.list
        done
    done
done < $dirlist/class-1.chr_12.bed | awk 'BEGIN{print "chr","geno","id","n.het","n.sites","n"
```

Plot heterozygosity for AA, AB, and BB.

```
Rscript $dirsripts/plot_het.R --dir $dirout
```



FST, dXY and