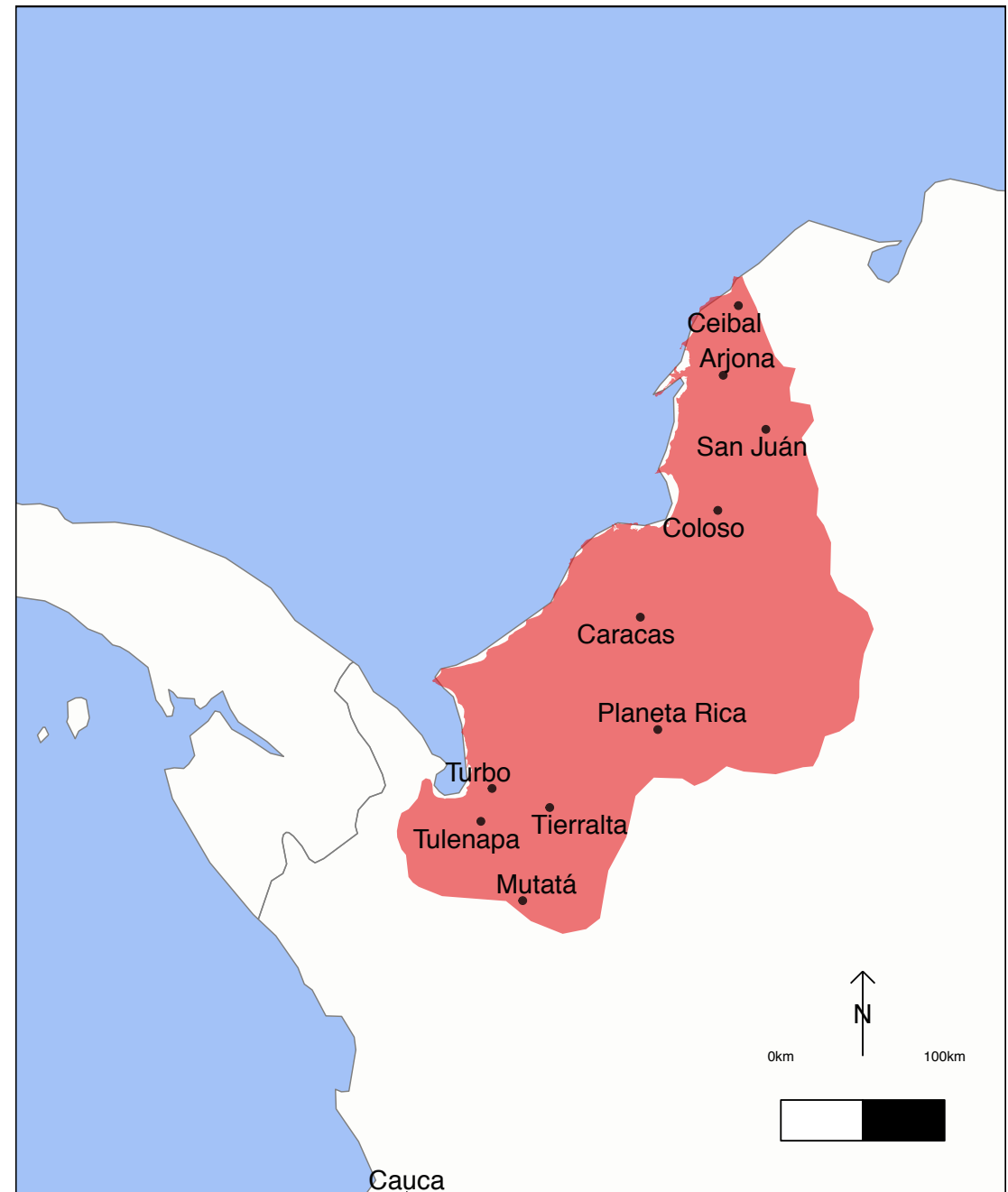# Cotton-Top Tamarins

Linett Rasmussen
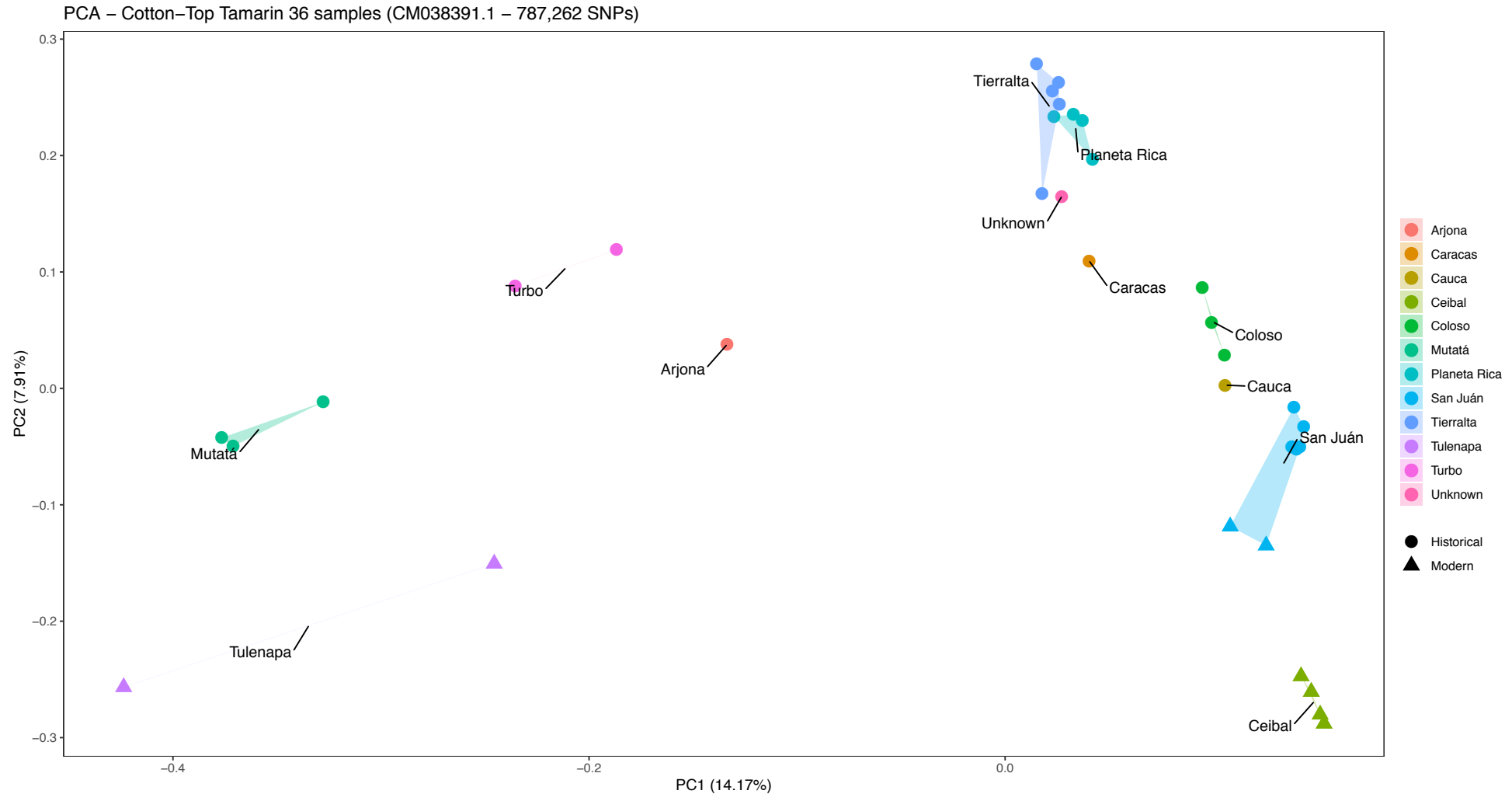
Claudia Fontsere

Map
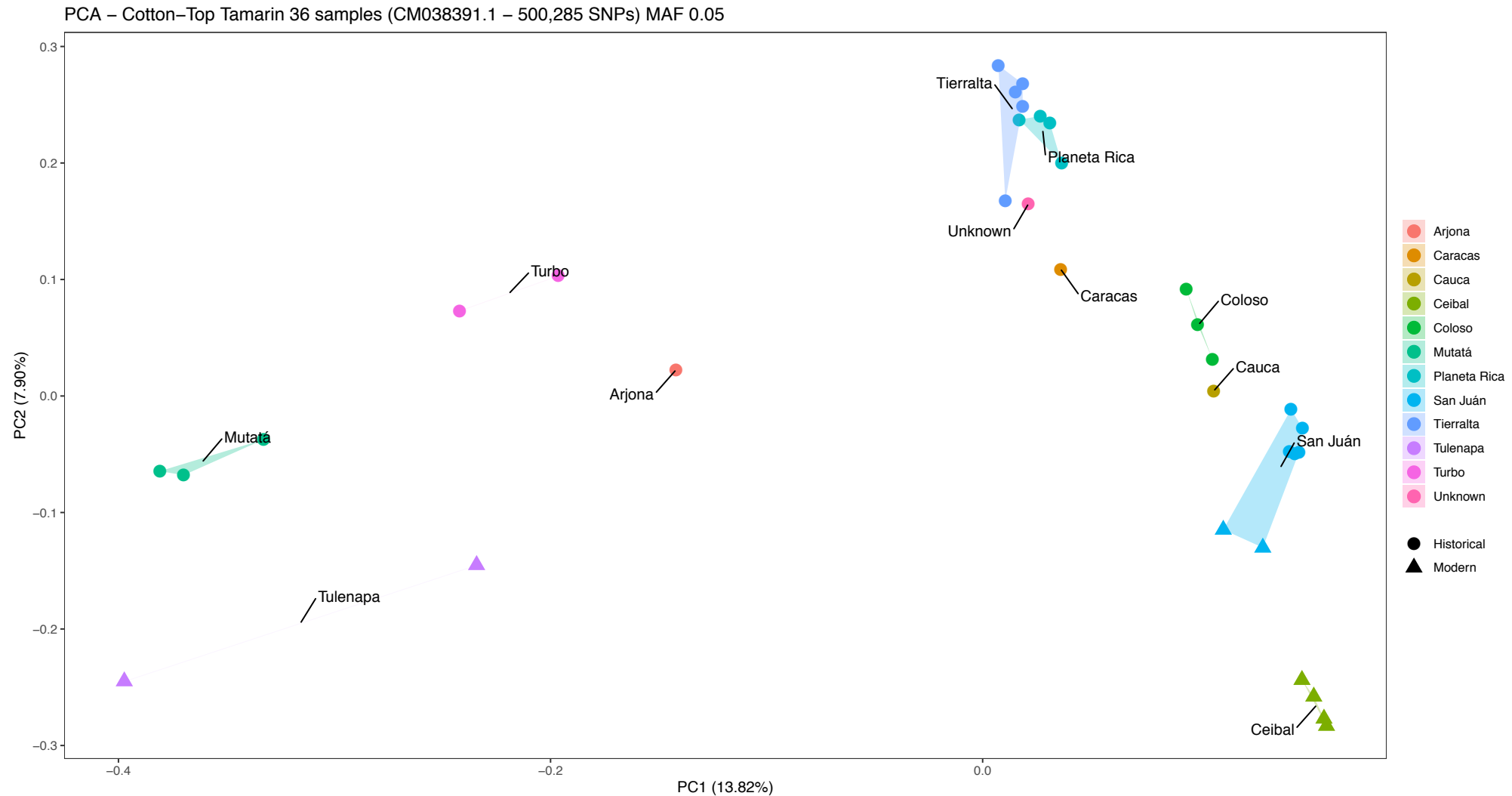
PCA using ANGSD with different parameters

-maf 0.01

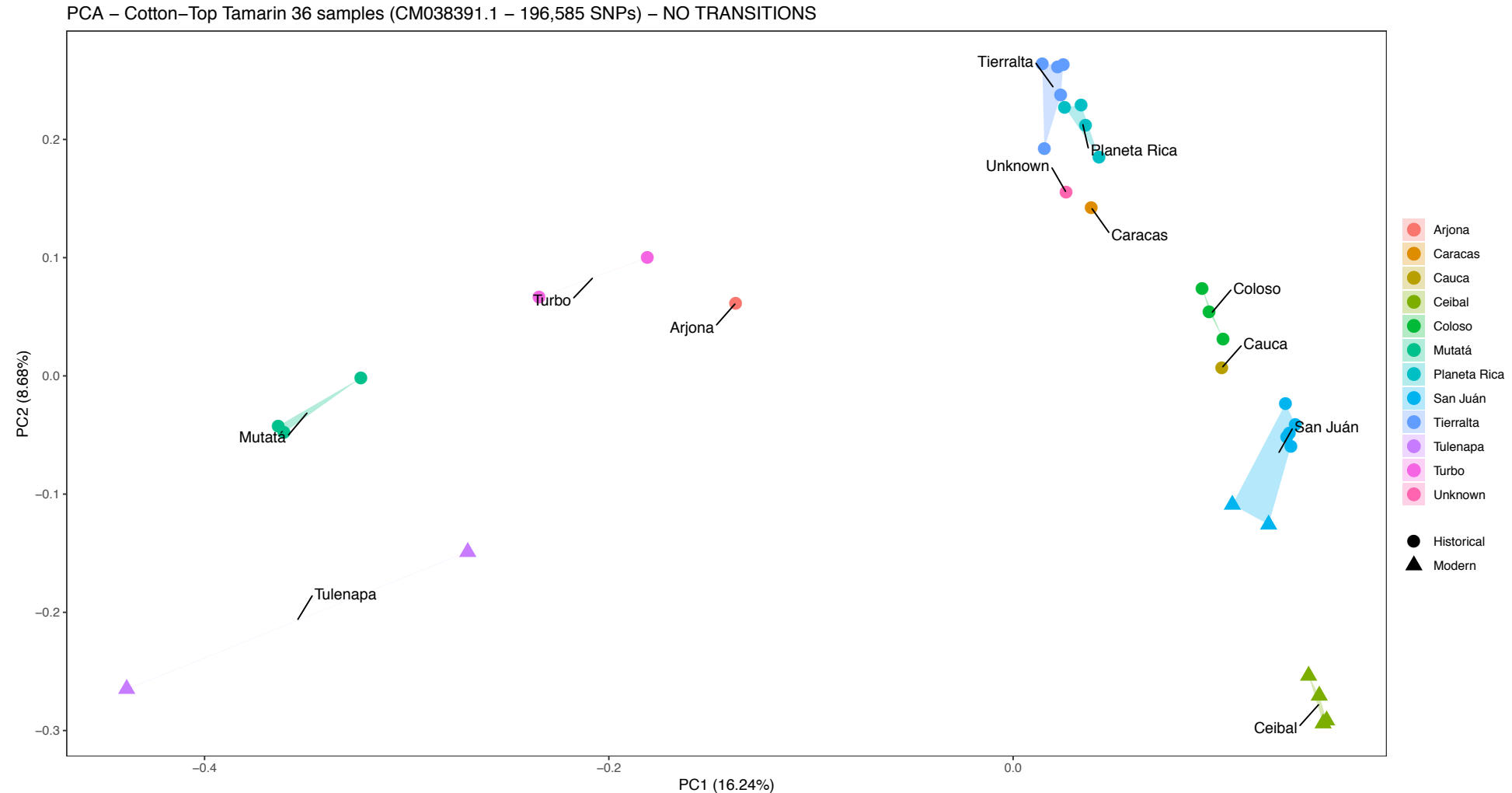PCA – Cotton–Top Tamarin 36 samples (CM038391.1 – 787,262 SNPs)

PCA using ANGSD with different parameters

-maf 0.05

PCA – Cotton–Top Tamarin 36 samples (CM038391.1 – 500,285 SNPs) MAF 0.05

PCA using ANGSD with different parameters

-maf 0.01 -rmTrans 1

PCA – Cotton–Top Tamarin 36 samples (CM038391.1 – 196,585 SNPs) – NO TRANSITIONS

PCA using ANGSD with different parameters

-maf 0.01 -trim 10
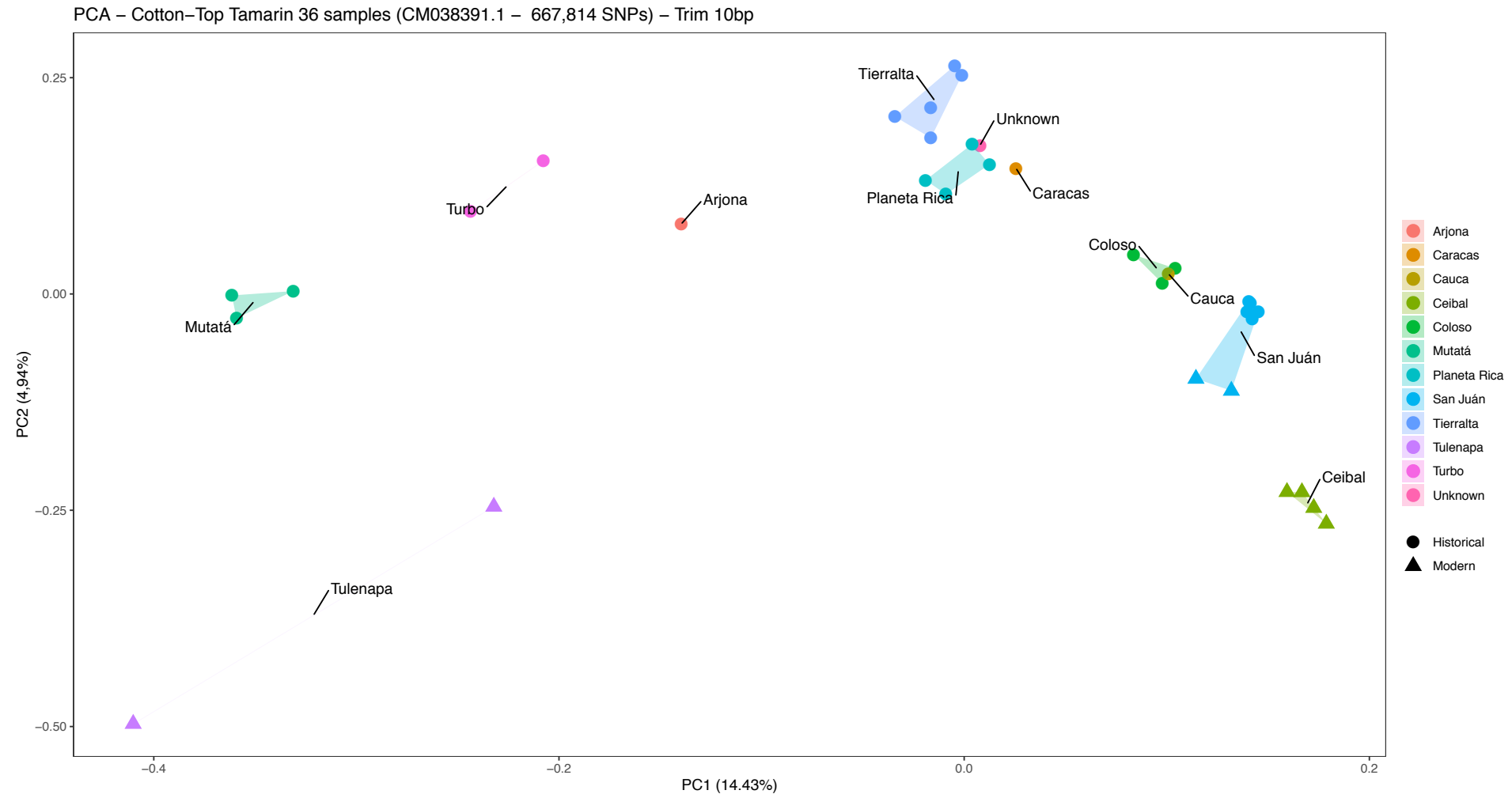
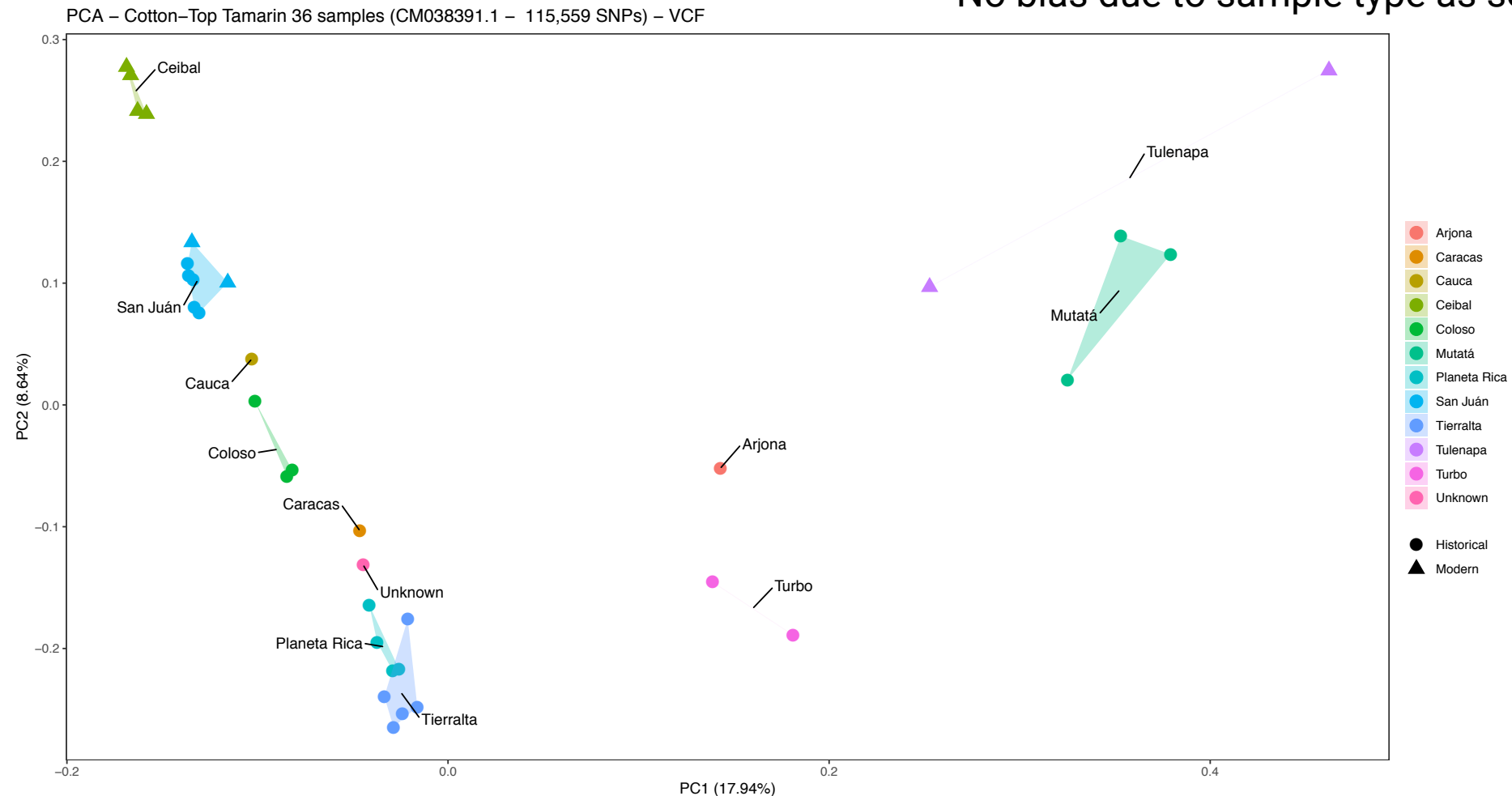
PCA – Cotton–Top Tamarin 36 samples (CM038391.1 – 667,814 SNPs) – Trim 10bp

PCA using SNP dataset (snpAD)

Vcftools --max-alleles 2 --max-missing 0.7 --minDP 3 --maxDP 50 --minQ 30 --remove-indels
Plink –pca -maf 0.05

No bias due to sample type as seen in San Juán

PCA – Cotton–Top Tamarin 36 samples (CM038391.1 – 115,559 SNPs) – VCF

PCA using SNP dataset (snpAD)

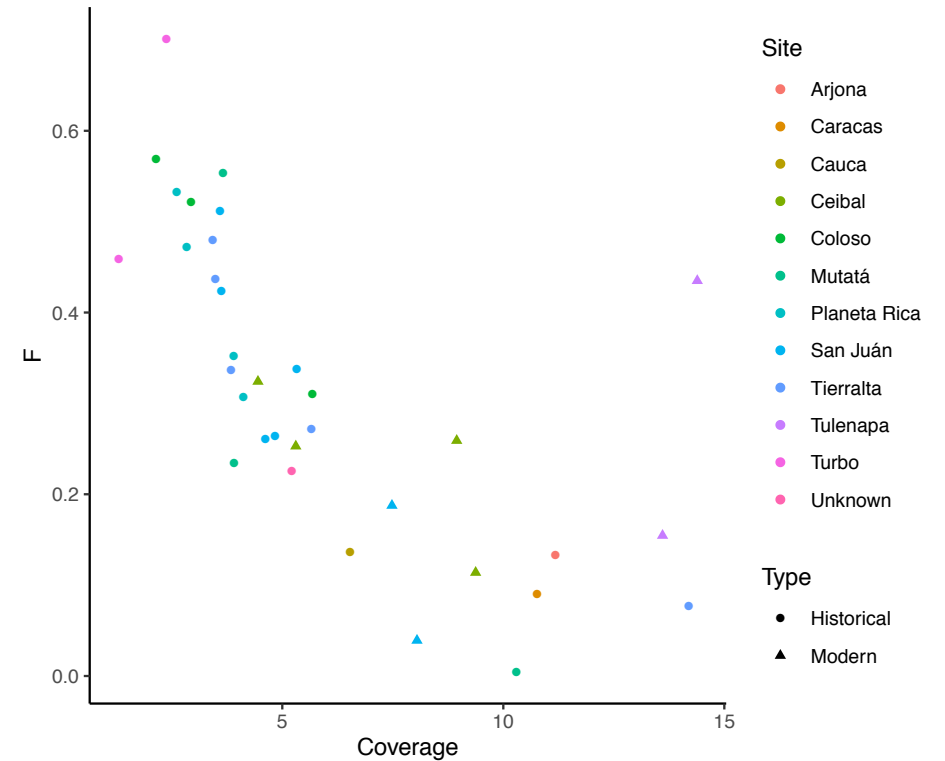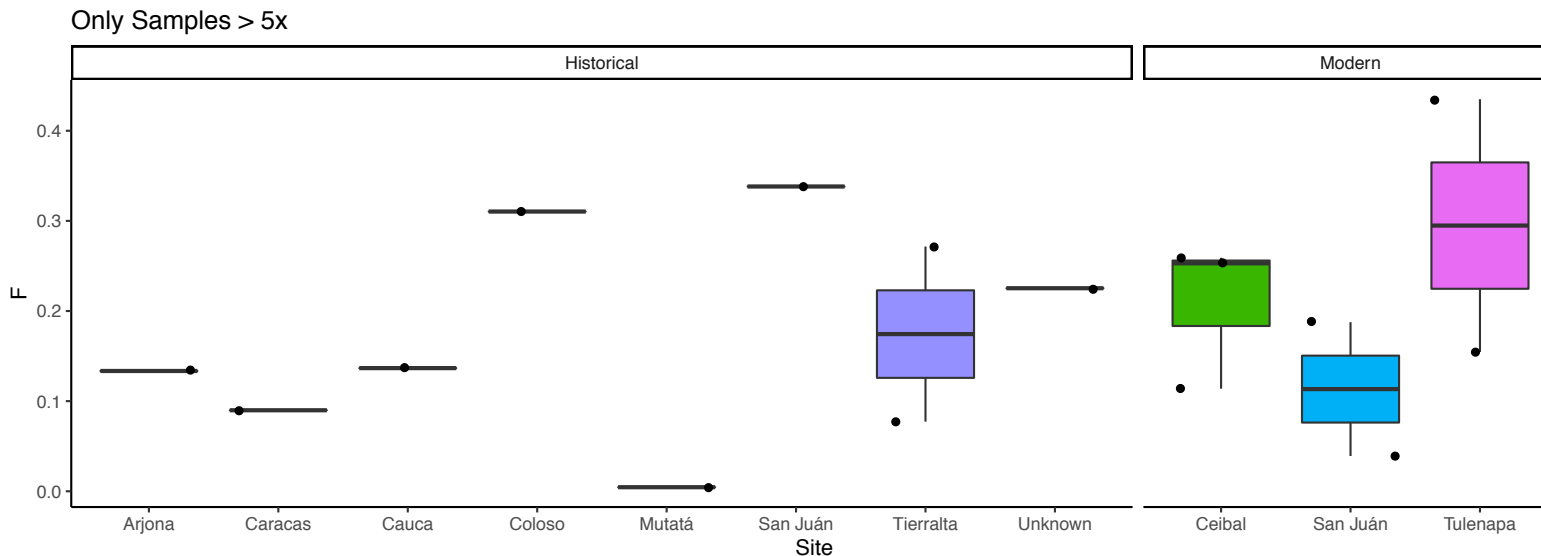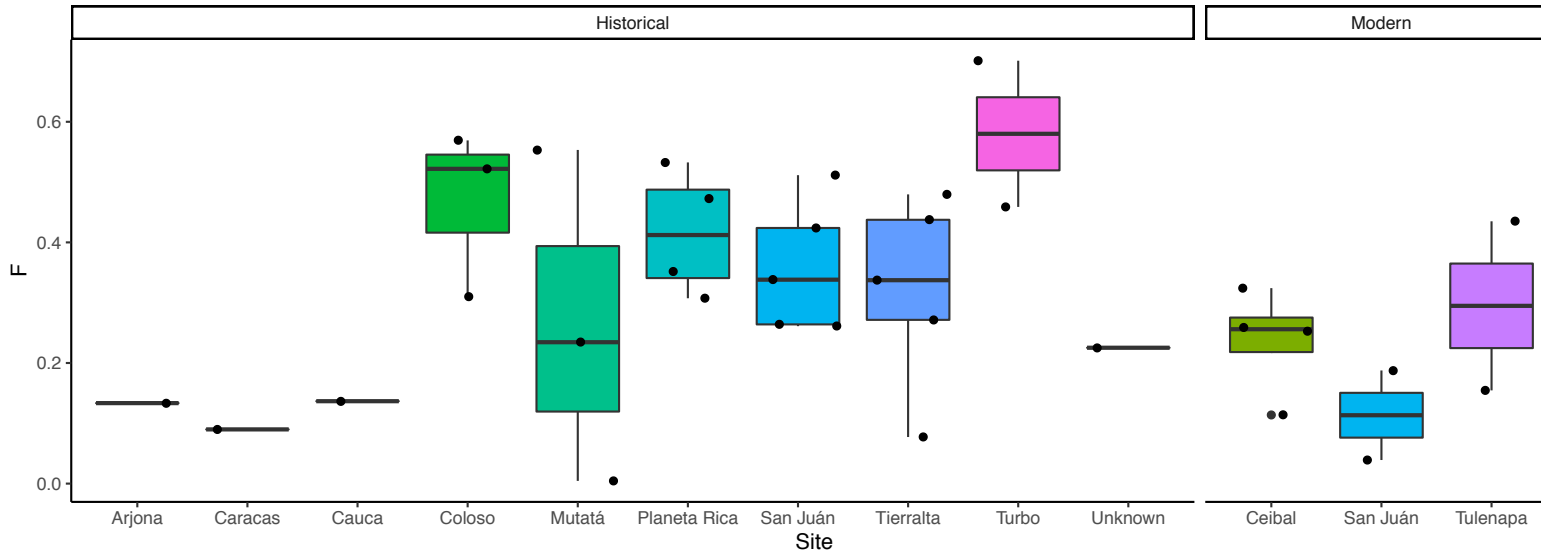Vcftools --max-alleles 2 --max-missing 0.7 --minDP 3 --maxDP 50 --minQ 30 --remove-indels
Projected PCA (PC built with historical and Modern projected) Plink –pca -maf 0.05 --pca-cluster-names
Historical --within clusters

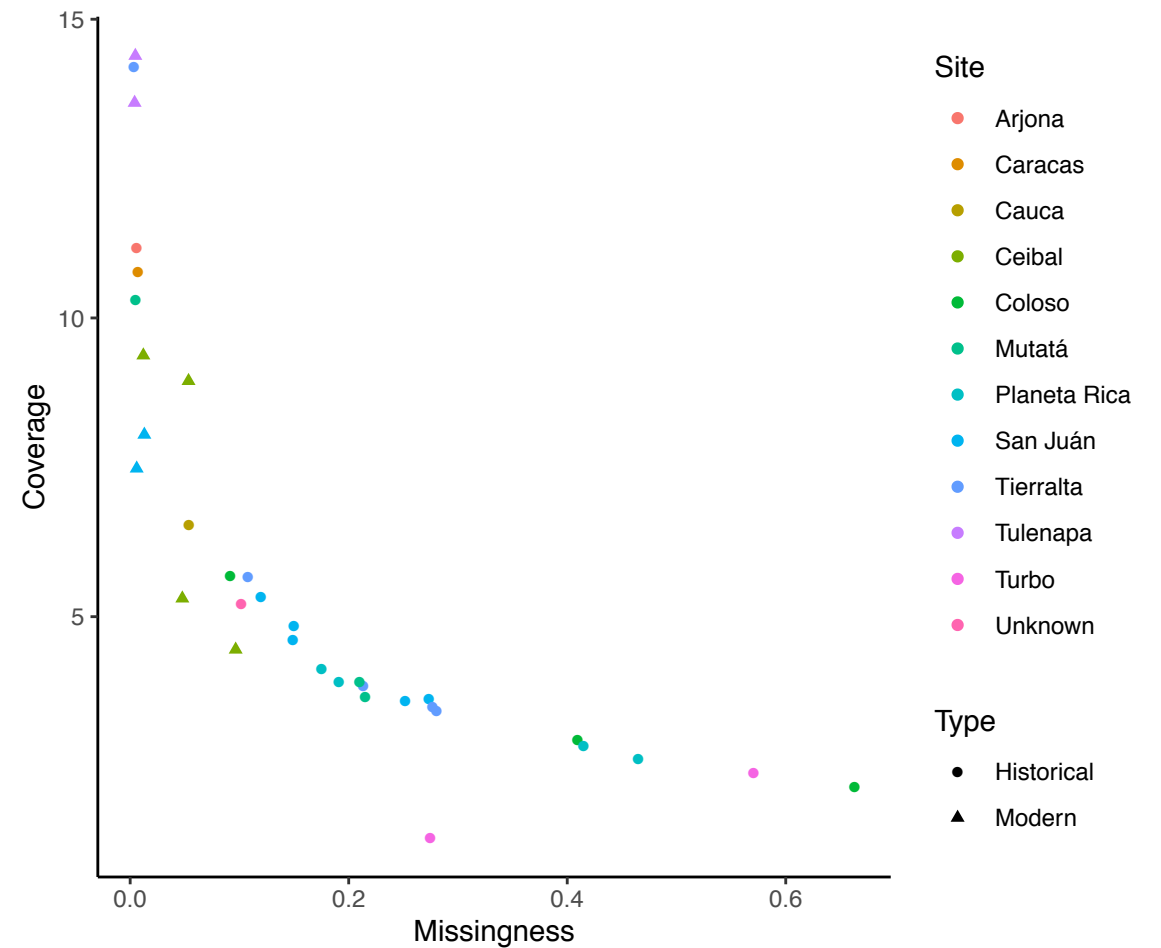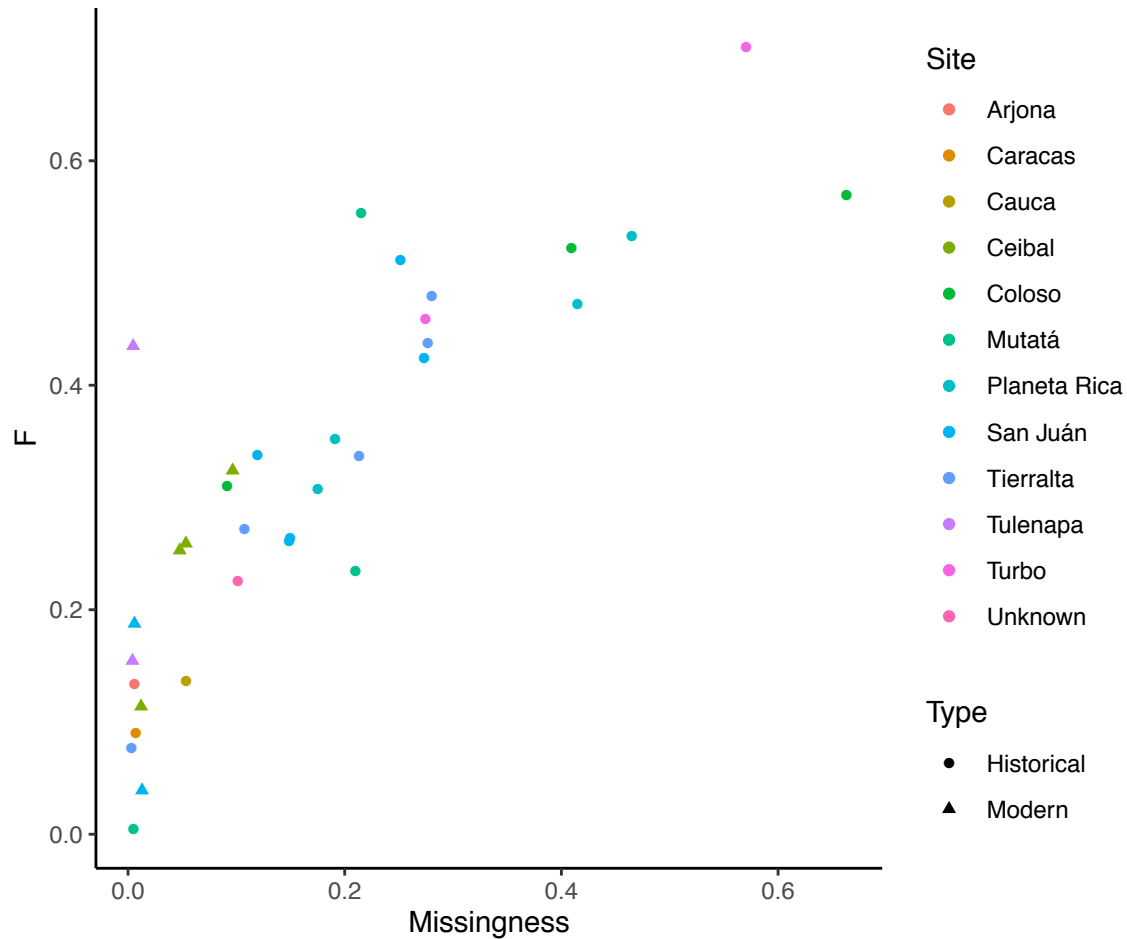PCA – Cotton–Top Tamarin 36 samples (CM038391.1 – 115,559 SNPs) – VCF
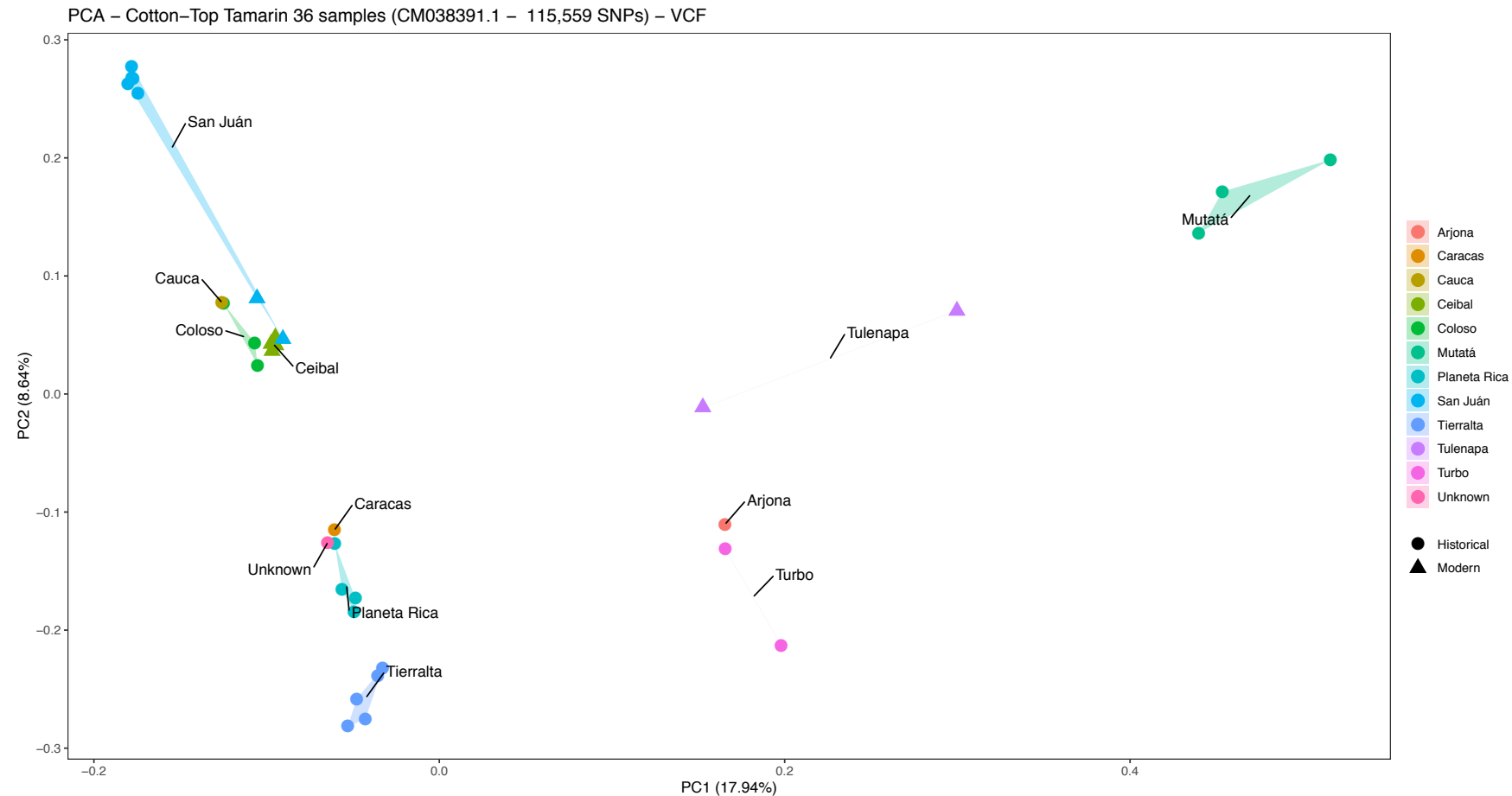
F inbreeding with --het (snpAD)

-maf 0.05 --max-alleles 2 --max-missing 0.7 --minDP 3 --maxDP 50 --minQ 30 --remove-indels
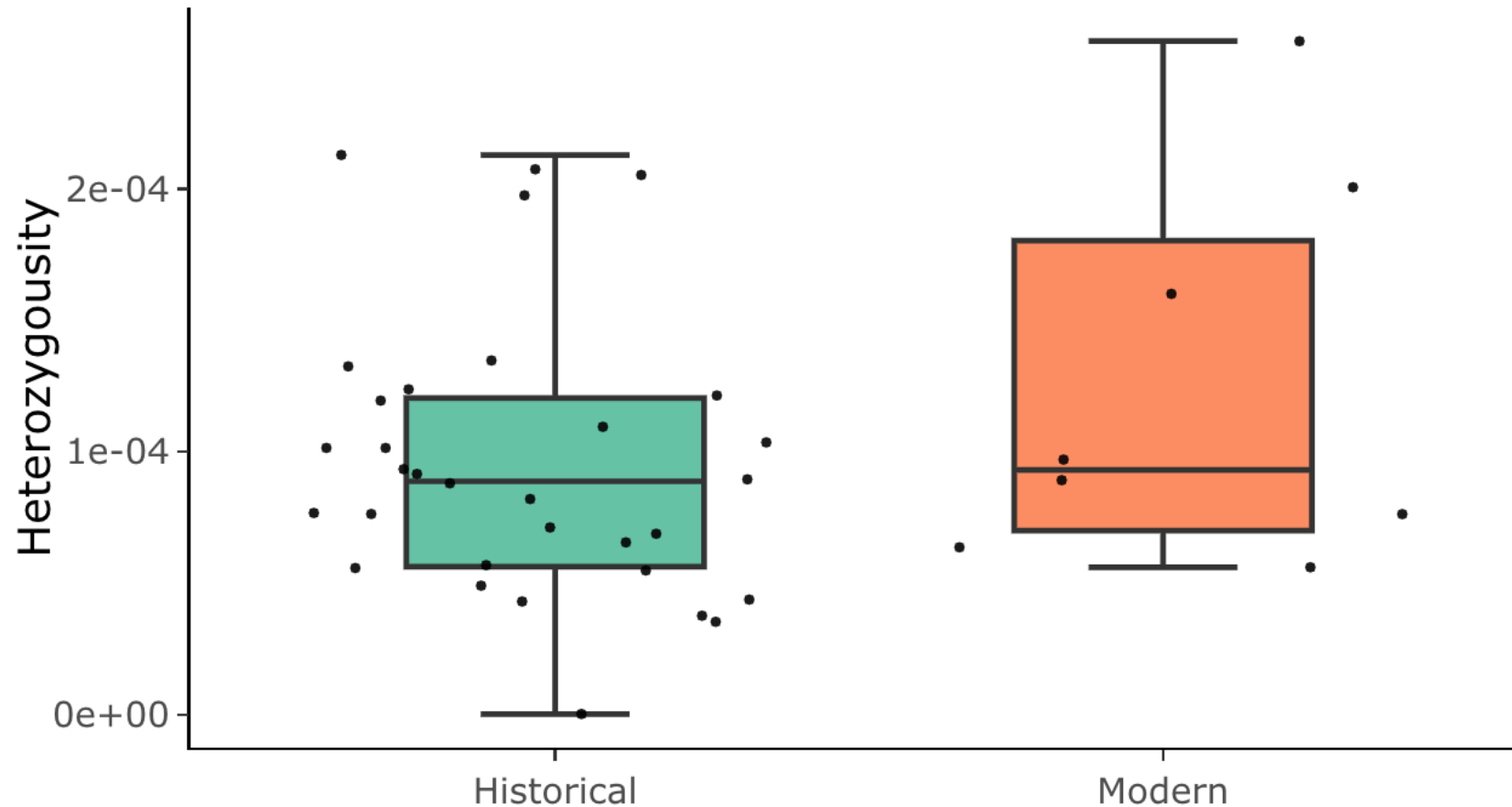

We should downsample to the same coverage to compare the values of F

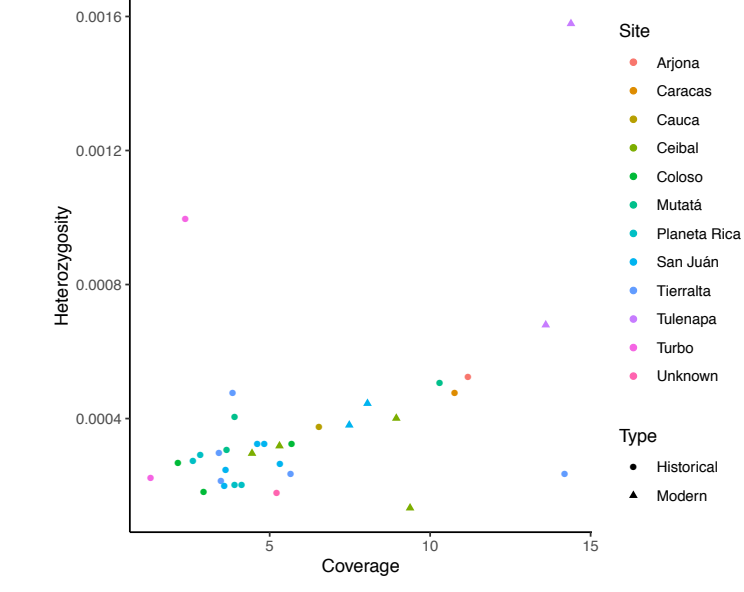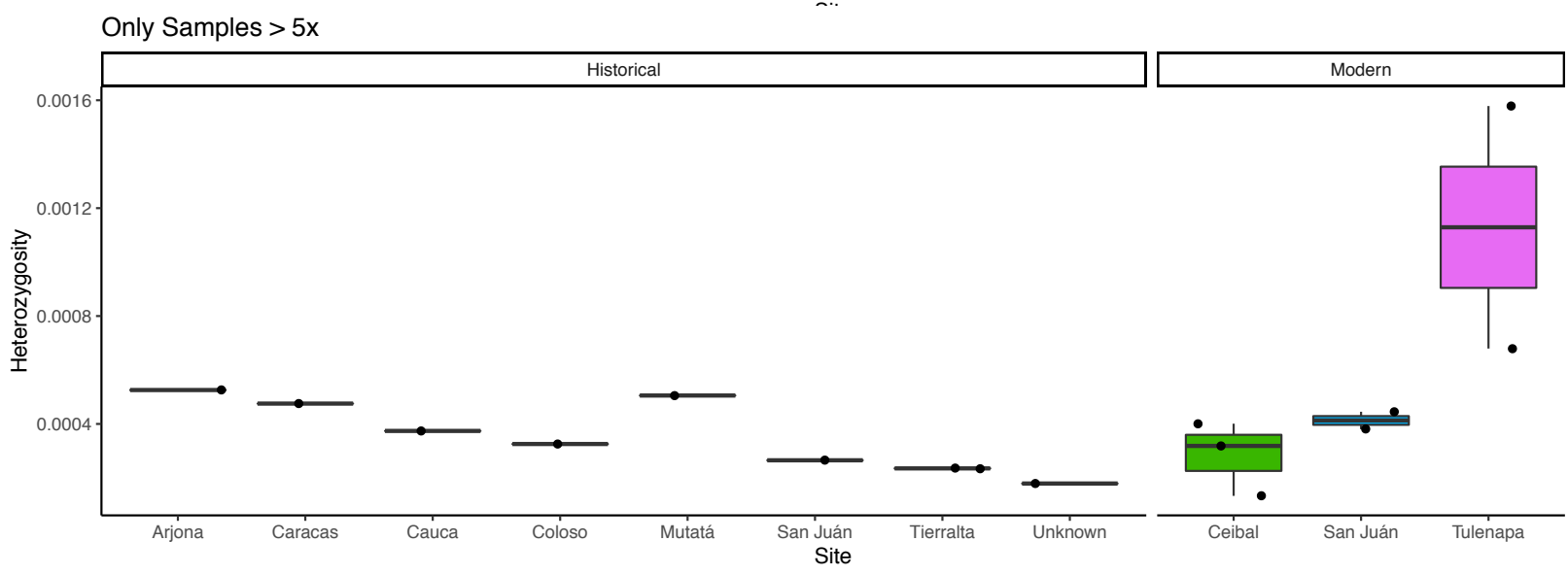# Inbreeding, coverage and missingness
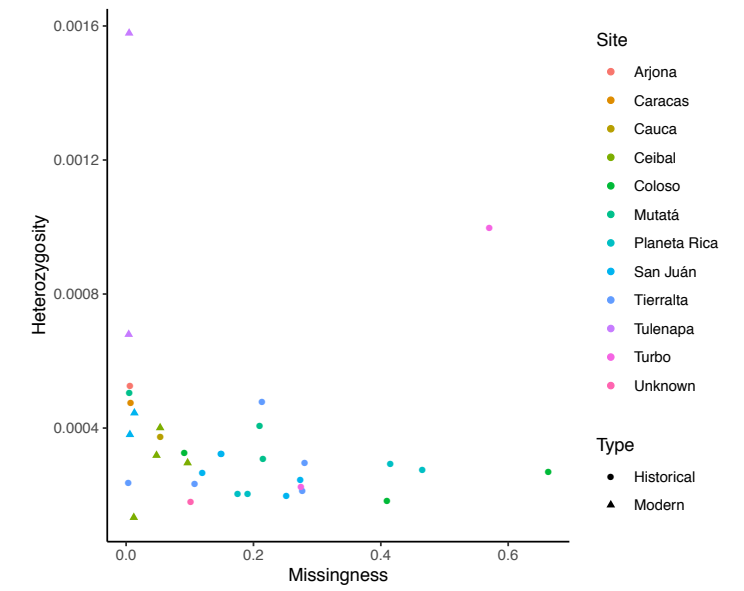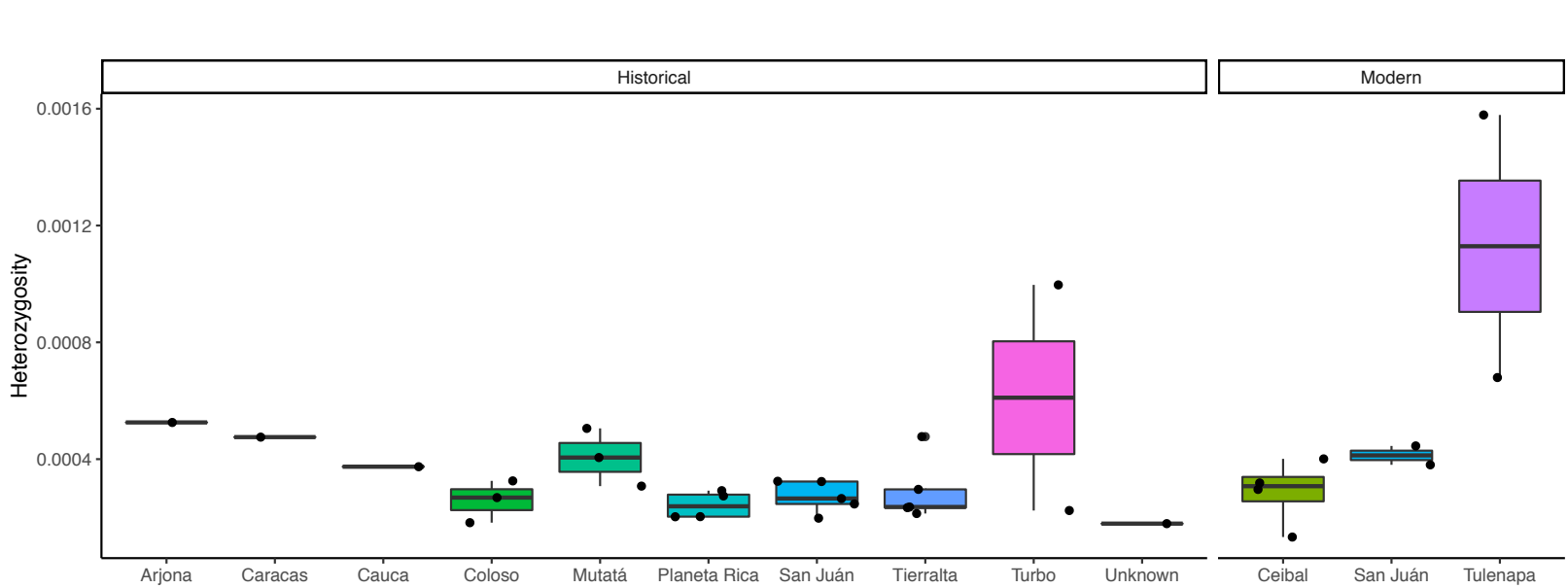
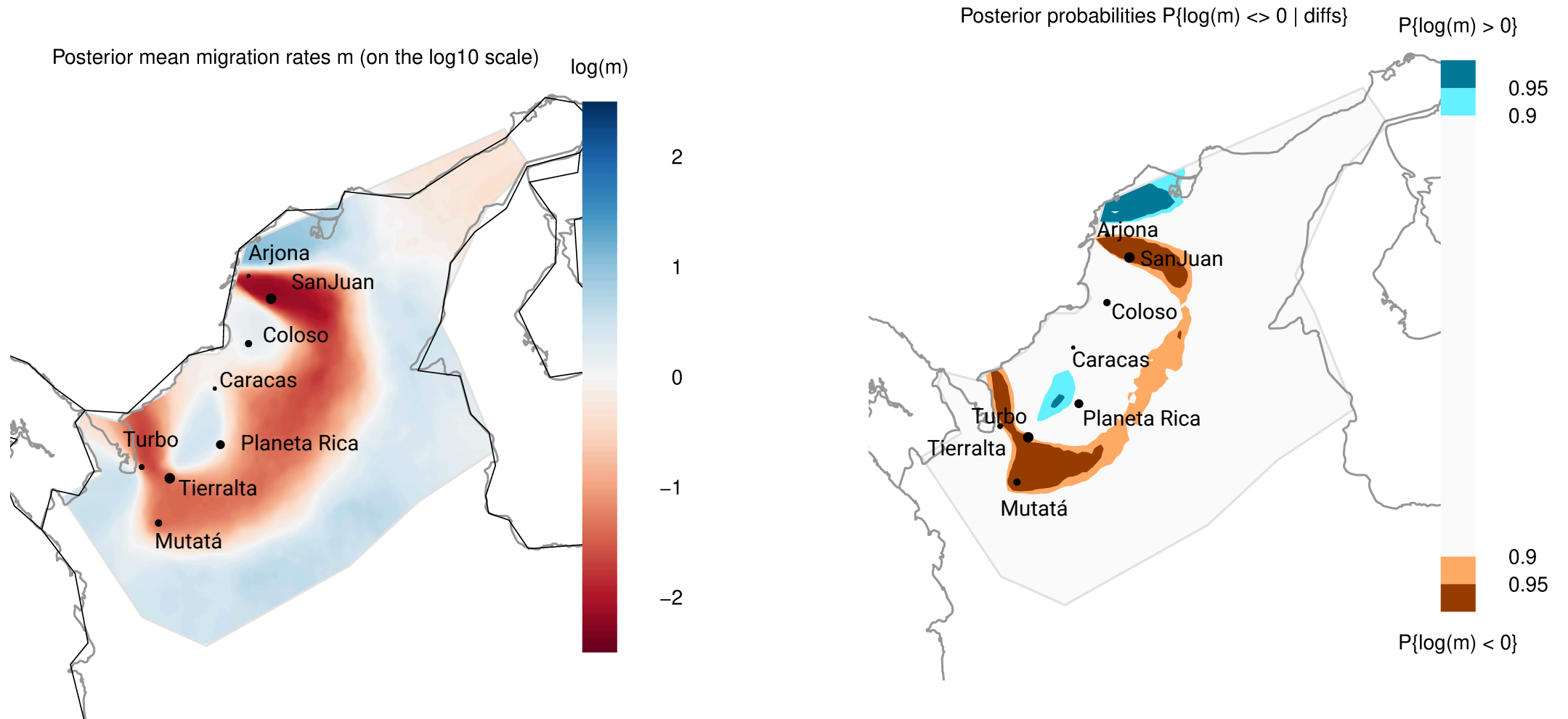# Heterozygosity (ANGSD) by Linett

Probable effect of coverage
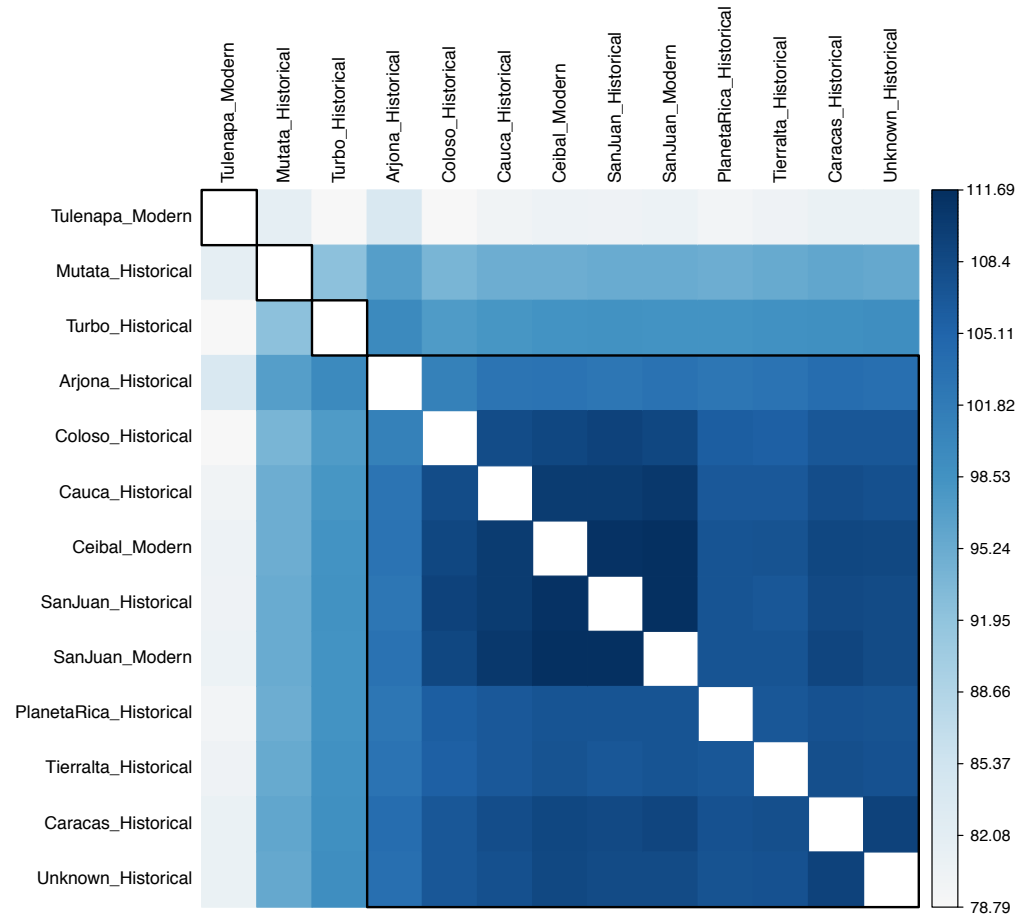
Heterozygosity (snpAD) by Claudia

EEMS only historical (excluded Cauca + Unknown)

-maf 0.05 --minGQ 20 --max-missing 0.8 --max-alleles 2 --minDP 3 --maxDP 50 --minQ 30 --remove-indels

Posterior mean migration rates m (on the log10 scale)

Posterior probabilities P{log(m) <> 0 | diffs}

F3 outgroup

--maf 0.00001 --minGQ 20 --max-missing 0.8 --max-alleles 2 --minDP 3 --maxDP 50 --minQ 30 --remove-indels

As outgroup I am using the Reference (it is the same species? To which pop it is more similar?) Consider adding a diferent individual