**Pipeline → Natural Selection**

[**Data Information:**](#_7aonoskcc42w) **5**

[CEDAR:](#_exz5qeotgjix) 5

[GRAHAN:](#_mhaelp7b5k9x) 5

[LDGH:](#_pgaxmootqv1k) 5

[India:](#_mxp044vtvqgt) 5

[India-1KG-SGDP-SAS-final:](#_9g9gienjwn2a) 5

[**Update Pop Name:**](#_x2eml5cbx5mh) **6**

[**SmartCleaning:**](#_u5xl7frpa9ek) **6**

[India-1KG-SGDP-SAS-final:](#_ldggt9run601) 6

[India:](#_yc3vkgwb6z8m) 6

[**Quality Control PBS specific**](#_e6js6pikayq1) **6**

[maf 0.05](#_jp8oza5yqqso) 6

[relatdness](#_sci2v9h3ihij) 6

[Try Other configurations](#_ia1fgqxctw1p) 7

[India X CEU X YRI](#_tediacwbq719) 7

[Caste/Tribe X Tribe/Caste X YRI](#_o7u5pa9fvo00) 7

[India1KGP X CEU X YRI](#_npqntl9onnlu) 7

[**Methods based on population differentiation**](#_ydz2bf9zzoua) **8**

[Separate Target Pairs](#_10q71v9sxs99) 8

[All X All X YRI/CEU](#_hi06tkujq6me) 8

[India X CEU X YRI](#_hce1cvse3s2v) 8

[Caste/Tribe X CEU X YRI](#_culrj71zm7d3) 8

[India1KGP X CEU X YRI](#_9ng30khv5jxj) 9

[**Calculate fst**](#_pfq5tqq6dtyq) **9**

[**Calculate PBS**](#_9ijwib46bhnl) **10**

[All x All CEU as outgroup](#_33jcr7qxhhca) 10

[All x All YRI as outgroup](#_5qnk5baly33t) 11

[India x CEU x YRI as outgroup](#_htckt6wlk7ko) 11

[Caste/Tribe X Tribe/Caste X YRI](#_y7acimky32cg) 11

[Caste X CEU X YRI](#_63obnflsbszf) 12

[Caste X CEU X YRI](#_8l8c39g2kn5l) 12

[Indian1kgp X CEU X YRI](#_f14tflwntxo8) 12

[**Order PBS values**](#_5ryfcme5wuxb) **13**

[All x All x CEU](#_2r7h9mzd3nuw) 13

[India x CEU x YRI](#_9zakqkbw06ha) 13

[Caste/Tribe x Tribe/Caste x YRI](#_dzmsovt2rnvx) 13

[**Calculate PBS for bins**](#_u51vhus2a6nj) **13**

[Split for chromosomes](#_5dn0g5n0eaxj) 13

[All x All x CEU](#_6or5wmi3yrnv) 13

[India x CEU x YRI](#_s0tsyqmvohm) 13

[Caste/Tribe x Tribe/Caste x YRI](#_wqdjn2sz9rtp) 14

[Caste/Tribe x CEU x YRI](#_50d1lrmpcg6u) 14

[Indian1kgp x CEU x YRI](#_5ctxzb1xoyo8) 14

[Run sliding window for chromosome](#_ays6p4wxzfpf) 14

[All x All x CEU](#_wv6b6kborfa2) 14

[India x CEU x YRI](#_kf2wfx2rhkpp) 14

[Caste/Tribe x Tribe/Caste x YRI](#_gel3s0gm7u) 15

[Caste/Tribe x CEU x YRI](#_z0mg4mq97q7o) 15

[Indian1kgp x CEU x YRI](#_idifeeaj5pik) 15

[Correr bins](#_2xb9hjpiww5n) 15

[All xAll x CEU/YRI](#_la5hg9u8a8un) 15

[India x CEU x YRI](#_tf5xyqjgp1ja) 15

[Caste/Tribe x Tribe/Caste x YRI](#_7v227y1olffr) 16

[Caste/Tribe x CEU x YRI](#_s0nlue9g2m1a) 16

[Indian1kgp x CEU x YRI](#_iy07ijb5sr7g) 16

[Merge chromosome files](#_u90i37dob9ak) 16

[All x All x CEU/YRI](#_ie93gg666jbf) 16

[India x CEU x YRI](#_hote4otcra9) 17

[Caste/Tribe x Tribe/Caste x YRI](#_imit0xre5om9) 17

[Caste/Tribe x CEU x YRI](#_4cq2ccgctqmd) 17

[Indian1kgp x CEU x YRI](#_9gegij6kevaj) 17

[**Methods based on LD**](#_czm9fhfs0n99) **17**

[Phasing Process](#_30mqqr682e2a) 17

[All X All X CEU/YRI](#_2vyqdkje187m) 17

[Check DNA strand](#_agjtzxb0bdwi) 17

[Swap the reference and alternate alleles](#_r22wixdr5fqc) 18

[Sort file](#_493qrreqi75) 18

[Second Check DNA strand](#_959qd4j8wwdb) 18

[Just India](#_abti6hwwpasd) 19

[Select Just India](#_7haikmemqfq1) 19

[Swap the reference and alternate alleles](#_tid6kw576nss) 19

[Sort file](#_p7wfh179u5ux) 19

[Second Check DNA strand](#_z8fn8txj35bp) 19

[CasteTribe x CEU](#_bx21nqd06z1x) 20

[Swap the reference and alternate alleles](#_f2pi42gwqti9) 20

[Sort file](#_z7m1c4hiylux) 20

[Second Check DNA strand](#_46stv5906p6b) 20

[Indian1lgp x CEU x YRI](#_3ajjo91ojbig) 20

[Check DNA strand](#_3x19t74slk3z) 20

[Swap the reference and alternate alleles](#_7ho6cfr21v85) 21

[Sort file](#_q2lodqpcinnb) 21

[Second Check DNA strand](#_w3qkt4ptum2n) 21

[All x All](#_on2gtyvsbhqn) 22

[Fit family/ID Names](#_mf3lccornezv) 22

[India x CEU](#_an90jhvdnfu8) 22

[Split for Population](#_uaoec3yijgrv) 22

[Split for chromosomes](#_5alovvny4fy0) 22

[Caste X CEU](#_39b2ioovbqy8) 22

[Split for Population](#_jbx57qrkn9dm) 23

[Split for chromosomes](#_as4cifo43lvw) 23

[Indian1kgp X CEU](#_bikvmkd0i1q8) 23

[Fit family/ID Names](#_3tcqmm4n5ea3) 23

[Split for Population](#_1xqwv0h8ack6) 24

[Split for chromosomes](#_5qzb4qyiweae) 24

[Make .map file](#_qlpqwbfz31y7) 24

[xpEHH](#_bpk0gzt0xv3k) 24

[Run xpEHH](#_v6rs6x8d7f4z) 24

[India X CEU](#_dnhkxshjrtmc) 24

[Tribe X CEU](#_mqq3u11v45n9) 25

[Caste X CEU](#_fns30rpa57b8) 25

[Indian1kgp X CEU](#_65yhwln8gsrp) 25

[Normalize xpEHH](#_az5xt08rn0aa) 25

[India X CEU](#_3j5svb5e678x) 25

[Caste X CEU](#_5cwbfkzcs2lj) 26

[Tribe X CEU](#_f2urtbmkp7uq) 26

[Indian1kgp X CEU](#_ucxi4ww58yg7) 26

[Addiction chromosome column](#_lwo68x5bd53c) 26

[India X CEU](#_fz2pqvvl1pee) 26

[CasteTribe X CEU](#_xl5cbgkf4hyf) 26

[Indian1kgp X CEU](#_u9nht8cueaqc) 26

[Merge Chromossome Files](#_61n943k22igz) 27

[India X CEU](#_93bjqcjyspsm) 27

[CasteTribe X CEU](#_uyz80ac44xpw) 27

[India1kgp X CEU](#_kl6cdj8vn19) 27

[iHS](#_gqvyrjymsoul) 27

[Define allele ancestral and derived](#_59ggw446e4zj) 27

[Split by Chromossome and Transform in haps sample](#_g4m5v1k1ych7) 27

[Just India](#_sc6wqr1jezbh) 27

[CasteTribe](#_ujv9chvzlrvq) 27

[India1kgp](#_37uqrtli9h1k) 28

[Format file from correspondence](#_vc9uv0d361tj) 28

[Just India](#_9zmfm33u0unx) 28

[CasteTribe](#_wf7pc9qp02y9) 28

[India1kgp](#_495tim5zpgkj) 28

[Make table chr/pos/ancestral allele](#_g0c4o0zbidd9) 29

[Just India](#_rj16d2swsbi2) 29

[CasteTribe](#_6qvkpfk8iujq) 29

[India1kgp](#_6aymfsoulcvv) 30

[Make header](#_6f51ol6s6rmz) 30

[Just India](#_1z9u934ogmjq) 30

[Caste/Tribe](#_vodv1vdsul25) 30

[India1kgp](#_mgnccsjy69at) 30

[Make vcf with Ancestral Allele Correspondence](#_cbye0i2dvnil) 31

[Just India](#_4bljma3onk78) 31

[Caste/Tribe](#_u95o7ztd20n8) 32

[India1kgp](#_g5vavih44atu) 34

[Run iHS](#_bjsjg7wws7oc) 35

[Just India](#_cnnrnvc331aq) 35

[Caste/Tribe](#_cktcyyiyw042) 35

[India1kgp](#_nwgtm2tpgru3) 36

[Normalize iHS](#_mivyg9hd4gei) 36

[Just India](#_k07pmt9dwf1s) 36

[Caste/Tribe](#_e62dj1t2uowf) 36

[India1kgp](#_9zmsuvtca1i6) 36

[Addiction Chromosome Information](#_s8h563qscrbs) 37

[Just India](#_kq26z3kppcnn) 37

[Caste/Tribe](#_kb26oupaqo1d) 37

[India1kgp](#_k449ddhznb97) 37

[Merge Chromosome Files](#_uxwfmcf94cd7) 37

[Just India](#_phbn0grb53w) 37

[Caste/Tribe](#_iosjsnr6g83i) 37

[India1kgp](#_tyxr8z9e3kcn) 37

[**Methods based on the Site Frequency Spectrum**](#_95aascqxpmic) **38**

[Tajima’sD](#_9r396wq79d3r) 38

[India X CEU](#_8orthppb6eyi) 38

[CasteTribe X CEU](#_jvur9v5vkb8u) 38

[India 1kgp Group](#_q2rjl4zhakz8) 38

[**Composite Tests**](#_iiamxpwegk0q) **38**

[Transfer Files to LDGH server](#_j38ks11ev85) 38

[India](#_2q5kknf5bnyz) 38

[Caste/Tribe](#_m6nuusq9wpy7) 39

[India 1KGP](#_3uk0rhtvp8i1) 39

[Run SweeD by Chromossome](#_tfrex615l8y) 39

[Just India](#_pysfk97f2tzr) 39

[Caste/Tribe](#_4782jjh94qhl) 39

[India1kgp](#_gmn9ynqsly43) 39

[Addiction Chromosome Information](#_r80b8hw6i4ro) 39

[JustIndia](#_uqx0kajk9462) 39

[Merge Chromosomes Files](#_f5cme7j3nqvq) 40

[JustIndia](#_m5yly7fmq1v2) 40

[India1kgp/caste/tribe](#_4hu31gmwtm4x) 40

[Sorted Files](#_dpyczymiqz8v) 40

# Data Information:

## *CEDAR*:

/project/6006002/marlama/Indians/India.bed

/project/6006002/marlama/Indians/India-1KG-SGDP-SAS-final.bed

## *GRAHAN*:

/project/6006002/marlama/Indians/India.bed

/project/6006002/marlama/Indians/India-1KG-SGDP-SAS-final.bed

## *LDGH*:

/media/LDGH/marlama/Indians/

## India:

~409K markers, including X-chromosome markers, for Manjari's samples (4 tribal groups and 2 castes).

409651 SNPs

456 Samples

## India-1KG-SGDP-SAS-final:

~385K markers (include X-chromosome markers), and includes Manjari's samples, 1KG samples and SGDP samples.

The marker rs names has been subsituted by chr-pos labels in order to do the merging. Markers with allele issues (more than 2 alleles) were eliminated in order to do the merging.

385180 SNPs

3223 samples

HG37

# Update Pop Name:

GRAHAN:

plink --bfile /project/6006002/marlama/Indians/India-1KG-SGDP-SAS-final --update-ids /project/6006002/marlama/Indians/ListaDeCorrespondencia\_Indians\_1KGP\_HGDP.txt --make-bed --out /project/6006002/marlama/Indians/India-1KG-SGDP\_Pop

# SmartCleaning:

## India-1KG-SGDP-SAS-final:

LDGH:

perl /home/DadosCongelados/FilesParaLimpeza/[SmartCleaning\_Aug2020.pl](http://smartcleaning2.1.pl/) --input /home/DadosCongelados/FilesParaLimpeza/InputSmartCleaning/Arquivo1\_Indians\_1KGP\_SGDP.txt --programs /home/DadosCongelados/FilesParaLimpeza/InputSmartCleaning/Arquivo2.txt --start 1 --end 26 --output /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/

Size of Bim file for each step

geno Indians\_1KGP\_SGDP 385180

mind Indians\_1KGP\_SGDP 385180

Chr0 Indians\_1KGP\_SGDP 385180

100% Hetero Indians\_1KGP\_SGDP 385180

ATCG Indians\_1KGP\_SGDP 370594

SDQC Indians\_1KGP\_SGDP 365152

## India:

perl /home/DadosCongelados/FilesParaLimpeza/[SmartCleaning\_Aug2020.pl](http://smartcleaning2.1.pl/) --input /home/DadosCongelados/FilesParaLimpeza/InputSmartCleaning/Arquivo1\_India\_Marla.txt --programs /home/DadosCongelados/FilesParaLimpeza/InputSmartCleaning/Arquivo2.txt --start 1 --end 26 --output /media/LDGH/marlama/Indians/DataClean\_India/

# Quality Control PBS specific

## maf 0.05

plink --bfile /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic --maf 0.05 --make-bed --out /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05

82291 variants removed due to minor allele threshold(s)

## relatdness

plink --bfile /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05 --genome --out /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_informations

perl /home/marlam/Programs/PLINK2NAToRA.pl --input /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_informations.genome --out /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Kinship --kinship

python /home/marlam/Programs/NAToRA\_3.1.py --input /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Kinship -c 0.1 -e 1 -l /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Natora\_Indians\_1KGP\_SGDP

120 individuals to remove

plink --bfile /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05 --remove /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Natora\_Indians\_1KGP\_SGDP/Remove.txt --make-bed --out /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated

## Try Other configurations

### India X CEU X YRI

plink --bfile /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated --update-ids /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/ListaDeCorrespondencia\_JustIndia.txt --make-bed --out /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup

### Caste/Tribe X Tribe/Caste X YRI

plink --bfile /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated --update-ids /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/ListaDeCorrespondencia\_TribeCaste.txt --make-bed --out /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_TribeCasteGroup

### India1KGP X CEU X YRI

plink --bfile /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated --update-ids /project/6006002/marlama/Indians/ListaDeCorrespondencia\_Indian1kgp.txt --make-bed --out /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup

### India1KGP\_HC X CEU X YRI

for chr in $(seq 1 22) ; do plink --bfile /home/marlama/scratch/1kgp\_HC/1kgp\_HC\_MosaicQC/1kgpHC\_chr${chr}\_Autosomal --update-ids /home/marlama/scratch/1kgp\_HC/PBS/OriginalFiles/India\_CEU\_YRI\_UpdatePopName.txt --make-bed --out /home/marlama/scratch/1kgp\_HC/PBS/OriginalFiles/1kgpHC\_chr${chr}\_India\_CEU\_YRI ; done

# Methods based on population differentiation

# 

## Separate Target Pairs

### All X All X YRI/CEU

for pop1 in YRI CEU BH DB KM KN PW WR ; do for pop2 in YRI CEU BH DB KM KN PW WR ; do awk -v pop1="$pop1" -v pop2="$pop2" '{if($1~ pop1 || $1~ pop2){print $1"\t"$2}}' /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated.fam > /media/LDGH/marlama/Indians/TargetPairs/${pop1}\_${pop2}.txt ; done ; done

for pop1 in YRI CEU BH DB KM KN PW WR ; do for pop2 in YRI CEU BH DB KM KN PW WR ; do plink --bfile /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated --keep /media/LDGH/marlama/Indians/TargetPairs/${pop1}\_${pop2}.txt --recode --out /media/LDGH/marlama/Indians/TargetPairs/${pop1}\_${pop2} ; done ; done

### India X CEU X YRI

for pop1 in YRI CEU India ; do for pop2 in YRI CEU India ; do awk -v pop1="$pop1" -v pop2="$pop2" '{if($1~ pop1 || $1~ pop2){print $1"\t"$2}}' /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup.fam > /media/LDGH/marlama/Indians/TargetPairs/${pop1}\_${pop2}.txt ; done ; done

for pop1 in YRI CEU India ; do for pop2 in YRI CEU India ; do plink --bfile /media/LDGH/marlama/Indians/DataClean\_Indians\_1KGP\_SGDP/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup --keep /media/LDGH/marlama/Indians/TargetPairs/${pop1}\_${pop2}.txt --recode --out /media/LDGH/marlama/Indians/TargetPairs/${pop1}\_${pop2} ; done ; done

### Caste/Tribe X CEU X YRI

for pop1 in Caste Tribe YRI CEU ; do for pop2 in Caste Tribe YRI CEU ; do awk -v pop1="$pop1" -v pop2="$pop2" '{if($1~ pop1 || $1~ pop2){print $1"\t"$2}}' /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_TribeCasteGroup.fam > /project/6006002/marlama/Indians/TargetPairs/${pop1}\_${pop2}.txt ; done ; done

for pop1 in Caste Tribe YRI CEU ; do for pop2 in Caste Tribe YRI CEU ; do plink --bfile /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_TribeCasteGroup --keep /project/6006002/marlama/Indians/TargetPairs/${pop1}\_${pop2}.txt --recode --out /project/6006002/marlama/Indians/TargetPairs/${pop1}\_${pop2} ; done ; done

### India1KGP X CEU X YRI

for pop1 in Indian1kgp ; do for pop2 in YRI CEU ; do awk -v pop1="$pop1" -v pop2="$pop2" '{if($1~ pop1 || $1~ pop2){print $1"\t"$2}}' /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup.fam > /project/6006002/marlama/Indians/TargetPairs/${pop1}\_${pop2}.txt ; done ; done

for pop1 in Indian1kgp ; do for pop2 in YRI CEU ; do plink --bfile /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup --keep /project/6006002/marlama/Indians/TargetPairs/${pop1}\_${pop2}.txt --recode --out /project/6006002/marlama/Indians/TargetPairs/${pop1}\_${pop2} ; done ; done

### Select just India GIH and ITU

cat Indian\_samples.txt CEU\_YRI.txt > India\_CEU\_YRI.txt

awk '{print $2"/t"$2}' India\_CEU\_YRI.txt > India\_CEU\_YRI\_NoFamilyID.txt

438 samples

for chr in $(seq 1 22) ; do plink --bfile /home/marlama/scratch/1kgp\_HC/PBS/OriginalFiles/1kgpHC\_chr${chr}\_India\_CEU\_YRI --maf 0.05 --make-bed --out /home/marlama/scratch/1kgp\_HC/PBS/OriginalFiles/1kgpHC\_chr${chr}\_India\_CEU\_YRI\_maf005 ; done

for chr in $(seq 1 22) ; do plink --bfile /home/marlama/scratch/1kgp\_HC/PBS/OriginalFiles/1kgpHC\_chr${chr}\_India\_CEU\_YRI\_maf005 --keep /home/marlama/scratch/1kgp\_HC/PBS/OriginalFiles/CEU\_YRI.txt --recode --out /home/marlama/scratch/1kgp\_HC/PBS/4P/CEU\_YRI/CEU\_YRI\_chr${chr} ; done

for i in YRI CEU ; do for chr in $(seq 1 22) ; do plink --bfile /home/marlama/scratch/1kgp\_HC/PBS/OriginalFiles/1kgpHC\_chr${chr}\_India\_CEU\_YRI\_maf005 --keep /home/marlama/scratch/1kgp\_HC/PBS/OriginalFiles/Indian1kgp\_${i}.txt --recode --out /home/marlama/scratch/1kgp\_HC/PBS/4P/Indian1kgp\_${i}/Indian1kgp\_${i}\_chr${chr} ; done

## Calculate fst

4P -i 0 -f BH\_DB.ped -m BH\_DB.map -n 163 -s **282861** -t 20

4P -i 0 -f BH\_KM.ped -m BH\_KM.map -n **136** -s **282861** -t 20

4P -i 0 -f BH\_KN.ped -m BH\_KN.map -n **170** -s **282861** -t 20

4P -i 0 -f BH\_PW.ped -m BH\_PW.map -n **164** -s **282861** -t 20

4P -i 0 -f BH\_WR.ped -m BH\_WR.map -n **169** -s **282861** -t 20

4P -i 0 -f BH\_CEU.ped -m BH\_CEU.map -n **186** -s **282861** -t 20

4P -i 0 -f BH\_YRI.ped -m BH\_YRI.map -n **195** -s **282861** -t 20

4P -i 0 -f DB\_KM.ped -m DB\_KM.map -n **125** -s **282861** -t 20

4P -i 0 -f DB\_KN.ped -m DB\_KN.map -n **159** -s **282861** -t 20

4P -i 0 -f DB\_PW.ped -m DB\_PW.map -n **153** -s **282861** -t 20

4P -i 0 -f DB\_WR.ped -m DB\_WR.map -n **158** -s **282861** -t 20

4P -i 0 -f DB\_CEU.ped -m DB\_CEU.map -n **175** -s **282861** -t 20

4P -i 0 -f DB\_YRI.ped -m DB\_YRI.map -n **184** -s **282861** -t 20

4P -i 0 -f KM\_KN.ped -m KM\_KN.map -n **132** -s **282861** -t 20

4P -i 0 -f KM\_PW.ped -m KM\_PW.map -n **126** -s **282861** -t 20

4P -i 0 -f KM\_WR.ped -m KM\_WR.map -n **131** -s **282861** -t 20

4P -i 0 -f KM\_CEU.ped -m KM\_CEU.map -n **148** -s **282861** -t 20

4P -i 0 -f KM\_YRI.ped -m KM\_YRI.map -n **157** -s **282861** -t 20

4P -i 0 -f KN\_PW.ped -m KN\_PW.map -n **126** -s **282861** -t 20

4P -i 0 -f KN\_WR.ped -m KN\_WR.map -n **165** -s **282861** -t 20

4P -i 0 -f KN\_CEU.ped -m KN\_CEU.map -n **182** -s **282861** -t 20

4P -i 0 -f KN\_YRI.ped -m KN\_YRI.map -n **191** -s **282861** -t 20

4P -i 0 -f PW\_WR.ped -m PW\_WR.map -n **159** -s **282861** -t 20

4P -i 0 -f PW\_CEU.ped -m PW\_CEU.map -n **176** -s **282861** -t 20

4P -i 0 -f PW\_YRI.ped -m PW\_YRI.map -n **185** -s **282861** -t 20

4P -i 0 -f WR\_CEU.ped -m WR\_CEU.map -n **181** -s **282861** -t 20

4P -i 0 -f WR\_YRI.ped -m WR\_YRI.map -n **190** -s **282861** -t 20

4P -i 0 -f India\_CEU.ped -m India\_CEU.map -n **553** -s **282861** -t 20

4P -i 0 -f India\_YRI.ped -m India\_YRI.map -n **562** -s **282861** -t 20

4P -i 0 -f Caste\_Tribe.ped -m Caste\_Tribe.map -n **454** -s **282861** -t 20

4P -i 0 -f Caste\_YRI.ped -m Caste\_YRI.map -n **233** -s **282861** -t 20

4P -i 0 -f Tribe\_YRI.ped -m Tribe\_YRI.map -n **437** -s **282861** -t 20

4P -i 0 -f Caste\_CEU.ped -m Caste\_CEU.map -n **224** -s **282861** -t 20

4P -i 0 -f Tribe\_CEU.ped -m Tribe\_CEU.map -n 428-s **282861** -t 20

4P -i 0 -f Tribe\_YRI.ped -m Tribe\_YRI.map -n 437-s **282861** -t 20

4P -i 0 -f CEU\_YRI.ped -m CEU\_YRI.map -n 207-s **282861** -t 20

4P -i 0 -f Indian1kgp\_CEU.ped -m Indian1kgp\_CEU.map -n 302-s **282861** -t 20

4P -i 0 -f Indian1kgp\_YRI.ped -m Indian1kgp\_YRI.map -n 311-s **282861** -t 20

4P -i 0 -f Indian1kgp\_YRI.ped -m Indian1kgp\_YRI.map -n 311-s **282861** -t 20

By CHR

for chr in $(seq 1 22) ; do for s in 518198 548418 474882 488002 425368 444798 392461 361638 291922 341706 328234 321794 245465 217296 195773 202095 185547 191100 160898 146962 97598 95179 ; do /home/marlama/programs/4p/src/4P -i 0 -f CEU\_YRI\_chr${chr}.ped -m CEU\_YRI\_chr${chr}.map -n 207 -s ${s} -t 100 ; awk '{if(NR>1){print}}' PAIR\_DIST\_CEU\_YRI.txt > PAIR\_DIST\_CEU\_YRI\_chr${chr}.txt ; done ; done

cat PAIR\_DIST\_CEU\_YRI\* > PAIR\_DIST\_CEU\_YRI.txt

awk '{if(NR==1){print}}' PAIR\_DIST\_CEU\_YRI.txt > PAIR\_DIST\_Header.txt

cat PAIR\_DIST\_Header.txt PAIR\_DIST\_CEU\_YRI.txt > PAIR\_DIST\_CEU\_YRI\_FinalFile.txt

for chr in $(seq 1 22) ; do for s in 518198 548418 474882 488002 425368 444798 392461 361638 291922 341706 328234 321794 245465 217296 195773 202095 185547 191100 160898 146962 97598 95179 ; do /home/marlama/programs/4p/src/4P -i 0 -f Indian1kgp\_CEU\_chr${chr}.ped -m Indian1kgp\_CEU\_chr${chr}.map -n 309 -s ${s} -t 100 ; awk '{if(NR>1){print}}' PAIR\_DIST\_India1kgp\_CEU.txt > PAIR\_DIST\_India1kgp\_CEU\_chr${chr}.txt ; done ; done

cat PAIR\_DIST\_India1kgp\_CEU\_chr\* > PAIR\_DIST\_India1kgp\_CEU.txt

cat /home/marlama/scratch/1kgp\_HC/PBS/4P/CEU\_YRI/PAIR\_DIST\_Header.txt PAIR\_DIST\_India1kgp\_CEU.txt > PAIR\_DIST\_India1kgp\_CEU\_FinalFile.tx

for chr in $(seq 1 22) ; do for s in 518198 548418 474882 488002 425368 444798 392461 361638 291922 341706 328234 321794 245465 217296 195773 202095 185547 191100 160898 146962 97598 95179 ; do /home/marlama/programs/4p/src/4P -i 0 -f Indian1kgp\_YRI\_chr${chr}.ped -m Indian1kgp\_YRI\_chr${chr}.map -n 318 -s ${s} -t 100 ; awk '{if(NR>1){print}}' PAIR\_DIST\_India1kgp\_YRI.txt > PAIR\_DIST\_India1kgp\_YRI\_chr${chr}.txt ; done ; done

cat PAIR\_DIST\_India1kgp\_YRI\_chr\* > PAIR\_DIST\_India1kgp\_YRI.txt

cat /home/marlama/scratch/1kgp\_HC/PBS/4P/CEU\_YRI/PAIR\_DIST\_Header.txt PAIR\_DIST\_India1kgp\_YRI.txt > PAIR\_DIST\_India1kgp\_YRI\_FinalFile.txt

## Calculate PBS

### All x All CEU as outgroup

Modify according to the order of the names of the populations in the archives.

for pop1 in WR ; do for pop2 in BH DB KM KN PW ; do perl /home/marlam/Scripts\_PBS/CalculatePBS.pl -file1 /media/LDGH/marlama/Indians/FST/PAIR\_DIST\_${pop2}\_${pop1}.txt -file2 /media/LDGH/marlama/Indians/FST/PAIR\_DIST\_${pop1}\_CEU.txt -file3 /media/LDGH/marlama/Indians/FST/PAIR\_DIST\_${pop2}\_CEU.txt -output /media/LDGH/marlama/Indians/PBS/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_CEU ; done ; done

### All x All YRI as outgroup

Modify according to the order of the names of the populations in the archives.

for pop1 in WR ; do for pop2 in BH DB KM KN PW ; do perl /home/marlam/Scripts\_PBS/CalculatePBS.pl -file1 /media/LDGH/marlama/Indians/FST/PAIR\_DIST\_${pop2}\_${pop1}.txt -file2 /media/LDGH/marlama/Indians/FST/PAIR\_DIST\_${pop1}\_YRI.txt -file3 /media/LDGH/marlama/Indians/FST/PAIR\_DIST\_${pop2}\_YRI.txt -output /media/LDGH/marlama/Indians/PBS/YRI\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_YRI ; done ; done

### India x CEU x YRI as outgroup

for pop1 in India ; do for pop2 in CEU ; do perl /home/marlam/Scripts\_PBS/CalculatePBS.pl -file1 /media/LDGH/marlama/Indians/FST/PAIR\_DIST\_${pop1}\_${pop2}.txt -file2 /media/LDGH/marlama/Indians/FST/PAIR\_DIST\_${pop1}\_YRI.txt -file3 /media/LDGH/marlama/Indians/FST/PAIR\_DIST\_${pop2}\_YRI.txt -output /media/LDGH/marlama/Indians/PBS/YRI\_outgroup/PBSNormalizado\_${pop1}\_${pop2}\_YRI ; done ; done

### Caste/Tribe X Tribe/Caste X YRI

for pop1 in Tribe ; do for pop2 in Caste ; do perl /home/marlam/Scripts\_PBS/CalculatePBS.pl -file1 /media/LDGH/marlama/Indians/FST/PAIR\_DIST\_${pop1}\_${pop2}.txt -file2 /media/LDGH/marlama/Indians/FST/PAIR\_DIST\_${pop1}\_YRI.txt -file3 /media/LDGH/marlama/Indians/FST/PAIR\_DIST\_${pop2}\_YRI.txt -output /media/LDGH/marlama/Indians/PBS/YRI\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_YRI ; done ; done

### Caste X CEU X YRI

for pop1 in Caste ; do for pop2 in CEU ; do perl /home/marlam/Scripts\_PBS/CalculatePBS.pl -file1 /home/marlam/Indians/Fst/PAIR\_DIST\_${pop1}\_${pop2}.txt -file2 /home/marlam/Indians/Fst/PAIR\_DIST\_${pop1}\_YRI.txt -file3 /home/marlam/Indians/Fst/PAIR\_DIST\_${pop2}\_YRI.txt -output /home/marlam/Indians/PBS/PBSNormalizado\_${pop1}\_${pop2}\_YRI ; done ; done

### Caste X CEU X YRI

for pop1 in Tribe ; do for pop2 in CEU ; do perl /home/marlam/Scripts\_PBS/CalculatePBS.pl -file1 /home/marlam/Indians/Fst/PAIR\_DIST\_${pop1}\_${pop2}.txt -file2 /home/marlam/Indians/Fst/PAIR\_DIST\_${pop1}\_YRI.txt -file3 /home/marlam/Indians/Fst/PAIR\_DIST\_${pop2}\_YRI.txt -output /home/marlam/Indians/PBS/PBSNormalizado\_${pop1}\_${pop2}\_YRI ; done ; done

### Indian1kgp X CEU X YRI

for pop1 in Indian1kgp ; do for pop2 in CEU ; do perl /home/marlam/Scripts\_PBS/CalculatePBS.pl -file1 /home/marlam/Indians/Fst/PAIR\_DIST\_${pop1}\_${pop2}.txt -file2 /home/marlam/Indians/Fst/PAIR\_DIST\_${pop1}\_YRI.txt -file3 /home/marlam/Indians/Fst/PAIR\_DIST\_${pop2}\_YRI.txt -output /home/marlam/Indians/PBS/PBSNormalizado\_${pop1}\_${pop2}\_YRI ; done ; done

### Indian1kgp\_HC X CEU X YRI

perl /home/marlama/programs/Scripts\_PBS/CalculatePBS.pl -file1 /home/marlama/scratch/1kgp\_HC/PBS/4P/Indian1kgp\_CEU/PAIR\_DIST\_India1kgp\_CEU\_FinalFile.txt -file2 /home/marlama/scratch/1kgp\_HC/PBS/4P/Indian1kgp\_YRI/PAIR\_DIST\_India1kgp\_YRI\_FinalFile.txt -file3 /home/marlama/scratch/1kgp\_HC/PBS/4P/CEU\_YRI/PAIR\_DIST\_CEU\_YRI\_FinalFile.txt -output /home/marlama/scratch/1kgp\_HC/PBS/PBSNormalizado\_India1kgp\_CEU\_YRI

## Order PBS values

### All x All x CEU

for pop1 in BH ; do for pop2 in BH DB KM KN PW WR ; do awk '{print |"sort -nk10 "}' /media/LDGH/marlama/Indians/PBS/CEU\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_CEU.txt > /media/LDGH/marlama/Indians/PBS/CEU\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_CEU\_sorted.txt ; done ; done

### India x CEU x YRI

for pop1 in India ; do for pop2 in CEU ; do awk '{print |"sort -nk10 "}' /media/LDGH/marlama/Indians/PBS/YRI\_outgroup/PBSNormalizado\_${pop1}\_${pop2}\_YRI.txt > /media/LDGH/marlama/Indians/PBS/YRI\_outgroup/PBSNormalizado\_${pop1}\_${pop2}\_YRI\_sorted.txt ; done ; done

### Caste/Tribe x Tribe/Caste x YRI

for pop1 in Tribe ; do for pop2 in Caste ; do awk '{print |"sort -nk10 "}' /media/LDGH/marlama/Indians/PBS/YRI\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_YRI.txt > /media/LDGH/marlama/Indians/PBS/YRI\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_YRI\_sorted.txt ; done ; done

## Calculate PBS for bins

### Split for chromosomes

#### All x All x CEU

for i in $(seq 1 22); do for pop1 in BH ; do for pop2 in DB KM KN PW WR ; do for pop3 in CEU ; do awk -v chr="${i}" '{if ($1 == chr) {print}}' /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}.txt > /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt ; done ; done ; done ; done

#### India x CEU x YRI

for i in $(seq 1 22) ; do for pop1 in India ; do for pop2 in CEU; do for pop3 in YRI ; do awk -v chr="${i}" '{if ($1 == chr) {print}}' /media/LDGH/marlama/Indians/PBS/YRI\_outgroup/PBSNormalizado\_${pop1}\_${pop2}\_YRI.txt > /media/LDGH/marlama/Indians/PBS/YRI\_outgroup/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt ; done ; done ; done ; done

#### Caste/Tribe x Tribe/Caste x YRI

for i in $(seq 1 22); do for pop1 in Caste Tribe ; do for pop2 in Tribe Caste ; do for pop3 in YRI ; do awk -v chr="${i}" '{if ($1 == chr) {print}}' /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}.txt > /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt ; done ; done ; done ; done

#### Caste/Tribe x CEU x YRI

for i in $(seq 1 22); do for pop1 in Caste Tribe ; do for pop2 in CEU ; do for pop3 in YRI ; do awk -v chr="${i}" '{if ($1 == chr) {print}}' /home/marlam/Indians/PBS/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}.txt > /home/marlam/Indians/PBS/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt ; done ; done ; done ; done

#### Indian1kgp x CEU x YRI

for i in $(seq 1 22); do for pop1 in Indian1kgp ; do for pop2 in CEU ; do for pop3 in YRI ; do awk -v chr="${i}" '{if ($1 == chr) {print}}' /home/marlam/Indians/PBS/Indian1kgpTarget/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}.txt > /home/marlam/Indians/PBS/Indian1kgpTarget/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt ; done ; done ; done ; done

#### Indian1kgp\_HC x CEU x YRI

for i in $(seq 1 22); do awk -v chr="${i}" '{if ($1 == chr) {print}}' /home/marlama/scratch/1kgp\_HC/PBS/PBSNormalizado\_India1kgp\_CEU\_YRI.txt > /home/marlama/scratch/1kgp\_HC/PBS/PBSNormalizado\_India1kgp\_CEU\_YRI\_chr${i}.txt ; done

### Run sliding window for chromosome

#### All x All x CEU

for i in $(seq 1 22) ; do for pop1 in WR ; do for pop2 in BH DB KM KN PW ; do for pop3 in CEU ; do Rscript /home/marlam/Scripts\_PBS/JanelasDeslizantesForChr.R /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_SlidingWindow.txt ; done ; done ; done ; done

#### India x CEU x YRI

for i in $(seq 1 22) ; do for pop1 in India ; do for pop2 in CEU ; do for pop3 in YRI ; do Rscript /home/marlam/Scripts\_PBS/JanelasDeslizantesForChr.R /media/LDGH/marlama/Indians/PBS/YRI\_outgroup/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt /media/LDGH/marlama/Indians/PBS/YRI\_outgroup/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_SlidingWindow.txt ; done ; done ; done ; done

#### Caste/Tribe x Tribe/Caste x YRI

for i in $(seq 1 22) ; do for pop1 in Caste Tribe ; do for pop2 in Tribe Caste ; do for pop3 in YRI ; do Rscript /home/marlam/Scripts\_PBS/JanelasDeslizantesForChr.R /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_SlidingWindow.txt ; done ; done ; done ; done

#### Caste/Tribe x CEU x YRI

for i in $(seq 1 22) ; do for pop1 in Caste Tribe ; do for pop2 in CEU ; do for pop3 in YRI ; do Rscript /home/marlam/Scripts\_PBS/JanelasDeslizantesForChr.R /home/marlam/Indians/PBS/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt /home/marlam/Indians/PBS/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_SlidingWindow.txt ; done ; done ; done ; done

#### Indian1kgp x CEU x YRI

for i in $(seq 1 22) ; do for pop1 in Indian1kgp ; do for pop2 in CEU ; do for pop3 in YRI ; do Rscript /home/marlam/Scripts\_PBS/JanelasDeslizantesForChr.R /home/marlam/Indians/PBS/Indian1kgpTarget/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt /home/marlam/Indians/PBS/Indian1kgpTarget/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_SlidingWindow.txt ; done ; done ; done ; done

#### Indian1kgp\_HG x CEU x YRI

for chr in $(seq 1 22) ; do cat /media/LDGH/marlama/India/1kgp\_HC/PBS/PBSNormalizado\_India1kgp\_CEU\_YRI\_Header.txt /media/LDGH/marlama/India/1kgp\_HC/PBS/PBSNormalizado\_India1kgp\_CEU\_YRI\_chr${chr}.txt > /media/LDGH/marlama/India/1kgp\_HC/PBS/PBSNormalizado\_India1kgp\_CEU\_YRI\_chr${chr}\_WithHeader.txt ; done

for i in $(seq 1 22) ; do Rscript /home/marlam/Scripts\_PBS/JanelasDeslizantesForChr.R /media/LDGH/marlama/India/1kgp\_HC/PBS/PBSNormalizado\_India1kgp\_CEU\_YRI\_chr${i}\_WithHeader.txt /media/LDGH/marlama/India/1kgp\_HC/PBS/PBSNormalizado\_India1kgp\_CEU\_YRI\_chr${i}\_SlidingWindow.txt ; done

## 

### Correr bins

#### All xAll x CEU/YRI

for i in $(seq 1 22) ; do for pop1 in PW ; do for pop2 in BH DB KM KN WR ; do for pop3 in CEU YRI ; do perl /home/marlam/Scripts\_PBS/Bins\_PBSnForChromosome.pl --bins /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/${pop2}\_close/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_SlidingWindow.txt --pbs /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/${pop2}\_close/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt --output /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/${pop2}\_close/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_Bins20\_5.txt ; done ; done ; done ; done

### 

#### India x CEU x YRI

for i in $(seq 1 22) ; do for pop1 in India ; do for pop2 in CEU ; do for pop3 in YRI ; do perl /home/marlam/Scripts\_PBS/Bins\_PBSnForChromosome.pl --bins /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_SlidingWindow.txt --pbs /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt --output /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_Bins20\_5.txt ; done ; done ; done ; done

#### Caste/Tribe x Tribe/Caste x YRI

for i in $(seq 1 22) ; do for pop1 in Caste Tribe ; do for pop2 in Caste Tribe ; do for pop3 in YRI ; do perl /home/marlam/Scripts\_PBS/Bins\_PBSnForChromosome.pl --bins /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_SlidingWindow.txt --pbs /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt --output /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_Bins20\_5.txt ; done ; done ; done ; done

#### Caste/Tribe x CEU x YRI

for i in $(seq 1 22) ; do for pop1 in Caste Tribe ; do for pop2 in CEU ; do for pop3 in YRI ; do perl /home/marlam/Scripts\_PBS/Bins\_PBSnForChromosome.pl --bins /home/marlam/Indians/PBS/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_SlidingWindow.txt --pbs /home/marlam/Indians/PBS/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt --output /home/marlam/Indians/PBS/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_Bins20\_5.txt ; done ; done ; done ; done

#### Indian1kgp x CEU x YRI

for i in $(seq 1 22) ; do for pop1 in Indian1kgp ; do for pop2 in CEU ; do for pop3 in YRI ; do perl /home/marlam/Scripts\_PBS/Bins\_PBSnForChromosome.pl --bins /home/marlam/Indians/PBS/Indian1kgpTarget/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_SlidingWindow.txt --pbs /home/marlam/Indians/PBS/Indian1kgpTarget/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}.txt --output /home/marlam/Indians/PBS/Indian1kgpTarget/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_chr${i}\_Bins20\_5.txt ; done ; done ; done ; done

#### Indian1kgp\_HG x CEU x YRI

for i in $(seq 1 22) ; do for pop1 in Indian1kgp ; do for pop2 in CEU ; do for pop3 in YRI ; do perl /home/marlam/Scripts\_PBS/Bins\_PBSnForChromosome.pl --bins /media/LDGH/marlama/India/1kgp\_HC/PBS/PBSNormalizado\_India1kgp\_CEU\_YRI\_chr${i}\_SlidingWindow.txt --pbs /media/LDGH/marlama/India/1kgp\_HC/PBS/PBSNormalizado\_India1kgp\_CEU\_YRI\_chr${i}\_WithHeader.txt --output /media/LDGH/marlama/India/1kgp\_HC/PBS/PBSNormalizado\_India1kgp\_CEU\_YRI\_chr${i}\_Bins20\_5.txt ; done ; done ; done ; done

### Merge chromosome files

#### All x All x CEU/YRI

for pop1 in DB KM KN WR PW ; do for pop2 in BH DB KM KN WR PW ; do for pop3 in CEU YRI ; do mkdir /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/${pop2}\_close/ ; done ; done ; done

for pop1 in DB KM KN WR PW ; do for pop2 in BH DB KM KN WR PW ; do for pop3 in CEU YRI ; do mv /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\* /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/${pop2}\_close/ ; done ; done ; done

for pop1 in BH DB KM KN WR PW ; do for pop2 in BH DB KM KN WR PW ; do for pop3 in CEU YRI ; do cat /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/${pop2}\_close/\*Bins20\_5.txt.txt > /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/${pop2}\_close/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_Bins.txt ; done ; done ; done

#### India x CEU x YRI

for pop1 in India ; do for pop2 in CEU ; do for pop3 in YRI ; do cat /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/\*Bins20\_5.txt.txt > /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_Bins.txt ; done ; done ; done

#### Caste/Tribe x Tribe/Caste x YRI

for pop1 in Caste Tribe ; do for pop2 in Caste Tribe ; do for pop3 in YRI ; do cat /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/\*Bins20\_5.txt.txt > /media/LDGH/marlama/Indians/PBS/${pop3}\_outgroup/${pop1}\_target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_Bins.txt ; done ; done ; done

#### Caste/Tribe x CEU x YRI

for pop1 in Caste Tribe ; do for pop2 in CEU ; do for pop3 in YRI ; do cat /home/marlam/Indians/PBS/${pop1}Target/\*Bins20\_5.txt.txt > /home/marlam/Indians/PBS/${pop1}Target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_Bins.txt ; done ; done ; done

#### Indian1kgp x CEU x YRI

for pop1 in Indian1kgp ; do for pop2 in CEU ; do for pop3 in YRI ; do cat /home/marlam/Indians/PBS/${pop1}Target/\*Bins20\_5.txt.txt > /home/marlam/Indians/PBS/${pop1}Target/PBSNormalizado\_${pop1}\_${pop2}\_${pop3}\_Bins.txt ; done ; done ; done

#### Indian1kgp\_HG x CEU x YRI

cat /media/LDGH/marlama/India/1kgp\_HC/PBS/\*Bins20\_5.txt > /media/LDGH/marlama/India/1kgp\_HC/PBS/PBSNormalizado\_India1kgp\_CEU\_YRI\_Bins.txt

# Methods based on LD

## Phasing Process

### All X All X CEU/YRI

#### Check DNA strand

plink --bfile /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated --recode vcf --out /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated

python /project/6006002/marlama/CheckVCF/checkVCF.py -r /project/6006002/marlama/CheckVCF/hs37d5.fa -o CPTP\_teste\_Strand\_India /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated.vcf

[ 66963 ] Inconsistent reference sites are outputted to [ CPTP\_teste\_Strand\_India.check.ref ]

#### Swap the reference and alternate alleles

/home/marlama/projects/def-eparra/marlama/bcftools/bcftools view -O u /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated.vcf > /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated.bcf

export BCFTOOLS\_PLUGINS=/home/marlama/projects/def-eparra/marlama/bcftools/plugins/

/home/marlama/projects/def-eparra/marlama/bcftools/bcftools +fixref /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated.bcf -Ob -o /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Swap.bcf -- -d -f /project/6006002/marlama/CheckVCF/hs37d5.fa -i /project/6006002/marlama/CheckVCF/All\_20180423.vcf.gz

#### Sort file

/home/marlama/projects/def-eparra/marlama/bcftools/bcftools sort /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Swap.bcf -Ob -o /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Swap\_Sorted.bcf

/home/marlama/projects/def-eparra/marlama/bcftools/bcftools view -O v /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Swap\_Sorted.bcf > /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Swap\_Sorted.vcf

#### Second Check DNA strand

python /project/6006002/marlama/CheckVCF/checkVCF.py -r /project/6006002/marlama/CheckVCF/hs37d5.fa -o /project/6006002/marlama/Indians/Second\_teste\_Strand\_India /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Swap\_Sorted.vcf

### Just India

#### Select Just India

plink --bfile /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup --keep /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/India\_FamilySamples.txt --recode vcf --out /project/6006002/marlama/Indians/JustIndians\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup

#### Swap the reference and alternate alleles

/home/marlama/projects/def-eparra/marlama/bcftools/bcftools view -O u /project/6006002/marlama/Indians/JustIndians\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup.vcf > /project/6006002/marlama/Indians/JustIndians\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup.bcf

export BCFTOOLS\_PLUGINS=/home/marlama/projects/def-eparra/marlama/bcftools/plugins/

/home/marlama/projects/def-eparra/marlama/bcftools/bcftools +fixref /project/6006002/marlama/Indians/JustIndians\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup.bcf -Ob -o /project/6006002/marlama/Indians/JustIndians\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup\_Swap.bcf -- -d -f /project/6006002/marlama/CheckVCF/hs37d5.fa -i /project/6006002/marlama/CheckVCF/All\_20180423.vcf.gz

#### Sort file

/home/marlama/projects/def-eparra/marlama/bcftools/bcftools sort /project/6006002/marlama/Indians/JustIndians\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup\_Swap.bcf -Ob -o /project/6006002/marlama/Indians/JustIndians\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup\_Swap\_Sorted.bcf

/home/marlama/projects/def-eparra/marlama/bcftools/bcftools view -O v /project/6006002/marlama/Indians/JustIndians\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup\_Swap\_Sorted.bcf > /project/6006002/marlama/Indians/JustIndians\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup\_Swap\_Sorted.vcf

#### Second Check DNA strand

python /project/6006002/marlama/CheckVCF/checkVCF.py -r /project/6006002/marlama/CheckVCF/hs37d5.fa -o /project/6006002/marlama/Indians/Second\_teste\_Strand\_IndiaGroup /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup\_Swap\_Sorted.vcf

### CasteTribe x CEU

#### Swap the reference and alternate alleles

for pop in Caste Tribe ; do /home/marlama/projects/def-eparra/marlama/bcftools/bcftools view -O u /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Just${pop}Group.vcf > /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Just${pop}Group.bcf ; done

export BCFTOOLS\_PLUGINS=/home/marlama/projects/def-eparra/marlama/bcftools/plugins/

for pop in Caste Tribe ; do /home/marlama/projects/def-eparra/marlama/bcftools/bcftools +fixref /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Just${pop}Group.bcf -Ob -o /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Just${pop}Group\_Swap.bcf -- -d -f /project/6006002/marlama/CheckVCF/hs37d5.fa -i /project/6006002/marlama/CheckVCF/All\_20180423.vcf.gz ; done

#### Sort file

for pop in Caste Tribe ; do /home/marlama/projects/def-eparra/marlama/bcftools/bcftools sort /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Just${pop}Group\_Swap.bcf -Ob -o /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Just${pop}Group\_Swap\_Sorted.bcf ; done

for pop in Caste Tribe ; do /home/marlama/projects/def-eparra/marlama/bcftools/bcftools view -O v /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Just${pop}Group\_Swap\_Sorted.bcf > /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Just${pop}Group\_Swap\_Sorted.vcf ; done

#### Second Check DNA strand

for pop in Caste Tribe ; do python /project/6006002/marlama/CheckVCF/checkVCF.py -r /project/6006002/marlama/CheckVCF/hs37d5.fa -o /project/6006002/marlama/Indians/Second\_teste\_Strand\_${pop}Group /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Just${pop}Group\_Swap\_Sorted.vcf ; done

### Indian1lgp x CEU x YRI

#### Check DNA strand

plink --bfile /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup --recode vcf --out /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup

python /project/6006002/marlama/CheckVCF/checkVCF.py -r /project/6006002/marlama/CheckVCF/hs37d5.fa -o CPTP\_teste\_Strand\_India1kgp /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup.vcf

#### Swap the reference and alternate alleles

/home/marlama/projects/def-eparra/marlama/bcftools/bcftools view -O u /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup.vcf > /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup.bcf

export BCFTOOLS\_PLUGINS=/home/marlama/projects/def-eparra/marlama/bcftools/plugins/

/home/marlama/projects/def-eparra/marlama/bcftools/bcftools +fixref /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup.bcf -Ob -o /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup\_Swap.bcf -- -d -f /project/6006002/marlama/CheckVCF/hs37d5.fa -i /project/6006002/marlama/CheckVCF/All\_20180423.vcf.gz

#### Sort file

/home/marlama/projects/def-eparra/marlama/bcftools/bcftools sort /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup\_Swap.bcf -Ob -o /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup\_Swap\_Sorted.bcf

/home/marlama/projects/def-eparra/marlama/bcftools/bcftools view -O v /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup\_Swap\_Sorted.bcf > /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup\_Swap\_Sorted.vcf

#### Second Check DNA strand

python /project/6006002/marlama/CheckVCF/checkVCF.py -r /project/6006002/marlama/CheckVCF/hs37d5.fa -o /project/6006002/marlama/Indians/Second\_teste\_Strand\_India1kgpGroup /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup\_Swap\_Sorted.vcf

### Indian1lgpHC x CEU x YRI

Prepare to Phase

for chr in $(seq 1 22) ; do bgzip /home/marlama/scratch/1kgp\_HC/PBS/OriginalFiles/1kgp\_HC\_chr${chr}\_India\_CEU\_YRI.vcf ; done

for chr in $(seq 1 22) ; do tabix -p vcf /home/marlama/scratch/1kgp\_HC/PBS/OriginalFiles/1kgp\_HC\_chr${chr}\_India\_CEU\_YRI.vcf.gz ; done

Phasing process shapeit4

for chr in $(seq 1 22) ; do shapeit4 --input /home/marlama/scratch/1kgp\_HC/PBS/OriginalFiles/1kgp\_HC\_chr${chr}\_India\_CEU\_YRI.vcf.gz --map /home/marlama/projects/def-eparra/marlama/GeneticMap/chr${chr}.b37.gmap.gz --thread 40 --region ${chr} --mcmc-iterations 10b,1p,1b,1p,1b,1p,1b,1p,10m --output /home/marlama/scratch/1kgp\_HC/xpEHH/Indians\_CEU\_YRI\_Shapeit4\_chr${chr}.vcf.gz ; done

for chr in $(seq 1 22) ; do shapeit4 --input /home/marlama/scratch/1kgp\_HC/1kgp\_HC\_chr${chr}\_JustGIH\_ITU.vcf.gz --map /home/marlama/projects/def-eparra/marlama/GeneticMap/chr${chr}.b37.gmap.gz --thread 20 --region ${chr} --mcmc-iterations 10b,1p,1b,1p,1b,1p,1b,1p,10m --output /home/marlama/scratch/1kgp\_HC/Indians\_1kgpHC\_Shapeit4\_chr${chr}.vcf.gz ; done

### All x All

#### Fit family/ID Names

plink2 --vcf /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/all\_IndiansEachPop.vcf --export haps --out /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/IndiansAllPop\_Phased

Fix population name .haps = exel

plink2 --vcf /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/all\_IndiansEachPop.vcf --update-ids /project/6006002/marlama/Indians/ListaDeReferencia\_EachPop\_IndiaGroup.txt --make-bed --out /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/all\_IndiaGroup

### India x CEU

#### Split for Population

for pop1 in India CEU ; do awk -v pop1="$pop1" '{if($1~ pop1){print $1"\t"$2}}' /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/IndiaGroup\_Phased.sample > /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/${pop1}\_FamilySamples.txt ; done

for pop1 in India CEU ; do plink2 --bfile /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/all\_IndiaGroup --keep /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/${pop1}\_FamilySamples.txt --recode vcf --out /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/${pop1}Group\_Phased ; done

## 

#### Split for chromosomes

for pop1 in India CEU ; do for i in $(seq 1 22) ; do plink2 --vcf /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/${pop1}Group\_Phased.vcf --chr ${i} --recode vcf --out /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/${pop1}Group\_Phased\_chr${i} ; done ; done

### Caste X CEU

plink --bfile /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/all\_IndiaGroup --update-ids /project/6006002/marlama/Indians/ListaDeCorrespondencia\_India\_TribeCaste.txt --make-bed --out /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/CasteTribeGroup/CasteTribeGroup\_Phased

#### Split for Population

for pop1 in Tribe Caste ; do awk -v pop1="$pop1" '{if($1~ pop1){print $1"\t"$2}}' /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/CasteTribeGroup/CasteTribeGroup\_Phased.fam > /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/CasteTribeGroup/${pop1}\_FamilySamples.txt ; done

for pop1 in Tribe Caste ; do plink --bfile /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/CasteTribeGroup/CasteTribeGroup\_Phased --keep /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/CasteTribeGroup/${pop1}\_FamilySamples.txt --recode vcf --out /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/CasteTribeGroup/${pop1}Group\_Phased ; done

## 

#### Split for chromosomes

for pop1 in Tribe Caste ; do for i in $(seq 1 22) ; do plink --vcf /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/CasteTribeGroup/${pop1}Group\_Phased.vcf --chr ${i} --recode vcf --out /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/CasteTribeGroup/${pop1}Group\_Phased\_chr${i} ; done ; done

### Indian1kgp X CEU

#### Fit family/ID Names

for i in $(seq 1 22) ; do plink2 --vcf /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/all.vcf --export haps --out /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/Indians1kgpAllPop\_Phased\_chr${i} ; done

Fix population name .haps = exel

for chr in $(seq 2 22); do awk '{print $1, $2, $3, $4}' /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/Indians1kgpAllPop\_Phased\_chr${i}.sample > /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/Indians1kgpAllPop\_Phased\_chr${i}.sample ; done

plink2 --vcf /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/all.vcf --update-ids /project/6006002/marlama/Indians/ListaDeCorrespondencia\_Indian1kgp.txt --make-bed --out /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/Indians1kgpAllPop\_Phased

#### Split for Population

for pop1 in Indian1kgp CEU ; do awk -v pop1="$pop1" '{if($1~ pop1){print $1"\t"$2}}' /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/Indians1kgpAllPop\_Phased\_chr1.sample > /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/${pop1}\_FamilySamples.txt ; done

for pop1 in Indian1kgp CEU ; do plink2 --bfile /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/Indians1kgpAllPop\_Phased --keep /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/${pop1}\_FamilySamples.txt --recode vcf --out /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/${pop1}Group\_Phased ; done

#### Split for chromosomes

for pop1 in Indian1kgp CEU ; do for i in $(seq 1 22) ; do plink --vcf /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/${pop1}Group\_Phased.vcf --chr ${i} --recode vcf --out /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/${pop1}Group\_Phased\_chr${i} ; done ; done

### India1kgp\_HC x CEU

for i in $(seq 1 22) ; do plink2 --vcf /home/marlama/scratch/1kgp\_HC/xpEHH/Indians\_CEU\_YRI\_Shapeit4\_chr${i}.vcf.gz --make-bed --out /home/marlama/scratch/1kgp\_HC/xpEHH/Indians\_CEU\_YRI\_Shapeit4\_chr${i} ; done

for i in $(seq 1 22) ; do plink2 --bfile /home/marlama/scratch/1kgp\_HC/xpEHH/Indians\_CEU\_YRI\_Shapeit4\_chr${i} --update-ids /home/marlama/scratch/1kgp\_HC/PBS/OriginalFiles/India\_CEU\_YRI\_UpdatePopName\_Shapeit.txt --make-bed --out /home/marlama/scratch/1kgp\_HC/xpEHH/Indians\_CEU\_YRI\_Shapeit4\_chr${i}\_Updateids ; done

#### Split for Population

for pop1 in India1kgp CEU ; do awk -v pop1="$pop1" '{if($1~ pop1){print $1"\t"$2}}' /home/marlama/scratch/1kgp\_HC/xpEHH/Indians\_CEU\_YRI\_Shapeit4\_chr22\_Updateids.fam > /home/marlama/scratch/1kgp\_HC/xpEHH/${pop1}\_FamilySamples.txt ; done

Shapeit4

for i in $(seq 1 22) ; do plink --bfile /home/marlama/scratch/1kgp\_HC/1kgp\_HC\_MosaicQC/1kgpHC\_chr${i}\_Autosomal --keep /home/marlama/scratch/1kgp\_HC/1kgp\_HC\_MosaicQC/CEU\_samples.txt --recode vcf --out /home/marlama/scratch/1kgp\_HC/1kgp\_HC\_MosaicQC/1kgpHC\_chr${i}\_JustCEU ; done

for chr in $(seq 1 22) ; do bgzip /home/marlama/scratch/1kgp\_HC/1kgp\_HC\_MosaicQC/1kgpHC\_chr${chr}\_JustCEU.vcf ; done

for chr in $(seq 1 22) ; do tabix -p vcf /home/marlama/scratch/1kgp\_HC/1kgp\_HC\_MosaicQC/1kgpHC\_chr${chr}\_JustCEU.vcf.gz ; done

module load StdEnv/2020 gcc/9.3.0 openmpi/4.0.3 shapeit4/4.2.1

for chr in $(seq 1 22) ; do shapeit4 --input /home/marlama/scratch/1kgp\_HC/1kgp\_HC\_MosaicQC/1kgpHC\_chr${chr}\_JustCEU.vcf.gz --map /home/marlama/projects/def-eparra/marlama/GeneticMap/chr${chr}.b37.gmap.gz --thread 200 --region ${chr} --mcmc-iterations 10b,1p,1b,1p,1b,1p,1b,1p,10m --output /home/marlama/scratch/1kgp\_HC/1kgp\_HC\_MosaicQC/1kgpHC\_chr${chr}\_JustCEU\_Shapeit4.vcf.gz ; done

### Make .map file

for chr in $(seq 1 22) ; do plink --vcf /home/marlama/scratch/1kgp\_HC/PhasedData/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.vcf --make-bed --out /home/marlama/scratch/1kgp\_HC/PhasedData/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4 ; done

for chr in $(seq 1 22) ; do awk -v chr="$chr" '{if($1 == chr){print $1"\t"$2"\t"$4"\t"$4}}' /home/marlama/scratch/1kgp\_HC/PhasedData/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.bim > /home/marlama/scratch/1kgp\_HC/PhasedData/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.map ; done

## xpEHH

CEDAR

### Run xpEHH

#### India X CEU

for chr in $(seq 1 22) ; do /project/6006002/marlama/selscan-master/src/selscan --xpehh --pmap --vcf /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/IndiaGroup/IndiaGroup\_Phased\_chr${chr}.vcf --vcf-ref /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/CEUGroup/CEUGroup\_Phased\_chr${chr}.vcf --map /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/GeneticMap\_Indians\_chr${chr}.map --out /project/6006002/marlama/Indians/xpEHH/xpEHH\_India\_CEU\_chr${chr} ; done

#### Tribe X CEU

for chr in $(seq 1 22) ; do /project/6006002/marlama/selscan-master/src/selscan --xpehh --pmap --vcf /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/TribeGroup/TribeGroup\_Phased\_chr${chr}.vcf --vcf-ref /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/CEUGroup/CEUGroup\_Phased\_chr${chr}.vcf --map /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/GeneticMap\_Indians\_chr${chr}.map --out /project/6006002/marlama/Indians/xpEHH/Tribe\_CEU/xpEHH\_Tribe\_CEU\_chr${chr} ; done

#### Caste X CEU

for chr in $(seq 1 22) ; do /project/6006002/marlama/selscan-master/src/selscan --xpehh --pmap --vcf /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/CasteGroup/CasteGroup\_Phased\_chr${chr}.vcf --vcf-ref /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/CEUGroup/CEUGroup\_Phased\_chr${chr}.vcf --map /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/GeneticMap\_Indians\_chr${chr}.map --out /project/6006002/marlama/Indians/xpEHH/Caste\_CEU/xpEHH\_Caste\_CEU\_chr${chr} ; done

#### Indian1kgp X CEU

for chr in $(seq 1 22) ; do /project/6006002/marlama/selscan-master/src/selscan --xpehh --pmap --vcf /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/Indian1kgpGroup\_Phased\_chr${i}.vcf --vcf-ref /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/CEUGroup\_Phased\_chr${i}.vcf --map /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/GeneticMap\_Indians\_chr${chr}.map --out /project/6006002/marlama/Indians/xpEHH/Indian1kgp\_CEU/xpEHH\_Indian1kgp\_CEU\_chr${chr} ; done

#### Indian1kgpHC X CEU

for chr in 1 ; do /project/6006002/marlama/selscan-master/src/selscan --xpehh --pmap --vcf /home/marlama/scratch/1kgp\_HC/PhasedData/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.vcf --vcf-ref /home/marlama/scratch/1kgp\_HC/PhasedData/1kgpHC\_chr${chr}\_JustCEU\_Shapeit4.vcf --map /home/marlama/scratch/1kgp\_HC/PhasedData/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.map --out /home/marlama/scratch/1kgp\_HC/xpEHH/xpEHH\_Indian1kgpHC\_CEU\_chr${chr} ; done

### Normalize xpEHH

#### India X CEU

for chr in $(seq 1 22) ; do /project/6006002/marlama/selscan-master/src/norm --xpehh --bins 20 --files /project/6006002/marlama/Indians/xpEHH/xpEHH\_India\_CEU\_chr${chr}.xpehh.out ; done

#### Caste X CEU

for chr in $(seq 1 22) ; do /project/6006002/marlama/selscan-master/src/norm --xpehh --bins 20 --files /project/6006002/marlama/Indians/xpEHH/Caste\_CEU/xpEHH\_Caste\_CEU\_chr${chr}.xpehh.out ; done

#### Tribe X CEU

for chr in $(seq 1 22) ; do /project/6006002/marlama/selscan-master/src/norm --xpehh --bins 20 --files /project/6006002/marlama/Indians/xpEHH/Tribe\_CEU/xpEHH\_Tribe\_CEU\_chr${chr}.xpehh.out ; done

#### Indian1kgp X CEU

for chr in $(seq 1 22) ; do /project/6006002/marlama/selscan-master/src/norm --xpehh --bins 20 --files /project/6006002/marlama/Indians/xpEHH/Indian1kgp\_CEU/xpEHH\_Indian1kgp\_CEU\_chr${chr}.xpehh.out ; done

### Addiction chromosome column

#### India X CEU

for chr in $(seq 1 22) ; do awk -v chr="$chr" '{print $1"\t"chr"\t"$2"\t"$8"\t"$9}' /project/6006002/marlama/Indians/xpEHH/xpEHH\_India\_CEU\_chr${chr}.xpehh.out.norm > /project/6006002/marlama/Indians/xpEHH/xpEHH\_India\_CEU\_chr${chr}.FinalInformations.txt ; done

#### CasteTribe X CEU

for pop in Caste Tribe ; do for chr in $(seq 1 22) ; do awk -v chr="$chr" '{print $1"\t"chr"\t"$2"\t"$8"\t"$9}' /project/6006002/marlama/Indians/xpEHH/${pop}\_CEU/xpEHH\_${pop}\_CEU\_chr${chr}.xpehh.out.norm > /project/6006002/marlama/Indians/xpEHH/${pop}\_CEU/xpEHH\_${pop}\_CEU\_chr${chr}.FinalInformations.txt ; done ; done

#### Indian1kgp X CEU

for chr in $(seq 1 22) ; do awk -v chr="$chr" '{print $1"\t"chr"\t"$2"\t"$8"\t"$9}' /project/6006002/marlama/Indians/xpEHH/Indian1kgp\_CEU/xpEHH\_Indian1kgp\_CEU\_chr${chr}.xpehh.out.norm > /project/6006002/marlama/Indians/xpEHH/Indian1kgp\_CEU/xpEHH\_Indian1kgp\_CEU\_chr${chr}.FinalInformations.txt ; done

### Run sliding window for chromosome

#### All x CEU

for i in $(seq 1 22) ; do for pop1 in India Caste Tribe Indian1kgp ; do Rscript /home/marlam/Scripts\_PBS/JanelasDeslizantesForChr.R /media/LDGH\_STORAGE/marlama/India/xpEHH/${pop1}\_CEU/xpEHH\_${pop1}\_CEU\_chr${i}.FinalInformations.txt /media/LDGH\_STORAGE/marlama/India/xpEHH/${pop1}\_CEU/xpEHH\_${pop1}\_CEU\_chr${i}.FinalInformations\_SlidingWindow.txt ; done ; done

for i in $(seq 1 22) ; do for pop1 in India Caste Tribe Indian1kgp ; do Rscript /home/marlam/Scripts\_PBS/JanelasDeslizantesForChr.R /media/LDGH\_STORAGE/marlama/India/xpEHH/${pop1}\_CEU/xpEHH\_${pop1}\_CEU\_chr${i}.FinalInformations.txt /media/LDGH\_STORAGE/marlama/India/xpEHH/${pop1}\_CEU/xpEHH\_${pop1}\_CEU\_chr${i}.FinalInformations\_SlidingWindow\_80SNP.txt ; done ; done

### Correr Bins

for i in $(seq 1 22) ; do for pop1 in India Caste Tribe Indian1kgp ; do perl /home/marlam/Scripts\_PBS/Bins\_PBSnForChromosome.pl --bins /media/LDGH\_STORAGE/marlama/India/xpEHH/${pop1}\_CEU/xpEHH\_${pop1}\_CEU\_chr${i}.FinalInformations\_SlidingWindow\_80SNP.txt --pbs /media/LDGH\_STORAGE/marlama/India/xpEHH/${pop1}\_CEU/xpEHH\_${pop1}\_CEU\_chr${i}.FinalInformations.txt --output /media/LDGH\_STORAGE/marlama/India/xpEHH/${pop1}\_CEU/xpEHH\_${pop1}\_CEU\_chr${i}\_Bins20\_5.txt ; done ; done

### Merge Chromossome Files

#### All X CEU

for pop1 in India Caste Tribe Indian1kgp ; do cat /media/LDGH\_STORAGE/marlama/India/xpEHH/${pop1}\_CEU/\*\_Bins20\_5.txt.txt > /media/LDGH\_STORAGE/marlama/India/xpEHH/${pop1}\_CEU/xpEHH\_${pop1}\_CEU\_.xpehh\_Bins20\_5.txt ; done

## iHS

CEDAR

### Define allele ancestral and derived

#### Split by Chromossome and Transform in haps sample

##### Just India

for chr in $(seq 1 22); do plink2 --vcf /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all.vcf --chr ${chr} --export haps --out /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all\_JustIndia\_chr${chr} ; done

fit .samples on exel (separate ID in “pop ID”)

for chr in $(seq 2 22); do awk '{print $1, $2, $3, $4}' /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all\_JustIndia\_chr1.sample > /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all\_JustIndia\_chr${chr}.sample ; done

##### CasteTribe

for pop in Caste Tribe ; do gunzip /project/6006002/marlama/Indians/Phasing${pop}.vcfs/all.vcf\* ; done

for pop in Caste Tribe ; do for chr in $(seq 1 22); do plink2 --vcf /project/6006002/marlama/Indians/Phasing${pop}.vcfs/all.vcf --chr ${chr} --export haps --out /project/6006002/marlama/Indians/Phasing${pop}.vcfs/all\_Just${pop}\_chr${chr} ; done ; done

fit .samples on exel (separate ID in “pop ID”)

for pop in Caste Tribe ; do for chr in $(seq 2 22); do awk '{print $1, $2, $3, $4}' /project/6006002/marlama/Indians/Phasing${pop}.vcfs/all\_Just${pop}\_chr1.sample > /project/6006002/marlama/Indians/Phasing${pop}.vcfs/all\_Just${pop}\_chr${chr}.sample ; done ; done

##### India1kgp

for pop in India1kgp ; do gunzip /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/all.vcf\* ; done

for pop in India1kgp ; do for chr in $(seq 1 22); do plink2 --vcf /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/all.vcf --chr ${chr} --export haps --out /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/Indians1kgpAllPop\_Phased\_chr${chr} ; done ; done

fit .samples on exel (separate ID in “pop ID”)

for chr in $(seq 2 22); do awk '{print $1, $2, $3, $4}' /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/Indians1kgpAllPop\_Phased\_chr1.sample > /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/Indians1kgpAllPop\_Phased\_chr${chr}.sample ; done

##### All Datasets second try

for pop in Caste Tribe India1kgp ; do for chr in $(seq 1 22); do /home/marlam/Programs/plink2 --vcf /media/LDGH\_STORAGE/marlama/India/Just${pop}/all\_Just${pop}.vcf --chr ${chr} --export haps --out /media/LDGH\_STORAGE/marlama/India/Just${pop}/all\_Just${pop}\_chr${chr} ; done ; done

for pop in Caste Tribe India1kgp ; do for chr in $(seq 2 22); do awk '{print $1, $2, $3, $4}' /media/LDGH\_STORAGE/marlama/India/Just${pop}/all\_Just${pop}\_chr1.sample > /media/LDGH\_STORAGE/marlama/India/Just${pop}/all\_Just${pop}\_chr${chr}.sample ; done ; done

#### Format file from correspondence

##### Just India

for chr in $(seq 1 22); do awk '{print $1, $2, "0", $3, $4, $5}' OFS="\t" /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all\_JustIndia\_chr${chr}.haps > /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all\_JustIndia\_chr${chr}.bim ; done

##### CasteTribe

for pop in Caste Tribe ; do for chr in $(seq 1 22); do awk '{print $1, $2, "0", $3, $4, $5}' OFS="\t" /project/6006002/marlama/Indians/Phasing${pop}.vcfs/all\_Just${pop}\_chr${chr}.haps > /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.bim ; done ; done

##### India1kgp

for chr in $(seq 1 22); do awk '{print $1, $2, "0", $3, $4, $5}' OFS="\t" /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/Indians1kgpAllPop\_Phased\_chr${chr}.haps > /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indians1kgpAllPop\_Phased\_chr${chr}.bim ; done

##### All Datasets second try

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do awk '{print $1, $2, "0", $3, $4, $5}' OFS="\t" /media/LDGH\_STORAGE/marlama/India/Just${pop}/all\_Just${pop}\_chr${chr}.haps > /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.bim ; done ; done

##### India1kgp\_HC

for chr in $(seq 1 22); do awk '{print $1, $2, "0", $3, $4, $5}' OFS="\t" /home/marlama/scratch/1kgp\_HC/PhasedData/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.haps > /home/marlama/scratch/1kgp\_HC/CLR/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.bim ; done

#### Make table chr/pos/ancestral allele

LDGH

##### Just India

for chr in $(seq 1 22); do awk 'NR==FNR{a[$1]=$2;next} {if($2 in a) {if(a[$2]==$5||a[$2]==$6) {if($5=="A"&&$6=="T"||$5=="T"&&$6=="A"||$5=="C"&&$6=="G"||$5=="G"&&$6=="C") {print $1,$4,tolower(a[$2])} else {print $1,$4,a[$2]}} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="A"&&$5=="T"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="A"&&$6 ="T") {print $1,$4,"T"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="T"&&$5=="A"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="T"&&$6= "A") {print $1,$4,"A"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="C"&&$5=="G"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="C"&&$6=="G") {print $1,$4,"G"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="G"&&$5=="C"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="G"&&$6=="C") {print $1,$4,"C"}} else print $1,$4,tolower($5)}' OFS="\t" /media/LDGH/Backup\_17Maio2019/home/home/natives\_project/Working\_datasets/variation\_ensembl\_ancAlleles.txt /media/LDGH/marlama/Indians/AncestralAllele/all\_JustIndia\_chr${chr}.bim > /media/LDGH/marlama/Indians/AncestralAllele/all\_JustIndia\_chr${chr}.tab ; done

##### CasteTribe

for pop in Caste Tribe ; do for chr in $(seq 1 22); do awk 'NR==FNR{a[$1]=$2;next} {if($2 in a) {if(a[$2]==$5||a[$2]==$6) {if($5=="A"&&$6=="T"||$5=="T"&&$6=="A"||$5=="C"&&$6=="G"||$5=="G"&&$6=="C") {print $1,$4,tolower(a[$2])} else {print $1,$4,a[$2]}} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="A"&&$5=="T"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="A"&&$6 ="T") {print $1,$4,"T"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="T"&&$5=="A"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="T"&&$6= "A") {print $1,$4,"A"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="C"&&$5=="G"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="C"&&$6=="G") {print $1,$4,"G"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="G"&&$5=="C"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="G"&&$6=="C") {print $1,$4,"C"}} else print $1,$4,tolower($5)}' OFS="\t" /media/LDGH\_STORAGE/Nativos/Backup\_22outubro2020/variation\_ensembl\_ancAlleles.txt /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.bim > /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.tab ; done ; done

##### India1kgp

scp [marlama@cedar.computecanada.ca](mailto:marlama@cedar.computecanada.ca):/project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indians1kgpAllPop\_Phased\_chr\* /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/India1kgp/

for chr in $(seq 1 22); do awk 'NR==FNR{a[$1]=$2;next} {if($2 in a) {if(a[$2]==$5||a[$2]==$6) {if($5=="A"&&$6=="T"||$5=="T"&&$6=="A"||$5=="C"&&$6=="G"||$5=="G"&&$6=="C") {print $1,$4,tolower(a[$2])} else {print $1,$4,a[$2]}} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="A"&&$5=="T"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="A"&&$6 ="T") {print $1,$4,"T"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="T"&&$5=="A"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="T"&&$6= "A") {print $1,$4,"A"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="C"&&$5=="G"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="C"&&$6=="G") {print $1,$4,"G"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="G"&&$5=="C"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="G"&&$6=="C") {print $1,$4,"C"}} else print $1,$4,tolower($5)}' OFS="\t" /media/LDGH\_STORAGE/Nativos/Backup\_22outubro2020/variation\_ensembl\_ancAlleles.txt /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/India1kgp/Indians1kgpAllPop\_Phased\_chr${chr}.bim > /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/India1kgp/Indians1kgpAllPop\_Phased\_chr${chr}.tab ; done

##### All Datasets second try

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do awk 'NR==FNR{a[$1]=$2;next} {if($2 in a) {if(a[$2]==$5||a[$2]==$6) {if($5=="A"&&$6=="T"||$5=="T"&&$6=="A"||$5=="C"&&$6=="G"||$5=="G"&&$6=="C") {print $1,$4,tolower(a[$2])} else {print $1,$4,a[$2]}} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="A"&&$5=="T"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="A"&&$6 ="T") {print $1,$4,"T"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="T"&&$5=="A"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="T"&&$6= "A") {print $1,$4,"A"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="C"&&$5=="G"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="C"&&$6=="G") {print $1,$4,"G"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="G"&&$5=="C"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="G"&&$6=="C") {print $1,$4,"C"}} else print $1,$4,tolower($5)}' OFS="\t" /media/LDGH\_STORAGE/Nativos/Backup\_22outubro2020/variation\_ensembl\_ancAlleles.txt /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.bim > /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.tab ; done ; done

##### India1kgp\_HC

for chr in $(seq 1 22); do awk 'NR==FNR{a[$1]=$2;next} {if($2 in a) {if(a[$2]==$5||a[$2]==$6) {if($5=="A"&&$6=="T"||$5=="T"&&$6=="A"||$5=="C"&&$6=="G"||$5=="G"&&$6=="C") {print $1,$4,tolower(a[$2])} else {print $1,$4,a[$2]}} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="A"&&$5=="T"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="A"&&$6 ="T") {print $1,$4,"T"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="T"&&$5=="A"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="T"&&$6= "A") {print $1,$4,"A"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="C"&&$5=="G"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="C"&&$6=="G") {print $1,$4,"G"} else if (a[$2]!=$5&&a[$2]!=$6&&a[$2]=="G"&&$5=="C"||a[$2]!=$5&&a[$2]!=$6&&a[$2]=="G"&&$6=="C") {print $1,$4,"C"}} else print $1,$4,tolower($5)}' OFS="\t" /home/marlama/projects/def-eparra/marlama/Ensembl\_AncestralAllele/variation\_ensembl\_ancAlleles.txt /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.bim > /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.tab ; done

#### Make header

CEDAR

##### Just India

for chr in $(seq 1 22); do echo "##INFO=<ID=AA,Number=1,Type=String,Description=Ancestral Allele>" > /project/6006002/marlama/Indians/AncestralAlleleFiles/all\_JustIndia\_chr${chr}.hdr ; done

##### Caste/Tribe

for pop in Caste Tribe ; do for chr in $(seq 1 22); do echo "##INFO=<ID=AA,Number=1,Type=String,Description=Ancestral Allele>" > /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.hdr ; done ; done

##### India1kgp

for chr in $(seq 1 22); do echo "##INFO=<ID=AA,Number=1,Type=String,Description=Ancestral Allele>" > /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indians1kgpAllPop\_Phased\_chr${chr}.hdr ; done

##### All Datasets second try

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do echo "##INFO=<ID=AA,Number=1,Type=String,Description=Ancestral Allele>" > /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.hdr ; done ; done

##### India1kgp\_HC

for chr in $(seq 1 22); do echo "##INFO=<ID=AA,Number=1,Type=String,Description=Ancestral Allele>" > /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.hdr ; done

#### Make vcf with Ancestral Allele Correspondence

CEDAR

##### Just India

for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/bgzip /project/6006002/marlama/Indians/AncestralAlleleFiles/all\_JustIndia\_chr${chr}.tab ; done

for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/tabix -s 1 -b 2 -e 2 /project/6006002/marlama/Indians/AncestralAlleleFiles/all\_JustIndia\_chr${chr}.tab.gz ; done

for chr in $(seq 1 22); do plink --vcf all\_JustIndia\_chr${chr}.vcf --update-ids ListadeCorrespondencia\_indiaID\_india\_id.txt --recode vcf --out all\_JustIndia\_chr${chr}\_popInfo ; done

for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/bgzip /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all\_JustIndia\_chr${chr}\_popInfo.vcf ; done

for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/tabix -p vcf /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all\_JustIndia\_chr${chr}\_popInfo.vcf.gz ; done

for chr in $(seq 1 22); do /project/6006002/marlama/bcftools/bcftools annotate -a /project/6006002/marlama/Indians/AncestralAlleleFiles/all\_JustIndia\_chr${chr}.tab.gz -h /project/6006002/marlama/Indians/AncestralAlleleFiles/all\_JustIndia\_chr${chr}.hdr -c CHROM,POS,AA /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all\_JustIndia\_chr${chr}\_popInfo.vcf.gz -o /project/6006002/marlama/Indians/AncestralAlleleFiles/all\_JustIndia\_AncestralAllele\_chr${chr}\_popInfo.vcf ; done

for chr in $(seq 1 22) ; do /project/6006002/marlama/htslib-1.10.2/bgzip -d /project/6006002/marlama/Indians/AncestralAlleleFiles/all\_JustIndia\_chr${chr}.tab.gz ; done

for chr in $(seq 1 22) ; do startCHR=$(head -n 1 /project/6006002/marlama/Indians/AncestralAlleleFiles/all\_JustIndia\_chr${chr}.tab | awk '{print $2}') ; done

for chr in $(seq 1 22) ; do endCHR=$(tail -n 1 /project/6006002/marlama/Indians/AncestralAlleleFiles/all\_JustIndia\_chr${chr}.tab | awk '{print $2}') ; done

for chr in $(seq 1 22) ; do contigLength=$((endCHR-startCHR)) ; done

for chr in $(seq 1 22) ; do sed "s/##contig=<ID=.\*/##contig=<ID=${chr},length=${contigLength}>/g" /project/6006002/marlama/Indians/AncestralAlleleFiles/all\_JustIndia\_AncestralAllele\_chr${chr}\_popInfo.vcf > /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all\_JustIndia\_AncestralAlleleFinal\_chr${chr}\_popInfo.vcf ; done

for chr in $(seq 1 22) ; do /project/6006002/marlama/htslib-1.10.2/bgzip /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all\_JustIndia\_AncestralAlleleFinal\_chr${chr}\_popInfo.vcf ; done

for chr in $(seq 1 22) ; do /project/6006002/marlama/htslib-1.10.2/tabix -p vcf /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all\_JustIndia\_AncestralAlleleFinal\_chr${chr}\_popInfo.vcf.gz ; done

for chr in $(seq 1 22) ; do startCHR=$(head -n 1 /project/6006002/marlama/Indians/AncestralAlleleFiles/all\_JustIndia\_chr${chr}.tab | awk '{print $2}') ; endCHR=$(tail -n 1 /project/6006002/marlama/Indians/AncestralAlleleFiles/all\_JustIndia\_chr${chr}.tab | awk '{print $2}') ; contigLength=$((endCHR-startCHR)) ; sed "s/##contig=<ID=.\*/##contig=<ID=${chr},length=${contigLength}>/g" /project/6006002/marlama/Indians/AncestralAlleleFiles/all\_JustIndia\_AncestralAllele\_chr${chr}\_popInfo.vcf > /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all\_JustIndia\_AncestralAlleleFinal\_chr${chr}\_popInfo.vcf ; done

##### Caste/Tribe

for pop in Caste Tribe ; do for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/bgzip /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.tab ; done; done

for pop in Caste Tribe ; do for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/tabix -s 1 -b 2 -e 2 /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.tab.gz ; done ; done

for pop in Caste Tribe ; do for chr in $(seq 1 22); do plink --vcf /project/6006002/marlama/Indians/Phasing${pop}.vcfs/all.vcf --update-ids /project/6006002/marlama/Indians/Phasing${pop}.vcfs/ListaDeCorrespondencia\_${pop}ID\_Pop\_ID.txt --chr ${chr} --recode vcf --out /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_popInfo ; done ; done

for pop in Caste Tribe ; do for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/bgzip /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_popInfo.vcf ; done ; done

for pop in Caste Tribe ; do for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/tabix -p vcf /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_popInfo.vcf.gz ; done ; done

for pop in Caste Tribe ; do for chr in $(seq 1 22); do /project/6006002/marlama/bcftools/bcftools annotate -a /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.tab.gz -h /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.hdr -c CHROM,POS,AA /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_popInfo.vcf.gz -o /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_AncestralAllele\_chr${chr}\_popInfo.vcf ; done ; done

for pop in Caste Tribe ; do for chr in $(seq 1 22) ; do /project/6006002/marlama/htslib-1.10.2/bgzip -d /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_popInfo.vcf.gz ; done ; done

for pop in Caste Tribe ; do for chr in $(seq 1 22) ; do /project/6006002/marlama/htslib-1.10.2/bgzip -d /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.tab.gz ; done ; done

for pop in Caste Tribe ; do for chr in $(seq 1 22) ; do startCHR=$(head -n 1 /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.tab | awk '{print $2}') ; done ; done

for pop in Caste Tribe ; do for chr in $(seq 1 22) ; do endCHR=$(tail -n 1 /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.tab | awk '{print $2}') ; done ; done

for pop in Caste Tribe ; do for chr in $(seq 1 22) ; do contigLength=$((endCHR-startCHR)) ; done ; done

for pop in Caste Tribe ; do for chr in $(seq 1 22) ; do sed "s/##contig=<ID=.\*/##contig=<ID=${chr},length=${contigLength}>/g" /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_popInfo.vcf > /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_AncestralAllele\_chr${chr}\_popInfo.vcf ; done ; done

for pop in Caste Tribe ; do for chr in $(seq 1 22) ; do /project/6006002/marlama/htslib-1.10.2/bgzip /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_AncestralAllele\_chr${chr}\_popInfo.vcf ; done ; done

for pop in Caste Tribe ; do for chr in $(seq 1 22) ; do /project/6006002/marlama/htslib-1.10.2/tabix -p vcf /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_AncestralAllele\_chr${chr}\_popInfo.vcf.gz ; done ; done

##### India1kgp

for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/bgzip /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indians1kgpAllPop\_Phased\_chr${chr}.tab ; done

for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/tabix -s 1 -b 2 -e 2 /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indians1kgpAllPop\_Phased\_chr${chr}.tab.gz ; done

for chr in $(seq 1 22); do plink --vcf /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/Indian1kgpGroup\_Phased\_chr${chr}.vcf --update-ids /project/6006002/marlama/Indians/PhasingIndians1kgp.vcfs/ListaDeCorrespondencia\_Indians1kgpID\_Pop\_ID.txt --recode vcf --out /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indian1kgpGroup\_Phased\_chr${chr}\_popInfo ; done

for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/bgzip /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indian1kgpGroup\_Phased\_chr${chr}\_popInfo.vcf ; done

for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/tabix -p vcf /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indian1kgpGroup\_Phased\_chr${chr}\_popInfo.vcf.gz ; done

for chr in $(seq 1 22); do /project/6006002/marlama/bcftools/bcftools annotate -a /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indians1kgpAllPop\_Phased\_chr${chr}.tab.gz -h /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indians1kgpAllPop\_Phased\_chr${chr}.hdr -c CHROM,POS,AA /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indian1kgpGroup\_Phased\_chr${chr}\_popInfo.vcf.gz -o /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indian1kgpGroup\_Phased\_AncestralAlleleFile\_chr${chr}\_popInfo.vcf ; done

for chr in $(seq 1 22) ; do /project/6006002/marlama/htslib-1.10.2/bgzip -d /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indians1kgpAllPop\_Phased\_chr${chr}.tab.gz ; done

for chr in $(seq 1 22) ; do startCHR=$(head -n 1 /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indians1kgpAllPop\_Phased\_chr${chr}.tab | awk '{print $2}') ; done

for chr in $(seq 1 22) ; do endCHR=$(tail -n 1 /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indians1kgpAllPop\_Phased\_chr${chr}.tab | awk '{print $2}') ; done

for chr in $(seq 1 22) ; do contigLength=$((endCHR-startCHR)) ; done

for chr in $(seq 1 22) ; do sed "s/##contig=<ID=.\*/##contig=<ID=${chr},length=${contigLength}>/g" /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indian1kgpGroup\_Phased\_AncestralAlleleFile\_chr${chr}\_popInfo.vcf > /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indian1kgpGroup\_Phased\_AncestralAlleleFile\_Final\_chr${chr}\_popInfo.vcf ; done

for chr in $(seq 1 22) ; do /project/6006002/marlama/htslib-1.10.2/bgzip /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indian1kgpGroup\_Phased\_AncestralAlleleFile\_Final\_chr${chr}\_popInfo.vcf ; done

for chr in $(seq 1 22) ; do /project/6006002/marlama/htslib-1.10.2/tabix -p vcf /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indian1kgpGroup\_Phased\_AncestralAlleleFile\_Final\_chr${chr}\_popInfo.vcf.gz ; done

##### All Datasets second try

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do bgzip /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.tab ; done ; done

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do tabix -s 1 -b 2 -e 2 /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.tab.gz ; done ; done

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do bgzip /media/LDGH\_STORAGE/marlama/India/Just${pop}/all\_Just${pop}\_chr${chr}.vcf ; done ; done

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do tabix -p vcf /media/LDGH\_STORAGE/marlama/India/Just${pop}/all\_Just${pop}\_chr${chr}.vcf.gz ; done ; done

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do bcftools annotate -a /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.tab.gz -h /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.hdr -c CHROM,POS,AA /media/LDGH\_STORAGE/marlama/India/Just${pop}/all\_Just${pop}\_chr${chr}.vcf.gz -o /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_AncestralAllele.vcf ; done ; done

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do bgzip -d /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.tab.gz ; done ; done

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do startCHR=$(head -n 1 /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.tab | awk '{print $2}') ; endCHR=$(tail -n 1 /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}.tab | awk '{print $2}'); contigLength=$((endCHR-startCHR)) ; sed "s/##contig=<ID=.\*/##contig=<ID=${chr},length=${contigLength}>/g" /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_AncestralAllele.vcf > /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_AncestralAllele\_Final.vcf ; done ; done

##### India1kgp\_HC

for chr in $(seq 1 22); do bgzip /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.tab ; done

for chr in $(seq 1 22); do tabix -s 1 -b 2 -e 2 /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.tab.gz ; done

for chr in $(seq 1 22); do cp /home/marlama/scratch/1kgp\_HC/PhasedData/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.vcf /home/marlama/scratch/1kgp\_HC/AncestralAllele/ ; done

for chr in $(seq 1 22); do bgzip /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.vcf ; done

for chr in $(seq 1 22); do tabix -p vcf /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.vcf.gz ; done

for chr in $(seq 1 22); do bcftools annotate -a /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.tab.gz -h /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.hdr -c CHROM,POS,AA /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.vcf.gz -o /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4\_AncestralAllele.vcf ; done

for chr in $(seq 1 22); do bgzip -d /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.tab.gz ; done

for chr in $(seq 1 22) ; do startCHR=$(head -n 1 /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.tab | awk '{print $2}') ; endCHR=$(tail -n 1 /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4.tab | awk '{print $2}') ; contigLength=$((endCHR-startCHR)) ; sed "s/##contig=<ID=.\*/##contig=<ID=${chr},length=${contigLength}>/g" /home/marlama/scratch/1kgp\_HC/AncestralAllele/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4\_AncestralAllele.vcf > /home/marlama/scratch/1kgp\_HC/CLR/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4\_AncestralAlleleFinal.vcf ; done

### Make Genetic Map

##### All Datasets second try

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do plink --vcf /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_AncestralAllele\_Final.vcf --cm-map /media/LDGH\_STORAGE/Nativos/Working\_Dataset/GeneticMap1.9M/genetic\_map\_chr${chr}\_combined\_b37.txt ${chr} --make-bed --out /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_AncestralAllele\_Final\_interpolated ; done ; done

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do awk '{print $1"\t"$2"\t"$3"\t"$4}' /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_AncestralAllele\_Final\_interpolated.bim > /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_input\_genetic\_map.selscan ; done ; done

### Run iHS

##### Just India

for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/bgzip -d /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all\_JustIndia\_AncestralAlleleFinal\_chr${chr}\_popInfo.vcf.gz ; done

for chr in $(seq 1 22) ; do /project/6006002/marlama/selscan-master/src/selscan --ihs --pmap --vcf /project/6006002/marlama/Indians/Phasing\_JustIndia.vcfs/all\_JustIndia\_AncestralAlleleFinal\_chr${chr}\_popInfo.vcf.gz --map /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/GeneticMap\_Indians\_chr${chr}.map --out /project/6006002/marlama/Indians/iHS/iHS\_all\_JustIndia\_chr${chr} ; done

plink --vcf Pelotas\_AncestralAllele\_Final\_chr2\_popInfo.vcf.gz --cm-map ../../Genetic\_maps/genetic\_map\_chr2\_combined\_b37.txt 2 --make-bed --out Pelotas\_chr2\_interpolated

awk ‘{print $1”\t”$2”\t”$3”\t”$4}’ Pelotas\_chr2\_interpolated.bim > chr\_2\_input\_genetic\_map.selscan

for chr in 2 ; do selscan --ihs --map chr\_2\_input\_genetic\_map.selscan --vcf Pelotas\_AncestralAllele\_Final\_chr${chr}\_popInfo.vcf.gz --out iHS\_Pelotas\_chr${chr} --threads 40; done

rs4988235

##### Caste/Tribe

for pop in Caste Tribe ; do for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/bgzip -d /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_AncestralAllele\_chr${chr}\_popInfo.vcf.gz ; done ; done

for pop in Caste Tribe ; do for chr in $(seq 1 22) ; do /project/6006002/marlama/selscan-master/src/selscan --ihs --pmap --vcf /project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/all\_Just${pop}\_AncestralAllele\_chr${chr}\_popInfo.vcf --map /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/GeneticMap\_Indians\_chr${chr}.map --out /project/6006002/marlama/Indians/iHS/${pop}/iHS\_all\_${pop}\_chr${chr} ; done ; done

##### India1kgp

for chr in $(seq 1 22); do /project/6006002/marlama/htslib-1.10.2/bgzip -d /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indian1kgpGroup\_Phased\_AncestralAlleleFile\_Final\_chr${chr}\_popInfo.vcf.gz ; done

for chr in $(seq 1 22) ; do /project/6006002/marlama/selscan-master/src/selscan --ihs --pmap --vcf /project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/Indian1kgpGroup\_Phased\_AncestralAlleleFile\_Final\_chr${chr}\_popInfo.vcf --map /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/GeneticMap\_Indians\_chr${chr}.map --out /project/6006002/marlama/Indians/iHS/India1kgp/iHS\_all\_India1kgp\_chr${chr} ; done

##### All Datasets second try

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do /home/marlam/Programs/selscan/src/selscan --ihs --map /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_input\_genetic\_map.selscan --vcf /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_chr${chr}\_AncestralAllele\_Final.vcf --out /media/LDGH\_STORAGE/marlama/India/IHS/${pop}/iHS\_${pop}\_chr${chr}\_SecondTry --threads 20 ; done ; done

### Normalize iHS

##### Just India

for chr in $(seq 1 22) ; do /project/6006002/marlama/selscan-master/src/norm --ihs --bins 20 --files /project/6006002/marlama/Indians/iHS/iHS\_all\_JustIndia\_chr${chr}.ihs.out ; done

##### Caste/Tribe

for pop in Caste Tribe ; do for chr in $(seq 1 22) ; do /project/6006002/marlama/selscan-master/src/norm --ihs --bins 20 --files /project/6006002/marlama/Indians/iHS/${pop}/iHS\_all\_${pop}\_chr${chr}.ihs.out ; done ; done

##### India1kgp

for chr in $(seq 1 22) ; do /project/6006002/marlama/selscan-master/src/norm --ihs --bins 20 --files /project/6006002/marlama/Indians/iHS/India1kgp/iHS\_all\_India1kgp\_chr${chr}.ihs.out ; done

##### All Datasets second try

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do /home/marlam/Programs/selscan/bin/linux/norm --ihs --bins 20 --files /media/LDGH\_STORAGE/marlama/India/IHS/${pop}/iHS\_${pop}\_chr${chr}\_SecondTry.ihs.out ; done ; done

### Addiction Chromosome Information

##### Just India

for chr in $(seq 1 22) ; do awk -v chr="$chr" '{print $1"\t"chr"\t"$2"\t"$3"\t"$4"\t"$5"\t"$6"\t"$7}' /project/6006002/marlama/Indians/iHS/iHS\_all\_JustIndia\_chr${chr}.ihs.out.20bins.norm > /project/6006002/marlama/Indians/iHS/iHS\_all\_JustIndia\_chr${chr}.FinalInformations.txt ; done

##### Caste/Tribe

for pop in Caste Tribe ; do for chr in $(seq 1 22) ; do awk -v chr="$chr" '{print $1"\t"chr"\t"$2"\t"$3"\t"$4"\t"$5"\t"$6"\t"$7}' /project/6006002/marlama/Indians/iHS/${pop}/iHS\_all\_${pop}\_chr${chr}.ihs.out.20bins.norm > /project/6006002/marlama/Indians/iHS/${pop}/iHS\_all\_${pop}\_chr${chr}.FinalInformations.txt ; done ; done

##### India1kgp

for pop in Caste Tribe ; do for chr in $(seq 1 22) ; do awk -v chr="$chr" '{print $1"\t"chr"\t"$2"\t"$3"\t"$4"\t"$5"\t"$6"\t"$7}' /project/6006002/marlama/Indians/iHS/India1kgp/iHS\_all\_India1kgp\_chr${chr}.ihs.out.20bins.norm > /project/6006002/marlama/Indians/iHS/India1kgp/iHS\_all\_India1kgp\_chr${chr}.FinalInformations.txt ; done ; done

##### All Datasets second try

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22); do awk -v chr="$chr" '{print $1"\t"chr"\t"$2"\t"$3"\t"$4"\t"$5"\t"$6"\t"$7"\t"$8}' /media/LDGH\_STORAGE/marlama/India/IHS/${pop}/iHS\_${pop}\_chr${chr}\_SecondTry.ihs.out.20bins.norm > /media/LDGH\_STORAGE/marlama/India/IHS/${pop}/iHS\_${pop}\_chr${chr}\_SecondTry.FinalInformations.txt ; done ; done

### Run sliding window for chromosome

#### All

for i in $(seq 1 22) ; do for pop1 in India Caste Tribe India1kgp ; do Rscript /home/marlam/Scripts\_PBS/JanelasDeslizantesForChr.R /media/LDGH\_STORAGE/marlama/India/IHS/${pop1}/iHS\_${pop1}\_chr${i}\_SecondTry.FinalInformations.txt /media/LDGH\_STORAGE/marlama/India/IHS/${pop1}/iHS\_${pop1}\_chr${i}\_SecondTry.FinalInformations\_SlidingWindow.txt ; done ; done

### Correr Bins

for i in $(seq 1 22) ; do for pop1 in India Caste Tribe India1kgp ; do perl /home/marlam/Scripts\_PBS/Bins\_PBSnForChromosome.pl --bins /media/LDGH\_STORAGE/marlama/India/IHS/${pop1}/iHS\_${pop1}\_chr${i}\_SecondTry.FinalInformations\_SlidingWindow.txt --pbs /media/LDGH\_STORAGE/marlama/India/IHS/${pop1}/iHS\_${pop1}\_chr${i}\_SecondTry.FinalInformations.txt --output /media/LDGH\_STORAGE/marlama/India/IHS/${pop1}/iHS\_${pop1}\_chr${i}\_SecondTry\_Bins20\_5.txt ; done ; done

### Merge Chromosome Files

for pop1 in India Caste Tribe India1kgp ; do cat /media/LDGH\_STORAGE/marlama/India/IHS/${pop1}/\*\_Bins20\_5.txt.txt > /media/LDGH\_STORAGE/marlama/India/IHS/${pop1}/iHS\_${pop1}\_iHS\_Bins20\_5.txt ; done

##### Just India

cat /project/6006002/marlama/Indians/iHS/\*FinalInformations.txt > /project/6006002/marlama/Indians/iHS/iHS\_JustIndia.txt

##### Caste/Tribe

for pop in Caste Tribe ; do cat /project/6006002/marlama/Indians/iHS/${pop}/\*FinalInformations.txt > /project/6006002/marlama/Indians/iHS/${pop}//iHS\_${pop}.txt ; done

##### India1kgp

cat /project/6006002/marlama/Indians/iHS/India1kgp/\*FinalInformations.txt > /project/6006002/marlama/Indians/iHS/India1kgp/iHS\_India1kgp.txt

##### All Datasets second try

for pop in India Caste Tribe India1kgp ; do cat /media/LDGH\_STORAGE/marlama/India/IHS/${pop}/\*FinalInformations.txt > /media/LDGH\_STORAGE/marlama/India/IHS/${pop}/iHS\_${pop}\_SecondTry.txt ; done

# Methods based on the Site Frequency Spectrum

## Tajima’sD

### India X CEU

vcftools --vcf /project/6006002/marlama/Indians/JustIndians\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup.vcf --TajimaD 25000 --out /project/6006002/marlama/Indians/TajimasD/TajimasD\_AllIndians

for window in 100000 200000 ; do vcftools --vcf /project/6006002/marlama/Indians/JustIndians\_QC\_Autossomic\_maf0.05\_Unrelated\_IndiaGroup.vcf --TajimaD ${window} --out /project/6006002/marlama/Indians/TajimasD/TajimasD\_AllIndians\_${window}bp ; done

### 

### CasteTribe X CEU

for pop in Caste Tribe ; do plink --bfile /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_TribeCasteGroup --keep /project/6006002/marlama/Indians/Phasing\_Indians\_EachPop.vcfs/${pop}Group/${pop}\_FamilySamples.txt --recode vcf --out /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Just${pop}Group ; done

for pop in Caste Tribe ; do vcftools --vcf /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Just${pop}Group.vcf --TajimaD 25000 --out /project/6006002/marlama/Indians/TajimasD/TajimasD\_Just${pop} ; done

### India 1kgp Group

vcftools --vcf /project/6006002/marlama/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Indian1kgpGroup.vcf --TajimaD 25000 --out /project/6006002/marlama/Indians/TajimasD/TajimasD\_JustIndian1kgp

for pop in Caste Tribe Indian1kgp ; do for window in 100000 200000 ; do vcftools --vcf /home/marlam/Indians/Indians\_1KGP\_SGDP\_QC\_Autossomic\_maf0.05\_Unrelated\_Just${pop}Group.vcf --TajimaD ${window} --out /home/marlam/Indians/TajimasD/TajimasD\_Just${pop}\_${window}bp ; done ; done

### India 1kgpHC

for chr in $(seq 1 22) ; do vcftools --gzvcf /home/marlama/scratch/1kgp\_HC/1kgp\_HC\_MosaicQC/1kgpHC\_chr${chr}\_JustIndians\_MosaicQC.vcf.gz --TajimaD 100000 --out /home/marlama/scratch/1kgp\_HC/TajimasD/TajimasD\_Indian1kgpHC\_chr${chr} ; done

cat \*.Tajima.D > TajimasD\_Indian1kgpHC.txt

awk '{if(NR==1){print}}' TajimasD\_Indian1kgpHC.txt > TajimasD\_Indian1kgpHC\_Header.txt

awk '{if($1 != "CHROM"){print}}' TajimasD\_Indian1kgpHC.txt > TajimasD\_Indian1kgpHC\_NoHeader.txt

cat TajimasD\_Indian1kgpHC\_Header.txt TajimasD\_Indian1kgpHC\_NoHeader.txt > TajimasD\_Indian1kgpHC\_FinalFile.txt

# Composite Tests

SweeD: <http://exelixis-lab.org/software.html>

## 

For each chromosome CLR was calculated with a resolution of 10 000 bins, assuring that the density of bins is much higher than the density of SNPs in each chromosome. All CLR scans were run on unfolded spectra using the polarized data.

## Transfer Files to LDGH server

### India

scp [marlama@cedar.computecanada.ca](mailto:marlama@cedar.computecanada.ca):/project/6006002/marlama/Indians/AncestralAlleleFiles/JustIndia/\*\_popInfo.vcf /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/JustIndia/

### Caste/Tribe

for pop in Caste Tribe ; do scp [marlama@cedar.computecanada.ca](mailto:marlama@cedar.computecanada.ca):/project/6006002/marlama/Indians/AncestralAlleleFiles/${pop}/\*\_popInfo.vcf /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/ ; done

### India 1KGP

scp [marlama@cedar.computecanada.ca](mailto:marlama@cedar.computecanada.ca):/project/6006002/marlama/Indians/AncestralAlleleFiles/India1kgp/\*\_popInfo.vcf /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/India1kgp/

## Run SweeD by Chromossome

cd /home/marlam/Programs/Sweed/

### Just India

for chr in $(seq 1 22) ; do ./SweeD-P -name CLR\_JustIndia\_chr${chr} -input /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/JustIndia/all\_JustIndia\_AncestralAllele\_chr${chr}\_popInfo.vcf -grid 10000 -threads 10 ; done

### Caste/Tribe

for chr in $(seq 1 22) ; do for pop in India Caste Tribe ; do ./SweeD-P -name CLR\_Just${pop}\_chr${chr} -input /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_AncestralAllele\_chr${chr}\_popInfo.vcf -grid 10000 -threads 10 ; done ; done

### India1kgp

for chr in $(seq 1 22) ; do ./SweeD-P -name CLR\_JustIndia1kgp\_chr${chr} -input /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/India1kgp/Indian1kgpGroup\_Phased\_AncestralAlleleFile\_Final\_chr${chr}\_popInfo.vcf -grid 10000 -threads 10 ; done

### Second Run 200kb

for chr in 1 ; do for pop in India Caste Tribe India1kgp ; do /home/marlam/Programs/Sweed/SweeD-P -name CLR\_Just${pop}\_chr${chr}\_200kb -input /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/all\_Just${pop}\_AncestralAllele\_chr${chr}\_popInfo.vcf -grid 1242 -threads 40 ; done ; done

for chr in 1 ; do for pop in India1kgp ; do /home/marlam/Programs/Sweed/SweeD-P -name CLR\_Just${pop}\_chr${chr}\_200kb -input /media/LDGH\_STORAGE/marlama/India/AncestralAlleleFiles/${pop}/Indian1kgpGroup\_Phased\_AncestralAlleleFile\_Final\_chr${chr}\_popInfo.vcf -grid 1242 -threads 40 ; done ; done

### 1kgp\_HC 200kb

cd /home/marlama/scratch/1kgp\_HC/AncestralAllele/

for chr in 2 ; do /home/marlama/programs/Sweed/SweeD-P -name CLR\_India1kgpHC\_chr${chr}\_200kb -input 1kgpHC\_chr${chr}\_JustIndians\_MosaicQC\_Shapeit4\_AncestralAlleleFinal.vcf -grid 1211 -threads 100 ; done

for chr in $(seq 2 22) ; do sbatch CLR\_${chr}.sh ; done

## Addiction Chromosome Information

### JustIndia

for chr in $(seq 1 22) ; do awk -v chr="$chr" '{print chr"\t"$1"\t"$2"\t"$3}' /media/LDGH\_STORAGE/marlama/India/CLR/JustIndia/SweeD\_Report.CLR\_JustIndia\_chr${chr}\_grid10000 > /media/LDGH\_STORAGE/marlama/India/CLR/JustIndia/SweeD\_Report.CLR\_JustIndia\_chr${chr}\_grid10000.FinalInformations.txt ; done

### India1kgp/caste/tribe

for pop in Caste Tribe India1kgp ; do for chr in $(seq 1 22) ; do awk -v chr="$chr" '{print chr"\t"$1"\t"$2"\t"$3}' /media/LDGH\_STORAGE/marlama/India/CLR/Just${pop}/SweeD\_Report.CLR\_Just${pop}\_chr${chr} > /media/LDGH\_STORAGE/marlama/India/CLR/Just${pop}/SweeD\_Report.CLR\_Just${pop}\_chr${chr}.FinalInformations.txt ; done ; done

### Second Run 200kb

for pop in India Caste Tribe India1kgp ; do for chr in $(seq 1 22) ; do awk -v chr="$chr" '{print chr"\t"$1"\t"$2"\t"$3}' /media/LDGH\_STORAGE/marlama/India/CLR/SecondRun\_200kb/Just${pop}/SweeD\_Report.CLR\_Just${pop}\_chr${chr}\_200kb > /media/LDGH\_STORAGE/marlama/India/CLR/SecondRun\_200kb/Just${pop}/SweeD\_Report.CLR\_Just${pop}\_chr${chr}\_200kb.FinalInformations.txt ; done ; done

### India1kgp\_HC

for chr in $(seq 1 22) ; do awk -v chr="$chr" '{print chr"\t"$1"\t"$2"\t"$3}' /home/marlama/scratch/1kgp\_HC/CLR/SweeD\_Report.CLR\_India1kgpHC\_chr${chr}\_200kb > /home/marlama/scratch/1kgp\_HC/CLR/SweeD\_Report.CLR\_India1kgpHC\_chr${chr}\_200kb.FinalInformations.txt ; done

## Merge Chromosomes Files

### JustIndia

cat /media/LDGH\_STORAGE/marlama/India/CLR/JustIndia/\*\_grid10000.FinalInformations.txt > /media/LDGH\_STORAGE/marlama/India/CLR/JustIndia/CLR\_JustIndia\_grid10000.txt

### India1kgp/caste/tribe

for pop in Caste Tribe India1kgp ; do cat /media/LDGH\_STORAGE/marlama/India/CLR/Just${pop}/\*FinalInformations.txt > /media/LDGH\_STORAGE/marlama/India/CLR/Just${pop}/CLR\_${pop}.txt ; done

### Second Run 200kb

for pop in India Caste Tribe India1kgp ; do cat /media/LDGH\_STORAGE/marlama/India/CLR/SecondRun\_200kb/Just${pop}/\*FinalInformations.txt > /media/LDGH\_STORAGE/marlama/India/CLR/SecondRun\_200kb/Just${pop}/CLR\_${pop}.txt ; done

### JustIndia

### India1kgpHC

for chr in $(seq 1 22) ; do awk '{if(NR>3){print}}' /home/marlama/scratch/1kgp\_HC/CLR/SweeD\_Report.CLR\_India1kgpHC\_chr${chr}\_200kb.FinalInformations.txt > /home/marlama/scratch/1kgp\_HC/CLR/SweeD\_Report.CLR\_India1kgpHC\_chr${chr}\_200kb.FinalInformations\_NoHeader.txt ; done

cat /home/marlama/scratch/1kgp\_HC/CLR/\*FinalInformations\_NoHeader.tx > /home/marlama/scratch/1kgp\_HC/CLR/CLR\_India1kgpHC\_200kb\_FinalFile.txt

### Sorted Files

sort -k 3 /media/LDGH\_STORAGE/marlama/India/CLR/JustIndia/CLR\_JustIndia\_grid10000.txt > /media/LDGH\_STORAGE/marlama/India/CLR/JustIndia/CLR\_JustIndia\_grid10000\_sorted.txt

sort -k 3 /media/LDGH\_STORAGE/marlama/India/CLR/JustIndia1kgp/CLR\_India1kgp.txt > /media/LDGH\_STORAGE/marlama/India/CLR/JustIndia1kgp/CLR\_India1kgp\_sorted.txt

for pop in Caste Tribe ; do sort -k 3 /media/LDGH\_STORAGE/marlama/India/CLR/Just${pop}/CLR\_${pop}.txt > /media/LDGH\_STORAGE/marlama/India/CLR/Just${pop}/CLR\_${pop}\_sorted.txt ; done

# Search shared signals

## 1kgp Browser

for met in CLR iHS TajimasD ; do for pop in CEU CHB YRI ; do awk '{print $1"\t"$2"\t"$3"\t"$4"\t"$5"\t"$6}' /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/${met}\_${pop}.whole\_genome.pvalues > /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/${met}\_${pop}.whole\_genome\_tab.pvalues ; done ; done

for met in CLR iHS TajimasD ; do for pop in CEU CHB YRI ; do awk '{if($6 >= "2.3"){print $1"\t"$2"\t"$3"\t"$4"\t"$5"\t"$6}}' /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/1kgpBrowser/${met}\_${pop}.whole\_genome\_tab.pvalues > /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/1kgpBrowser/${met}\_${pop}.whole\_genome\_tab\_maior2.3.pvalues ; done ; done

python SearchSharedSignals.py

for met in XPEHH ; do for pop in CEU CHB YRI; do for pop2 in CEU CHB YRI ; do awk '{print $1"\t"$2"\t"$3"\t"$4"\t"$5}' /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/${met}\_${pop}\_vs\_${pop2}.whole\_genome.pvalues > /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/${met}\_${pop}\_vs\_${pop2}.whole\_genome\_tab.pvalues ; done ; done ; done

for met in XPEHH ; do for pop in CEU CHB YRI ; do for pop2 in CEU CHB YRI ; do awk '{if($5 >= "2.3"){print $1"\t"$2"\t"$3"\t"$4"\t"$5}}' /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/1kgpBrowser/${met}\_${pop}\_vs\_${pop2}.whole\_genome\_tab.pvalues > /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/1kgpBrowser/${met}\_${pop}\_vs\_${pop2}.whole\_genome\_tab\_maior2.3.pvalues ; done ; done ; done

for met in iHS; do for pop in CEU CHB YRI ; do awk '{print $1"\t"$2"\t"$3"\t"$4"\t"$5}' /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/${met}\_${pop}.whole\_genome.pvalues > /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/${met}\_${pop}.whole\_genome\_tab.pvalues ; done ; done

for met in iHS ; do for pop in CEU CHB YRI ; do awk '{if($5 >= "2.3"){print $1"\t"$2"\t"$3"\t"$4"\t"$5}}' /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/1kgpBrowser/${met}\_${pop}.whole\_genome\_tab.pvalues > /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/1kgpBrowser/${met}\_${pop}.whole\_genome\_tab\_maior2.3.pvalues ; done ; done

## PopHuman Browser

for pop in YRI PJL STU TSI MSL LWK KHV JPT ITU GIH FIN ESN GBR CHS CEU ASW BEB CDX ACB ; do awk -v pop="$pop" '{if(NR > 1){print pop"\t"$1"\t"$2"\t"$3"\t"$4"\t"$5"\t"$6"\t"$7"\t"$8"\t"$9}}' /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/PopHumanBrowser/${pop}.ihs.norm > /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/PopHumanBrowser/${pop}.ihs\_pop.norm ; done

cat \*\_pop.norm > All1kgpPop\_iHS.txt

awk '{if($10 == "1"){print $1"\t"$2"\t"$3"\t"$4"\t"$5"\t"$6"\t"$7"\t"$8"\t"$9"\t"$10}}' /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/PopHumanBrowser/All1kgpPop\_iHS.txt > /media/LDGH\_STORAGE/marlama/India/Browser\_NaturalSelection/PopHumanBrowser/All1kgpPop\_iHS\_ExtremeValues.txt