Positive selection in the chromosome 3 region 3p21.31

The 3p21.31 region on the human chromosome 3 spans about five megabases where positive selection seems to act recurrently. Previously published papers suggest that genes in the region have been under selection on multiple occasions in both African humans, in the ancestors of humans and chimpanzees, and more generally across primates. Why strong selection so often affects this region and which genes this selection affects is not not really known. With your newly aquired skills, you can apply the the most advanced population genomic methods and produce an updated inference of selection in Africans. For this project you have phased genotypes for chr3:46000000-54000000 individuals from the following populations:

```
YRI    Yoruba      Yoruba in Ibadan, Nigeria
LWK    Luhya       Luhya in Webuye, Kenya
GWD    Gambian     Gambian in Western Division, The Gambia
MSL    Mende       Mende in Sierra Leone
ESN    Esan        Esan in Nigeria
```

Make yourself familiar with the study populations. Where in Africa are they? How are they related?

Added all locations to favorites on google maps:

1. Gambian = GWD
    a. Western Division is on the west coast of the Gambia (most west)
2. Mende = MSL
    a. Sierra Leone us south of The Gambia (Second from West Coast
3. Yoruba = YRI
    a. Ibadan is in South West Nigeria (3rd from the west coast)
4. Esan = ESN
    a. Esan is in South Central Nigeria (East of Ibadan)
5. Luhya = LWK
    a. Webuye is in Western Kenya (furthest east)

o

Investigate the following

Note: The data I am given is phased VCF files for all individuals and for each African population separately, so I do not need to do the week 3 exercise of using Beagle to to call the genotype phase.

1. Copy Data to folder on cluster
#NOTE= using Cluster not Ucloud because I do not trust Ucloud.

Scp data
scp ~/populationgenomics/project_data/chr3region/chr3_460_540_phased.vcf.gz .
~/populationgenomics/project_data/chr3region/chr3_ESN_460_540_phased.vcf.gz
~/populationgenomics/project_data/chr3region/chr3_GWD_460_540_phased.vcf.gz
~/populationgenomics/project_data/chr3region/chr3_LWK_460_540_phased.vcf.gz
~/populationgenomics/project_data/chr3region/chr3_MSL_460_540_phased.vcf.gz
~/populationgenomics/project_data/chr3region/chr3_YRI_460_540_phased.vcf.gz
~/populationgenomics/project_data/chr3region/all_inds.txt
~/populationgenomics/project_data/chr3region/ESN_inds.txt
~/populationgenomics/project_data/chr3region/GWD_inds.txt
~/populationgenomics/project_data/chr3region/LWK_inds.txt
~/populationgenomics/project_data/chr3region/MSL_inds.txt
~/populationgenomics/project_data/chr3region/YRI_inds.txt
~/populationgenomics/project_data/chr3region/20140520.chr3.strict_mask.fasta.gz
~/populationgenomics/project_data/chr3region/human_ancestor_3.fa

.

Try using vcftools

```
ESN    Esan         Esan in Nigeria
GWD    Gambian      Gambian in Western Division, The Gambia
LWK    Luhya        Luhya in Webuye, Kenya
MSL    Mende        Mende in Sierra Leone
YRI    Yoruba       Yoruba in Ibadan, Nigeria
```

Remove non-relevant individuals from text file:

vcftools --gzvcf vcf/chr3_460_540_phased.vcf.gz --keep txt/all_inds.txt --recode --out filtered_inds

NEW vcf file to use is called

vcf/filtered_inds.recode.vcf.gz

1. ESN as compare
    a. vcftools --gzvcf vcf/filtered_inds.recode.vcf.gz   --weir-fst-pop txt/ESN_inds.txt --weir-fst-pop txt/GWD_inds.txt --out output/ESN_GWD
    b. vcftools --gzvcf vcf/filtered_inds.recode.vcf.gz   --weir-fst-pop txt/ESN_inds.txt --weir-fst-pop txt/LWK_inds.txt --out output/ESN_LWK
    c. vcftools --gzvcf vcf/filtered_inds.recode.vcf.gz   --weir-fst-pop txt/ESN_inds.txt --weir-fst-pop txt/MSL_inds.txt --out output/ESN_MSL
    d. vcftools --gzvcf vcf/filtered_inds.recode.vcf.gz   --weir-fst-pop txt/ESN_inds.txt --weir-fst-pop txt/YRI_inds.txt --out output/ESN_YRI
        i.   grep -v "\-nan" ESN_YRI.weir.fst | awk '!/NaN/' > ESN_YRI_filtered
            1. Removes -nan values
2. GWD as compare
    a. vcftools --gzvcf vcf/filtered_inds.recode.vcf.gz   --weir-fst-pop txt/GWD_inds.txt --weir-fst-pop txt/LWK_inds.txt --out output/GWD_LWK
    b. vcftools --gzvcf vcf/filtered_inds.recode.vcf.gz   --weir-fst-pop txt/GWD_inds.txt --weir-fst-pop txt/MSL_inds.txt --out output/GWD_MSL
    c. vcftools --gzvcf vcf/filtered_inds.recode.vcf.gz   --weir-fst-pop txt/GWD_inds.txt --weir-fst-pop txt/YRI_inds.txt --out output/GWD_YRI
3. LWK as compare

a. vcftools --gzvcf vcf/filtered_inds.recode.vcf.gz   --weir-fst-pop txt/LWK_inds.txt --weir-fst-pop txt/MSL_inds.txt --out output/LWK_MSL

b. vcftools --gzvcf vcf/filtered_inds.recode.vcf.gz   --weir-fst-pop txt/LWK_inds.txt --weir-fst-pop txt/YRI_inds.txt --out output/LWK_YRI

4. MSL as compare

a. vcftools --gzvcf vcf/filtered_inds.recode.vcf.gz   --weir-fst-pop txt/MSL_inds.txt --weir-fst-pop txt/YRI_inds.txt --out output/MSL_YRI

https://cloud.sdu.dk/app/files?path=%2F567654%2FPopgen%2FExercises%2F02_F_statistics

*"**F-statistics** (also known as **fixation indices**) describe the statistically expected level of heterozygosity in a population; more specifically the expected degree of (usually) a reduction in heterozygosity when compared to Hardy–Weinberg expectation."*

./vcftools —vcf

B. Use Relate on all the individuals and visualize trees (using Relate or tskit) to get an impression of the relationship between the populations. How does this relate to your Fst results?

scp ~/populationgenomics/project_data/chr3region/20140520.chr3.strict_mask.fasta.gz .
scp ~/populationgenomics/data/relate_data/genetic_map_chr2_combined_b37.txt
scp ~/populationgenomics/data/relate_data/human_ancestor_2.fa
scp ~/populationgenomics/data/relate_data/60_inds.txt
scp ~/populationgenomics/data/relate_data/chr2_130_145_phased.vcf.gz

Remove non-relevant individuals from vcf file

vcftools --gzvcf vcf/chr3_460_540_phased.vcf.gz --keep txt/all_inds.txt --recode --out filtered_inds

Relate does not accept the standard VCF file format, but instead uses a haps/sample format. You can read up on in the Relate documentation. The authors have been so kind as to supply a script to transform it. First, the vcf is converted to another file format (haplotype file format). If you want to know how it is structured, you can read about it here.

~/populationgenomics/software/relate/bin/RelateFileFormats --mode ConvertFromVcf --haps chr3.haps --sample chr3.sample -i filtered_inds.recode

Then, repetitive (unreliably sequenced) regions must be masked to exclude them from our analysis. We also need to assign each variant as either ancestral or derived using the chimpanzee genome. We do both with this command:

NOTE: needed to gunzip all files first.

~/populationgenomics/software/relate/scripts/PrepareInputFiles/PrepareInputFiles.sh --haps chr3.haps --sample chr3.sample --ancestor human_ancestor_3.fa --mask 20140520.chr3.strict_mask.fasta -o prep.chr3

Now, the input is fully prepared, and Relate can be run.

Initial

```
~/populationgenomics/software/relate/bin/Relate --mode All -m 1.25e-8 -N 30000
--haps prep.chr3.haps.gz --sample prep.chr3.sample.gz --map
genetic_map_chr3_combined_b37.txt -o chr3_relate
```

ESN

```
~/populationgenomics/software/relate/bin/Relate --mode All -m 1.25e-8 -N 30000
--haps prep.chr3_ESN.haps.gz --sample prep.chr3_ESN.sample.gz --map
genetic_map_chr3_combined_b37.txt -o chr3_relate
```

Relate outputs estimated mutation rate and coalescence times along the region

in which population size is estimated, and the population size is re-estimates branch lengths:

~/populationgenomics/software/relate/scripts/EstimatePopulationSize/EstimatePopulationSize.sh -i chr3_relate -m 1.25e-8 --num_iter 1 --poplabels all_inds.txt -o popsize_oneiter --threshold 0

~/populationgenomics/software/relate/scripts/EstimatePopulationSize/EstimatePopulationSize.sh -i chr3_relate -m 1.25e-8 --num_iter 2 --poplabels all_inds.txt -o popsize_twoiter --threshold 0

~/populationgenomics/software/relate/scripts/EstimatePopulationSize/EstimatePopulationSize.sh -i chr3_relate -m 1.25e-8 --num_iter 3 --poplabels all_inds.txt -o popsize_threeiter --threshold 0

When you are ready, you should have a command prompt from srun. Use that terminal to activate the `pg-relate` environment.

```
conda activate pg-relate
```

The Relate command below detects positive selection. At this point, we are detecting selection based on 5 distinct populations.

```
~/populationgenomics/software/relate/scripts/DetectSelection/DetectSelection.sh
-i popsize_oneiter -m 1.25e-8 --poplabels all_inds.txt -o
selection_relate_oneiter

~/populationgenomics/software/relate/scripts/DetectSelection/DetectSelection.sh

-i popsize -m 1.25e-8 --poplabels 60_inds.txt -o selection_relate
```

PLAN: USE MY ONE_ITER TO ID REGION THAT I WILL BUILD TREES ON

relate/one_iter/selection_relate_oneiter.sele
```
-n 10
```

```
sort -k35r relate/one_iter/selection_relate_oneiter.sele | cut -f 1,2,35 -d ' ' | head
-n 10
```

```
Output:

pos rs_id when_mutation_has_freq2
46707244 rs17079122 -9.93441e-05
52381322 rs563971947 -7.16044e-05
52386137 rs76451298 -7.16044e-05
52210700 rs352152 -6.73561
46374436 rs35465430 -6.69821
46051363 rs58139750 -6.40215
46051532 rs55743140 -6.40215
51836739 rs113885732 -6.08449
52196241 rs6786592 -5.9645
```

Try plotting some of the sites with the following command. Use -o to determine name, and bp_of_interest for the position (remember to chagne both each time you run the command).

```
~/populationgenomics/software/relate/scripts/TreeView/TreeView.sh --haps relate/chr3.haps
--sample relate/chr3.sample --anc relate/chr3_relate.anc --mut
relate/one_iter/popsize_oneiter.mut --poplabels relate/all_inds.txt --years_per_gen 28 -o
tree_52196241 --bp_of_interest 52196241

~/populationgenomics/software/relate/scripts/TreeView/TreeView.sh --haps relate/chr3.haps
--sample relate/chr3.sample --anc relate/chr3_relate.anc --mut
relate/one_iter/popsize_oneiter.mut --poplabels relate/all_inds.txt --years_per_gen 28 -o
tree_468300001 --bp_of_interest 46830000

46830000
```

C. Use Relate on each population separately to infer positive selection.

```
~/populationgenomics/software/relate/bin/RelateFileFormats --mode ConvertFromVcf --haps chr3.haps --sample
chr3.sample -i filtered_inds.recode
```

ESN:
1. ~/populationgenomics/software/relate/bin/Relate --mode All -m 1.25e-8 -N 30000 --haps prep.chr3_ESN.haps.gz --sample prep.chr3_ESN.sample.gz --map genetic_map_chr3_combined_b37.txt -o chr3_relate_ESN
2. ~/populationgenomics/software/relate/scripts/EstimatePopulationSize/EstimatePopulationSize.sh -i chr3_relate_ESN -m 1.25e-8 --poplabels all_inds.txt -o popsize --threshold 0
3. ~/populationgenomics/software/relate/scripts/DetectSelection/DetectSelection.sh -i popsize_ESN -m 1.25e-8 --poplabels ESN_inds.txt -o selection_relate_ESN
4. sort -k35r selection_relate_ESN.sele | cut -f 1,2,35 -d ' ' | head -n 10 > top_SNPs_Relate_ESN

LWK:
1. ~/populationgenomics/software/relate/bin/Relate --mode All -m 1.25e-8 -N 30000 --haps prep.chr3_LWK.haps.gz --sample prep.chr3_LWK.sample.gz --map genetic_map_chr3_combined_b37.txt -o chr3_relate_LWK
2. ~/populationgenomics/software/relate/scripts/EstimatePopulationSize/EstimatePopulationSize.sh -i chr3_relate_LWK -m 1.25e-8 --poplabels LWK_inds.txt -o popsize_LWK --threshold 0
3. ~/populationgenomics/software/relate/scripts/DetectSelection/DetectSelection.sh -i popsize_LWK -m 1.25e-8 --poplabels LWK_inds.txt -o selection_relate_LWK
4. sort -k35r selection_relate_LWK.sele | cut -f 1,2,35 -d ' ' | head -n 10 > top_SNPs_Relate_LWK

GWD:
1. ~/populationgenomics/software/relate/bin/Relate --mode All -m 1.25e-8 -N 30000 --haps prep.chr3_GWD.haps.gz --sample prep.chr3_GWD.sample.gz --map genetic_map_chr3_combined_b37.txt -o chr3_relate_GWD
2. ~/populationgenomics/software/relate/scripts/EstimatePopulationSize/EstimatePopulationSize.sh -i chr3_relate_GWD -m 1.25e-8 --poplabels all_inds.txt -o popsize_GWD --threshold 0
3. ~/populationgenomics/software/relate/scripts/DetectSelection/DetectSelection.sh -i popsize_GWD -m 1.25e-8 --poplabels GWD_inds.txt -o selection_relate_GWD
4. sort -k35r selection_relate_GWD.sele | cut -f 1,2,35 -d ' ' | head -n 10 > top_SNPs_Relate_GWD

MSL
1. ~/populationgenomics/software/relate/bin/Relate --mode All -m 1.25e-8 -N 30000 --haps prep.chr3_MSL.haps.gz --sample prep.chr3_MSL.sample.gz --map genetic_map_chr3_combined_b37.txt -o chr3_relate_MSL
2. ~/populationgenomics/software/relate/scripts/EstimatePopulationSize/EstimatePopulationSize.sh -i chr3_relate_MSL -m 1.25e-8 --poplabels all_inds.txt -o popsize_MSL --threshold 0

3. ~/populationgenomics/software/relate/scripts/DetectSelection/DetectSelection.sh -i popsize_MSL -m 1.25e-8 --poplabels MSL_inds.txt -o selection_relate_MSL
4. sort -k35r selection_relate_MSL.sele | cut -f 1,2,35 -d ' ' | head -n 10 > top_SNPs_Relate_MSL

YRI
1. ~/populationgenomics/software/relate/bin/Relate --mode All -m 1.25e-8 -N 30000 --haps prep.chr3_YRI.haps.gz --sample prep.chr3_YRI.sample.gz --map genetic_map_chr3_combined_b37.txt -o chr3_relate_YRI
2. ~/populationgenomics/software/relate/scripts/EstimatePopulationSize/EstimatePopulationSize.sh -i chr3_relate_YRI -m 1.25e-8 --poplabels all_inds.txt -o popsize_YRI --threshold 0
3. ~/populationgenomics/software/relate/scripts/DetectSelection/DetectSelection.sh -i popsize_YRI -m 1.25e-8 --poplabels YRI_inds.txt -o selection_relate_YRI
4. sort -k35r selection_relate_YRI.sele | cut -f 1,2,35 -d ' ' | head -n 10 > top_SNPs_Relate_YRI

```
~/populationgenomics/software/relate/scripts/TreeView/TreeView.sh --haps
relate/ESN/chr3_ESN.haps --sample relate/ESN/chr3_ESN.sample --anc
relate/ESN/chr3_relate_ESN.anc --mut relate/ESN/chr3_relate_ESN.mut --poplabels
relate/ESN/ESN_inds.txt --years_per_gen 28 -o tree_46659530_ESN --bp_of_interest
46659530
```

https://github.com/standard-aaron/clues/wiki/Sampling-coalescence-times-with-Relate

D. Run one or more additional methods for selection inference. If possible this should be another tree sequence based method such as CLUES. Compare the results to those obtained using Relate.

ESN SNP -> 46659530

~/populationgenomics/software/relate/scripts/SampleBranchLengths/SampleBranchLengths.sh -i chr3_relate_ESN -o chr3_ESN_relate_resample_46659530_100 -m 1.25e-8 --coal popsize_ESN.coal --format b --num_samples 100 --first-bp 46659530 --last-bp 46659530

LWK SNP -> 46659530

~/populationgenomics/software/relate/scripts/SampleBranchLengths/SampleBranchLengths.sh -i chr3_relate_LWK -o chr3_LWK_relate_resample_46659530_100 -m 1.25e-8 --coal popsize_LWK.coal --format b --num_samples 100 --first-bp 46659530 --last-bp 46659530

GWD SNP -> 46659530

~/populationgenomics/software/relate/scripts/SampleBranchLengths/SampleBranchLengths.sh -i chr3_relate_GWD -o chr3_GWD_relate_resample_46659530_100 -m 1.25e-8 --coal popsize_GWD.coal --format b --num_samples 100 --first-bp 46659530 --last-bp 46659530

YRI SNP -> 46659530

~/populationgenomics/software/relate/scripts/SampleBranchLengths/SampleBranchLengths.sh -i chr3_relate_YRI -o chr3_YRI_relate_resample_46659530_100 -m 1.25e-8 --coal popsize_YRI.coal --format b --num_samples 100 --first-bp 46659530 --last-bp 46659530

MSL SNP -> 46659530

~/populationgenomics/software/relate/scripts/SampleBranchLengths/SampleBranchLengths.sh -i chr3_relate_MSL -o chr3_MSL_relate_resample_46659530_100 -m 1.25e-8 --coal popsize_MSL.coal --format b --num_samples 100 --first-bp 46659530 --last-bp 46659530

52196241

ESN SNP -> 52196241

~/populationgenomics/software/relate/scripts/SampleBranchLengths/SampleBranchLengths.sh -i chr3_relate_ESN -o chr3_ESN_relate_resample_52196241_100 -m 1.25e-8 --coal popsize_ESN.coal --format b --num_samples 100 --first-bp 52196241 --last-bp 52196241

python inference.py --times ~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/ESN/chr3_ESN_relate_resample_52196241_100 --coal ~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/ESN/popsize_ESN.coal --burnin 60 --out ~/populationgenomics/students/tessdiv/FinalProjectPopGenome/CLUES/selection_clues_ESN_52196241_100_burnin_coal

Log 10 p value = -2.72972

LWK SNP -> 52196241

~/populationgenomics/software/relate/scripts/SampleBranchLengths/SampleBranchLengths.sh -i chr3_relate_LWK -o chr3_LWK_relate_resample_52196241_100 -m 1.25e-8 --coal popsize_LWK.coal --format b --num_samples 100 --first-bp 52196241 --last-bp 52196241

python inference.py --times
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/LWK/chr3_LWK_relate_resample_52196241_100 --coal
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/LWK/popsize_LWK.coal
--burnin 60 --out
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/CLUES/selection_clues_LWK_52196241_100_burnin_coal

GWD SNP -> 52196241

~/populationgenomics/software/relate/scripts/SampleBranchLengths/SampleBranchLengths.sh -i chr3_relate_GWD -o chr3_GWD_relate_resample_52196241_100 -m 1.25e-8 --coal popsize_GWD.coal --format b --num_samples 100 --first-bp 52196241 --last-bp 52196241

YRI SNP -> 52196241

~/populationgenomics/software/relate/scripts/SampleBranchLengths/SampleBranchLengths.sh -i chr3_relate_YRI -o chr3_YRI_relate_resample_52196241_100 -m 1.25e-8 --coal popsize_YRI.coal --format b --num_samples 100 --first-bp 52196241 --last-bp 52196241

python inference.py --times
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/YRI/chr3_YRI_relate_resample_52196241_100 --coal
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/YRI/popsize_YRI.coal
--burnin 60 --out

~/populationgenomics/students/tessdiv/FinalProjectPopGenome/CLUES/selection_clues_YRI_
52196241_100_burnin_coal

-1.5974

MSL SNP -> 52196241

~/populationgenomics/software/relate/scripts/SampleBranchLengths/SampleBranchLeng
ths.sh -i chr3_relate_MSL -o chr3_MSL_relate_resample_52196241_100 -m
1.25e-8 --coal popsize_MSL.coal --format b --num_samples 100 --first-bp
52196241--last-bp 52196241

MSL
python inference.py --times
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/MSL/chr3_MSL_relate_r
esample_52196241_100 --coal
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/MSL/popsize_MSL.coal
--burnin 60 --out
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/CLUES/selection_clues_MSL_
52196241_100_burnin_coal

Log 10 p- value = -2.65713

Run CLUES from CLUES folder

ESN

python inference.py --times
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/ESN/chr3_ESN_relate_r
esample_46659530_100 --coal
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/ESN/popsize_ESN.coal
--burnin 60 --out
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/CLUES/selection_clues_ESN_
46659530_100_burnin_coal

YRI
python inference.py --times
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/YRI/chr3_YRI_relate_res
ample_46659530_100 --coal

~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/YRI/popsize_YRI.coal
--burnin 60 --out
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/CLUES/selection_clues_YRI_4
6659530_100_burnin_coal


```
LWK
```

python inference.py --times
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/LWK/chr3_LWK_relate_r
esample_46659530_100 --coal
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/LWK/popsize_LWK.coal
--burnin 60 --out
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/CLUES/selection_clues_LWK_
46659530_100_burnin_coal

GWD
python inference.py --times
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/GWD/chr3_GWD_relate
_resample_46659530_100 --coal
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/GWD/popsize_GWD.coa
l --burnin 60 --out
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/CLUES/selection_clues_GWD
_46659530_100_burnin_coal

MSL
python inference.py --times
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/MSL/chr3_MSL_relate_r
esample_46659530_100 --coal
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/relate/MSL/popsize_MSL.coal
--burnin 60 --out
~/populationgenomics/students/tessdiv/FinalProjectPopGenome/CLUES/selection_clues_MSL_
46659530_100_burnin_coal

```
python plot_traj.py selection_clues_ESN_46659530_100_burnin_coal plots/ESN_traj.png --ext png

python plot_traj.py selection_clues_GWD_46659530_100_burnin_coal plots/GWD_traj.png --ext png

python plot_traj.py selection_clues_LWK_46659530_100_burnin_coal plots/LWK_traj.png --ext png


python plot_traj.py selection_clues_MSL_46659530_100_burnin_coal plots/MSL_traj.png --ext png
python plot_traj.py selection_clues_YRI_46659530_100_burnin_coal plots/YRI_traj.png --ext png
```

https://github.com/standard-aaron/clues/wiki

Papers

Patterns of Ancestry, Signatures of Natural Selection, and Genetic Association with Stature in Western African Pygmies

Perhaps: An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data