



Queen Mary
University of London

NIXOS SOFTWARE

ELLIE, CERI, CIAN, NATALIA

PROJECT & STRATEGY

plan

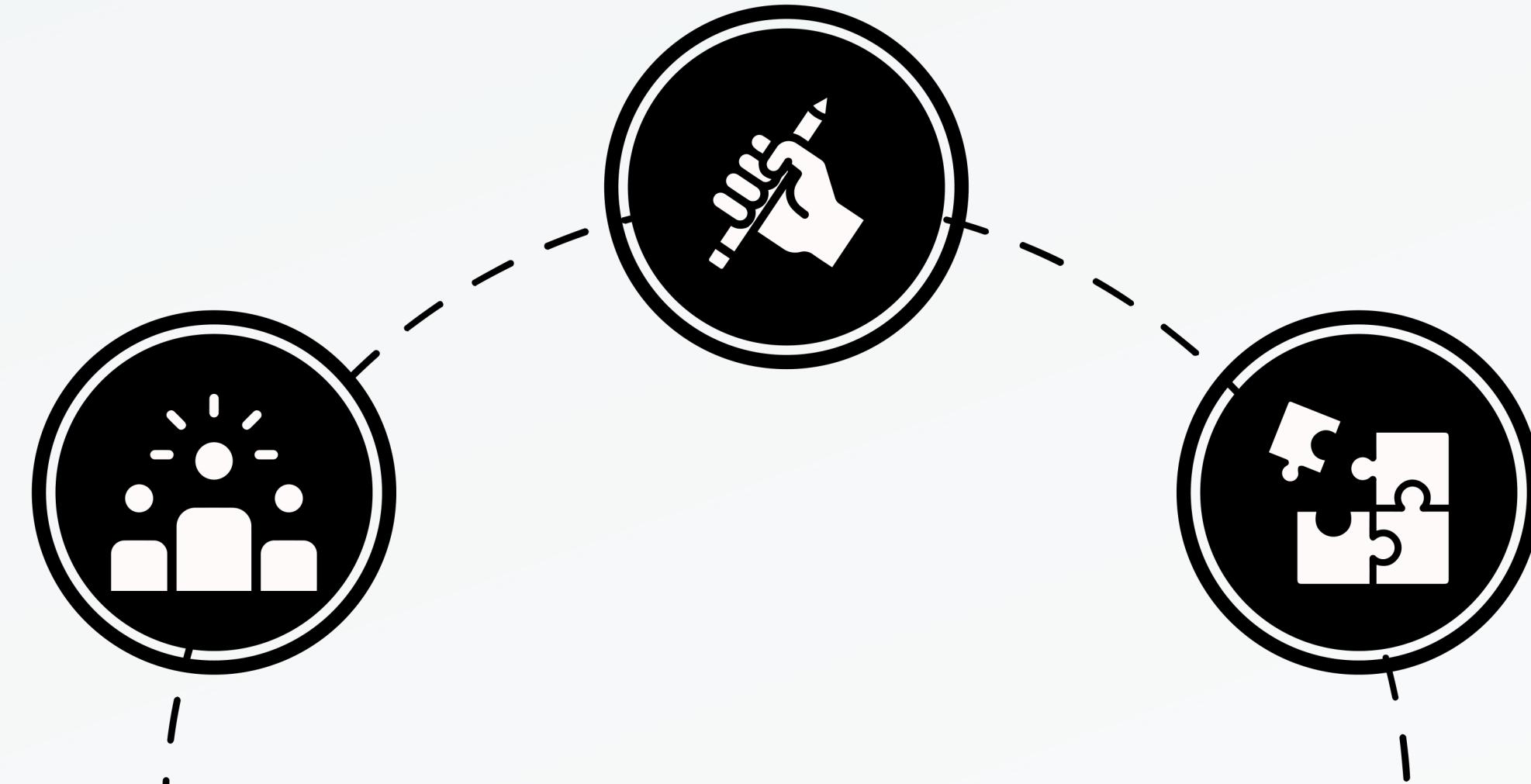
delegate roles based on individual preference and skill set. Create timeline and what the project aims to achieve.

build

Start the project build phase, figuring out what methods and softwares are better suited for the project.e.g make database, choose analysis techniques.

test

confirm analysis decisions, trial everything is integrated properly, brief is followed and have a running software



CLUSTERING ANALYSIS

- Decided on PCA
- pre calculated values better for user experience
- PLINK in UNIX and analysed in R mapped to populations
- pve and pca added to sql
- Interactive plot

Admixture analysis

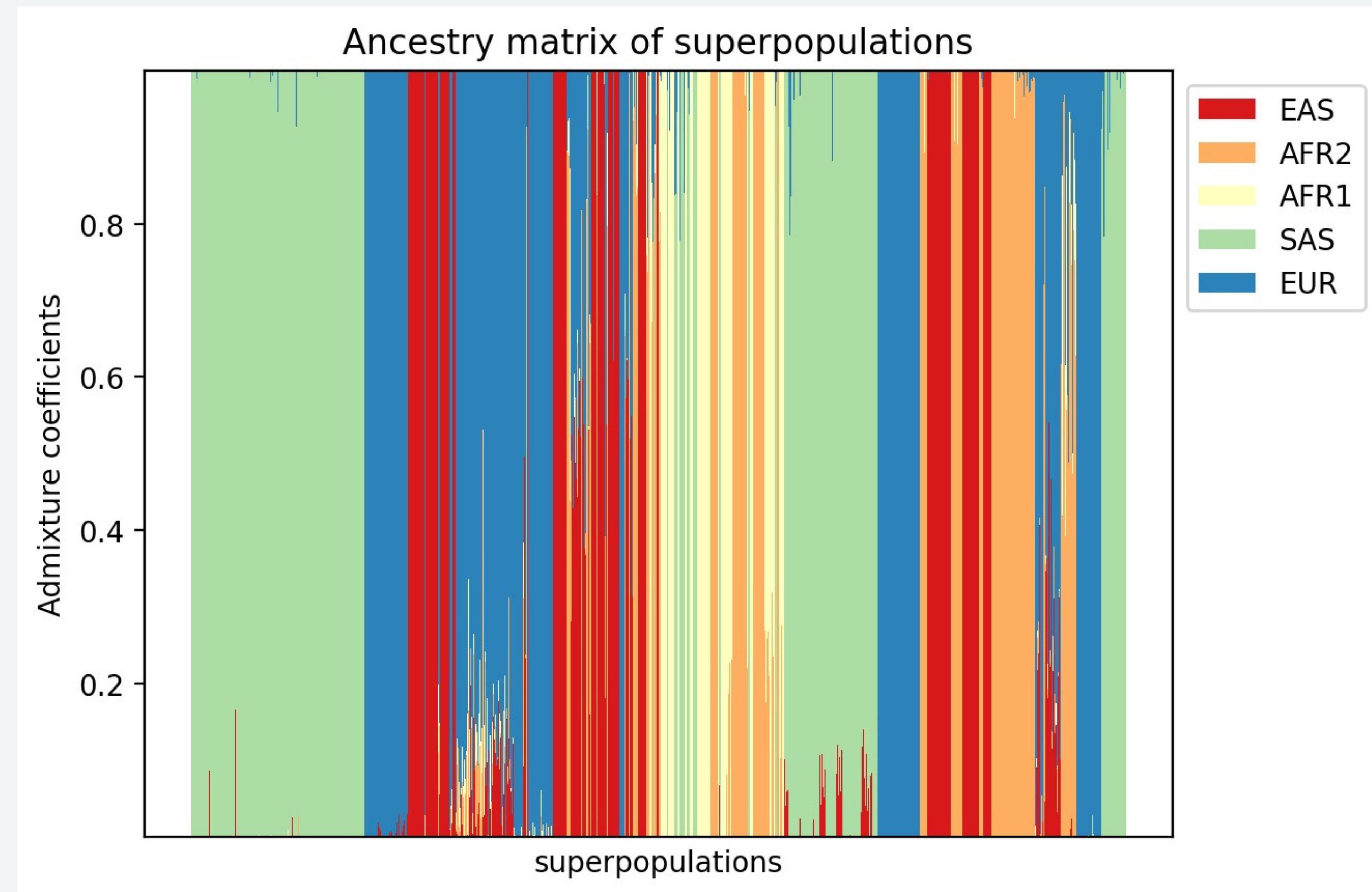
conducted with ADMIXTURE

Ancestry coefficients

| | A | B | C | D | E | F | G | H |
|----|---------|---------|---------|---------|---------|------|------------|------------------|
| 1 | EUR | SAS | AFR1 | AFR2 | EAS | id | population | Super population |
| 2 | 0.00001 | 0.99996 | 0.00001 | 0.00001 | 0.00001 | 99_1 | SIB | SAS |
| 3 | 0.00001 | 0.99996 | 0.00001 | 0.00001 | 0.00001 | 99_2 | SIB | SAS |
| 4 | 0.00001 | 0.99996 | 0.00001 | 0.00001 | 0.00001 | 99_4 | SIB | SAS |
| 5 | 0.00001 | 0.99996 | 0.00001 | 0.00001 | 0.00001 | 98_1 | SIB | SAS |
| 6 | 0.00001 | 0.99996 | 0.00001 | 0.00001 | 0.00001 | 98_2 | SIB | SAS |
| 7 | 0.00001 | 0.99996 | 0.00001 | 0.00001 | 0.00001 | 98_4 | SIB | SAS |
| 8 | 0.00001 | 0.99996 | 0.00001 | 0.00001 | 0.00001 | 98_3 | SIB | SAS |
| 9 | 0.00001 | 0.99996 | 0.00001 | 0.00001 | 0.00001 | 98_5 | SIB | SAS |
| 10 | 0.00001 | 0.99996 | 0.00001 | 0.00001 | 0.00001 | 97_2 | SIB | SAS |
| 11 | 0.00001 | 0.99996 | 0.00001 | 0.00001 | 0.00001 | 97_1 | SIB | SAS |
| 12 | 0.00001 | 0.99996 | 0.00001 | 0.00001 | 0.00001 | 97_4 | SIB | SAS |

Genetic samples

Admixture plot



GENOTYPE & ALLELE FREQUENCY

| | Sample | Genotype | ID | population |
|----|--------|----------|------------|------------|
| 1 | 1_1 | 0 | rs750087 | SIB |
| 2 | 1_1 | 0 | rs7514366 | SIB |
| 3 | 1_1 | 1 | rs1774780 | SIB |
| 4 | 1_1 | 0 | rs12408124 | SIB |
| 5 | 1_1 | 0 | rs4847281 | SIB |
| 6 | 1_1 | 0 | rs17130397 | SIB |
| 7 | 1_1 | 0 | rs1926289 | SIB |
| 8 | 1_1 | 2 | rs4143772 | SIB |
| 9 | 1_1 | 0 | rs12726592 | SIB |
| 10 | 1_1 | 1 | rs11809207 | SIB |
| 11 | 1_1 | 0 | rs4845057 | SIB |
| 12 | 1_1 | 0 | rs547976 | SIB |

GENOTYPE & ALLELE FREQUENCY

| ID | Population | Genotype | Count | TotalSamples | Frequency |
|----|------------|----------|-------|--------------|-------------|
| 1 | ACB | 0 | 37 | 116 | 0.318965517 |
| 2 | ACB | 1 | 55 | 116 | 0.474137931 |
| 3 | ACB | 2 | 24 | 116 | 0.206896552 |
| 4 | ASW | 0 | 22 | 74 | 0.297297297 |
| 5 | ASW | 1 | 35 | 74 | 0.472972973 |
| 6 | ASW | 2 | 17 | 74 | 0.229729730 |
| 7 | BEB | 0 | 45 | 131 | 0.343511450 |
| 8 | BEB | 1 | 53 | 131 | 0.404580153 |
| 9 | BEB | 2 | 33 | 131 | 0.251908397 |
| 10 | CDX | 0 | 1 | 93 | 0.010752688 |
| 11 | CDX | 1 | 14 | 93 | 0.150537634 |
| 12 | CDX | 2 | 78 | 93 | 0.838709677 |

GENOTYPE & ALLELE FREQUENCY

| ID | Population | Genotype_0 | Genotype_1 | Genotype_2 | RefFrequency | AltFrequency | |
|----|------------|------------|-------------|------------|--------------|--------------|-------------|
| 1 | rs1000313 | ACB | 0.318965517 | 0.47413793 | 0.206896552 | 0.55603448 | 0.443965517 |
| 2 | rs1000313 | ASW | 0.297297297 | 0.47297297 | 0.229729730 | 0.53378378 | 0.466216216 |
| 3 | rs1000313 | BEB | 0.343511450 | 0.40458015 | 0.251908397 | 0.54580153 | 0.454198473 |
| 4 | rs1000313 | CDX | 0.010752688 | 0.15053763 | 0.838709677 | 0.08602151 | 0.913978495 |
| 5 | rs1000313 | CEU | 0.636871508 | 0.31843575 | 0.044692737 | 0.79608939 | 0.203910615 |
| 6 | rs1000313 | CHB | 0.019417476 | 0.21359223 | 0.766990291 | 0.12621359 | 0.873786408 |
| 7 | rs1000313 | CHS | 0.012269939 | 0.16564417 | 0.822085890 | 0.09509202 | 0.904907975 |
| 8 | rs1000313 | CLM | 0.553030303 | 0.37121212 | 0.075757576 | 0.73863636 | 0.261363636 |
| 9 | rs1000313 | ESN | 0.174496644 | 0.44295302 | 0.382550336 | 0.39597315 | 0.604026846 |
| 10 | rs1000313 | FIN | 0.555555556 | 0.40404040 | 0.040404040 | 0.75757576 | 0.242424242 |
| 11 | rs1000313 | GBR | 0.725274725 | 0.27472527 | NA | 0.86263736 | 0.137362637 |
| 12 | rs1000313 | GIH | 0.330097087 | 0.52427184 | 0.145631068 | 0.59223301 | 0.407766990 |

PAIRWISE POPULATION GENETIC DIFFERENTIATION

| | ID | population | TotalSamples | Ref | Alt |
|----|-----------|------------|--------------|-----|-----|
| 1 | rs1000313 | ACB | 232 | 129 | 103 |
| 2 | rs1000313 | ASW | 148 | 79 | 69 |
| 3 | rs1000313 | BEB | 262 | 143 | 119 |
| 4 | rs1000313 | CDX | 186 | 16 | 170 |
| 5 | rs1000313 | CEU | 358 | 285 | 73 |
| 6 | rs1000313 | CHB | 206 | 26 | 180 |
| 7 | rs1000313 | CHS | 326 | 31 | 295 |
| 8 | rs1000313 | CLM | 264 | 195 | 69 |
| 9 | rs1000313 | ESN | 298 | 118 | 180 |
| 10 | rs1000313 | FIN | 198 | 150 | 48 |
| 11 | rs1000313 | GBR | 182 | 157 | 25 |
| 12 | rs1000313 | GIH | 206 | 122 | 84 |

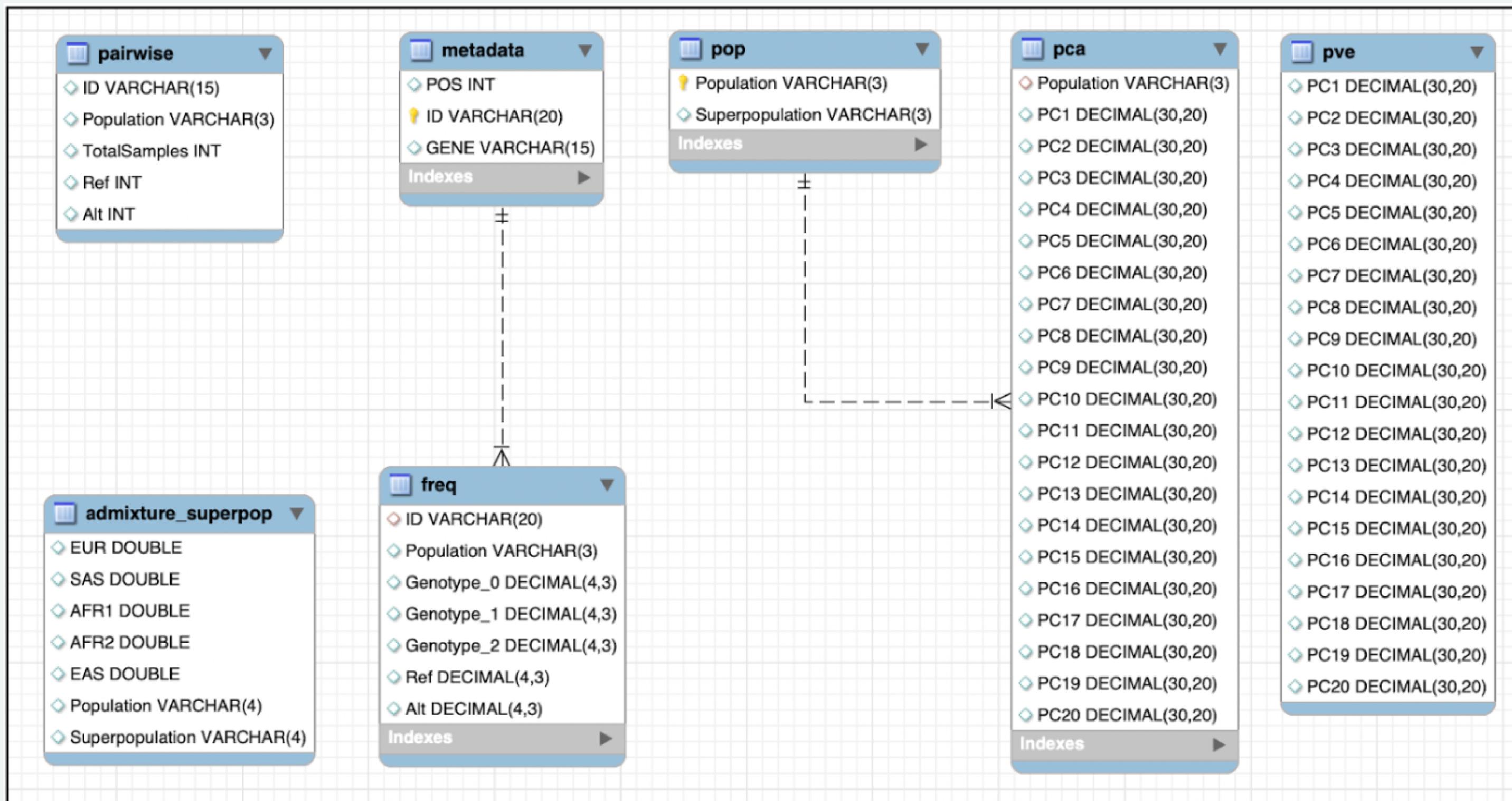
$$Fst = \frac{Ht - Hs}{Ht}$$

Ht = total genetic diversity

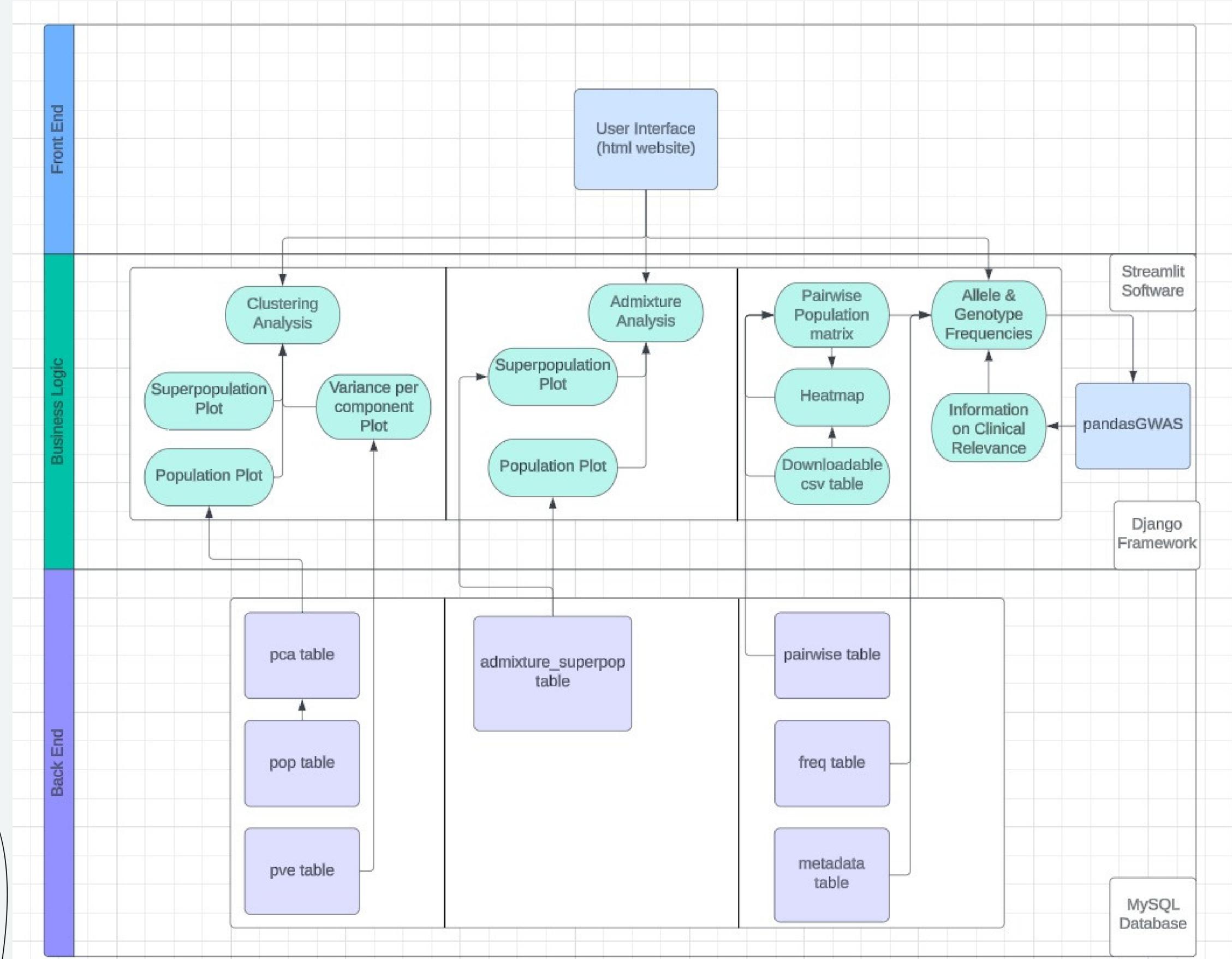
Hs = average genetic diversity within
subpopulations

| | SIB | GBR | ACB |
|-----|----------|----------|----------|
| SIB | 0.000000 | 0.005862 | 0.025470 |
| GBR | 0.005862 | 0.000000 | 0.108427 |
| ACB | 0.025470 | 0.108427 | 0.000000 |

SCHEMA



ARCHITECTURE



Homepage

Home Page x +

127.0.0.1:8000

NixOS Home Cluster Analysis Admixture Analysis Allele/Genotype Frequencies

Welcome to the NixOS Population Genetics Analysis Platform

Explore the intricacies of human population genetics with our user-friendly tool. Explore the fascinating world of genetic diversity and analyse data from nearly 40,000 Single Nucleotide Polymorphisms (SNPs) on chromosome 1 across 28 populations.

Key Features:

- 1. Comprehensive Data Set**

Our database encompasses data from 26 populations sourced from the renowned 1000 Genomes Project. Additionally, we provide unique insights into the genetic makeup of Siberia, adding a distinct dimension to your analyses.
- 2. Powerful Analyses**

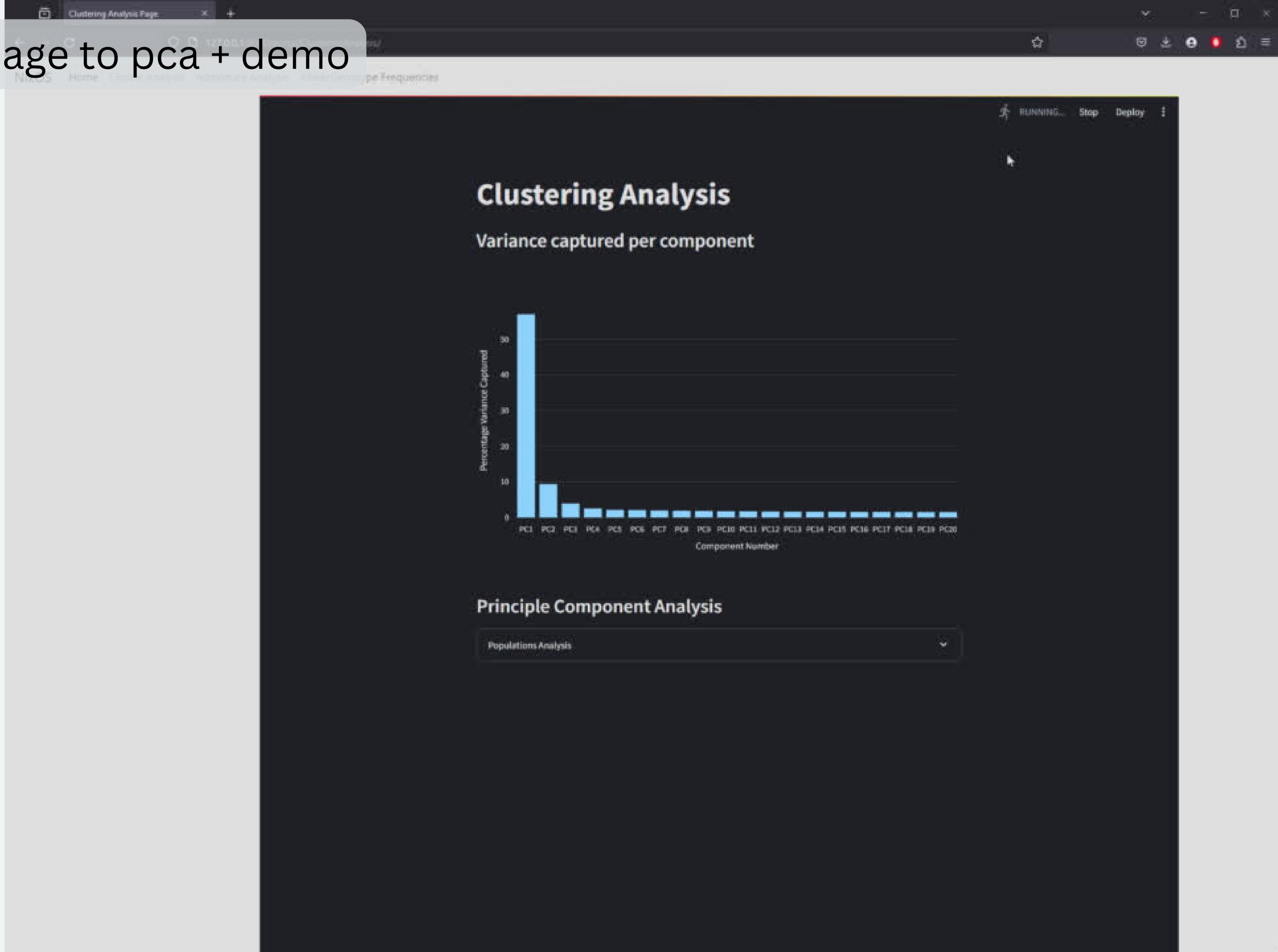
Uncover the potential of our platform with a range of analyses:

 - **Principal Component Analysis (PCA):** Discern patterns and relationships between populations, shedding light on intricate genetic variations.
 - **Admixture Analysis:** Investigate the nuanced ancestral composition of populations, unravelling the complex threads of human genetic history.
 - **Gene and Allele Frequencies:** Discover more about gene and allele frequencies of any population.
 - **Population Genetic Differentiation (FST):** Evaluate the degree of genetic divergence between populations, offering insights into their unique evolutionary paths.

Population Data

| Superpopulation Code | Superpopulation | Population Code | Population |
|----------------------|----------------------|-----------------|---|
| AFR | African Ancestry | ACB | African Caribbean in Barbados |
| AFR | African Ancestry | ASW | African Ancestry in Southwest US |
| AFR | African Ancestry | ESN | Esan in Nigeria |
| AFR | African Ancestry | GWD | Gambian in Western Division, The Gambia - Mandinka |
| AFR | African Ancestry | LWK | Luhya in Webuye, Kenya |
| AFR | African Ancestry | MSL | Mende in Sierra Leone |
| AFR | African Ancestry | YRI | Yoruba in Ibadan, Nigeria |
| AMR | American Ancestry | CLM | Colombian in Medellin, Colombia |
| AMR | American Ancestry | MXL | Mexican Ancestry in Los Angeles, California |
| AMR | American Ancestry | PEL | Peruvian in Lima, Peru |
| AMR | American Ancestry | PUR | Puerto Rican in Puerto Rico |
| EAS | East Asian Ancestry | CDX | Chinese Dai in Xishuangbanna, China |
| EAS | East Asian Ancestry | CHB | Han Chinese in Beijing, China |
| EAS | East Asian Ancestry | CHS | Southern Han Chinese |
| EAS | East Asian Ancestry | JPT | Japanese in Tokyo, Japan |
| EAS | East Asian Ancestry | KHV | Kinh in Ho Chi Minh City, Vietnam |
| EUR | European Ancestry | CEU | Utah residents (CEPH) with Northern and Western European ancestry |
| EUR | European Ancestry | FIN | Finnish in Finland |
| EUR | European Ancestry | GBR | British in England and Scotland |
| EUR | European Ancestry | IBS | Iberian populations in Spain |
| EUR | European Ancestry | TSI | Toscani in Italy |
| SAS | South Asian Ancestry | BEB | Bengali in Bangladesh |
| SAS | South Asian Ancestry | GIH | Gujarati Indians in Houston, TX |
| SAS | South Asian Ancestry | ITU | Indian Telugu in the UK |

homepage to pca + demo



pca to admixture

The screenshot shows a web application window titled "Admixture Analysis" running in a browser. The URL in the address bar is 127.0.0.1:8000/admixtureAnalysis/. The top navigation bar includes links for "NixOS", "Home", "Cluster Analysis", "Admixture Analysis", and "Allele/Genotype Frequencies". A toolbar at the top right features icons for "RUNNING...", "Stop", "Deploy", and a help icon.

The main content area is titled "Admixture Analysis" and contains two sections:

- Superpopulations Analysis**: A dropdown menu titled "Select Superpopulations" currently displays "SAS X".
- Populations Analysis**: A dropdown menu currently displays "Populations Analysis".

admix and list of IDs

NixOS Home Cluster Analysis Admixture Analyzer Allele/Genotype Frequencies

RUNNING Stop Deploy i

Allele & Genotype Frequencies

Select input method:

- List of IDs
- Genomic Coordinates
- List of Genes

Enter start genomic coordinate (e.g., 111784500):

Enter end genomic coordinate (e.g., 111787000):

genomic coordinates

NixOS Home Cluster Analysis Admixture Analysis Allele/Genotype Frequencies

Deploy :

Allele & Genotype Frequencies

Select input method:

List of IDs
 Genomic Coordinates
 List of Genes

Enter start genomic coordinate (e.g., 111784500):
111784500

Enter end genomic coordinate (e.g., 111787000):

[]

list of genes

127.0.0.1:8000/portal/AlleleGenotypeFrequencies/

Home Cluster Analysis Admixture Analysis Allele/Genotype Frequencies

RUNNING Stop Deploy :

Allele & Genotype Frequencies

Select input method:

- List of IDs
- Genomic Coordinates
- List of Genes

Enter a comma-separated list of Genes (e.g., KAZN,WASH7P,MACO1):

SNP IDs found within Gene:

Select Population(s) (e.g., FIN, BEB & GBR)

FIN × BEB × GBR ×

: First 10 SNPs found

**THANK'S FOR
LISTENING**

