# Project scope and objectives

## Background

Due to their nature, forensic samples experience genetic degradation that prohibits complete genotyping of short-tandem-repeats (STRs). This same problem does wholly exist for single nucleotide variants (SNVs). Incorporating these into analyses will therefore repair the fidelity lost by incomplete STR profiles.

## Aims

Identify STR-SNP markers relevant to the African populations.

Objectives

1. Collate STRs from existing databases and select those that can be in an African context (penta- & tetranucleotides, heterozygosity)
2. Identify SNPs within 200 base pairs (bp) of relevant STRs by searching public databases and select for those with a population frequency greater than 15%.
3. Further annotate the selected SNPs

# Materials and Methods

## List STRs

Public databases have the option to download their entire set and some of the relevant databases do include population statistics for their STRs. Where these statistics are not available, the data will be combined with other datasets that allow population-specific interrogation. The starting list for these databases include STRBase, PopSTR and 1000Genomes.

The STRs will then be given rank scores on multiple criteria, such as number of alleles, rate of mutation, ability to be amplified and finally by how standard they are in current use. These rank scores will then be used to select those STRs that can successfully be used discriminatorily within an African context.

The coordinates of the selected STRs will then be used to query SNP databases to select those that are within 200 bp of the STRs. Initially, dbSNP, 1000 Genomes and GnomAD will be used for this purpose, but the list may expand based on available data. These SNPs will then undergo annotation using ANNOVAR. This annotation will then allow elimination of low frequency SNPs (< 0.15) as well as indicate their association with phenotypes.

# Project admin

## Milestones and timeline

|  |  |  |  |
| --- | --- | --- | --- |
| Category/Aim | Objective | Working hours | Projected completion |
| Admin | Meetings | 5 | 2024-07-29 |
| Analytical study plan | 5 | 2024-07-09 |
| Reports/feedback | 10 |  |
| Aim 1 | Collate African STRs | 30 | 2024-07-29 |
| Identify SNPs | 10 | 2024-07-29 |
| SNP phenotypes | 10 | 2024-07-29 |
| Total |  | 70 |  |