Link of the Study:

This study analyzes the worldwide human relationships based on genome-wide patterns of variation. It studied 938 unrelated individuals from 51 populations of the Human Genome Diversity Panel at 650,000 common single-nucleotide polymorphism loci. Based on principal component analysis, the result primary genome factors support the hypothesis of a serial founder effect with a single origin in sub-Saharan Africa. Although previous studies on geographical structuring of human populations have been done at the continent level, they are limited to a small fraction of the genome and to limited population. This study has allows a detailed characterization of worldwide genetic variation.

Researchers first studied genetic ancestry of each individual without using his/her population identity; instead, they suppose that each individual was originated from K ancestral (K coefficients in experiment). The result reputes a widely accepted theory that mixed ancestries arise from recent admixture among multiple founder populations. It shows that the estimated mixed ancestry is from either recent admixture or shared ancestry before the divergence of the two populations but without subsequent gene flow between them.

After that, researchers calculated the Wright's fixation indices Fsts among the 51 populations from the population allele frequencies across all autosomal SNPs. Principal Component Analysis was then performed on the Fst matrix. To identify the components of the overall genetic variation, researchers carried out analysis of molecular variance (AMOVA) and derived 3 categories: within-population (WP), among-population-within-group (AP/WG), among geographical region (AG).

In this study, PCA is important for finding similar and distinct traits between populations from different continents. It shows the genetic distribution of different populations based on two principle components as shown below. PC1 primarily describes the contrast between sub-Saharan Africans and non-Africans; PC2 primarily describes the East-West difference in Eurasia. The variances show that Africa and America, East Asia and Middle East are relatively genetically distinctive.

Two Component PCA
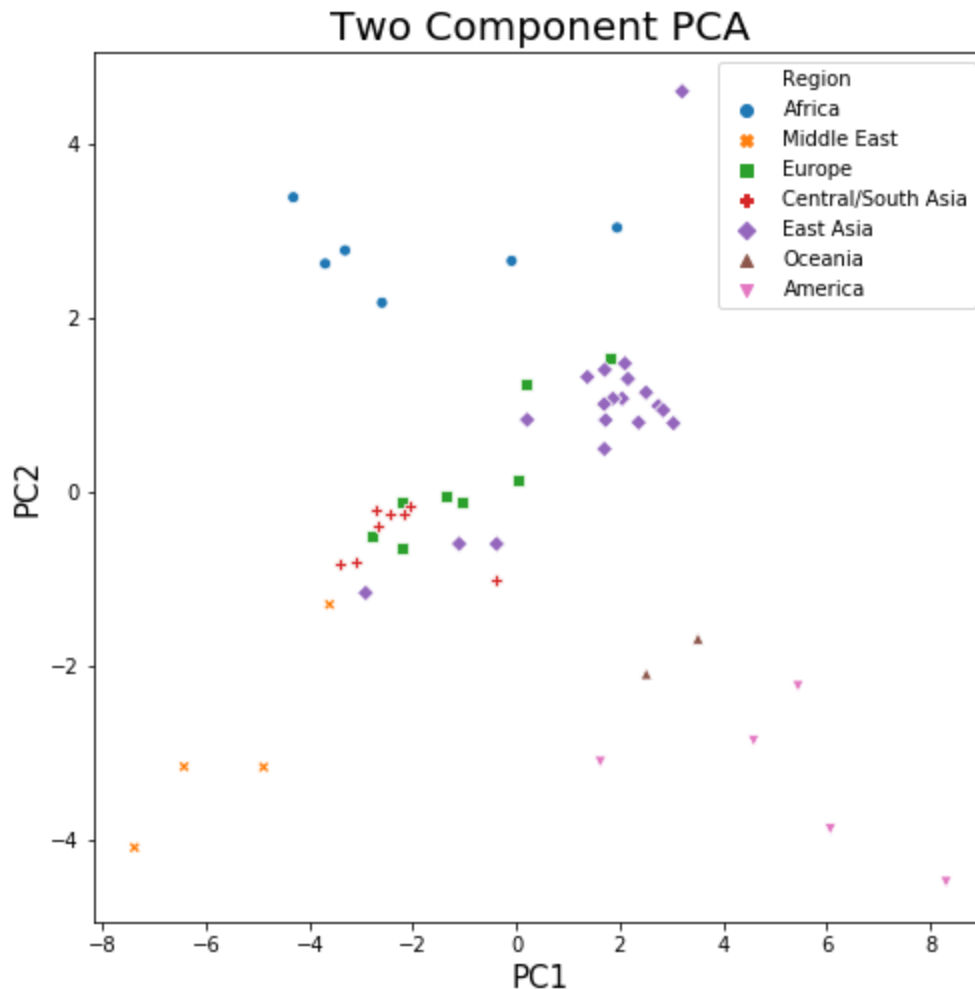
The original data has 33 columns. Important features include:

N_Male, N_Female: number of males and number of females among the population

c1-c7: the ancestry coefficients for the seven components as shown in K=7 Fig., averaged over all individuals within each population.

Geo.dist(km): geometric distance in kilometer

Ne(ChrX): Ne based on 20 segments in Chrx

BranchLength_NJ: branch lengths as measured from the root, from the Neighbor Joining Tree

……

The reason I choose this study is that it analyses the genetic relations at continent level comprehensively. Since there are many treatments that only work efficiently for people in certain areas, it is crucial to understand the genetic relations first in order to better understand both diseases and treatments.