



7TH ANNUAL INTERNATIONAL

ARAB WOMEN IN
COMPUTING
AND TECHNOLOGY
CONFERENCE

ArabWIC 2021



ArabWIC

Understanding Population Genomics Using Data Science

Shahad Hardan, Khalifa University Alumni
Applied Mathematics and Statistics Graduate
shahdhardan@gmail.com

7TH ANNUAL INTERNATIONAL
ARAB WOMEN IN COMPUTING AND TECHNOLOGY CONFERENCE



ArabWIC

Environment

Genes



7TH ANNUAL INTERNATIONAL
ARAB WOMEN IN COMPUTING AND TECHNOLOGY CONFERENCE

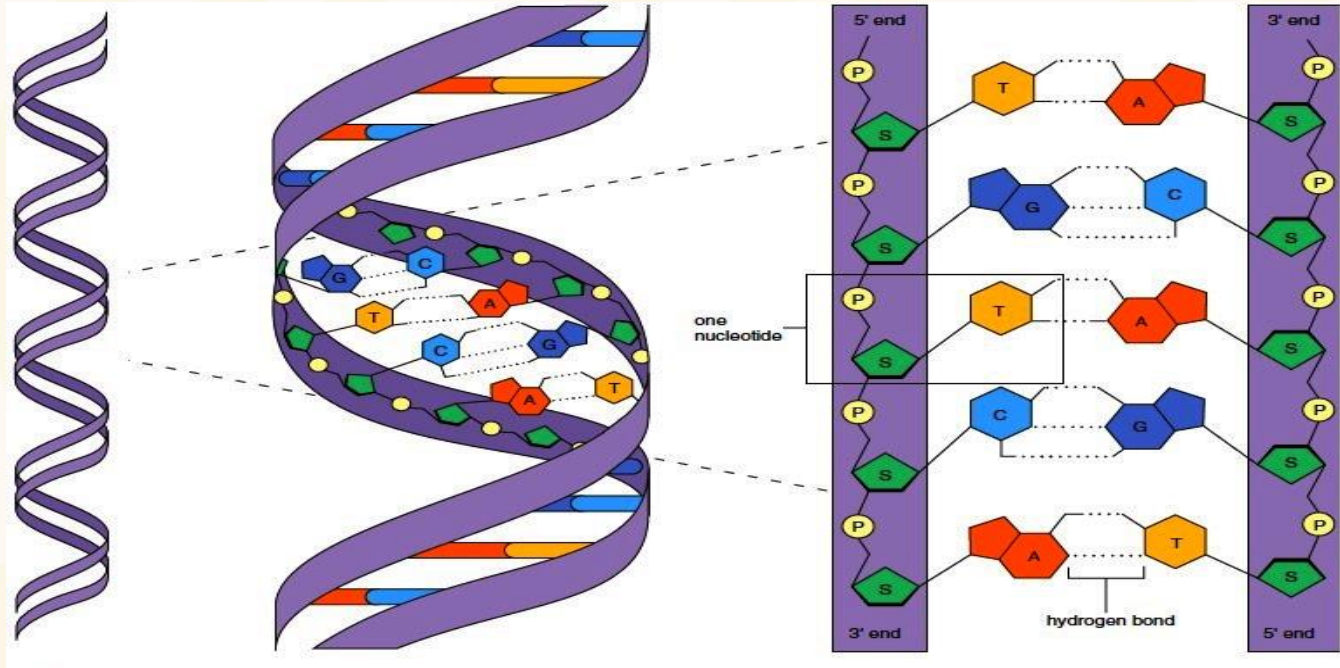




Nucleotides: The building blocks of our DNA



A T
C G

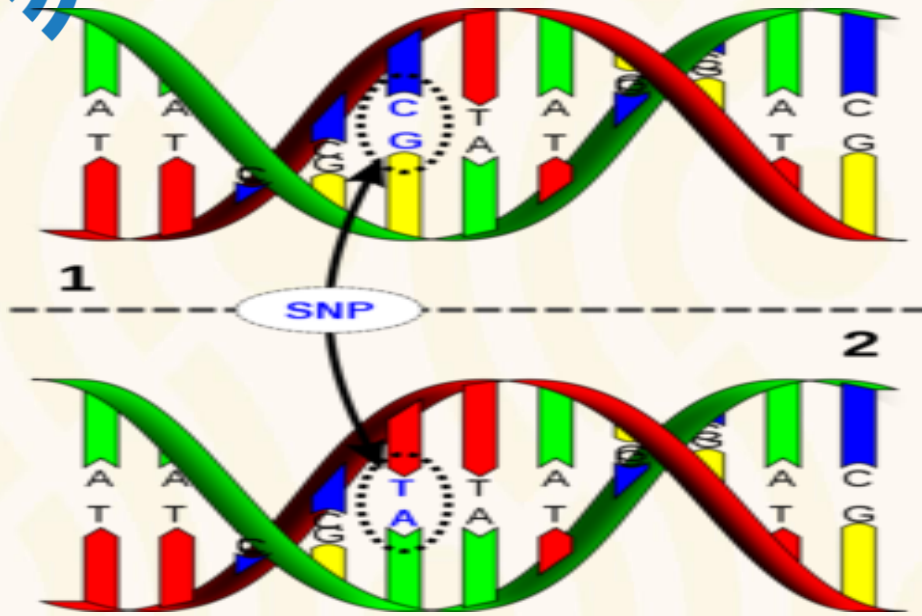




What are SNPs?



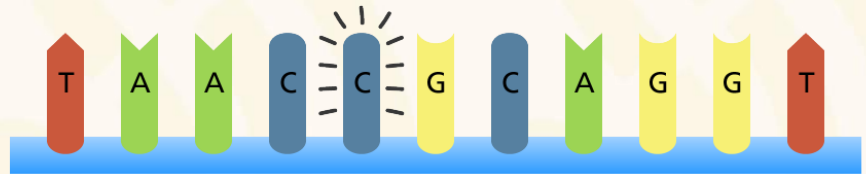
ArabWIC



Original sequence



Point mutation





The Curse of Dimensionality



ArabWIC

Around 10 million
SNPs in the human
genome

The curse
of dimensionality

Data is huge and
cannot be easily
processed

Dimensionality reduction
methods are needed.
ex: PCA



Objective: understand high dimensional genomic data and unlock insights about populations



ArabWIC

Data: SNP Data

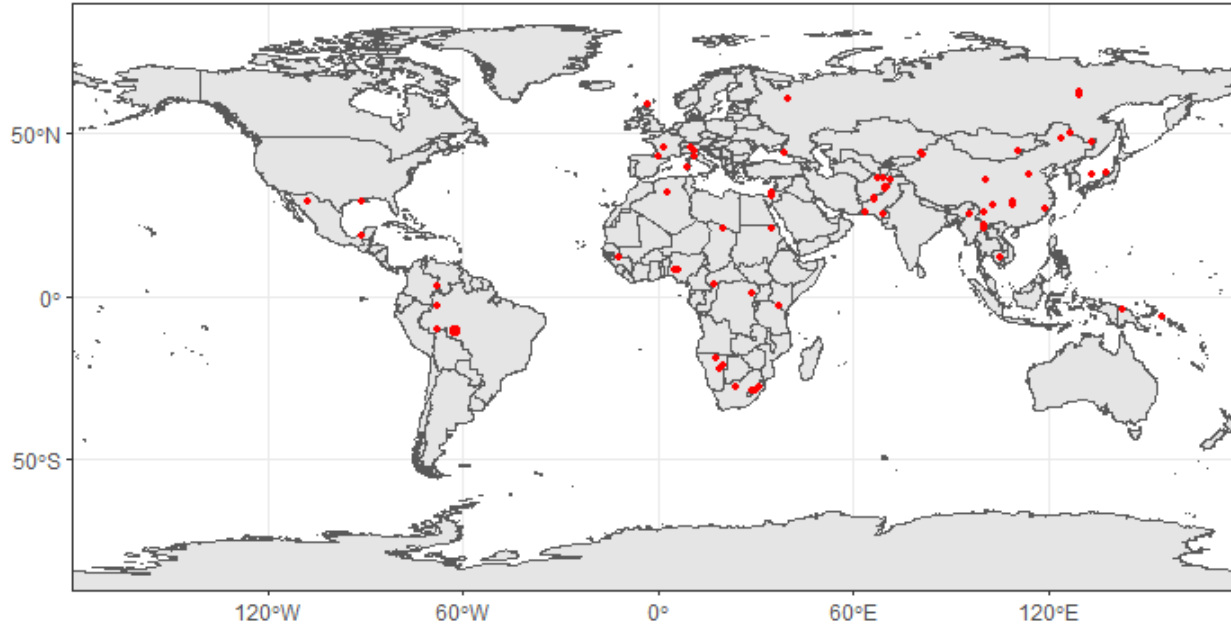
Method : Principal Component Analysis



ArabWIC

Our Data

Locations of the Subjects of the HGDP-CEPH Data



Source: The Human Genome Diversity Panel

Size: 660K SNPs, 1043 subjects from 7 population groups



Our Data



ArabWIC

	rs9439462	rs3766178	rs2031709	rs3128342	rs880051
HGDP00448	TT	AG	GG	TT	TC
HGDP00479	TC	GG	GG	TT	TT
HGDP00985	TT	GG	GG	TT	CC
HGDP01094	TT	GG	GG	TT	CC
HGDP00982	TT	GG	TG	TT	TC
HGDP00911	TC	GG	GG	TT	TT
HGDP01202	TC	GG	TG	TT	TT
HGDP00927	TC	GG	GG	TT	TT
HGDP00461	TT	GG	GG	TT	TC

Source: The Human Genome Diversity Panel

Size: 660K SNPs, 1043 subjects from 7 population groups

7TH ANNUAL INTERNATIONAL

ARAB WOMEN IN COMPUTING AND TECHNOLOGY CONFERENCE



What is PCA? A technique that transforms a large set of variables into a smaller one that still contains most of the information in the large set.

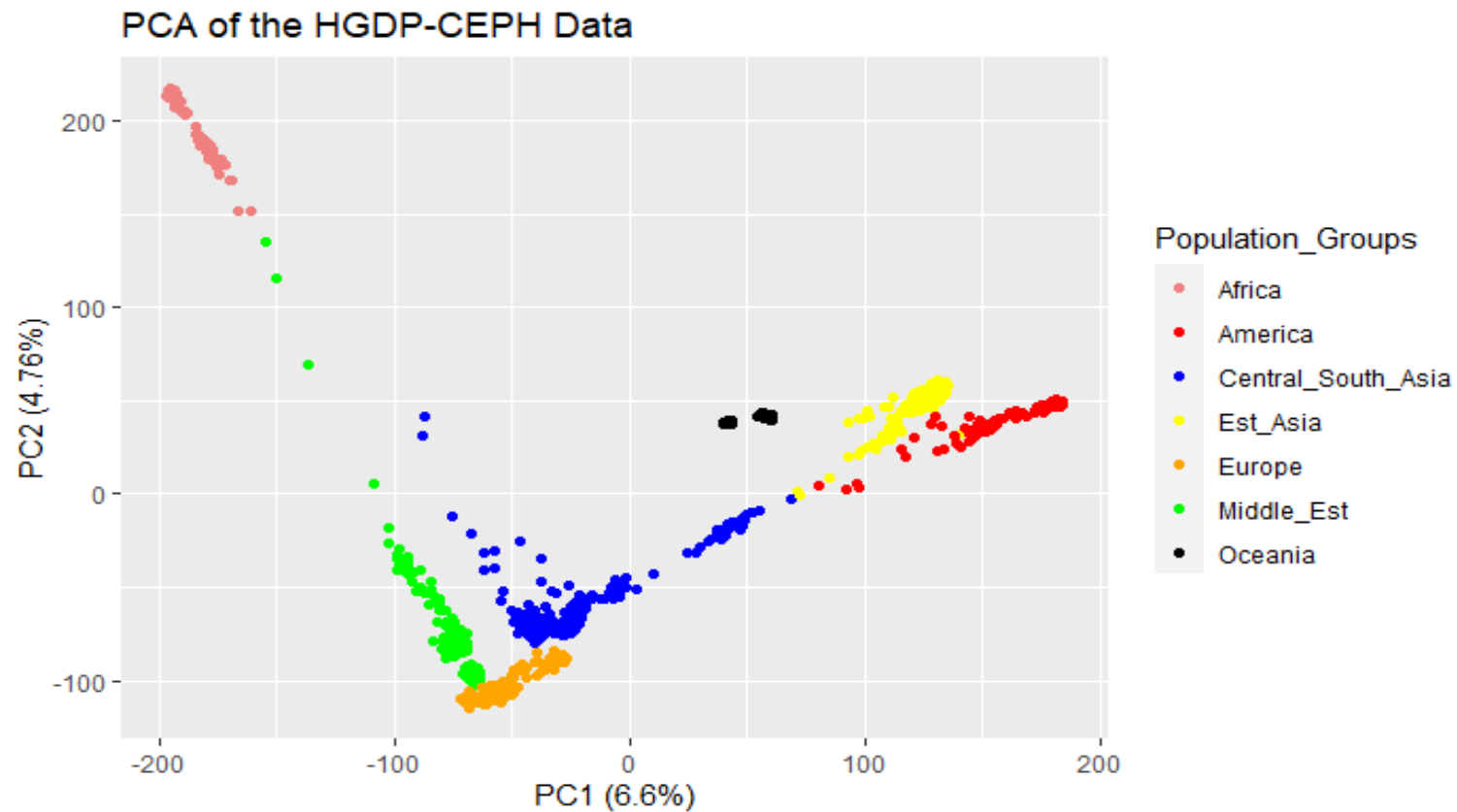


Application: convert our data size from
 $1043 \times 660K \rightarrow 1043 \times 2$

Why? To investigate the population structure and get value from data



Results





7TH ANNUAL INTERNATIONAL
ARAB WOMEN IN COMPUTING AND TECHNOLOGY CONFERENCE



Thank You!

Acknowledgments:

Dr. Samuel Feng, Khalifa University

Khalifa University Center for Biotechnology

Eng. Gebrehiwot Kalayu, Khalifa University