

Genetic Structure of Iraqi Populations

Ammar M. Shakir

Freelance researcher

Email: ammar_m_alani@yahoo.com

Abstract

Background: The field of genetics has become more and more crucial to our comprehension of human evolution. Genetic developments have given us new means of verifying or disproving historical theories. Our ancestors mingled with prehistoric humans, who seem to have more genetic variation in common, for example with modern non-African humans than with modern sub-Saharan African humans.

Method: A laboratory specializing in family tree construction evaluated 2238 samples in total, drawn at random from a group of Iraqis.

Results: It was demonstrated that the Neolithic (Caucasian), Natufian (local Jazri Levantine), and Anatolian Iranians make up the majority of Iraqis. Examining a single sample of pure Jazri ancestry, the findings showed that the sample was 60% Natufian (local Jazri Levantine), 28% Neolithic.

Conclusions: The genetic composition has no relation to the paternal genealogy as it differs from one computer to another, as each computer uses a different Algorithm. Generally, the genetic composition differs according to the comparison samples, when the purer samples the more accurate results, and to the genetic distances, where the closer distance to zero, the more precise results.

Keywords: Genetic structure, family tree, Iraqi population

Introduction

Patterns of human genetic diversity suggest that all modern humans originated from a small population in Africa that expanded rapidly 50,000 years ago to occupy the whole world. While moving into new environments, genetic drift and natural selection affected populations differently, creating genetic structure. By understanding the genetic structure of human populations, we can reconstruct human history and understand the genetic basis of diseases (Tishkoff *et al.*, 2012).

Traditionally, information about the origin of human populations came from fossil and archeological data. But over the last half century, genetics has played an increasingly important role in our understanding of human evolution. Advances in genetics have provided new tools for approving or rejecting hypotheses about our past. Today, one can expect new discoveries in human history and evolution to emerge from a genetics lab as from archeological sites. Indeed, fascinating new discoveries about our origins appeared recently through studying human genomes which allowed resolving questions about human evolution previously deemed inscrutable. One thrilling example, which substantially revised previous models on our evolution, is the finding that our ancestors have admixed with archaic humans which appear to share more genetic variants with present-day non-African humans than with present-day humans in sub-Saharan Africa. Such findings were made possible through rapid

advancement of sequence and genotyping technologies which facilitate deficient survey of archaic and modern human populations. The flow of data allowed the construction of more accurate mathematical models that can infer past processes from modern population diversity. The implications are vast, not only to understand how we came to be the way we are but also to speculate about the future evolution of our species (Sebastiani *et al.*, 2012). The genetic markers used to study populations have changed considerably over the last 40 years. The major aim remains to construct a realistic understanding of human population changes in space and time (Wood *et al.*, 2014). Genetic markers showed differences between human populations first by using blood groups and protein types (Cann *et al.*, 2002). These “classical markers” preceded DNA based markers and comparisons were based only on allele frequencies since the molecular basis of the polymorphisms was unknown (Jobling *et al.*, 2004). With the advance of polymerase chain reaction (PCR) methods, molecular markers were introduced and allowed, in addition of allele frequencies measure, an assessment of evolutionary distances between alleles at a locus (Julian C. Knight, 2009). The first molecular markers focused on the mitochondrial DNA (mtDNA) and the non-recombining region of the Y chromosome (NRY). mtDNA is inherited maternally and transmitted from a mother to her descendants while the NRY is inherited paternally passing down from father to son only (Wallace and Chalkia, 2013). These markers pass intact from generation to generation, escaping recombination and the chromosome shuffling process (Jobling and Tyler-Smith, 2017). They change only by mutation and therefore present a simple record of their history and allow an easy construction of unique phylogenies (Hartl and Clark, 2019). The development of genome-wide SNP chips, and later advances in whole-genome sequencing, allowed hundreds of thousands of polymorphisms to be surveyed collectively (Moreno-Mayar *et al.*, 2018). This permitted the assessment of evolutionary processes without a particular sex bias (Collins and Varmus, 2015). In addition, the effective population size of genome-wide markers is expected to be four times that of mtDNA and NRY which pushes back the time of coalescence, giving more in-depth view of the population history. Also, this makes the genome-wide markers less prone to genetic drift and when analyzed collectively the markers should present a selectively neutral overall picture (Telenti and Goldstein, 2018). Rapid technological advancements have made genetic data available to scientists to apply evolutionary concepts to a wide range of problems that we could not have imagined one day. Evolutionary approaches have helped reconstruct the origin and history of humans via studying the genomes of Central Asians, Near Easterners, and North Africans (Lazaridis *et al.*, 2014). It has been shown that many aspects of modern human diversity can be explained by developments during the warming period after the last ice age (Malmström *et al.*, 2019). The favorable climate conditions and the spread of agriculture resulted in a human population explosion from few millions to seven billions in less than 10,000 years, triggering an extraordinary evolutionary trial and error on the genome of modern humans (Haber and Tyler-Smith, 2017). This gives us a unique opportunity to understand how evolution shapes the diversity of populations and enables our understanding of the kinds of mechanisms that can affect the future evolutionary trajectory of humans.

Method

The 119 Samples were prepared by cheek swab; the DNA samples were genotyped for 630,000 autosomal SNPs by Family Tree DNA (FTDNA-Gene By Gene, Ltd, Houston, TX,

www.familytreedna.com). We integrated the genotyped data of 119 modern Iraqis By PLINK 1.9 (Auton *et al.*, 2015) with modern and ancient data from East and West, South Asia, Europe and Africa From 1240K dataset from David Reich’s Lab website, Harvard University Alexander *et al.* (2009) and the dataset generated from the integration of 2238 samples after applying filtering and quality control with Plink was used for Determining the genetic structure, we entered ADMIXTURE 1.3.0 (Alexander *et al.*, 2009) and it run for k=5 to k=14. The lowest amount of CV error was related to k=9. This cross sectional study was conducted on a sample of Iraqis, whom were selected randomly. The study involved taking 2238 samples from the study group were examined in a specialized company (DNA Family Tree), a company specialized in family tree construction. The software program used was the Admixture ⁽¹⁾, which is a computer software used for the analysis of genomic details that are linked to population genetic variability. It aims at understanding and analyzing the genetic backgrounds of individuals and different human groups. The program uses the Expectation-Maximization (EM) Algorithm to estimate the size and share of each ethnic group in an individual's genome (Ma, 2001). The implementation of the program depends on the available genome data, which usually represented as DNA sequences in different individuals. The program counts the statistical distances between individuals based on the available DNA sequences. Accordingly, the program software determines the ethnic groups. The program was operated using different sorting or dismantling values/degrees. The software sorts the peoples and samples according to 7 genetic components.

The data of some nucleic acid chromosomes were used to test 2238 samples from different contemporary and very ancient counties countries and races. A total of 2119 samples were taken from the results of previous studies conducted at and published by Harvard University ⁽²⁾. The other 119 samples were from Iraq from different strains, ethnic groups, religions, and country provinces that were collected with the help of the Genetic Project of Iraq.

The researcher was trained to effectively use the program and a checking mechanism was followed to validate the results. Table (1) shows the distribution of the samples taken from different Iraqi ethnic/racial/religious groups. The highest number was taken from Arabs, while the least were taken from Mi'dans and Gypsies.

Table (1): Ethnic/racial/religious distribution of the samples from Iraq

Ethnicity/race/religion	No.	Ethnicity/race/religion	No.
Arabs	62	Marsh Arabs	5
Christians	10	Faili	5
Kurds	9	Yezidis	4
Turkmans	7	Mi'dans	3
Jewish	6	Gypsies	3
Manda'is	5		
Total	119		

(1) <https://dalexander.github.io/admixture/publications.html>

(2) https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data?fbclid=IwAR2ZgMMV3gGPOw5lsZJN7CjqO0nZyO5Xp-gq87YPxiHlvKMLoren_fyu98M

The geographical distribution of the samples from Iraq is shown in Table (2). The highest number of samples was taken in Ninawa and the lowest was taken in Baghdad, Duhok, and Erbil.

Table (2): Geographical distribution of the samples from Iraq

Province	No.	Province	No.
Ninawa	18	Al-Najaf Al-Ashraf	5
Misan	14	Babylon	5
Anbar	10	Holy Karbala	5
Thi-Qar	9	Diyala	5
Kirkuk	7	Salahu Al-Deen	5
Basrah	6	Al-Muthanna	4
Diwaniya	6	Baghdad	3
Suleimaniya	6	Duhok	3
Wasit	5	Erbil	3
Total	119		

Results

The most prevalent constituent of the Iraqis is the Neolithic (Caucasian) Iranian, the Natufian (local Jazri⁽³⁾ Levantine), and the Anatolian ones. When a pure Jazri origin Bedwings was examined as a solitary sample, its results were 60% Natufian (local Jazri Levantine), 28% Neolithic (Caucasian) Iranian, and 9% Neolithic Turkish.

Generally, the constituent, South Asia (India) does not exceed the 5% among Iraqis, however, its rate is as high as 38% among Gypsies Iraqis and may be higher (44%) in pure Gypsies samples (in isolated samples examination). But, it decreases in Christian and Jewish samples (which may be zero in isolated samples from them).

The African constituent is very few among Iraqis in general; it slightly increases among Arabs, especially marsh Arabs (through examining the isolated samples), then the Gypsies, and it is negligible among others.

The Jazri Levantine local constituent is widely prevalent among all Iraqis, even those who do not speak Afro-Asian languages, like Kurds and Turkmans; it forms the highest proportion among Arabs (34%). This percentage increases among Al-Badiya dwellers to reach 60% (examination of isolated samples).

Generally, the Anatolian Neolithic constituent was found in all Iraqis, mostly high (26%) among Jewish and Christians. The European effect was found to be approximately similar among all Iraqi groups in a proportion ranges between 4-6%. The Iranian Neolithic (Caucasian) constituent seems to hold the highest representation among all Iraqi samples (more than one third of the genetic composition. It forms the highest percentage among Failies, Kurds, and Yazidis in a percentage approaches the half of the genetic composition; then the Manda'is and Christians come next.

3. Derived from "Al-Jazeera"

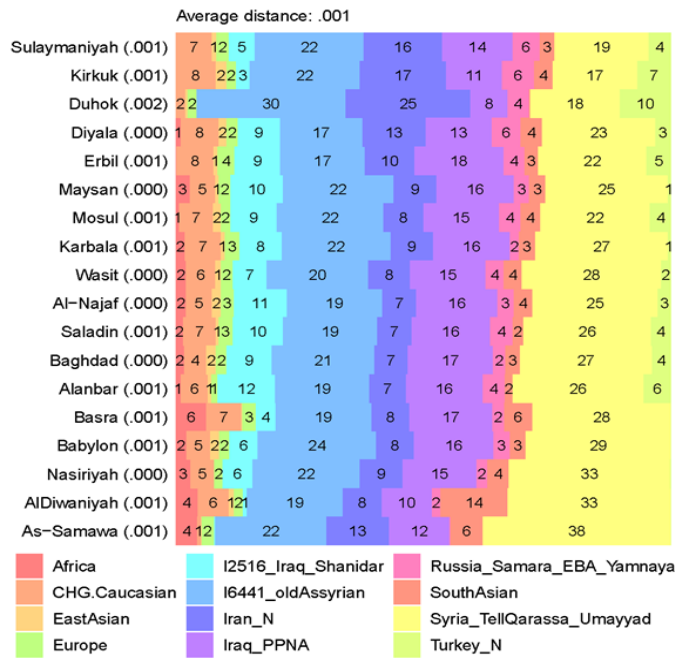


Figure (1): The genetic component according to ethnicity/race and religion (explanation of the colors/numbers shown in the figure)

The Eastern-Asian effect was found to be very little; it expressed itself in Turkmans, while it was nearly zero in Manda'is and Christians. The variability among Kurds and Failis was nearly identical, which ascertains that they constitute the same population group. The Yezidis had similar composition to Kurds, but the general variability was not completely identical, when the proportion of the Jazri Levantine local constituent reaches down to 18% and the Anatolian Neolithic proportion goes high to a level similar to the other non-Muslim minorities like Manda'is.

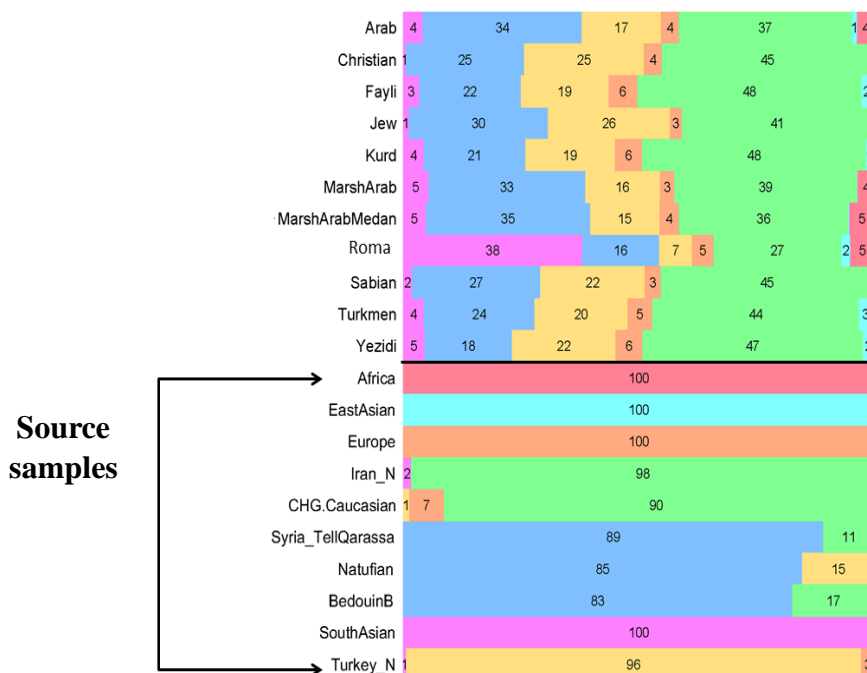


Figure (2): Percentage of similarity according to the genetic distance of the provinces

The highest proportion, which mimics Caucasian sample, was in Kurds and Failis. The highest proportion, which mimics Chandar sample, was in the Gypsies.

That is because Chandar sample has a high level Zagros/Caucasian constituent and Gypsies hold high ancient Iranian constituents.

The highest similarity to the Assyrian sample was found to be to Manda'is followed by Jewish people.

The highest similarity to the Neolithic Iranian (Caucasian) sample was found to be to Manda'is. The highest similarity to the Neolithic Iraqi sample was found to be to Arabs followed by Kurds, which goes down in Manda'is.

The highest genetic similarity percentage to the Southern Asian (Indian) sample was found to be to Gypsies. The highest similarity to the Syrian (Qarasa Hill) sample was found to be to Arabs. The highest similarity to the Neolithic Turkish (Anatolian) sample was found to be to Jewish, Christians, and Yezidis.

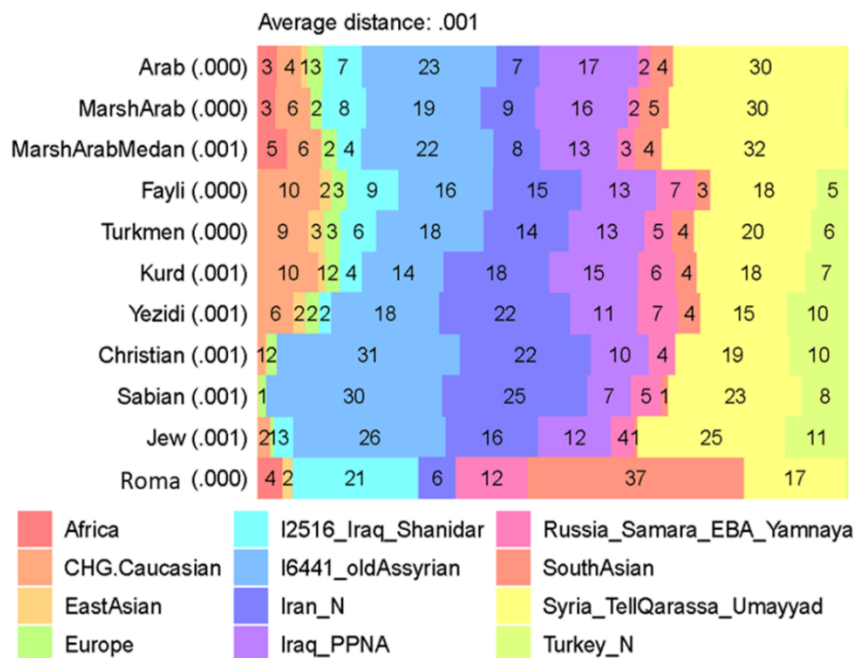


Figure (3): Percentage of similarity according to genetic distance for ethnicities and nationalities

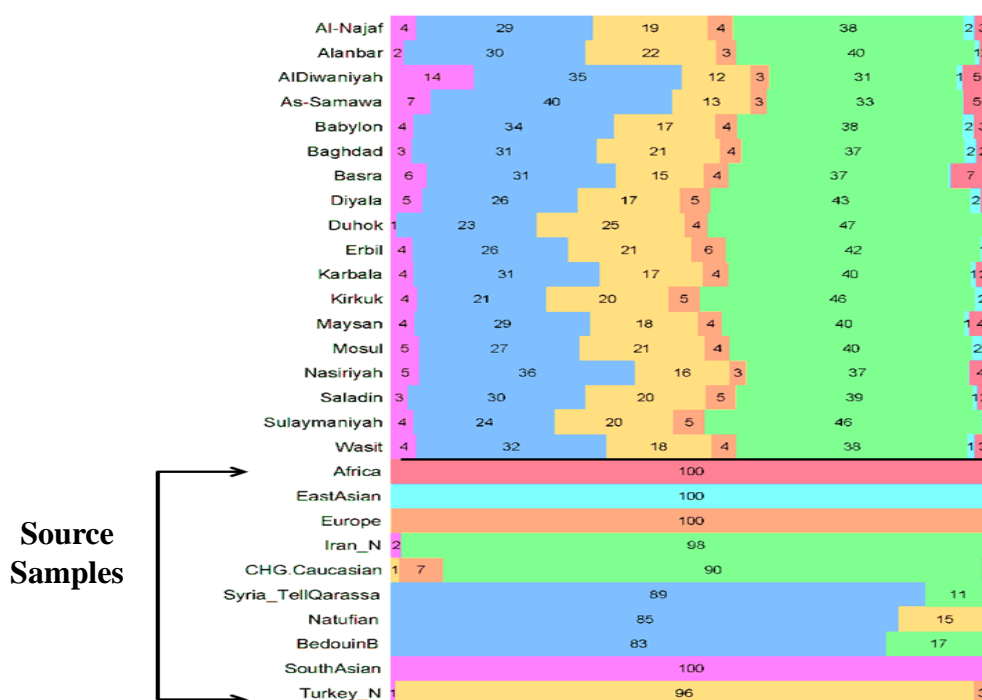
The most important constituents of the Iraqi provinces were the Neolithic Iranian (Caucasian), the Jazri Levantine local, and the Anatolian. The Southern-Asian constituent expressed itself in small proportions and nearly equal in all provinces; the reason behind its high expression in Al-Diwniya and Ninawa provinces was because of the presence of the Gypsies samples.

The Jazri Levantine local constituent was the main constituent in all provinces of the country; its highest percentage was recorded in Al-Muthanna, when it reached up to 40%, because of its vicinity to the Arab Peninsula. It shows less expression in the Northern provinces. The highest proportion (25%) of the Anatolian constituent was expressed in samples from Duhok.

This belongs to the fact that most of the participants had overlapping with the Christian samples.

The European constituent was nearly equal in all provinces in percentage of 4-6%. The Caucasian Iranian constituent forms more than one-third of the genetic constituents in the provinces; its proportion increases in the north of Iraq. The Eastern-Asian constituent was very little in all provinces. The African constituent was little in the southern and middle provinces and about to be zero in the northern ones. The highest Chandar-similar sample's proportion was in Anbar and Al-NajafAshraf, which was followed by Salahul-Deen, Misan, Diyala, and Erbil. In Duhok and Al-Muthanna, it was zero.

The highest similarity proportion to Assyrian sample was found in Duhok. Similarity to the Neolithic Iraqi sample was, almost close between the governorates in percentage ranges from 10% to 17%, except in Duhok, which was 8%.



Discussion

The study of genetics has grown more important to our understanding of human evolution. Genetic advancements have provided us with new tools for validating or disproving historical hypotheses. Scientists may now apply evolutionary notions to a wide range of issues that were previously unimaginable thanks to the rapid improvements in technology, including genetic data. The genomes of Central Asians, Near Easterners, and North Africans have been studied using evolutionary techniques to help reconstruct human origins and history. This research sought to evaluate genetic diversity and organization among Iraqi governorates. These people inhabit a vast region of central, eastern, and southern Iraq.

According to results of the study, the Neolithic (Caucasian) Iranian is the most common ethnic group among Iraqis. Neolithic sites in Mesopotamia, part of the Fertile Crescent, revealed square mud homes, furnaces, and clay basins. They produced bone tools, needles, spindle whorls, and domesticated animals, indicating agriculture (Stansfield, 2016).

The element, South Asia (India), does not surpass 5% among Iraqis; nevertheless, it can reach 38% among Gypsies Iraqis and may be higher (44%) in pure Gypsy samples. Migration models suggest that gene flow to Iraq originates from East Africa, with the Levantine corridor being the most probable exit route from Africa (Lazim *et al.*, 2020). The African component is quite little among Iraqis in general; it is slightly higher among Arabs, particularly marsh Arabs. Since the Abbasid Empire enslaved Black people from the Zanj area in the ninth century, there has been a Black Iraqi community in Iraq (Dabbas *et al.*, 2011). In general, the Anatolian Neolithic ingredient was found in all Iraqis, with the highest percentage (26%) among Jewish and Christians. The constituents registered in the old comparison references do not represent the current ethnic composition, although it still represents part of it. For example, the Iranian Neolithic-Caucasus Hunter Gatherer (CHG) does not represent Persian ethnicity, it is formed from constituents most of them is Caucasian. Also, the Iranian Neolithic-Caucasus Hunter Gatherer (CHG) Turkish sample does not represent Turks, but Anatolia people. Again, the Natufian sample does not represent Arabs. If the constituent present in the comparison references was contemporary, it would represent the current ethnic composition.

Kurdish and Faili variability was almost the same, indicating that they belong to the same demographic group. The Yezidis had a similar makeup to Kurds, but the general variability was not entirely the same when the proportion of the Jaziri Levantine local constituent reaches down to 18% and the Anatolian Neolithic proportion goes high to a level equivalent to the other non-Muslim minorities like Manda'is.

The genetic composition has no relation to the paternal genealogy. For example, an Arab person's genetic strain can be Arabic, but his genetic composition is 90% Indian because the marriage of his ancestral to Indian ladies. The genetic composition differs from a computer to another, as each computer uses a different Algorithm. The genetic composition differs according to the comparison samples, when the purer samples the more accurate results.

The genetic composition differs according to the genetic distances, where the closer distance to zero, the more precise results. Some of the genetic references, used for comparison are not 100% pure. Therefore, it is wrong to consider it representative to a specific entity, like the Assyrians, which is a mixture of Levantines, Anatolians, and Caucasians.

References

- [1] Al-Dabbas, M.; Abbas, M.A.; Al-Khafaji, R. The mineralogical and micro-organisms effects of regional dust storms over Middle East region. *Int. J. Water Resour. Arid. Environ.* 2011, 1, 129–141.
- [2] Ancient human genomes suggest three ancestral populations for present-day Europeans" by Lazaridis et al. (2014).
- [3] Auton, A.; Abecasis, G. R.; Altshuler, D. M.; Durbin, R. M.; Bentley, D. R.; Chakravarti, A.; Clark, A. G.; Donnelly, P.; Eichler, E. E.; Flicek, P.; Gabriel, S. B.; Gibbs, R. A.; Green, E. D.; Hurles, M. E.; Knoppers, B. M.; Korbel, J. O.; Lander, E. S.; Lee, C.; Lehrach, H, Schloss, J. A. (2015). A global reference for human genetic variation. *In Nature*, 526, 7571.
- [4] D.H. Alexander, J. Novembre, and K. Lange. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19:1655–1664, 20

- [5] Genetic markers and human evolution" by Jobling et al. (2004).
- [6] Genetic markers: progress and perspectives" by Cann et al. (2002).
- [7] Genetic signatures of exceptional longevity in humans by Sebastiani et al. (2012).
- [8] Genome-wide data from a single assay reveals multiple waves of migration into southern South America" by Moreno-Mayar et al. (2018).
- [9] Human Genetic Diversity and the Spread of Modern Humans by Wood et al. (2014).
- [10] Human Genetic Diversity: Functional Consequences for Health and Disease" by Julian C. Knight (2009).
- [11] Human population genomics in the Middle East" by Haber and Tyler-Smith (2017).
- [12] Lazim, H., Almohammed, E.K., Hadi, S. et al. Population genetic diversity in an Iraqi population and gene flow across the Arabian Peninsula. *Sci Rep* 10, 15289 (2020).
- [13] Ma, Q., Wang, J.T., Shasha, D. and Wu, C.H., (2001) 'DNA sequence classification via an expectation maximization algorithm and neural networks: A case study', *IEEE*
- [14] Molecular markers in human population genetics" in the book "Principles of Population Genetics" by Hartl and Clark (2019).
- [15] Stansfield, G. (2016). *Iraq: People, History, Politics* (2nd ed.). Malden, MA: Polity Press.
- [16] The Genomic History of Eastern Europe" by Malmström et al. (2019).
- [17] The Genomic History of Human Populations in Africa by Tishkoff et al. (2012).
- [18] The Human Y Chromosome: An Evolutionary Marker Comes of Age" by Jobling and Tyler-Smith (2017).
- [19] The Power of Mitochondrial DNA in Human Genetic Studies" by Wallace and Chalkia (2013).
- [20] The Precision Medicine Initiative Cohort Program: Building a Research Foundation for 21st Century Medicine" by Collins and Varmus (2015).
- [21] *Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 31(4), pp. 468–475. doi: 10.1109/5326.983930.
- [22] Whole genome sequencing for understanding human biology" by Telenti and Goldstein (2018).