Forensic Science, Medicine and Pathology
https://doi.org/10.1007/s12024-025-01115-3

POPULATION DATA

Deciphering the genetic diversity in the Arabian Peninsula and Africa: insights from Y-STR data

Abdullah Hadi[1] · **Shams Hadi**[2] · **Eida Khalaf Almohammed**[3] · **Hayder Lazim**[4]

Accepted: 11 October 2025

Abstract

Middle Eastern and African populations make up a significant portion of the global population and exhibit substantial genetic diversity. However, genetic studies on these populations have been largely underrepresented compared to other populations. This study analysed published Y-STR data for 186 populations and regions, including 14,504 individuals from 52 Middle Eastern and 134 African populations. The highest genetic diversity was found at the DYS458 locus in the Middle East and North Africa, and at the DYS385b locus in other African regions. Multidimensional scaling (MDS) analysis and genetic distance calculations between Middle Eastern and African populations revealed five distinct clusters. The Arabian Peninsula countries formed two small clusters, while most African countries formed two mains centrally located clusters. The most common haplogroups in the Middle Eastern populations were J1a (29.4%), while in the African populations, E1b1a (43.2%) was the most prevalent. This study examined two allelic richness parameters: distinct and private alleles. Central Africa showed the highest levels of distinct alleles, with the Middle East having the third-highest level. The prevalence of private alleles in the Middle East was moderate, lower than South Africa but higher than North Africa. A population Q-matrix graph was constructed, yielding 10 clusters (K = 10) that identified population clusters in the Y-STR data corresponding to specific geographical regions and revealed stronger sub-grouping of countries within each population.

Keywords Y-Chromosome; Middle East; African gene flow; Y-Haplogroups; Jewish population; Ancestry variability

Abdullah Hadi and Shams Hadi Joint first authors.

✉ Hayder Lazim
alazawihayder@yahoo.com; lazimh@edgehill.ac.uk

Abdullah Hadi
h.abdullah819@yahoo.com

Shams Hadi
shamshayder77@yahoo.com

Eida Khalaf Almohammed
eida-k-al@hotmail.com

[1] Royal Preston Hospital, Sharoe Green Lane, Fulwood, Preston PR2 9HT, UK

[2] School of Medicine, University of Lancashire, 135A Adelphi St, Preston PR1 7BH, UK

[3] Ministry of Interior of Qatar, Doha, Qatar, Department of Biomedical Sciences, College of Health Sciences, QU Health, Qatar University, Doha, Qatar

[4] Faculty of Health, Social Care and Medicine (FHSCM), School of Medicine, Edge Hill University, Ormskirk L39 4QP, UK

Abbreviations

AMOVA	Analysis of molecular variance
DNA	Deoxyribonucleic acid
FST	Is the proportion of the total genetic variance contained in a subpopulation
GD	Gene diversity
Ia	The amount of information gained about ancestry coefficients from observation of a single randomly chosen allele at a locus.
In	The amount of information gained about population assignment from observation of a single randomly chosen allele at a locus.
MDS	Multidimensional scaling
ORCA	This is the optimal rate of correct assignment for a locus if a randomly chosen allele in a pooled collection of populations is assigned to its most likely source population.
PiXY	The average pairwise differences between individuals from two different populations (X and Y) providing insights into their genetic

Published online: 06 November 2025

	divergence
PiX	Is the average pairwise difference within population X
PiY	Is the average pairwise difference within population Y
R_{ST}	A pairwise population genetic distance measures that incorporate the stepwise mutation process in microsatellites
STR	Short tandem repeat
STRAF	STR analysis for forensics

Introduction

Y-chromosome analysis is crucial in population genetics due to its non-recombining, unipaternally inherited nature [1]. Y chromosome Short Tandem Repeat (Y-STR) markers are particularly valuable in forensic cases, especially sexual assaults with mixed DNA samples containing minimal male DNA alongside abundant female DNA [2, 3]. They also aid in studying paternal evolutionary history [4] and molecular anthropology [5]. Since males are hemizygous for Y-chromosome markers, profiles typically show single alleles per locus. Multiple alleles indicate contributions from multiple males. Beyond establishing paternal relationships, Y-STR analysis provides geographical ancestry information for male DNA contributors, proving essential in missing person investigations [6].

Genetic studies focusing on the Y chromosome Short Tandem Repeat (Y-STR) analysis have provided valuable insights into the population history, genetic diversity, and migration patterns of populations in Africa and the Middle East. These studies have utilized various analytical approaches such as admixture analysis, phylogenetic analysis, network analysis, and haplogroup analysis to unravel the complex genetic landscape of these regions. By examining the Y chromosome researchers have been able to trace the paternal lineage of populations and better understand the genetic relationships between different groups. Africa is widely recognized as the birthplace of Homo sapiens. This conclusion is supported by anthropological and DNA evidence, indicating that populations worldwide can be traced back to an African origin approximately 200,000 years ago [7–10].

According to the Out-of-Africa model, anatomically modern humans initially expanded into the Arabian Peninsula and the Levant during the terminal Middle Pleistocene, with ongoing interaction with Africans until the Late Pleistocene. This led to two waves of modern human dispersal into Eurasia, offering a potential resolution for current archaeological, genetic, and paleontological evidence [11].

The Middle East, situated at the crossroads of Africa, Europe, and South Asia, has garnered recognition for harbouring some of the most ancient evidence of modern human presence outside of Africa, notably in the regions of the Levant and Northwest Arabia, with established dates of at least 177 thousand years ago (kya) and approximately 85 kya. Consequently, the Middle East assumes a pivotal role in elucidating the intricacies of human evolution, historical developments, and migratory patterns [12, 13].

In this study, we present a comprehensive database comprising 17 Y-STR loci from Middle Eastern and African populations to examine genetic and phylogenetic relationships between different African regions and the Middle East. We analyzed ancestry variability, allelic richness, and informativeness of genetic markers across populations. Additionally, we investigated migration patterns and routes within Africa using Y-STR data to provide insights into the demographic history and population structure of these regions. This analysis contributes to the understanding of paternal lineage diversity and evolutionary relationships in populations that remain underrepresented in genomic databases.

Material and methods

Published Y-STR data from 186 populations and regions, comprising 14,504 individuals, were used in this study: 52 Middle Eastern populations (5,568 individuals) [14] and 134 African populations (8,936 individuals) [15–47]. These populations represent 13 Middle Eastern countries and 30 African countries.

Three main kits were used to generate the published data: the AmpFLSTR™ Yfiler™ PCR Amplification Kit, the PowerPlex® Y23 System, and the Yfiler™ Plus PCR Amplification Kit. These kits analyze 17, 23, and 27 Y-STR markers, respectively. For this study, we used the set of 17 markers from the AmpFLSTR™ Yfiler™ PCR Amplification Kit that are common to all three kits. This approach allowed us to include the largest possible number of populations in our analysis.

Statistical analyses

Population genetic and phylogenetic analysis

The population genetic structure within our dataset was assessed using the analysis of molecular variance (AMOVA) approach. Calculations were performed with the Arlequin v3.5.2.2 software [48, 49], which enabled us to determine the average pairwise differences both between populations (PiXY) and within populations (PiX). We also calculated

the corrected average pairwise difference between populations as PiXY − (PiX + PiY)/2.

To further quantify genetic distances between populations, we calculated pairwise genetic distances (R_{ST}) based on the Y-STR data. To investigate genetic similarities and visualize the variance in genetic differences among populations, we conducted a multidimensional scaling (MDS) analysis. For phylogenetic tree construction, we used allele frequency data with the POPTREE2 online tool [50] in conjunction with the FigTree software [51].

Allelic richness in different population groups

To gain a more comprehensive understanding of genetic diversity and population relationships, this study investigated allele distributions across six distinct regions: the Middle East, North Africa, East Africa, West Africa, South Africa, and African populations residing outside of Africa. The analysis focused on assessing both the number of distinct alleles present within each population and the number of alleles exclusive to a particular population (referred to as private alleles), which are not observed in any other populations. These two fundamental characteristics serve as valuable indicators when examining populations at a specific locus, especially when analyzing highly variable multiallelic markers such as microsatellites. The Allelic Diversity Analyzer (ADZE)v1.0 software was used to accurately quantify the counts of distinct and private alleles [52, 53].

Informativeness statistics for genetic markers

This study also used the program Infocalc, which calculates informativeness statistics for genetic markers used in ancestry inference. Three parameters were assessed: first, the informativeness for assignment (In), which measures the amount of information gained about population assignment from observing a single, randomly chosen allele at a locus; second, the informativeness for ancestry coefficients (Ia), which quantifies the information gained about ancestry coefficients from such an observation; and finally, the optimal rate of correct assignment (ORCA), which represents the highest possible rate at which a randomly chosen allele from a pooled collection of populations can be correctly assigned to its most likely source population [54–56].

Haplogroups and median-joining networks

The haplogroups of the study populations were determined using NevGen Genealogy Tools v1.1. [57]. Additionally, the NevGen Probability Calculator for Time to Most Recent Common Ancestor (TMRCA) was used to estimate the TMRCA for the J1a haplogroup in Arabia and to identify the most frequent value as the ancestral haplotype. The TMRCA calculations in NevGen take into account backward Y-STR mutations, and also multiple-step mutations.

Median-joining networks were constructed for the most common haplogroups using Network 10.2.0.0 and Network Publisher [58, 59]. In cases involving intermediate alleles, repeat numbers were rounded to the nearest integer. Additionally, for the purpose of network construction, constitutively duplicated loci (DYS385a,b) were excluded from the analysis.

Population structure

This study examined the population structure of 186 Middle Eastern and African populations, encompassing a total of 14,504 individuals. The analysis was conducted using STRUCTURE v2.3.7 with an admixture model [60, 61].

To process the output and evaluate probability values across a wide range of K values, the STRUCTURE HARVESTER program was used. This program also helped identify the optimal number of genetic clusters that best fit the data [62, 63]. To consolidate the findings, multiple iterative analyses were performed on each dataset, and the results were aligned using CLUMPP [64, 65]. These aligned results were then used to generate the population Q-matrix graph with Distruct [64, 66].

Ancestry variability analysis

The ancestry variability in this study examined the variation in membership coefficients among individuals assigned to the designated clusters. The FSTruct program was used to investigate differences in membership coefficient variability between admixed and non-admixed populations [67, 68].

Migration and gene flow in Africa

The Migrate program was used to examine gene flow between different geographical regions of North, East, Central, West, and South Africa [69]. Three models were applied: Model 1 involved unidirectional gene flow from one population to another; Model 2 accounted for divergence from a common ancestral population; and Model 3 incorporated both divergence from the ancestral population and ongoing immigration. Gene flow calculations were performed using pairs of regions, with all three models applied in both directions. This approach enabled comprehensive coverage of the entire African continent.

Results

Population genetic and phylogenetic analysis

Genetic diversity for the 17 Y-STR loci was calculated among Middle Eastern populations, the five African geographical regions (North, West, East, Central, and South Africa), and African populations residing outside Africa. The locus DYS458 exhibited the highest genetic diversity in Middle Eastern and North African populations. In all other African populations, the locus DYS385b showed the highest genetic diversity. The lowest genetic diversity in the Middle East and East Africa was observed at locus DYS392, while DYS437 showed the lowest genetic diversity in North, Central, and West Africa. In both South Africa and African populations outside Africa, the locus DYS391 exhibited the lowest genetic diversity. The results are presented in Fig. 1 and Supplementary Table S1.

A pairwise matrix plot of R_{ST} distances was generated to compare 43 populations—13 from the Middle East and 30 from Africa. Among Middle Eastern populations, the closest pair was Israel and Palestine (R_{ST}=0.01365), while the most distant were Qatar and Turkey (R_{ST}=0.12883). In Africa, the closest populations were from the Bahamas and Haiti, with an R_{ST} of 0.01138; this pair was also the closest overall across both Africa and the Middle East. Conversely, the most distant populations were Tanzania and Sudan, with an R_{ST} of 0.19113, making them the most genetically distant pair across both regions. The results are presented in Supplementary Table S2 and Fig. 2.

The average pairwise differences were examined to estimate the corrected genetic differences among 43 populations. These differences were specifically analysed in three contexts: between the populations as a whole, within each individual population, and between different populations using Nei's distance.

For the average number of pairwise differences between populations, the lowest value was observed between Jordan and Israel (164.61448), while the highest was recorded between the Central African Republic and Turkey (272.2404). Within the Middle East, the lowest value was again between Jordan and Israel, and the highest was between Turkey and the UAE (238.23841). In Africa, the lowest value was found between Kenya and Djibouti (168.01477), while the highest was between the Central African Republic and Chad (271.70886).

For within-population calculations across both the Middle East and Africa, Djibouti had the lowest value (160.13153), while the Central African Republic had the highest (241.80979). In the Middle East specifically, Jordan had the lowest value (161.72676), and Turkey had the highest (239.42689).

Regarding Nei's distance, for both the entire dataset and the African populations, the lowest value was found between the Bahamas and Haiti (2.13661), while the highest was observed between Sudan and Tanzania (50.67652). In the Middle East, Palestine had both the lowest and highest

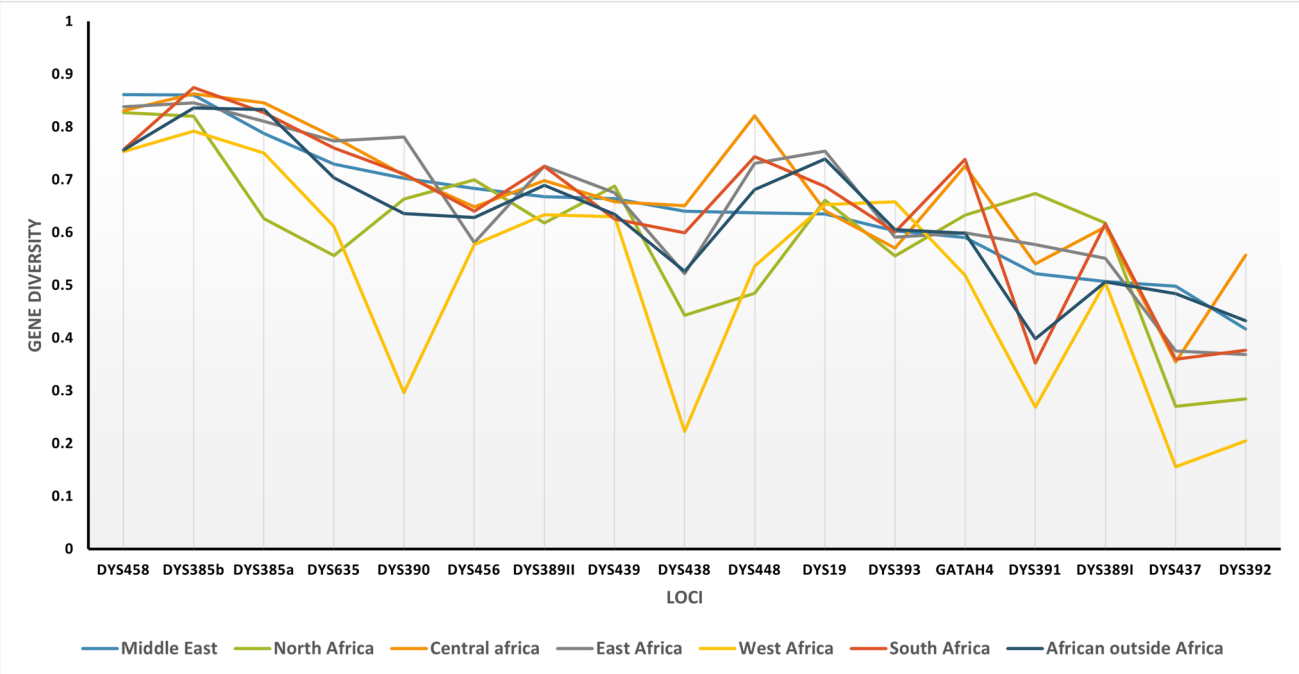

Fig. 1 The genetic diversities of the 17 Y-STR loci in the Middle Eastern populations, the five African geographical regions (North, West, East, Central and South) and Africans outside Africa

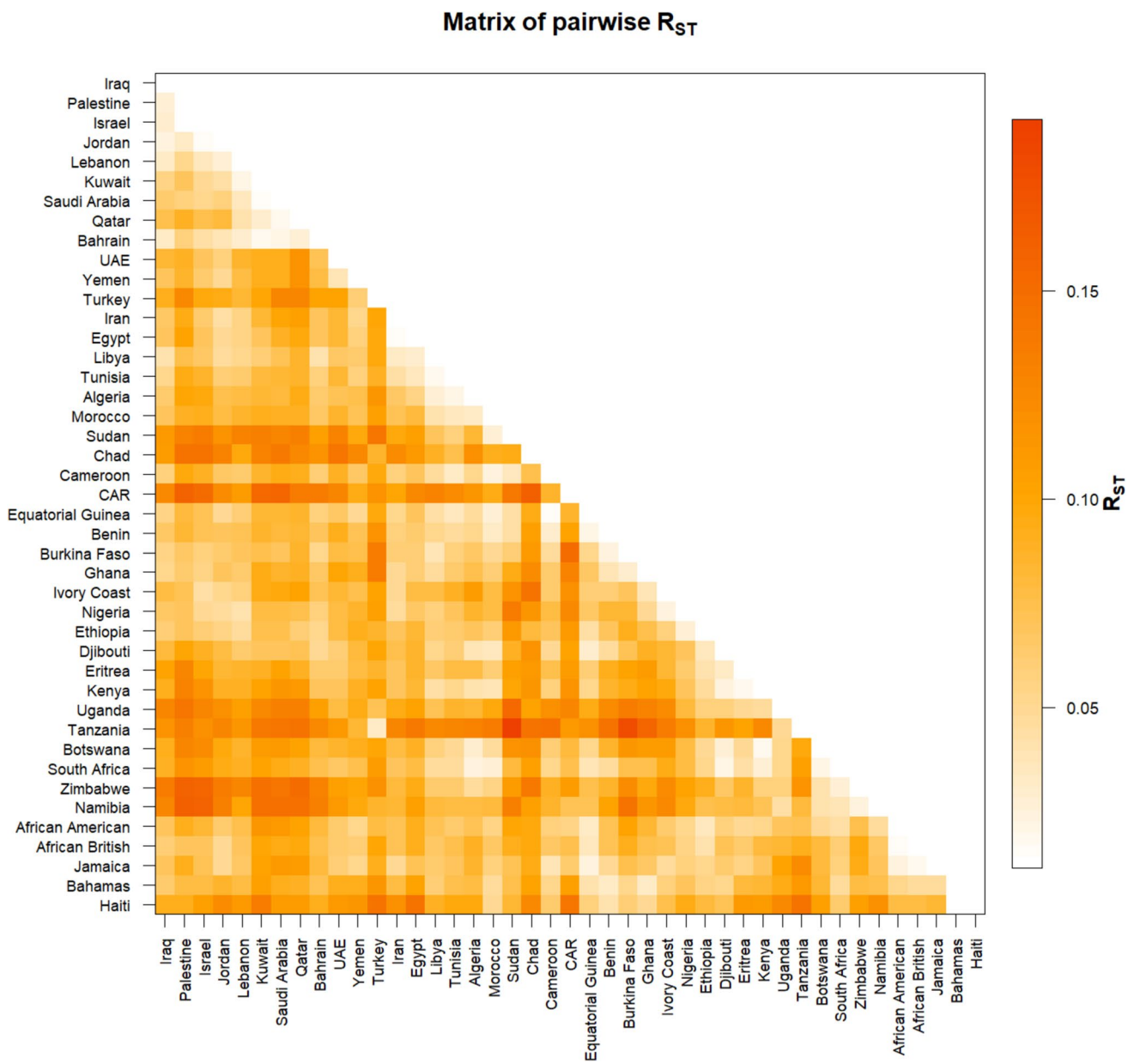

Fig. 2 The matrix of pairwise genetic distance R_{ST} of Y-STR between the Middle Eastern and African populations based on 17 Y-STR markers. This matrix was generated using Arlequin v3.5.2.2 software. CAR: Central African Republic, UAE: United Arab Emirates

Nei's distance values: the lowest with Israel (2.3199) and the highest with Turkey (28.22702). The results are shown in Supplementary Table S3 and Fig. 3.

Multidimensional scaling (MDS) and genetic distance analyses were performed among all Middle Eastern and African populations. Five clusters were identified and distributed as follows: countries of the Arabian Peninsula were located in the lower left quadrant. Most African countries formed two main, centrally located clusters. The first cluster, composed of North African countries, was situated in the central upper left quadrant; this cluster also included four Middle Eastern countries—Yemen, UAE, Iran, and Turkey. The second cluster appeared in the central lower right quadrant, where African populations outside Africa merged with most other African countries. The remaining African countries formed a smaller fourth cluster in the lower right quadrant. The fifth cluster, located in the upper right quadrant, consisted of three African countries: Sudan, Chad, and the Central African Republic. The MDS results are shown in Fig. 4.

A three-dimensional MDS was generated to get a better idea about these five clusters; it could be found as an interactive open-source which can be viewed in Supplementary Figure S1.

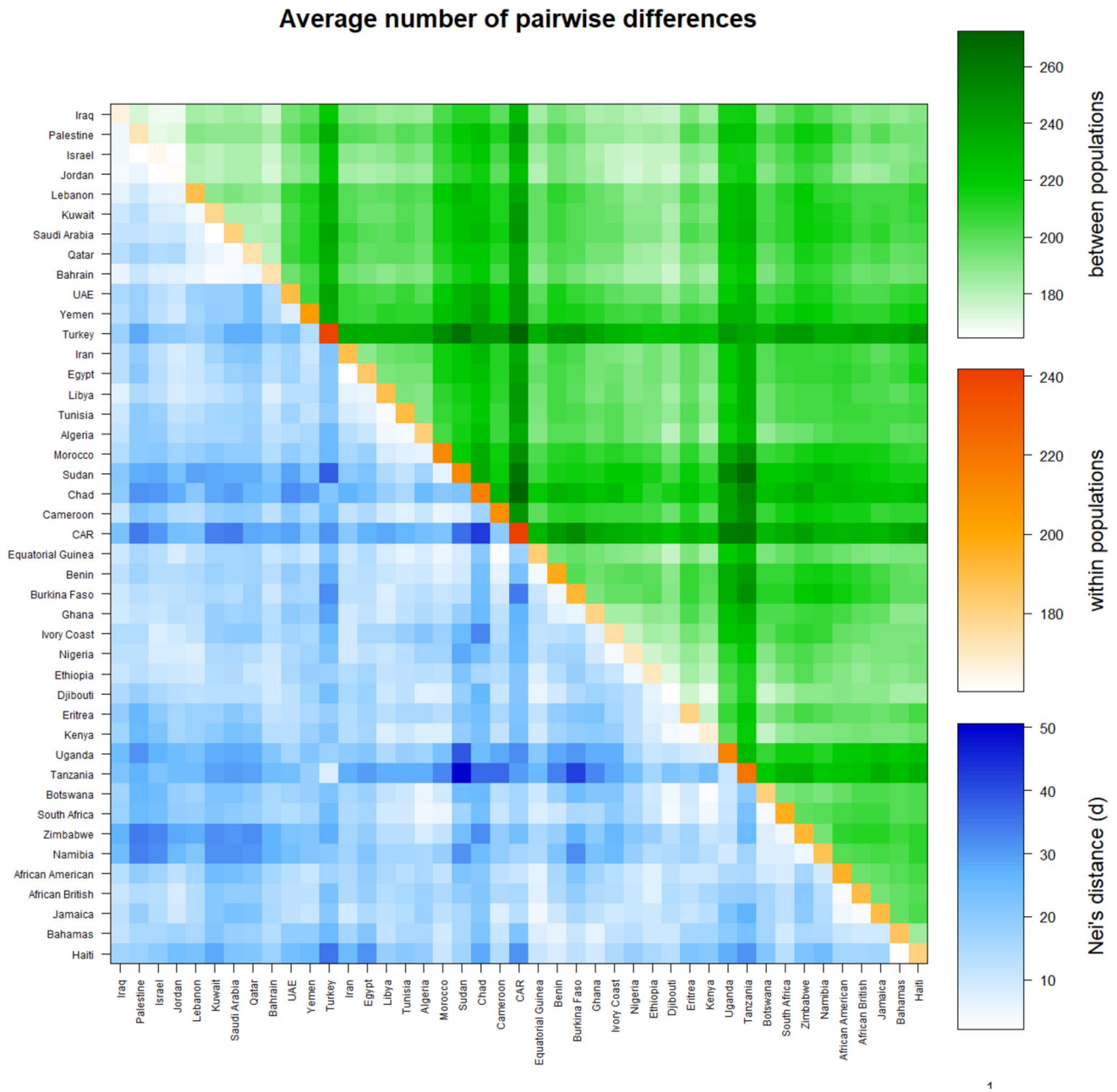

Fig. 3 Matrix plot showing population average pairwise differences based on 17 loci. The area above the diagonal (green) shows the average number of pairwise differences between populations; the diagonal (orange) shows the average number of pairwise differences within population; and below the diagonal (blue) shows the corrected average pairwise difference. The scale of differences is shown on the right side of the matrix

To assess the diversity between African and Middle Eastern populations, a phylogenetic tree was constructed. Five subpopulations (K=5) represented the optimal clustering for the 43 populations analyzed (Fig. 5). The Middle Eastern populations formed two small clusters, with Sudan—an African country—joining one of these clusters. The North African countries formed a third cluster, which also included Palestine and Jordan. The fourth and fifth clusters encompassed the majority of the African countries, with the fourth being the largest cluster.

Allelic richness in Middle East and Africa

This study examined two measures of allelic richness, distinct alleles and private alleles, across the Middle East and five regions of Africa (North, Central, East, West, and South). The findings showed that Central Africa had the highest level of distinct alleles, while West Africa had the lowest. The Middle East ranked third in the number of distinct alleles (Fig. 6A and Supplementary Table S4).

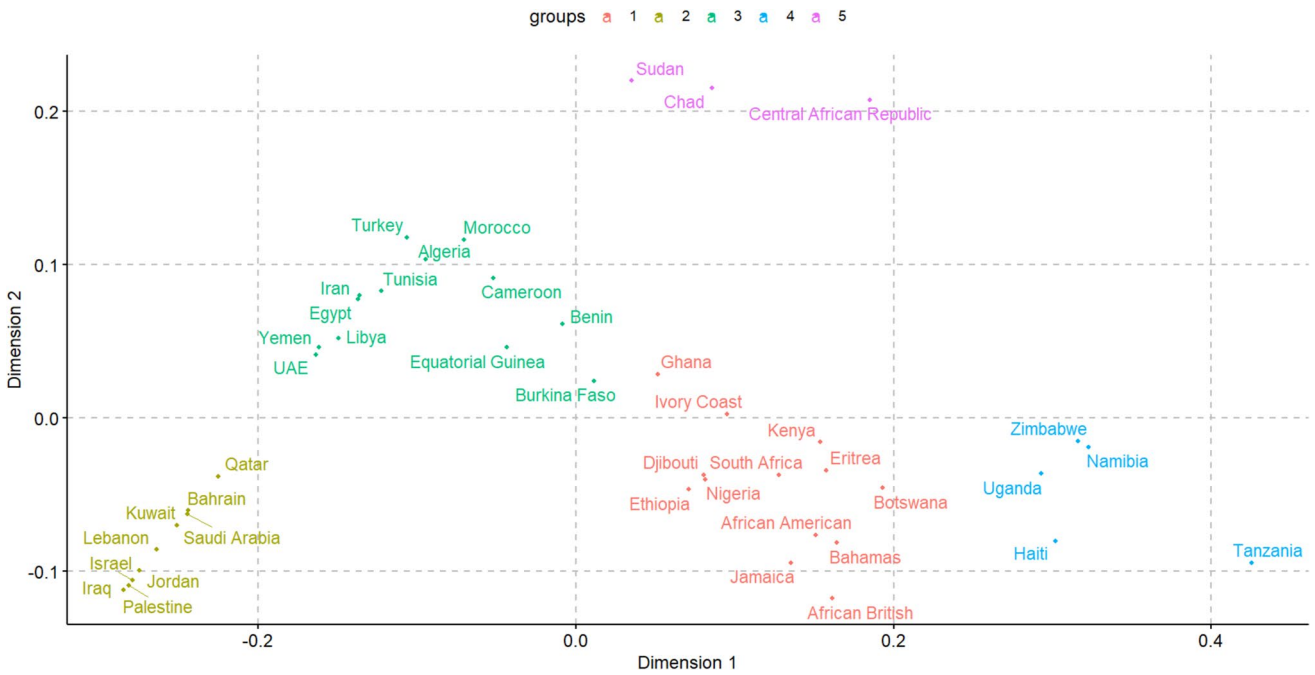

Fig. 4 Multidimensional scaling (MDS) plots comparing the Middle Eastern and the African populations based on 17 Y-STR markers. Five clusters could be identified

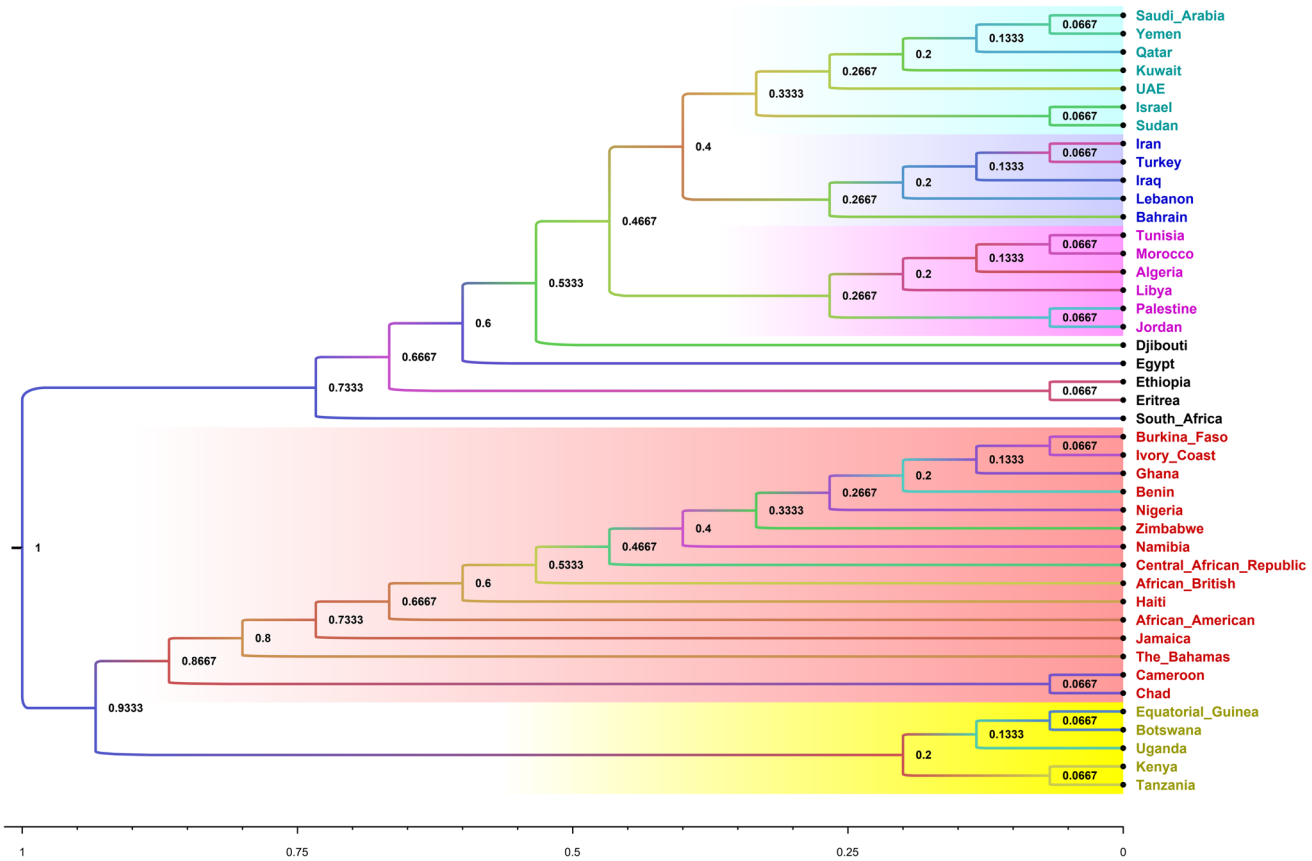

Fig. 5 Phylogenetic tree of genetic relationships among the 43 Middle Eastern and African populations. Five clusters (K = 5) were created. The red-coloured cluster is the largest (15 populations), followed by the cyan-coloured cluster (7 populations), and the purple-coloured cluster (6 populations). The remaining two clusters each have five populations. This phylogenetic tree was generated using POPTREE2 software

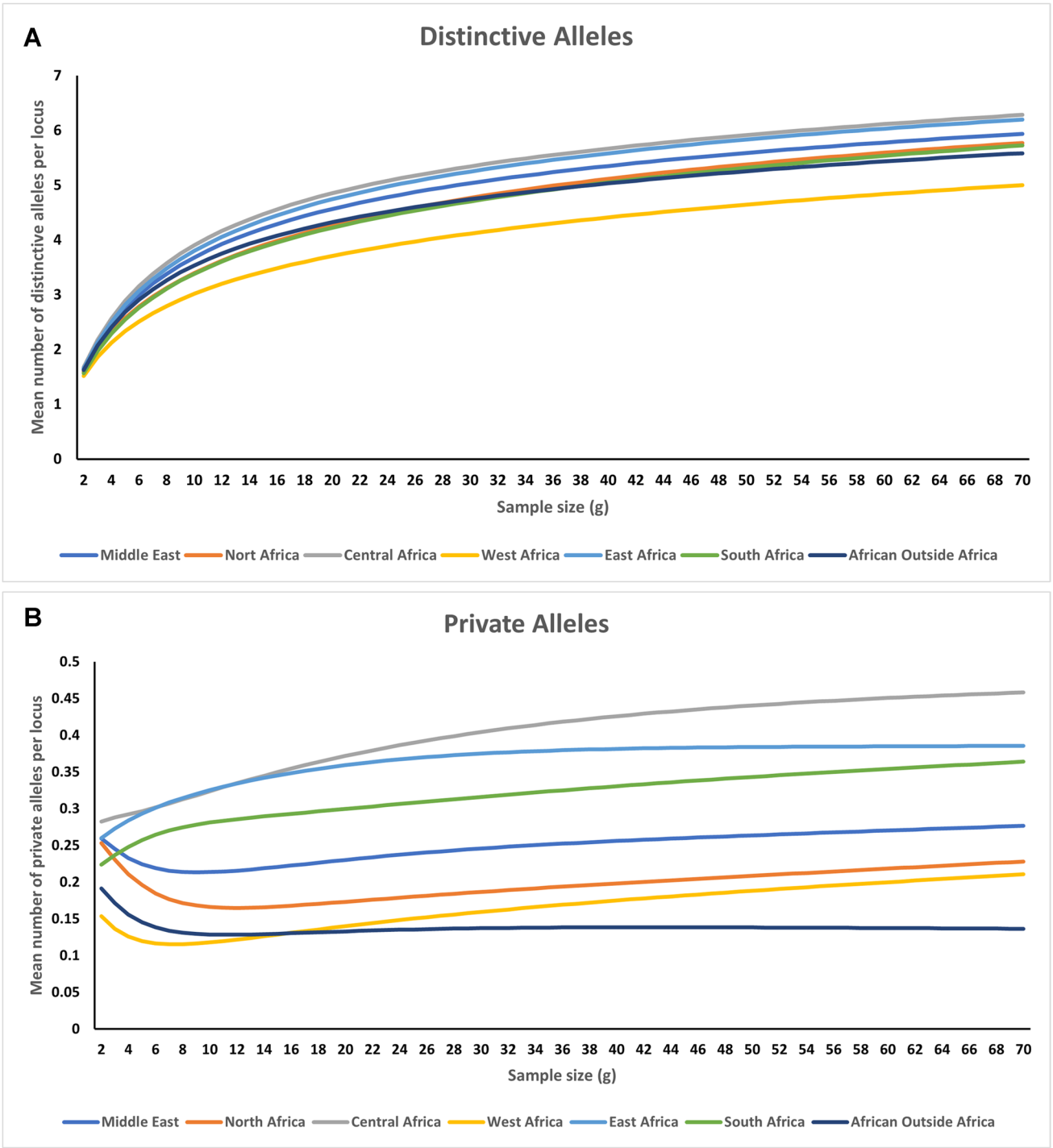

Fig. 6 The mean number of (**A**) distinct alleles per locus and (**B**) private alleles per locus of the seven regions: Middle East, North Africa, Central Africa, West Africa, East Africa, South Africa and Africans outside Africa

Additionally, the analysis revealed that Central Africa exhibited the highest levels of private alleles, whereas populations of African descent living outside Africa had the lowest levels. The Middle East was intermediate, ranking below South Africa but above North Africa (Fig. 6B and Supplementary Table S5).

Informativeness statistics for genetic markers

To assess the informativeness (In) of genetic markers for distinguishing between Africa and the Middle East, two markers, DYS438 and DYS390, stood out with high values of 0.25 and 0.24, respectively (Supplementary table

S6a). Within the Middle Eastern population specifically, the marker DYS458 showed the highest informativeness with a value of 0.34, followed by DYS456 with a value of 0.14 (Supplementary table S6b).

Haplogroups and median-joining networks

The most common haplogroups among Middle Eastern populations are J1a (29.4%, 1637/5568 individuals), J2a (17.9%, 999/5568), E1b1b (11%, 613/5568), R1a (9.8%, 551/5568), and G2a (4.2%, 239/5568). Median-joining networks were constructed for each of these haplogroups to show their distribution among the Middle Eastern populations (Supplementary Figure S2).

The most common haplogroups among African populations are E1b1a (42.2%, 3,774/8936 individuals), E1b1b (21.6%, 1935/8936), and R1b (8.5%, 761/8936).

NevGen probability calculator of Time to Most Recent Ancestor (TMRCA) was used to calculate the generations and most frequent value as ancestor haplotype of main Middle Eastern haplogroup J1a. Iraq was the highest with 167 generations followed by Yemen and Palestine 166 and 163 generations respectively. The lowest generation number was noticed in Saudi Arabia and Israel 96 and 95 generations respectively. The differences between the ancestor haplotypes was noticed to involve nine markers: DYS456, DYS390, DYS389II, DYS458, DYS385a, DYS385b, DYS391, DYS439 and DYS635. Table 1 shows the calculations of TMRCA for the haplogroup J1a of the countries in Arabia.

Population structure

The Y-STR graph of the populations' Q-matrix (Fig. 7 and Supplementary Table S7) revealed 10 clusters, based on 17 STR markers from 186 populations and regions (14,504 individuals), including 52 Middle Eastern and 134 African populations. The population clusters identified from the Y-STR data corresponded to specific geographical regions and showed a clear sub-grouping of countries within each cluster.

The Middle Eastern populations formed a distinct cluster. North Africa also showed a unique cluster, except for Egypt, which was more closely related to the Eastern Africa cluster. Central Africa formed its own cluster, with the exception of the Central African Republic and Equatorial Guinea, both of which overlapped with the West Africa cluster. East Africa had its own cluster as well, except for Kenya (Bantu and Luhya), which shared elements with West Africa. South Africa formed a separate cluster, except for Zimbabwe and Namibia, which showed some shared elements with West Africa. Populations of African descent living outside Africa formed their own cluster, which shared characteristics with both Central and West African countries.

Ancestry variability analysis

The F_{ST}/F_{ST}^{max} ratios were used to compare the variability of Q matrices across seven regions in order to study ancestry variability among the populations. The analysis revealed that North Africa had the highest F_{ST}/F_{ST}^{max} ratio (0.67636), while West Africa had the lowest (0.49807). The results are presented in Fig. 8 and Supplementary Table S8.

We analyzed ancestry variability for countries within each of the seven regions to gain a clearer understanding of regional differences. First, the Middle Eastern populations exhibited a distinct range of F_{ST}/F_{ST}^{max} ratios, with Kuwait showing the highest value (0.58169) and Iran the

Table 1 The calculations of TMRCA, the generations and most frequent value as ancestral haplotype of the countries in Arabia for the haplogroup J1a. The order of the loci in the most frequent values for the haplotypes are: DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a, DYS385b, DYS393, DYS391, DYS439, DYS635, DYS392, GATAH4, DYS437, DYS438, DYS448. The differences between the ancestral haplotypes were highlighted in yellow

Country	Generations	Probability %	Sum Probability %	Most frequent value as ancestral haplotype
Iraq	167	7.16715	56.2814	14 13 23 30 18 14 13 18 12 11 11 21 11 11 14 10 20
Yemen	166	4.25957	51.7964	14 13 23 30 18 14 **14** 18 12 **10** 11 21 11 11 14 10 20
Palestine	163	3.53043	53.5234	**13** 13 23 **29 20** 14 13 18 12 11 11 21 11 11 14 10 20
Jordan	163	4.86133	50.583	14 13 **22 29 19** 14 13 18 12 **10** 11 21 11 11 14 10 20
Lebanon	160	5.44218	52.6427	**15** 13 23 30 18 14 13 18 12 **10** 11 21 11 11 14 10 20
Bahrain	133	6.41968	56.0834	14 13 23 30 18 14 13 18 12 **10** 11 21 11 11 14 10 20
UAE	124	5.14747	54.5235	14 13 23 30 18 14 13 18 12 **10** 11 21 11 11 14 10 20
Qatar	117	8.27285	53.3473	14 13 23 **29** 18 14 13 18 12 11 11 21 11 11 14 10 20
Kuwait	102	6.79695	53.0673	14 13 23 30 18 14 13 **19** 12 11 11 21 11 11 14 10 20
Saudi Arabia	96	13.5289	52.8574	14 13 23 30 18 14 13 18 12 11 11 21 11 11 14 10 20
Israel	95	2.55379	50.3233	**16** 13 23 30 18 14 13 18 12 11 **10 20** 11 11 14 10 19

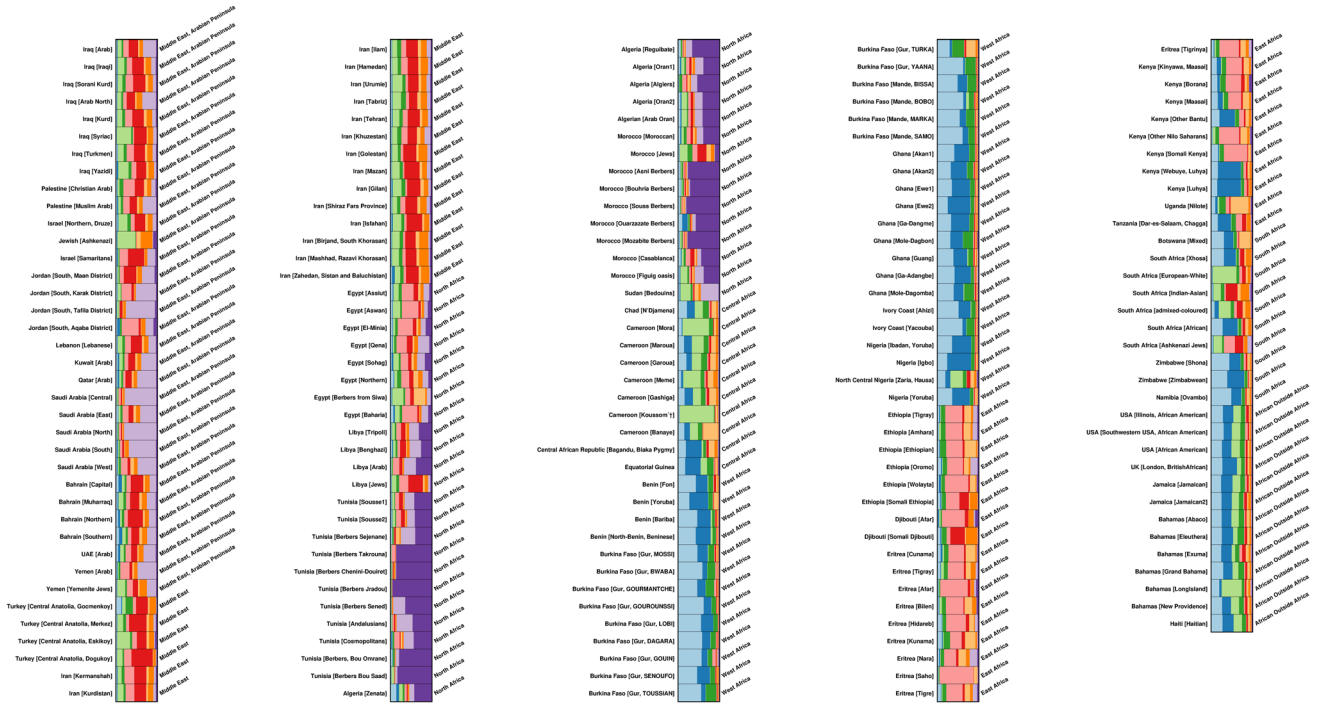

Fig. 7 The graph of population Q- matrix of the Y-STR haplotypes using 17 STR markers from 186 Middle Eastern and African populations (14504 individuals) showing 10 clusters (K = 10)

lowest (0.42201). The results also revealed that, among the Arab populations in the Middle East, the Yemeni population has the lowest F_{ST}/F_{ST}^{max} ratio (0.47690). The results are presented in Fig. 9 and Supplementary Table S9.

Similarly, in North Africa, Tunisia and Sudan registered the highest and lowest F_{ST}/F_{ST}^{max} ratios, with values of 0.77727 and 0.58417, respectively. In the sub–Saharan Africa, the lowest ratios were found in Djibouti and Eritrea with values 0.28213 and 0.39647, respectively. While the highest ratios were found in Zimbabwe and Kenya with values 0.58631 and 0.61103, respectively. Figure 10 and supplementary tables S10 and S11 show the F_{ST}/F_{ST}^{max} ratios of different African countries in different regions of Africa.

Finally, the six Jewish populations in this study were analysed individually, revealing that the Jews from Morocco had the highest F_{ST}/F_{ST}^{max} ratio (0.57191), and the Yemeni Jews had the lowest (0.41759). The F_{ST}/F_{ST}^{max} ratio observed in the Ashkenazi Jewish population was notably low, nearly reaching the baseline, with a recorded value of 0.00114. Figure 11 and supplementary table S12 show the F_{ST}/F_{ST}^{max} ratios of different Jewish populations.

Gene flow in Africa

Across Africa, Model 1 was the most probable gene-flow route for three corridors—East to Central (0.8625), South to East (0.9385), and West to South (0.8899)—with the sole exception of the Central to North corridor, where Model 2 was overwhelmingly favoured (0.9999). The gene flow results are shown in Fig. 12 and supplementary table S13.

Discussion

The genetic diversity calculations for the 17 Y-STR markers for the Middle Eastern and the five regions in Africa showed very distinctive patterns. In the Middle East and North Africa, the locus DYS358 showed the highest genetic diversity. In the Middle East, this is due to the presence of microvariant alleles in this locus which is also associated with the haplogroup J1 [70, 71]. The high diversity of this locus in North Africa it might be due to the impact the Arab expansion to North Africa [72].

Haplogroup J-M304 exhibits a notable prevalence in the Arabian Peninsula and Mesopotamia. This particular haplogroup undergoes a division into two subgroups:

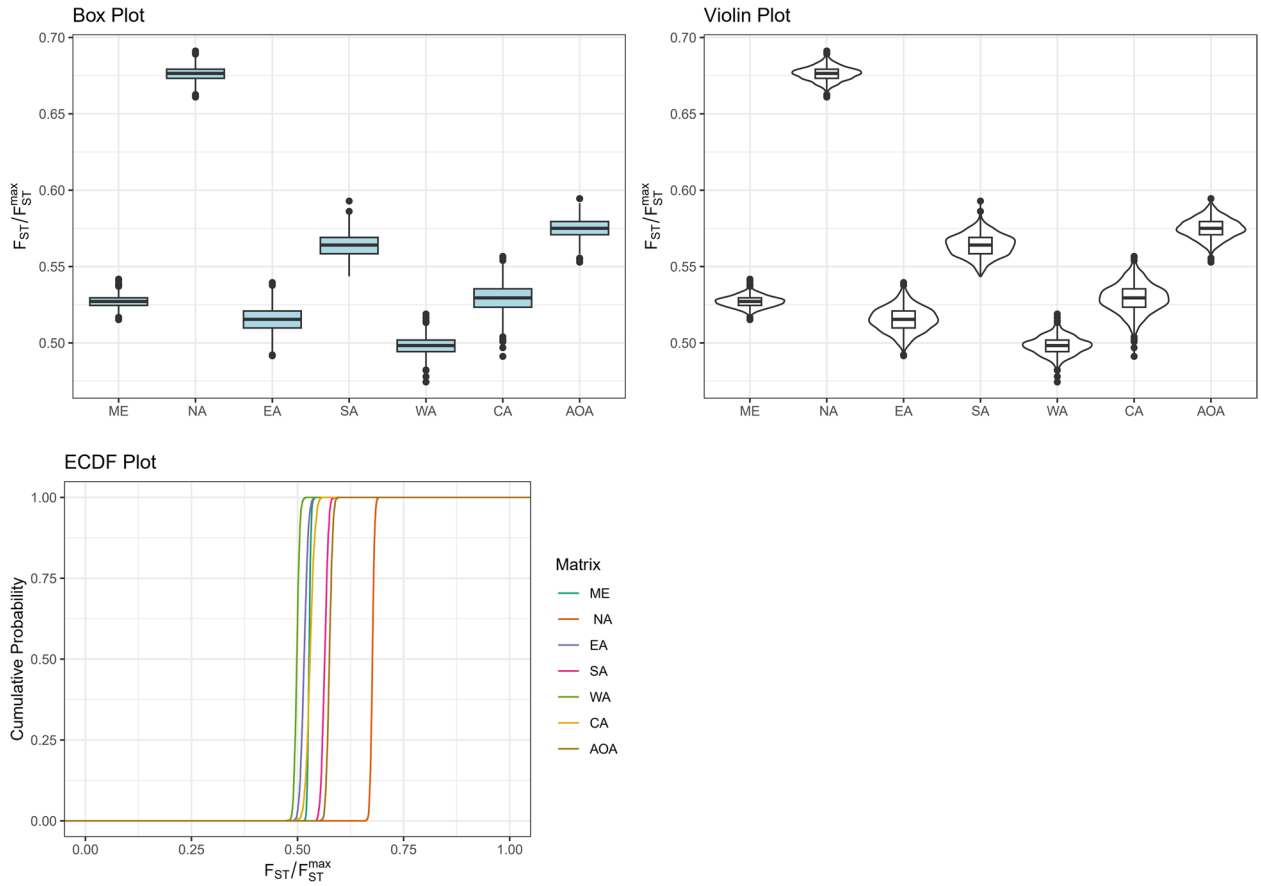

Fig. 8 Box plot, violin plot, and empirical cumulative distribution function (ECDF) plot of the bootstrap distribution of F_{ST}/F_{ST}^{max} for each Q matrix in the STRUCTURE analysis for the Middle Eastern and African populations. ME: Middle East, NA: North Africa, EA: East Africa, SA: South Africa, WA: West Africa, C: Central Africa, AOA: Africans outside Africa. West and East Africa had the lowest ratio, and North Africa had the highest

J1-M267, which has been linked to the dissemination of pastoral economies within the arid regions of West Asia, and J2-M172, which demonstrates a stronger association with agricultural practices in the northern latitudes of West Asia [73].

TMRCA calculations for the Middle Eastern population indicate that the Levant and northern Arabia are older regions compared to the southern part of Arabia, with the exception of Yemen. This finding supports the proposed migration route within Arabia, suggesting that the Levantine corridor was likely the main passageway out of Africa [71].

Studying the TMRCA in Middle Eastern population showed that 9 out of the 17 markers exhibited mutations. Six of these markers ranked in the top half for marker informativeness, and their occurrence was not related to the markers' genetic diversity rankings.

The abundance of distinctive and private alleles in Africa supports the hypothesis that human evolution originated in Africa and subsequently spread to other parts of the world through a series of founder events. The regions of Africa and the Middle East, which are geographically connected, exhibit the highest diversity of alleles. As the migration of humans from Africa occurred in stages, it is likely that many alleles in the founding population only migrated partially outside of Africa. The results of this study on allelic richness and ancestry variability align with the expectations of models proposing an African origin of humanity, which involve successive founder effects during outward migrations [74]. This finding aligns with a previous study conducted on the X chromosome, which specifically examined the parameter of allelic richness [75].

The Q-matrices derived from Y-STRs have been previously analysed in the genetic landscapes of Africa and the Middle East, encompassing 135 populations and regions, and comprising 11,305 individuals. Within this dataset, 97 African and 38 Middle Eastern populations

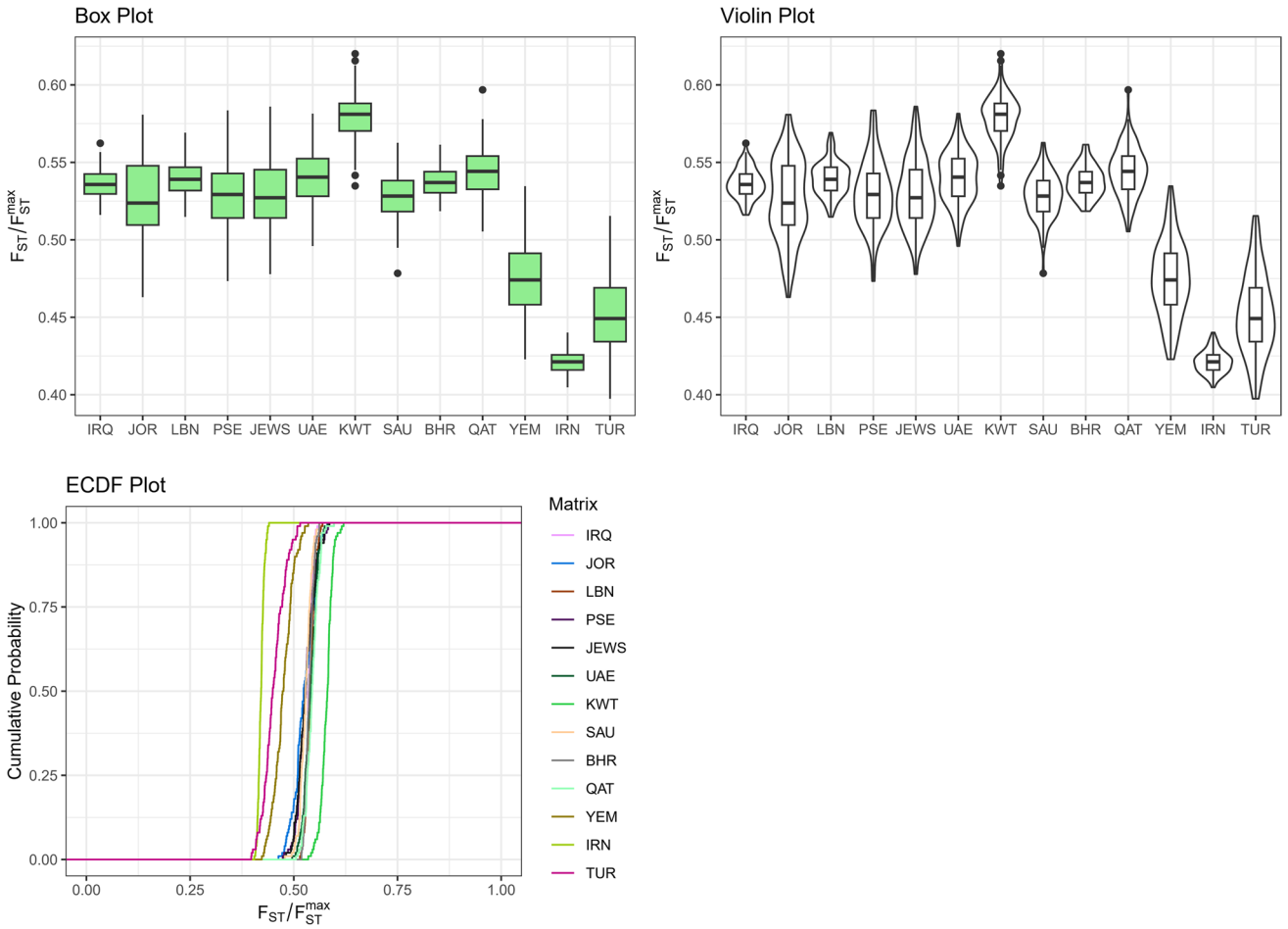

Fig. 9 Box plot, violin plot, and empirical cumulative distribution function (ECDF) plot of the bootstrap distribution of F_{ST}/F_{ST}^{max} for each Q matrix in the STRUCTURE analysis for the Middle Eastern populations. IRQ: Iraq, JOR: Jordan, LBN: Lebanon, PSE: Palestine, UAE: United Arab Emirates, KWT: Kuwait, SAU: Saudi Arabia, BHR: Bahrain, QAT: Qatar, YEM: Yemen, IRN: Iran, TUR: Turkey

were investigated, leading to the identification of 8 genetic clusters (K = 8) [70]. Building on this groundwork, the current study broadened its scope to encompass 186 populations (14,504 individuals), including 52 Middle Eastern and 134 African populations leading to the identification of 10 genetic clusters (K = 10). The increased precision and accuracy of the Q-matrix obtained can be attributed to the enhanced demographic and ethnic diversity captured within the expanded survey, emphasising the importance of broader population sampling in elucidating the complex genetic structure of diverse human populations.

The findings of this study revealed that African populations outside Africa (AOA), such as Africans in the UK, USA, Jamaica, the Bahamas, and Haiti, are genetically related to Western and Central African populations. In addition, the phylogenetic analysis showed that these populations merged with the main African cluster. However, the study also highlighted that these populations have formed a distinct genetic cluster that is slightly different from their ancestral groups. This assertion is further substantiated by the comparatively low genetic variability observed within the AOA population, in contrast to other African populations. Additionally, the allelic richness of the AOA population was notably lower than that of other African populations, indicating a deviation from the founder effect commonly observed in African populations. This could be explained by the fact that environmental stresses and social transitions have acted as major selective forces, which may have reshaped the genetic makeup of the African populations inhabiting these regions [76–78].

The aforementioned findings are consistent with the outcomes of a previous study, which revealed that the genetic ancestry of all African Americans is distinguished by admixture in their African elements, primarily originating from West and West-Central Africa. Additionally,

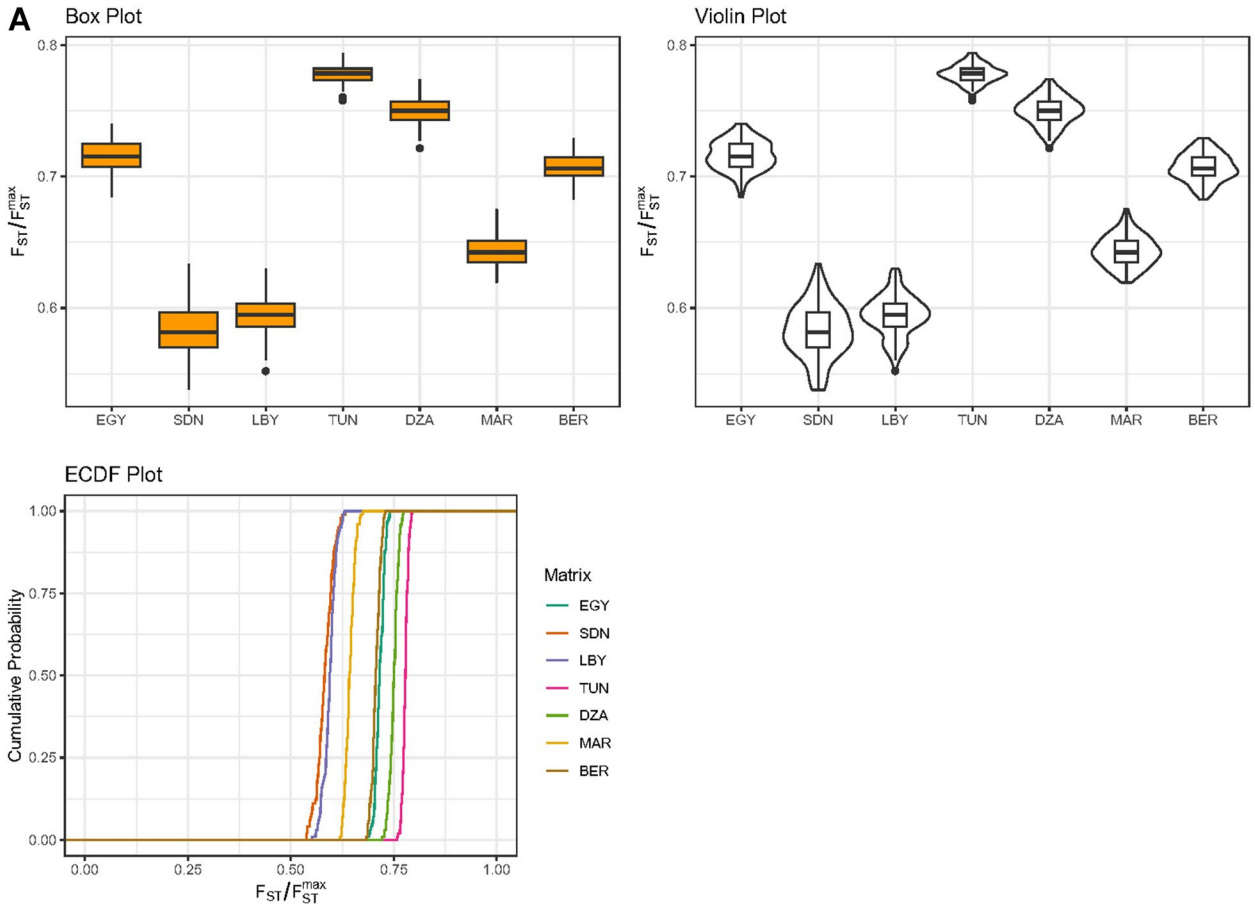

Fig. 10 Box plot, violin plot, and empirical cumulative distribution function (ECDF) plot of the bootstrap distribution of F_{ST}/F_{ST}^{max} for each Q matrix in the STRUCTURE analysis for: A. North Africa: Egypt (EGY), Sudan (SDN), Libya (LBY), Tunisia (TUN), Algeria (DZA), Morocco (MAR), Berber (BER). B. Sub-Saharan Africa: East Africa [Ethiopia (ETH), Eritrea (ERI), Kenya (KEN), Djibouti (DJI), Uganda (UGA)], Central Africa [Equatorial Guinea (GNQ), Central Africa republic (CAF), Cameroon (CMR), Chad (TCD)], West Africa [Benin (BEN), Nigeria (NGA), Ivory coast (CIV), Ghana (GHA), Burkina Faso (BFA)], South Africa [Tanzania (TZA), South Africa (ZAF), Namibia (NAM), Botswana (BWA), Zimbabwe (ZWE)], Africans outside Africa (AOA)

the study observed limited variability in these proportions of African heritage among individuals within this demographic [79].

This investigation revealed that the Jewish communities in Libya, Morocco, Yemen, Samaritans, and South Africa exhibit similar genetic structures and ancestral variation to those of Middle Eastern populations. This similarity may be attributed to inter-ethnic marriages among diverse Jewish communities and genetic exchange with the respective host Diaspora populations [80, 81]. Notably, the Ashkenazi Jews stood out as an exception, as clearly demonstrated in the Q matrix and by their significantly low genetic variability. This distinctiveness could be linked to their emergence from migrations northward into the Rhineland from Mediterranean Jewish populations during the early Middle Ages, as well as their prolonged isolation from other communities [82].

This study conducted an analysis of 13 Berber populations located in Egypt, Tunisia, and Morocco. The findings revealed that despite the historical admixture with Arab populations in North Africa, the Berber communities exhibited a distinct and relatively consistent genetic structure, thereby substantiating the hypothesis that Arabs and Berbers possess separate gene pools [83–85]. Notably, an exception was observed in the genetic structure of the Berber population in Egypt, which closely resembled that of the Egyptian Arab population. Additionally, the study of ancestry variability among the Berbers indicated that there was no discernible difference between this population and other North African populations.

This study has demonstrated that the Yemeni population exhibits low ancestry variability in comparison to other Arab populations in the Middle East, suggesting that Yemen may be the origin of Arab populations in the

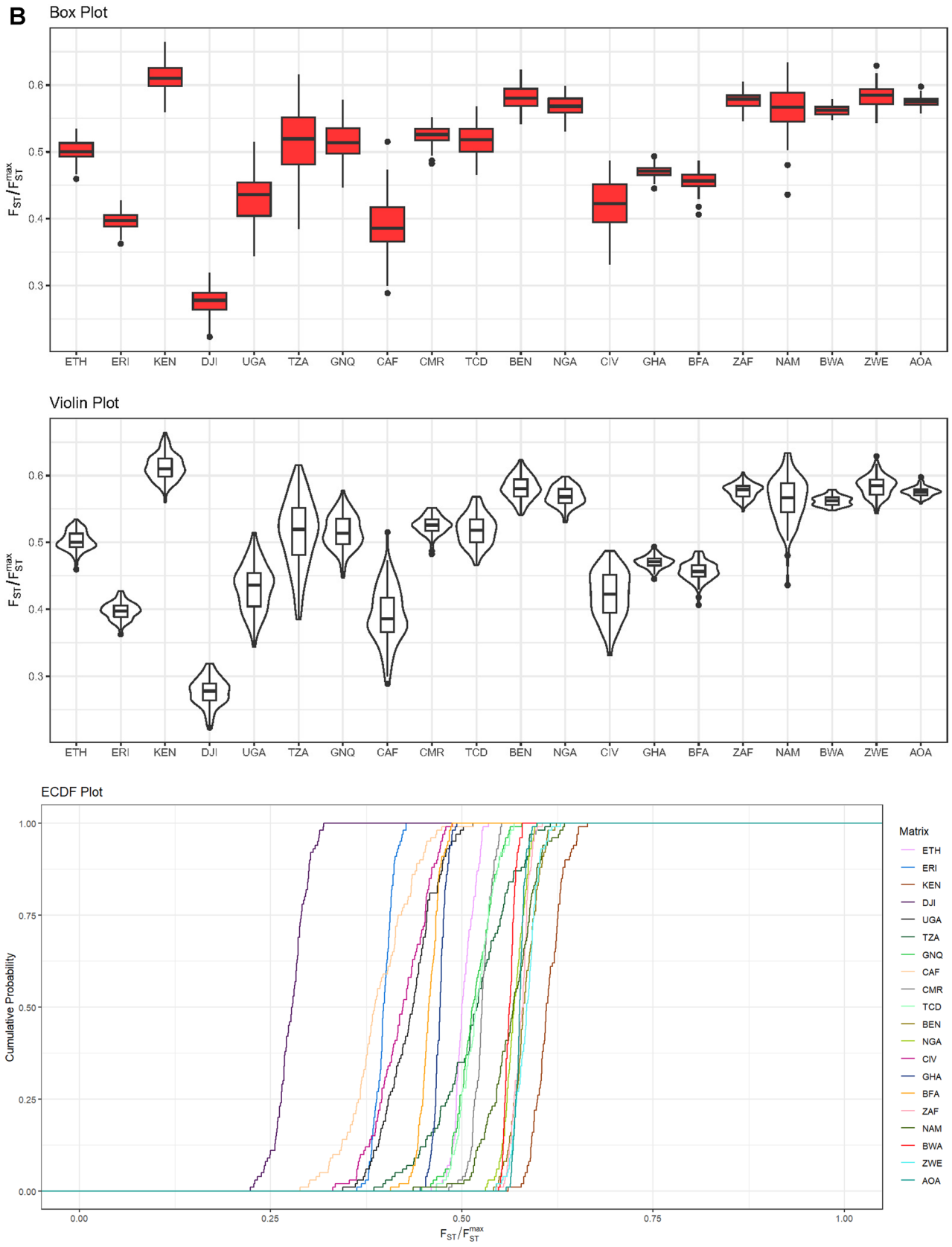

Fig. 10 (continued)

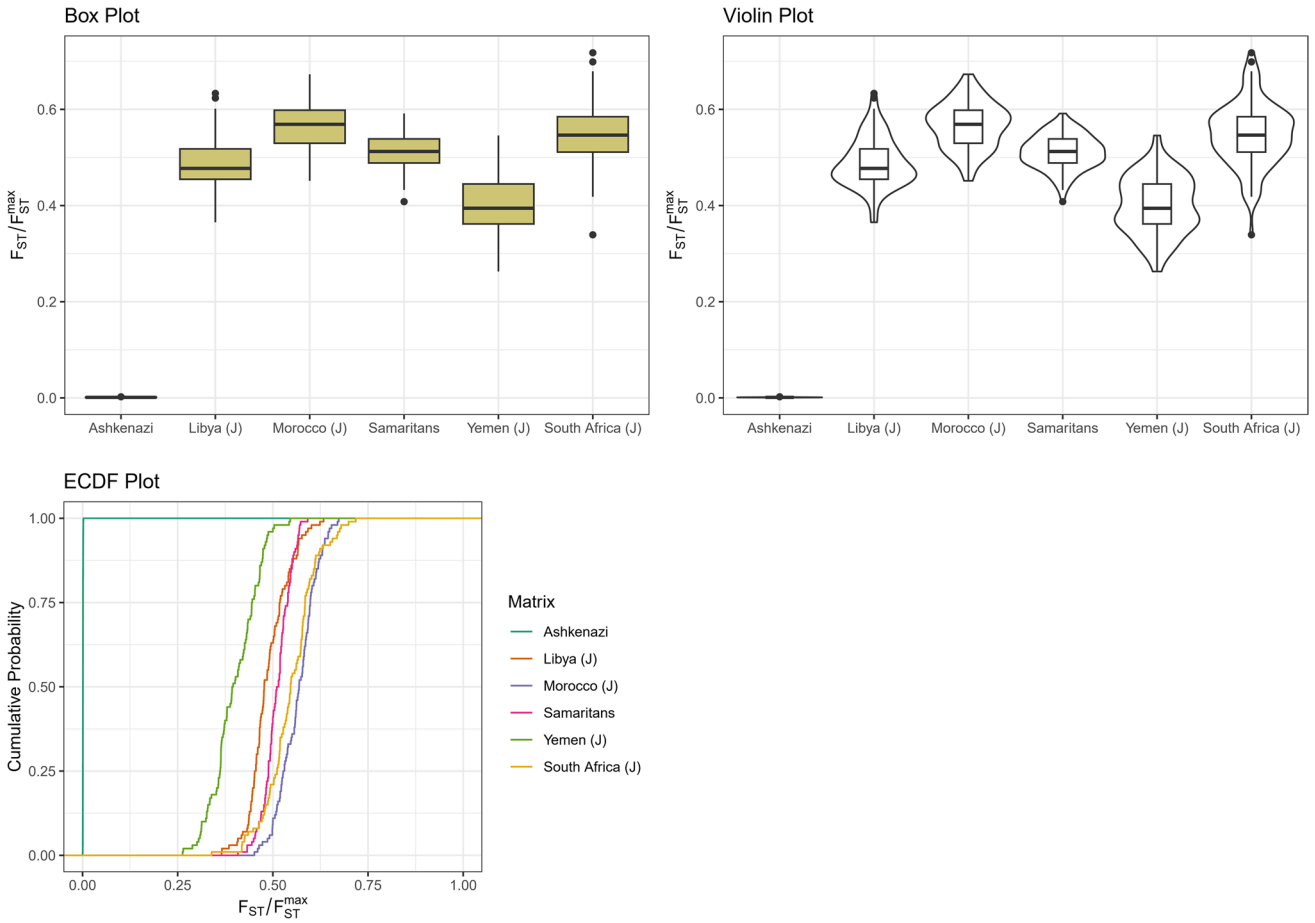

Fig. 11 Box plot, violin plot, and empirical cumulative distribution function (ECDF) plot of the bootstrap distribution of $\mathbf{F_{ST}/F_{ST}^{max}}$ for each Q matrix in the STRUCTURE analysis for the Jewish populations

Arabian Peninsula. This finding was highlighted in a previous study [14]; however, our current research, which incorporated more populations, ethnic groups, and regions, yielded ten genetic clusters instead of six and revealed striking differences in genetic ancestry. The use of a higher number of populations and ethnic groups likely allowed for a more detailed examination of genetic variation. This increased granularity captured more subtle differences in genetic ancestry, leading to clearer distinctions between different populations.

The analysis of gene flow has revealed predominant migration patterns within Africa, indicating a consistent trajectory from West Africa to the southern regions, followed by movement from the south to East Africa. Subsequently, gene flow suggests migration from East Africa to Central Africa, and from Central Africa to North Africa. A recent study highlighted the gene flow of Bantu population from West Africa spreading through the Congo rainforest to eastern and southern Africa [86]; however, this study did not trace the migrations of other African populations within the continent.

Key points

1. The most common haplogroup among Middle Eastern populations was J1a (29.4%), while in African populations, it was E1b1a (43.2%).
2. TMRCA analysis identified Iraq and Yemen as having the most ancient paternal lineages among Middle Eastern populations.
3. Berbers generally form their own distinct genetic cluster, with the notable exception of Berbers in Egypt, who show greater genetic similarity to Egyptian Arabs.
4. Eastern African populations represent the ancestral origin of African populations. They display lower ancestry variability compared to populations in other African regions.
5. Jewish populations generally exhibit genetic structures similar to other Middle Eastern populations, with the significant exception of Ashkenazi Jews. This is supported by both genetic structure analysis and ancestry variability analysis.

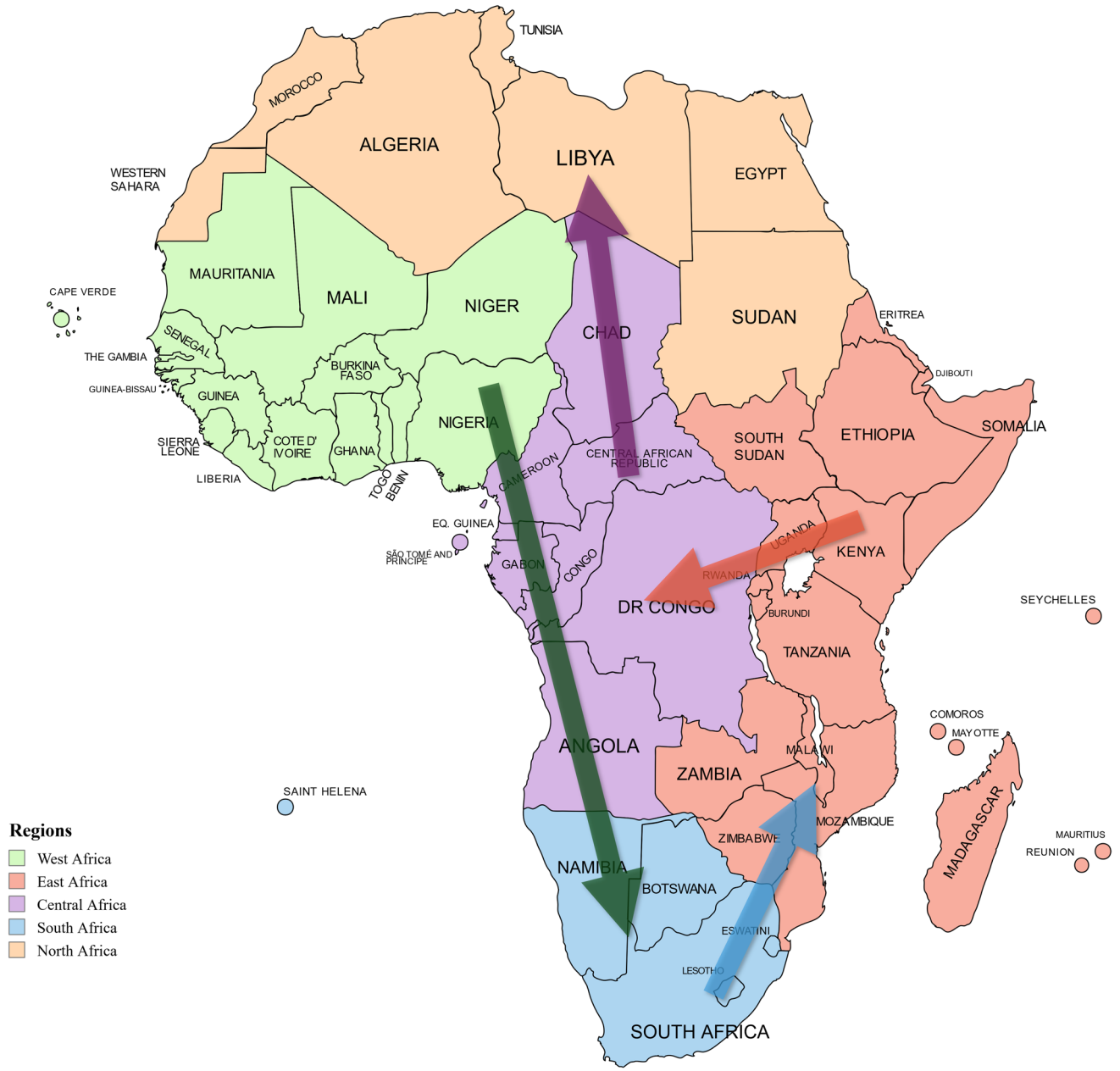

Fig. 12 The most dominant migration routes within the African continent. The numbers in front of the dominant migration routes represent the dominant migration model. The most probable routes were: west→south1, south→ east1, east→central1 and central→north2

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12024-025-01115-3.

Acknowledgements N/A

Authors' contributions Abdullah Hadi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing.
Shams Hadi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing.
Eida Khalaf Almohammed: Conceptualization, Investigation, Writing – review & editing.
Hayder Lazim: Conceptualization, Data curation, Formal analysis, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding None.

Data availability All relevant data are within the paper and its Supporting Information files.

Code availability N/A.

Declarations

Ethics approval N/A.

Consent to participate N/A.

Consent for publication N/A

Clinical Trial number N/A

Conflicts of interest/Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

1. Kumawat RK, Shrivastava P, Shrivastava D, Mathur GK. Molecular diversity of 23 Y-STR genetic markers in the population of Rajasthan, India. Meta Gene. 2020;24:100694. https://doi.org/10.1016/j.mgene.2020.100694.
2. Neyra-Rivera CD, Ticona Arenas A, Delgado Ramos E, Velasquez Reinoso MRE, Caceres Rey OA, Budowle B. Population data of 27 Y-chromosome STRs in Aymara population from Peru. Aust J Forensic Sci. 2022;54(5):596–610. https://doi.org/10.1080/00450618.2021.1882571.
3. Costa R, Fadoni J, Amorim A, Cainé L. Y-STR databases—application in sexual crimes. Genes. 2025;16(5):484. https://doi.org/10.3390/genes16050484.
4. Jin X, Xing G, Yang C, Zhang X, Cui W, Chen C, et al. Genetic polymorphisms of 44 Y chromosomal genetic markers in the Inner Mongolia Han population and its genetic relationship analysis with other reference populations. Forensic Sci Res. 2022;7(3):510–7. https://doi.org/10.1080/20961790.2020.1857509.
5. Boattini A, Sarno S, Mazzarisi AM, Viroli C, De Fanti S, Bini C, et al. Estimating Y-str mutation rates and tmrca through deep-rooting Italian pedigrees. Sci Rep. 2019;9(1):9032. https://doi.org/10.1038/s41598-019-45398-3.
6. Hodișan R, Zaha DC, Jurca CM, Petchesi CD, Bembea M. Genetic Diversity Based on Human Y Chromosome Analysis: A Bibliometric Review Between 2014 and 2023. Cureus. 2024;16(4):e58542. https://doi.org/10.7759/cureus.58542.
7. Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. Nature. 1987;325(6099):31–6. https://doi.org/10.1038/325031a0.
8. McDougall I, Brown FH, Fleagle JG. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. Nature. 2005;433(7027):733–6. https://doi.org/10.1038/nature03258.
9. White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, et al. Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. Nature. 2003;423(6941):742–7. https://doi.org/10.1038/nature01669.
10. Bergström A, Stringer C, Hajdinjak M, Scerri EML, Skoglund P. Origins of modern human ancestry. Nature. 2021;590(7845):229–37. https://doi.org/10.1038/s41586-021-03244-5.
11. Reyes-Centeno H. Out of Africa and into Asia: fossil and genetic evidence on modern human origins and dispersals. Quatern Int. 2016;416:249–62. https://doi.org/10.1016/j.quaint.2015.11.063.
12. Groucutt HS, Grün R, Zalmout IAS, Drake NA, Armitage SJ, Candy I, et al. *Homo sapiens* in Arabia by 85,000 years ago. Nat Ecol Evol. 2018;2(5):800–9. https://doi.org/10.1038/s41559-018-0518-2.
13. Hershkovitz I, Weber GW, Quam R, Duval M, Grün R, Kinsley L, et al. The earliest modern humans outside Africa. Science. 2018;359(6374):456–9. https://doi.org/10.1126/science.aap8369.
14. Al-Shoba K, Al-Hamadi N, Almohammed EK, Hadi S, Goodwin W, Lazim H. The Yemeni genetic structure revealed by the Y chromosome STRs. Forensic Sci Med Pathol. 2025. https://doi.org/10.1007/s12024-025-00975-z.
15. Aboukhalid R, Bouabdellah M, Abbassi M, Bentayebi K, Elmzibri M, Squalli D, et al. Haplotype frequencies for 17 Y-STR loci (AmpFlSTRY-filer) in a Moroccan population sample. Forensic Sci Int Genet. 2010;4(3):e73–4. https://doi.org/10.1016/j.fsigen.2009.06.004.
16. Carvalho M, Brito P, Bento AM, Gomes V, Antunes H, Costa HA, et al. Paternal and maternal lineages in Guinea-Bissau population. Forensic Sci Int Genet. 2011;5(2):114–6. https://doi.org/10.1016/j.fsigen.2010.10.007.
17. D'Atanasio E, Iacovacci G, Pistillo R, Bonito M, Dugoujon JM, Moral P, et al. Rapidly mutating Y-STRs in rapidly expanding populations: discrimination power of the Yfiler Plus multiplex in northern Africa. Forensic Sci Int Genet. 2019;38:185–94. https://doi.org/10.1016/j.fsigen.2018.11.002.
18. Della Rocca C, Cannone F, D'Atanasio E, Bonito M, Anagnostou P, Russo G, et al. Ethnic fragmentation and degree of urbanization strongly affect the discrimination power of Y-STR haplotypes in central Sahel. Forensic Sci Int Genet. 2020;49:102374. https://doi.org/10.1016/j.fsigen.2020.102374.
19. Elmrghni S, Coulson-Thomas YM, Kaddura M, Dixon RA, Williams DR. Population genetic data for 17 Y STR markers from Benghazi (East Libya). Forensic Sci Int Genet. 2012;6(2):224–7. https://doi.org/10.1016/j.fsigen.2011.05.001.
20 Fadhlaoui-Zid K, Chennakrishnaiah S, Zemni R, Grinberg S, Herrera RJ, Benammar-Elgaaied A. <scp>S</scp>ousse, <scp>T</scp>unisia: Tumultuous history and high <scp>Y</scp>-<scp>STR</scp>diversity. Electrophoresis. 2012;33(23):3555–63. https://doi.org/10.1002/elps.201200361.
21. Fadhlaoui-Zid K, Garcia-Bertrand R, Alfonso-Sánchez MA, Zemni R, Benammar-Elgaaied A, Herrera RJ. Sousse: extreme genetic heterogeneity in North Africa. J Hum Genet. 2015;60(1):41–9. https://doi.org/10.1038/jhg.2014.99.
22. Fortes-Lima C, Brucato N, Croze M, Bellis G, Schiavinato S, Massougbodji A, et al. Genetic population study of Y-chromosome markers in Benin and Ivory Coast ethnic groups. Forensic Sci Int Genet. 2015;19:232–7. https://doi.org/10.1016/j.fsigen.2015.07.021.
23 Gomes V, Alves C, Amorim A, Carracedo A, Sánchez-Diz P, Gusmão L. Nilotes from Karamoja, Uganda: haplotype data defined by 17 Y-chromosome STRs. Forensic Sci Int Genet. 2010;4(4):e83–6. https://doi.org/10.1016/j.fsigen.2009.07.001.
24. González M, Gomes V, López-Parra AM, Amorim A, Carracedo A, Sánchez-Diz P, et al. The genetic landscape of Equatorial Guinea and the origin and migration routes of the Y chromosome haplogroup R-V88. Eur J Hum Genet. 2013;21(3):324–31. https://doi.org/10.1038/ejhg.2012.167.
25. Haddish K, Chierto E, Di Vella G, Lacerenza D, Raddi S, Aneli S, et al. A reference database of forensic autosomal and gonosomal STR markers in the Tigray population of Ethiopia. Forensic Sci Int Genet. 2022;56:102618. https://doi.org/10.1016/j.fsigen.2021.102618.

26. Iacovacci G, D'Atanasio E, Marini O, Coppa A, Sellitto D, Trombetta B, et al. Forensic data and microvariant sequence characterization of 27 Y-STR loci analyzed in four Eastern African countries. Forensic Sci Int Genet. 2017;27:123–31. https://doi.org/10.1016/j.fsigen.2016.12.015.
27. Kofi AE, Hakim HM, Khan HO, Ismail SA, Ghansah A, David AA, et al. Population data of 23 Y chromosome STR loci for the five major human subpopulations of Ghana. Int J Legal Med. 2020;134(4):1313–5. https://doi.org/10.1007/s00414-019-02099-w.
28. Omran GA, Rutty GN, Jobling MA. Diversity of 17-locus Y-STR haplotypes in upper (southern) Egyptians. Forensic Sci Int Genet Suppl Ser. 2008;1(1):230–2. https://doi.org/10.1016/j.fsigss.2007.11.009.
29. Purps J, Siegert S, Willuweit S, Nagy M, Alves C, Salazar R, et al. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. Forensic Sci Int Genet. 2014;12:12–23. https://doi.org/10.1016/j.fsigen.2014.04.008.
30. Shonhai M, Nhiwatiwa T, Nangammbi T, Mazando S. Genetic analysis of 27 Y-chromosomal STR loci in a Zimbabwean Shona ethnic group. Leg Med Tokyo. 2020;43:101660. https://doi.org/10.1016/j.legalmed.2019.101660.
31. Triki-Fendri S, Sánchez-Diz P, Rey-González D, Ayadi I, Alfadhli S, Rebai A, et al. Population genetics of 17 Y-STR markers in West Libya (Tripoli region). Forensic Sci Int Genet. 2013;7(3):e59-61. https://doi.org/10.1016/j.fsigen.2013.02.002.
32. Wepeba P-P, Abaidoo CS, Goodwin WH. Haplogroup prediction in the Ghanaian population using haplotype data of 27 Yfiler® plus loci and TaqMan SNP genotyping. Forensic Sci Int: Genetics Supplement Series. 2022;8:147–8. https://doi.org/10.1016/j.fsigss.2022.10.015.
33. Xu H, Wang CC, Shrestha R, Wang LX, Zhang M, He Y, et al. Inferring population structure and demographic history using Y-STR data from worldwide populations. Mol Genet Genomics. 2015;290(1):141–50. https://doi.org/10.1007/s00438-014-0903-8.
34. Zeye MMJ, Li J, Ouedraogo SY, Zha L, Simpore J, Jifeng C. Population data and genetic structure analysis based on 29 Y-STR loci among the ethnolinguistic groups in Burkina Faso. Int J Legal Med. 2021;135(5):1767–9. https://doi.org/10.1007/s00414-021-02544-9.
35. Barbieri C, Hübner A, Macholdt E, Ni S, Lippold S, Schröder R, et al. Refining the Y chromosome phylogeny with southern African sequences. Hum Genet. 2016;135(5):541–53. https://doi.org/10.1007/s00439-016-1651-0.
36. Bekada A, Arauna LR, Deba T, Calafell F, Benhamamouch S, Comas D. Genetic heterogeneity in Algerian human populations. PLoS ONE. 2015;10(9):e0138453. https://doi.org/10.1371/journal.pone.0138453.
37. Bini C, Sarno S, Tangorra E, Iuvaro A, De Fanti S, Tseghereda YG, et al. Haplotype data and forensic evaluation of 23 Y-STR and 12 X-STR loci in eight ethnic groups from Eritrea. Int J Legal Med. 2021;135(2):449–53. https://doi.org/10.1007/s00414-020-02446-2.
38. Ennafaa H, Fregel R, Khodjet-el-khil H, González AM, Mahmoudi HAE, Cabrera VM, et al. Mitochondrial DNA and Y-chromosome microstructure in Tunisia. J Hum Genet. 2011;56(10):734–41. https://doi.org/10.1038/jhg.2011.92.
39. Fadhlaoui-Zid K, Martinez-Cruz B, Khodjet-el-khil H, Mendizabal I, Benammar-Elgaaied A, Comas D. Genetic structure of Tunisian ethnic groups revealed by paternal lineages. Am J Phys Anthropol. 2011;146(2):271–80. https://doi.org/10.1002/ajpa.21581.
40. Frigi S, Pereira F, Pereira L, Yacoubi B, Gusmão L, Alves C, et al. Data for Y-chromosome haplotypes defined by 17 STRs (AmpFLSTR Yfiler) in two Tunisian Berber communities. Forensic Sci Int. 2006;160(1):80–3. https://doi.org/10.1016/j.forsciint.2005.05.007.
41. Fujihara J, Yuasa I, Muro T, Iida R, Tsubota E, Nakamura H, et al. Allele frequencies and haplotypes for 28 Y-STRs in Ovambo population. Leg Med (Tokyo). 2009;11(4):205–8. https://doi.org/10.1016/j.legalmed.2009.03.009.
42. Laouina A, El Houate B, Yahia H, Azeddoug H, Boulouiz R, Chbel F. Allele frequencies and population data for 17 Y-STR loci (The AmpFlSTR® Y-filer™) in Casablanca resident population. Forensic Sci Int Genet. 2011;5(1):e1-3. https://doi.org/10.1016/j.fsigen.2010.10.016.
43. Palet L, Coudray C, Galey CE, Keyser C, Melhaoui M, Gagnor C, et al. Y-STR genetic diversity in Moroccans from the Figuig oasis. Forensic Sci Int Genet. 2010;4(5):e139–41.
44. Robino C, Crobu F, Di Gaetano C, Bekada A, Benhamamouch S, Cerutti N, et al. Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample. Int J Legal Med. 2008;122(3):251–5. https://doi.org/10.1007/s00414-007-0203-5.
45. Simms TM, Wright MR, Martinez E, Regueiro M, McCartney Q, Herrera RJ. Y-STR diversity and sex-biased gene flow among Caribbean populations. Gene. 2013;516(1):82–92. https://doi.org/10.1016/j.gene.2012.11.006.
46 Tau T, Davison S, D'Amato ME. Polymorphisms at 17 Y-STR loci in Botswana populations. Forensic Sci Int Genet. 2015;17:47–52. https://doi.org/10.1016/j.fsigen.2015.03.001.
47. Martinez B, Catelli L, Romero M, Okolie VO, Keshinro SO, Carvalho EF, et al. Forensic evaluation of 27 Y-STR haplotypes in a population sample from Nigeria. Forensic Sci Int Genet Suppl Ser. 2017;6:e289. https://doi.org/10.1016/j.fsigss.2017.09.138.
48. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10(3):564–7. https://doi.org/10.1111/j.1755-0998.2010.02847.x.
49. Arlequin ver 3.5.2.2. Available from: http://cmpg.unibe.ch/software/arlequin35/. Accessed 27 June 2025.
50. Takezaki N, Nei M, Tamura K. Poptree2: software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. Mol Biol Evol. 2009;27(4):747–52. https://doi.org/10.1093/molbev/msp312.
51. FigTree. Available from: http://tree.bio.ed.ac.uk/software/figtree/. Accessed 27 June 2025.
52. Szpiech ZA, Jakobsson M, Rosenberg NA. ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics. 2008;24(21):2498–504. https://doi.org/10.1093/bioinformatics/btn478.
53. ADZE. Available from: https://rosenberglab.stanford.edu/adze.html. Accessed 27 June 2025.
54. Rosenberg NA. Algorithms for selecting informative marker panels for population assignment. J Comput Biol. 2005;12(9):1183–201. https://doi.org/10.1089/cmb.2005.12.1183.
55. Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry*. Am J Hum Genet. 2003;73(6):1402–22. https://doi.org/10.1086/380416.
56. infocalc. Available from: https://rosenberglab.stanford.edu/infocalc.html. Accessed 27 June 2025.
57. NEVGEN GENEALOGY TOOLS V1.1. Available from: https://site.nevgen.org/nevgen-genealogy-tools-v1-1/. Accessed 27 June 2025.
58. Network Publisher. Available from: http://www.fluxus-engineering.com/sharenet.htm. Accessed 27 June 2025.
59. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16(1):37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036.
60. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945–59. https://doi.org/10.1093/genetics/155.2.945.

61. Structure Software. Available from: https://web.stanford.edu/group/pritchardlab/structure.html. Accessed 27 June 2025.
62. Earl DA, von Holdt BM. Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. Conserv Genet Resour. 2012;4(2):359–61. https://doi.org/10.1007/s12686-011-9548-7.
63. SRUCTURE HARVESTER. Available from: https://github.com/dentearl/structureHarvester. Accessed 27 June 2025.
64. Rosenberg NA. Distruct: a program for the graphical display of population structure. Mol Ecol Notes. 2004;4(1):137–8. https://doi.org/10.1046/j.1471-8286.2003.00566.x.
65. CLUMPP. Available from: https://rosenberglab.stanford.edu/clumpp.html. Accessed 27 June 2025.
66. Distruct. Available from: https://rosenberglab.stanford.edu/distruct.html. Accessed 27 June 2025.
67. Morrison ML, Alcala N, Rosenberg NA. Fstruct: an FST-based tool for measuring ancestry variation in inference of population structure. Mol Ecol Resour. 2022;22(7):2614–26. https://doi.org/10.1111/1755-0998.13647.
68. FSTruct. Available from: https://github.com/maikemorrison/fstruct. Accessed 27 June 2025.
69. MIGRATE. Available from: https://peterbeerli.com/migrate-html5/. Accessed 27 June 2025.
70. Almohammed EK, Hadi A, Al-Asmakh M, Lazim H. The qatari population's genetic structure and gene flow as revealed by the Y chromosome. PLoS ONE. 2023;18(9):e0290844. https://doi.org/10.1371/journal.pone.0290844.
71. Lazim H, Almohammed EK, Hadi S, Smith J. Population genetic diversity in an Iraqi population and gene flow across the Arabian Peninsula. Sci Rep. 2020. https://doi.org/10.1038/s41598-020-72283-1.
72. Vicente M, Schlebusch CM. African population history: an ancient DNA perspective. Curr Opin Genet Dev. 2020;62:8–15. https://doi.org/10.1016/j.gde.2020.05.008.
73. Sahakyan H, Margaryan A, Saag L, Karmin M, Flores R, Haber M, et al. Origin and diffusion of human Y chromosome haplogroup J1–M267. Sci Rep. 2021;11(1):6659. https://doi.org/10.1038/s41598-021-85883-2.
74. Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc Natl Acad Sci U S A. 2005;102(44):15942–7. https://doi.org/10.1073/pnas.0507611102.
75. B MA, Hadi S, Ismael B, Barqee A, Hadi A, Lazim H. An X-STRs analysis of the Iraqi Sorani Kurds. PLoS One. 2023;18(11):e0294973. https://doi.org/10.1371/journal.pone.0294973
76 Shi H, Su B. Molecular adaptation of modern human populations. Int J Evol Biol. 2010;2011:484769. https://doi.org/10.4061/2011/484769.
77. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet. 2008;9(1):403–33. https://doi.org/10.1146/annurev.genom.9.081307.164258.
78. Reed FA, Tishkoff SA. African human diversity, origins and migrations. Curr Opin Genet Dev. 2006;16(6):597–605. https://doi.org/10.1016/j.gde.2006.10.008.
79. Zakharia F, Basu A, Absher D, Assimes TL, Go AS, Hlatky MA, et al. Characterizing the admixed African ancestry of African Americans. Genome Biol. 2009;10(12):R141. https://doi.org/10.1186/gb-2009-10-12-r141.
80. Kedar-Barnes I, Rozen P. The Jewish people: their ethnic history, genetic disorders and specific cancer susceptibility. Fam Cancer. 2004;3(3–4):193–9. https://doi.org/10.1007/s10689-004-9544-0.
81. Ostrer H, Skorecki K. The population genetics of the Jewish people. Hum Genet. 2013;132(2):119–27. https://doi.org/10.1007/s00439-012-1235-6.
82. Costa MD, Pereira JB, Pala M, Fernandes V, Olivieri A, Achilli A, et al. A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. Nat Commun. 2013;4(1):2543. https://doi.org/10.1038/ncomms3543.
83. Arauna LR, Comas D. Genetic Heterogeneity between Berbers and Arabs. Ency Life Sci. 2017: 1–7. https://doi.org/10.1002/9780470015902.a0027485.
84. Arauna LR, Mendoza-Revilla J, Mas-Sandoval A, Izaabel H, Bekada A, Benhamamouch S, et al. Recent Historical Migrations Have Shaped the Gene Pool of Arabs and Berbers in North Africa. Mol Biol Evol. 2017;34(2):318–29. https://doi.org/10.1093/molbev/msw218.
85. Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, et al. Genomic ancestry of North Africans supports back-to-Africa migrations. PLoS Genet. 2012. https://doi.org/10.1371/journal.pgen.1002397.
86. Fortes-Lima CA, Burgarella C, Hammarén R, Eriksson A, Vicente M, Jolly C, et al. The genetic legacy of the expansion of Bantu-speaking peoples in Africa. Nature. 2024;625(7995):540–7. https://doi.org/10.1038/s41586-023-06770-6.