Article

Genomic Landscape of the Mitochondrial Genome in the United Arab Emirates Native Population

Fatma A. Aljasmi [1,†], Ranjit Vijayan [2,†], Naganeeswaran Sudalaimuthuasari [3], Abdul-Kader Souid [1], Noushad Karuvantevida [2], Raja Almaskari [2], Hidaya Mohammed Abdul Kader [2], Biduth Kundu [2], Khaled Michel Hazzouri [3] and Khaled M. A. Amiri [2,3,*]

[1] Pediatric Department, United Arab Emirates University, Al Ain, Abu Dhabi 15551, UAE;
    aljasmif@uaeu.ac.ae (F.A.A.); asouid@uaeu.ac.ae (A.-K.S.)
[2] Biology Department, United Arab Emirates University, Al Ain, Abu Dhabi 15551, UAE;
    ranjit.v@uaeu.ac.ae (R.V.); kvnoushad@me.com (N.K.); r.almeskari@uaeu.ac.ae (R.A.);
    201790053@uaeu.ac.ae (H.M.A.K.); biduth.k@uaeu.ac.ae (B.K.)
[3] Khalifa Center for Genetic Engineering and Biotechnology, United Arab Emirates University, Al Ain,
    Abu Dhabi 15551, UAE; naganeeswaran@uaeu.ac.ae (N.S.); khaled_hazzouri@uaeu.ac.ae (K.M.H.)
*   Correspondence: k.amiri@uaeu.ac.ae
†   Equally contributed.

check for updates

Abstract: In order to assess the genomic landscape of the United Arab Emirates (UAE) mitogenome, we sequenced and analyzed the complete genomes of 232 Emirate females mitochondrial DNA (mtDNA) within and compared those to Africa. We investigated the prevalence of haplogroups, genetic variation, heteroplasmy, and demography among the UAE native population with diverse ethnicity and relatively high degree of consanguinity. We identified 968 mtDNA variants and high-resolution 15 haplogroups. Our results show that the UAE population received enough gene flow from Africa represented by the haplogroups L, U6, and M1, and that 16.8% of the population has an eastern provenance, depicted by the U haplogroup and the M Indian haplogroup (12%), whereas western Eurasian and Asian haplogroups (R, J, and K) represent 11 to 15%. Interestingly, we found an ancient migration present through the descendant of L (N1 and X) and other sub-haplogroups (L2a1d and L4) and (L3x1b), which is one of the oldest evolutionary histories outside of Africa. Our demographic analysis shows no population structure among populations, with low diversity and no population differentiation. In addition, we show that the transmission of mtDNA in the UAE population is under purifying selection with hints of diversifying selection on ATP8 gene. Last, our results show a population bottleneck, which coincides with the Western European contact (1400 ybp). Our study of the UAE mitogenomes suggest that several maternal lineage migratory episodes liking African–Asian corridors occurred since the first modern human emerges out of Africa.

Keywords: Mitochondrial DNA; next generation sequencing; Selection; heteroplasmy; demography; single nucleotide polymorphism (SNP)

1. Introduction

There are two potential routes to the Arabian Peninsula, the northern and southern route, which is the first step out of Africa. This is the primary link between Africa and Eurasia. The maternally inherited mtDNA has been used as a marker to relate lineages across geographic origins culminating in African haplogroup L and the Eurasian M and N, which shared a common route with the African L3, and this radiation likely started the Eurasian colonization [1–3]. Furthermore, the star shape

radiation in the Indian and East Asian M lineage supports a fast southern dispersal [2]. Previous studies highlight the presence of autochthone M and N lineages along the southern route [3–7]. As a result, the M and N lineages have a unique migration trail [8,9] and the southern coastal trail was the only route for the western Eurasian colonization, which is an early sprout of the southern radiation in India [3,10]. Under these scenarios, the Arabian Peninsula, an obliged link between East Africa and South Asia, attracted a lot of attention. Indeed, several mtDNA studies have been published from this region [11–13] and the majority of these studies point to more recent African, Asian, or northern Neolithic origins. Kivisild et al. 2004 defined a new group, L6, with no match to African populations, suggesting an ancient migration from Africa to Yemen. This suggests that ancient migration via the southern route to the neighboring countries such as Oman and UAE is plausible.

Demographic history, such as effective population size changes, short and long distance migrations, as well as admixture, shape the genetic variation of modern African populations. This is in addition to selection on specific loci, combined with recombination and mutation. For instance, one of the migrations that impacts the genetic variation in modern African populations is the migration of agricultural Bantu speaking from West Africa throughout sub-Saharan Africa 4000 years ago followed by admixture with indigenous populations [14–18]. Compared to non-African populations, African populations have higher genetic diversity, population substructure, and low linkage disequilibrium (LD) [19]. In addition, they have evolved an adaptive response to various diets and climate change. Thus far, the evolutionary force(s) that shape the genetic variation and diversity of the UAE population and how they compare to the African population are not known.

Almost in all cases, nuclear DNA is used to describe the signature of selection, while ignoring its effect on the tissue, cells, and subcellular compartments. This is very crucial for mitogenome, which in contrast to the nucleus, serves as a powerhouse for the cell and are present in multiple copies in the cytoplasm that may vary in sequence (heteroplasmy) and quantity among tissues [20]. Each mitochondrion is maternally inherited and codes for enzymes that are mainly involved in cellular bioenergetics [21]. As it is a vital compartment for the generation of cellular metabolism, including ATP production, nucleotide biosynthesis, and other activities, any dysfunction will lead to tissue and systemic disorders [20]. Therefore, strong purifying negative selection acts to remove deleterious mutations, and in parallel, positive selection acts on the mitochondria to promote adaptation of cells, and in return the whole organism, to environmental and physiological changes [20,22,23]. Furthermore, due to the small mitochondria genome size (16.5 Kb), and with the advances in next-generation sequencing, it is facile to get high-throughput information from hundreds of individuals. The information generated from the human mitogenome data help in addressing population evolutionary history and quantify the genetic variation and its effect, which are relevant to metabolic, genetic, and forensic fields [24,25].

The human mitogenome is highly polymorphic, and most of its variants are benign [26]. However, deleterious variants have been reported in various diseases, including Leber hereditary optic nephropathy (LHON); the familial mitochondrial cytopathies; mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS); and many others [27]. The higher rate of accumulation of deleterious mutations in the mitogenome is due to the small effective population size associated with its haploid inheritance [28]. In addition, heteroplasmy levels of mutations are of importance. For instance, in the 1980s, the first study of heteroplasmy in the mitogenome showed different levels of mutations (e.g., deletions and point mutations) in affected patients [29,30]. Heteroplasmy also exists with no apparent functional consequences and with no mitochondrial diseases [31–36].

In this study, we analyzed the mitogenome landscape of the UAE population. The UAE is located in Western Asia on the Gulf, south of the Strait of Hormuz at the southeast end of the Arabian Peninsula. It borders Oman (east) and Saudi Arabia (south) and shares sea borders with Qatar (west) and Iran (north). The country is a federation of seven emirates: Abu Dhabi (capital), Dubai, Sharjah, Ajman, Umm Al-Quwain (UAQ), Ras Al-Khaimah (RAK), and Fujairah. Approximately 13% of the population (947,997 in 2010) is Emirati citizens. Al Ain, the site of the current study, is the largest inland city in UAE and is part of the emirate of Abu Dhabi. It is located east of the capital and south of Dubai and

has the highest proportion of Emirati nationals (~20%). Citizens of the UAE have diverse ethnicity that includes links to the Arabian Peninsula, Persia, Baluchistan, and East Africa. The society is mainly tribal, and intratribal marriages are fairly practiced. Consequently, founder variants and prevalence of inborn errors of metabolism and genetic disorders are exceptionally high [37–39].

We carried out complete whole mitogenome sequencing, assembly, and annotation of 232 female UAE citizens and highlighted the genetic population grouping (haplogrouping). We characterized variants and their effect and test for selective force acting on the UAE mitogenome. We estimated diversity, population structure, and differentiation among cities and regions. In addition, to characterize variants, we estimated heteroplasmy in the population. Finally, we used mtDNA to construct a portrait of the Holocene and late Pleistocene population size of the UAE population.

2. Materials and Methods

2.1. Ethics Statement, Sample Collection, and DNA Isolation

Al Ain Medical Human Research Ethics Committee approved this study according to the national regulations (#10/09). This study recruited national UAE female students (age range: 18 to 24 years) matriculated at the UAE University. The UAE University is a federal institution and students' diversity represents all of the seven Emirates. Venous blood samples (5 mL in EDTA-Vacutainers) were collected randomly from 248 consented female students. Total genomic DNA was extracted from blood lymphocytes using the DNeasyBlood and Tissue Kit (Qiagen, Hilden, Germany). DNA quality and concentration were confirmed with NanoDrop and agarose gel, and the samples were stored at −20 °C. We used already published 70 genomes from East Africa in our analysis (accession numbers JN655773–JN655842) for comparative analysis.

2.2. Human Mitogenome Enrichment and Sequencing

Mitogenome enrichment is a critical step to reduce nuclear DNA contamination and was accomplished using a specific long-range PCR amplification step directed by two sets of overlapping primers, where each pair of primers flanked 8500 base pairs (bp) [40]. Two PCR amplicons were used for the NGS library preparation. A 200-bp sequencing kit (Ion Xpress Plus Fragment Library Kit; Life Technologies, Carlsbad, CA, USA) was used to generate a short mitogenome fragment library. The samples were loaded onto sequencing chips, and they were sequenced using an Ion Torrent™ Personal Genome Machine™ (PGM) system platform (Life Technologies, Marsiling, Singapore). Each sequencing chip had the capacity to sequence mitogenomes from 10–20 subjects at a coverage of 250–500X. Primer pairs (F16441: 5′ACTCTCCTCGCTCCGGGCCC3′, R29: 5′TCTATCACCCTATTAACCAC3′) were used to amplify the control region of the mitogenomes. A Genetic Analyzer (model 3500; Applied Biosystems, software version 3.0, Applied Biosystems, Hitachi, Japan) was used to sequence the PCR amplified mitogenome control region (~120 bp), and gap regions were filled manually.

2.3. Data QC, Assembly, and Variant Identification

The quality of the raw Ion Torrent PGM fastq files was checked using FastQC. The Geneious software platform [41] was used for read trimming, reference mapping, and assembly process. The homopolymer quality reduction option of Geneious was used to manage homopolymer runs that are associated with Ion Torrent data. The raw fastq reads with a quality value less than 20 (Q < 20) were trimmed out. Preprocessed reads were mapped to the revised Cambridge Reference Sequence (rCRS) in Geneious version 9.0.4, using the Bowtie2 mapper (bowtie parameter: bowtie2-align-s -I 0 -X 800 -*p* 20 –sensitive -D 15 -R 2 -N 0 -L 22 -i S,1,1.15) [42]. Homopolymer quality reduction was set to 30% to account for such errors in Ion Torrent reads [43]. Geneious bowtie plugin was used for mitogenome assembly from the aligned BAM files. A variant calling file (VCF) for the merged BAMs was generated using GATK [44] using ploidy 1 for haploid genome. Variant annotation of the VCF

was carried out using HmtNote database [45]. The annotation was performed using data from hmtVar, which is a recently published database that collects information from several online databases as well as offering in-house pathogenicity predictions. The number of variants (SNPs, Indels, and multiple nucleotide polymorphisms (MNPs)) was counted per genes. Allele frequency was calculated as well as synonymous and nonsynonymous variants from the VCF file. Circos and bar plots were generated to summarize the variants and their distributions [46].

2.4. Tree Reconstruction and Haplogroup Prediction

A maximum parsimony tree was generated using coding region sequences from the 232 female mitogenomes (positions 576-16,023) separately, as well as combined with 70 African mitogenomes using MEGA6 [47]. Branch lengths are proportional to the number of mutations. Tree visualization was performed using Figtree v.1.4.0 [48]. Detected variants were manually assessed to ensure they were not assembly artifacts. Identified variations were exported and converted using in-house scripts for haplogroup determination. Haplogroup, based on PhyloTree Build 17, was determined using HaploGrep2 and mtHap [49]. Haplogroups were defined and their relative frequencies for the UAE population are represented as pie chart. The close geographical proximities of neighboring different Emirates can be essentially considered as one region, which allowed us to add their frequencies and place it on a UAE geographic map generated using R map [46].

2.5. Population Structure and Differentiation

We calculated pairwise population differentiation Fst with vcftools using haploid option. We also estimated Hs (Heterozygosity with structure) and Ht (Heterozygosity without structure) as well as Gst, G'st, and D statistic using R package adegenet [50]. The same package was used to describe population structure (K = 1 to 12) using discriminant analysis of principal components (DAPC) [51]. We also evaluated population structure using maximum likelihood phylogenetic tree using raxml [52]. We tested if distinct subpopulations (e.g., cities) are mixed together using analysis ADMIXURE [53] with the same number of clusters and the best k value was evaluated using CLUMP [54]. The analysis was performed on the major cities of the UAE: Al Ain (n = 86), Abu Dhabi (n = 18), Dubai (n = 15), Sharjah (n = 15), RAK (n = 39), UAQ (n = 3), and Fujairah (n = 43). Furthermore, Fst was also measured based on the UAE's geographical regions: Southwest (Al Ain and Abu Dhabi, n = 104), most-Northeast (Fujairah, n = 43), most-Northwest (RAK, n = 39), and mid-Northwest (Dubai, Sharjah, Ajman, and UAQ, n = 37).

2.6. Diversity, Kinship, and Selection

Coding sequences for the thirteen mitochondrial genes were codon-aligned in frame using pal2nal (version 14) for the UAE individuals. The number of polymorphic sites per population (S), nucleotide diversity (π), Watterson theta (θ), and Tajima's D [55] were calculated using MEGA6. This was conducted for different haplogroups as well as for each city in the UAE. We used NGSrelate (version 2) to estimate relatedness and plotted kinship matrix relationship among the different 232 female individuals. We inferred the strength of selection using dN/dS metric implemented in HyPhy using the method BUSTED.

2.7. Reconstruction of Demographic History

To reconstruct the demographic history in our samples, we used BEAST version 1.8.0 [56]. The program will estimate wide different model parameters, such as genealogical structure, substitution model, and effective population size given a set of genetic sequences. An uncorrelated relaxed clock model was used, which allows the rate to vary across branches in the genealogy. Demographic history was reconstructed using Bayesian skyline model [57]. The complete BEAST input file is available upon request.

2.8. Heteroplasmy and Structural Variation Identification

Heteroplasmy identification was carried out using mtDNA-Server [58]. The mtDNA-Server is optimized to analyze the Ion Torrent PGM-aligned reads. Heteroplasmy was plotted as proxy for heterozygosity and a cut-off of > 5% was used in the identification process. Another manual curation, including coverage, being away from indels, and MNPs, as well as the Ts/Tv ratio, was used in the filtering process. Long-range structural variation was run using eKLIPse [59].

3. Results

3.1. Mitogenome Assembly and Variant Annotation

Preprocessed NGS reads were aligned against the reference mitogenome using the Bowtie2 program. We identified 968 SNPs and 30 indels including 11 MNPs (Figure 1A). We found more synonymous than nonsynonymous mutations in our population. This is also the same for all coding genes except for ATP8 gene, where the numbers are the same. Allele frequency found to be skewed toward low-frequency polymorphism (Figure 1A). We found an average read coverage depth of ~400X from the alignments and only high quality variants per site proceeded for analysis (Figure 1B). The level of heteroplasmy was estimated as a proxy for heterozygosity (Figure 1C). From the alignment BAM files, 232 complete mitogenomes were assembled. Annotation of these mitogenomes resulted in 13 protein-coding genes, 2 ribosomal RNA (rRNA) genes, and 22 transfer RNA (tRNA) genes (Figure 1A). The remaining control regions (~120 bp) were separately sequenced and manually filled the mitogenome gap. The number of variants was calculated per genomic features, and we observe more variants in ND5 gene and D-loop (Figure 1D) compared to others. Sequenced mitogenomes (*n* = 232) were deposited in NCBI-Genbank database (accession numbers: MF437054–MF437285), and generated NGS reads were deposited in NCBI-SRA database (SRA ID: PRJNA566159).
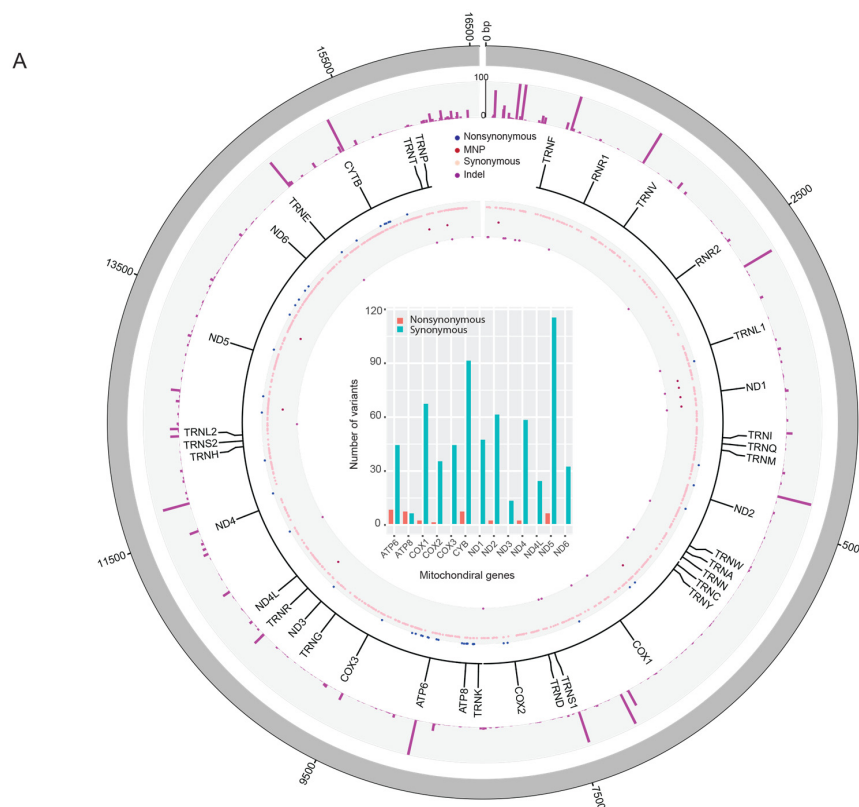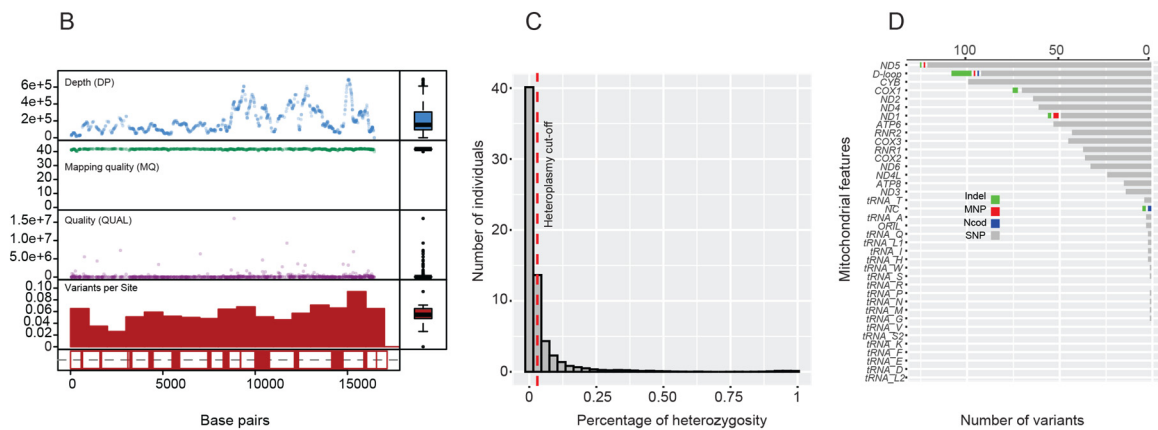

Figure 1. *Cont.*

Figure 1. Mitochondrial genomic variation within the UAE population. (**A**) Circos plot depicting the overall mitochondrial genomic variation. From the outer to inner rings: (1) variant allele frequency, (2) variant type, and (3) gene positions. The inner barplot shows the synonymous and nonsynonymous variants for each mitochondrial gene. (**B**) Quality control statistics such as depth of coverage (DP), mapping quality (MQ), and quality of variants (QUAL), as well as the number of variants per site. (**C**) Percentage of heterozygosity as a proxy for heteroplasmy for heterozygote variants depicted on the *x*-axis. The *y*-axis is the number of individuals with heteroplasmic locus. (**D**) The number of variants distribution per mitochondrial gene, RNR, tRNA, and noncoding.

3.2. Haplogroup Identification

Fifteen haplogroups were identified in the 232 samples (Table 1). A network relationship of different haplogroups is shown in Figure 2A. It presents the ancestral diverse haplogroups with long branch length (L0, L1, L2, L3), indicating more diversity compared to the remaining M and N and other derived haplogroups. A pie chart summarizing the frequencies of different haplogroups is highlighted in Figure 2B. Briefly, haplogroup U was predominant, representing 16.81% (39 samples) of the total studied samples. All of the sub-haplogroups of U (U1, U2, U3, U4, U5, U6, U7, and U9) were identified in the UAE samples. U2 (U2b1, U2b2, U2c1, and U2c1a), U3 (U3a2a1, U3b1, and U3b1a), and U4 (U4c1) constituted 15% of individuals in our population and they are also common haplogroups in India, while North African subclade U6a has a provenance primarily in Morocco. European subclade U2e (U2e1, U2e1b, and U2e3) and rare subclades U9a and U9b represent gene flows from the north (Table S1).

Table 1. UAE population haplogroup distribution.

Group [a]	F [b]	F [b]%	Unknown	NE ** Fujairah	NW ** RAK *	UAQ *	Mid-NW ** Dubai	Sharjah	Ajman	SW ** Al Ain	Abu Dhabi
E	2	0.86	0	0	0	0	2	0	0	0	0
F	1	0.43	0	0	0	0	0	0	0	1	0
H	19	8.18	0	5	3	0	1	3	0	7	0
HV	10	4.31	0	0	1	0	1	1	0	5	2
I	3	1.29	0	0	0	0	1	0	0	2	0
J	25	10.77	0	0	5	0	1	2	1	13	3
K	26	11.20	0	11	11	0	0	0	0	4	0
L	20	8.62	1	3	4	0	0	4	0	5	3
M	28	12.06	2	11	5	0	2	0	0	7	1
N	6	2.58	0	0	0	0	1	0	1	4	0
R	35	15.08	1	6	7	2	2	2	1	14	0

Table 1. *Cont.*

Group [a]	F [b]	F [b]%	Unknown	NE ** Fujairah	NW ** RAK *	UAQ *	Mid-NW ** Dubai	Sharjah	Ajman	SW ** Al Ain	Abu Dhabi
T	14	6.03	0	3	0	0	2	1	0	6	2
U	39	16.81	3	4	3	1	2	2	1	16	7
W	1	0.43	1	0	0	0	0	0	0	0	0
X	3	1.29	1	0	0	0	0	0	0	2	0
Total	232	100	9	43	39	3	15	15	4	86	18

** Region-wise classification: NE—Northeast region (*n* = 43), NW—Northwest region (*n* = 39); Mid-NW—mid-Northwest region (*n* = 37); SW—Southwest region (*n* = 104); * RAK—Ras al Khaimah; UAQ—Umm al-Quwain; a Haplogroup; b Frequency.
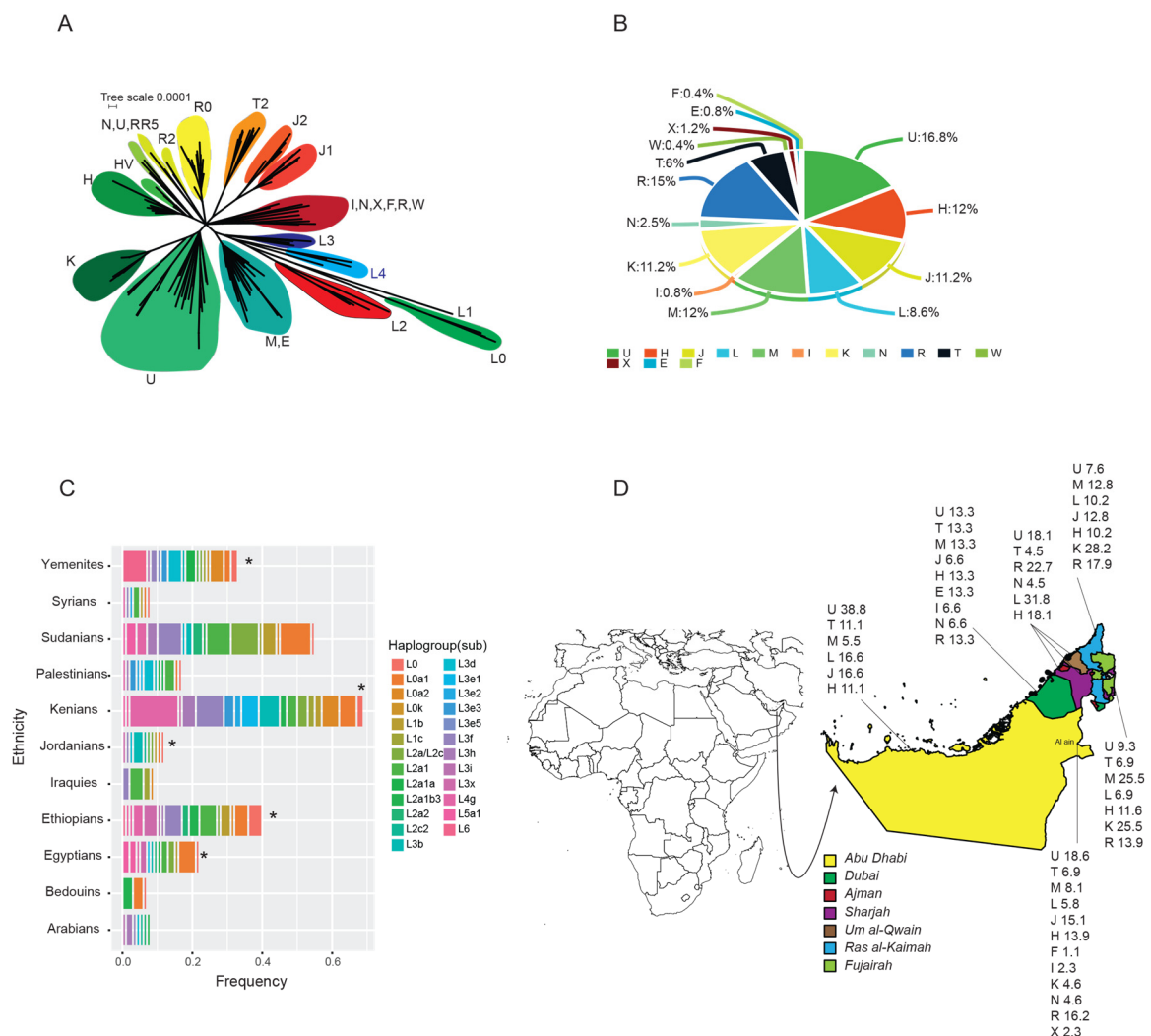

Figure 2. Haplogroup distribution and frequency. (**A**) Network haplogroup tree depicting the ancestral (L) haplogroups and the out of Africa haplogroups. (**B**) Total haplogroup frequencies for the UAE population. (**C**) Stacked bar plots for the frequencies of L3 haplogroup from different Levant and African countries. Stars highlight the presence of L3x1b sub-haplogroups in these countries. (**D**) A map of Africa and the Arabian Peninsula and a zoom on the UAE map with highlights of the different haplogroup frequencies for the different cities of the UAE.

Haplogroup R was the second most common haplogroup, found in 15.08% of samples. Sub-haplogroups of R (R0a2f, R0a2f1b, R0a2h, R2d, R30a1a, R30b1, and R5a2) were identified

in 35 samples. Haplogroup M accounted for 12.06% of the total; the generalized African subclade M1a (M1a1, M1a1b1b, and M1a1f) constituted 14% of the total haplogroup. Haplogroup K represented 11.20% of the total population. Haplogroup J was found in 25 samples (10.77%); sub-haplogroups J1b (J1b, J1b1b1, J1b1b3, and J1b2), J2a (J2a2a1a1, J2a2b, J2a2b1, and J2a2c1), and J2b (J2b1 and J2b1f) were detected in this study (Table S1).

Haplogroup L (Sub-Saharan Africa) accounted for 8.62% of the total mitogenome samples (L0, 1.29%; L1, 0.43%; L2, 3.01%; L3, 1.72%; and L4, 2.15%). We also identified one individual with L3x1b sub-haplogroups in the population. Sub-haplogroups L3x1b in L3 is one of the oldest evolutionary steps in the history of out of Africa, and it was previously reported in Kenya, Jordan, Yemen, Ethiopia, and Egypt (Figure 2C).

North Africa haplogroup HV was detected in 4.31% of the total, while Western Asia (the Near East) haplogroup H was detected in 8.18% of the total. Haplogroup T constituted 6.03% of the total, a clade that emanated from the Near East and was common among Iranians. The remaining identified haplogroups (E, F, I, N, W, and X) constituted approximately 7% of the total samples. The frequency of the different haplogroups per city is highlighted on the UAE map (Figure 2D). The frequency distribution of the different sub-haplogroups across the different cities is summarized in Table S1. Briefly, 120 sub-haplogroups were detected, where 58 sub-haplogroups (48.3%) were identified in Al Ain, 26 (21.6%) in Fujairah, 26 (21.6%) in RAK, 13 (10.8%) in Dubai, 12 (10%) in Sharjah, and 15 (12.5%) in Abu Dhabi (Table S1). A Venn diagram of the shared and unique sub-haplogroups across the different cities is summarized in Supplementary Figure S1.

3.3. Population Structure and Differentiation

The maximum likelihood phylogenetic tree of the different haplogroups colored by cities (Figure 3A) shows there is no clear population structure based on geographical areas. In addition, it is evident, in conjunction with 70 African mitogenomes, that the sub-haplogroup L3x1b clusters with an Ethiopian sample confirming the ancient step out in the region out of African. Furthermore, we observe admixture among geographical regions in the UAE, which is also observed in previous studies [12,60]; Figure 3B depicts admixture results where a clear admixture event at K = 7 number ancestral population can be seen. The lack of population structure is also addressed using a DAPC analysis (Figure 3C), which demonstrate non-clustering of different cities and geographical regions.

The pairwise population differentiation (Fst) among cities shows no great differentiation with F_{st} varying from 0.009 to 0.07 as it is shown in (Table S2) and the F_{st} boxplot (Figure 4A). We observe a slight differentiation in Alain compared to Abu Dhabi and the other cities. The H_s, H_t, G_{st}, G'_{st}, and D statistics (Tables S3 and S4) (Figure 4B) ($p < 0.05$) confirm that there is no structure and population differentiation among the different cities.
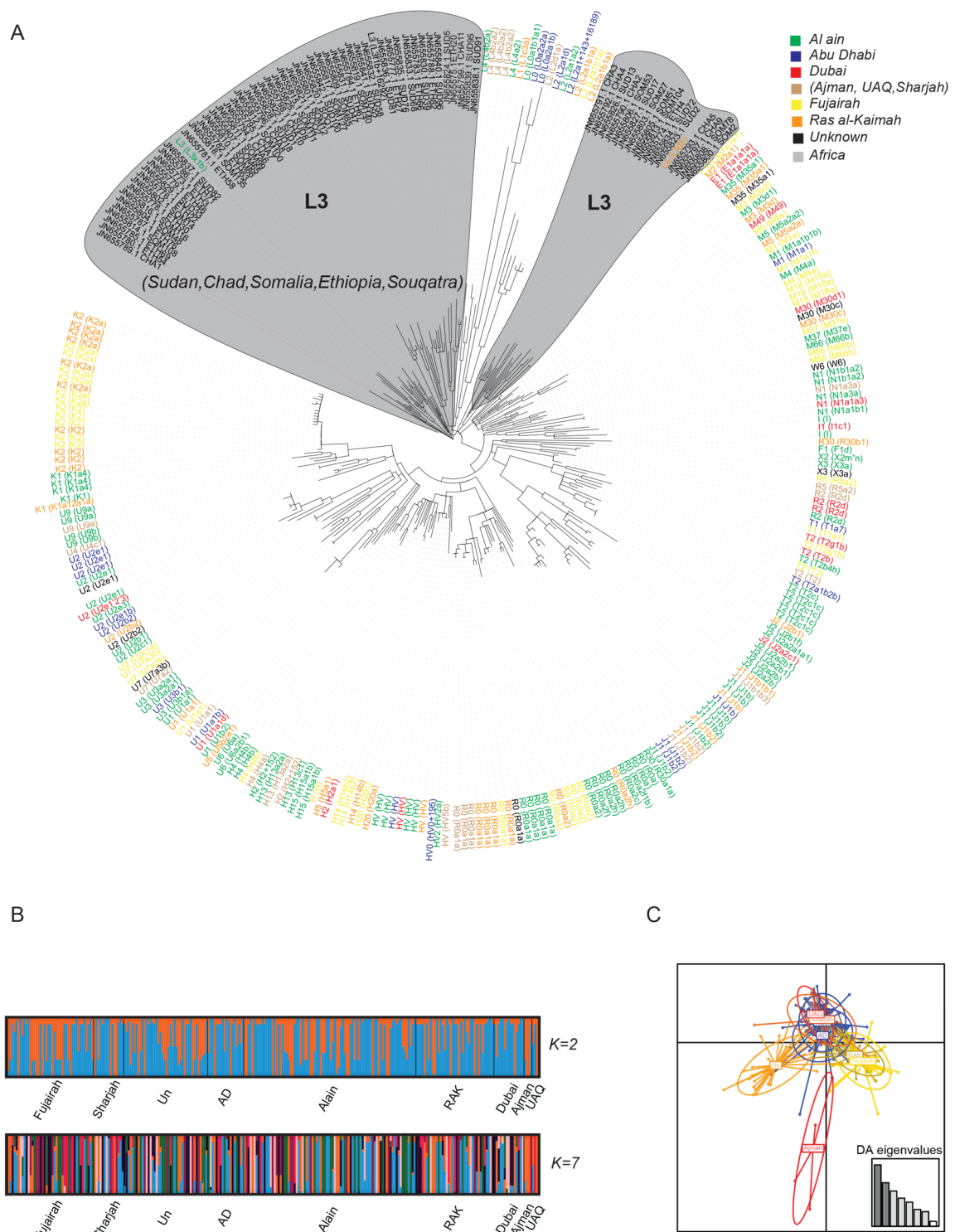
Figure 3. Population structure in UAE population. (**A**) Maximum likelihood phylogenetic tree depicting the network relationship among the different haplogroups (sub-haplogroups), where they are color-coded by city/region. (**B**) Admixture plot for K = 2 and K = 7 ancestral population. (**C**) Discriminant Analysis of Principal Component (DAPC) showing the clustering of the different cities of the UAE population.

3.4. Diversity, Kinship, and Selection

As it is expected, the pairwise diversity estimate (π) is higher in Africa compared to different cities in the UAE population. It is also higher in Abu Dhabi and Dubai compared to Alain, which is a province of Abu Dhabi. In addition, we observe the L3 haplogroup in Africa, and the UAE population ($\pi = 0.004$) has higher diversity compared to M and N haplogroup. It is also higher even when combined the L haplogroup together (Figure 4C). Kinship relationship using KING matrix shows that there is a tapestry of closely related individuals starting from distantly related individuals (DS), first cousins (C1), Half-sibs (HS1), Parent-offspring (PO), and full-sibs (FS) in the population.

Looking at the coding sequences of the 13 different mitochondrial genes, we observe higher Watterson theta (θ) compared to Pi (π), as well as a significant negative Tajima's D (Table 2). Only ATP8 gene shows the signature of selection. We looked at the strength and nature of selection from HyPhy BUSTED for all the coding sequences, ATP8 show signs of diversifying selection with d_N/d_S ratio > 1.
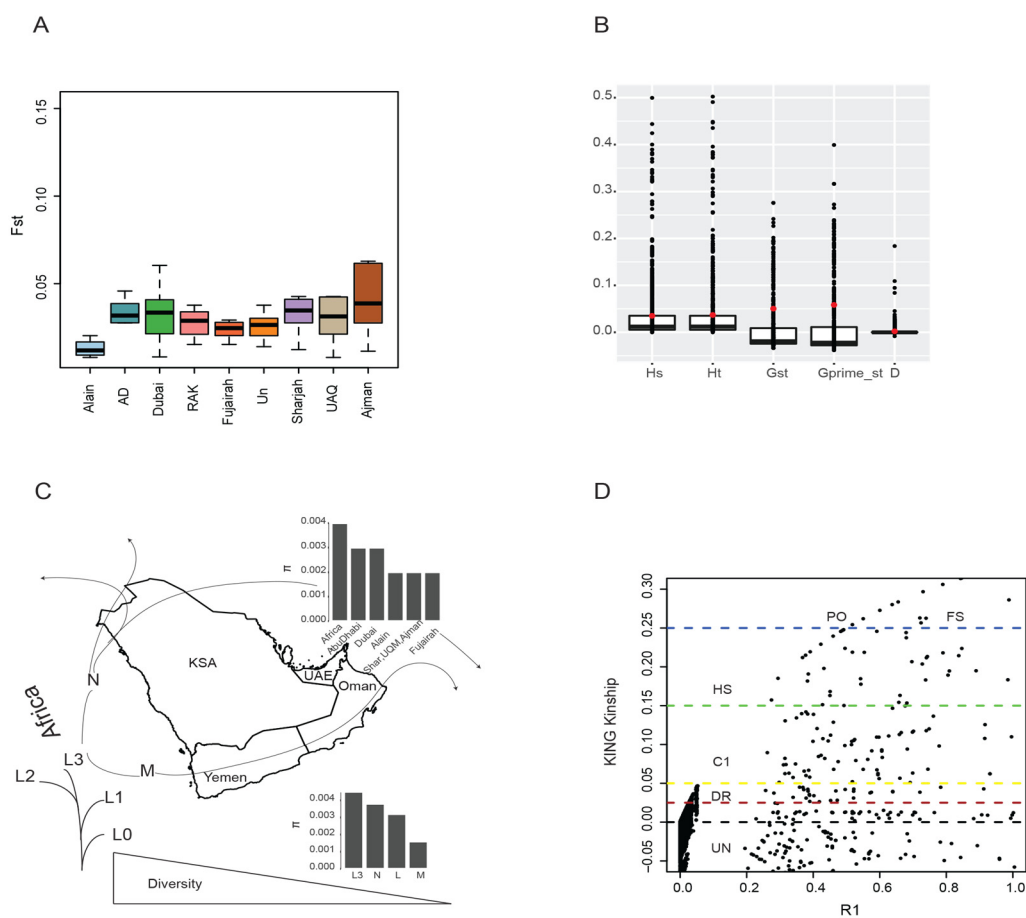

Figure 4. Population differentiation and diversity. (**A**) Population differentiation (Fst) per city of the UAE. (**B**) Population differentiation statistics depicted by Hs, Ht, Gst, G'st, and D. (**C**) A map showing the expected gradient of diversity decrease along the out of Africa passage. Diversity estimates for Africa in comparisons to other UAE cities as well as L haplogroup (and L3) to M and N haplogroups. (**D**) Consanguinity statistics plotted as a KING robust kinship matrix for the different individuals, where the degree of relatedness is shown as UN (unrelated), DS (distantly related), C 1(first cousin), HS (half-sibs), PO (parents-offspring), and FS (full-sibs).

Table 2. Population genetic parameters and the ratio of nonsynonymous to synonymous substitution rate for each mitochondrial gene.

Genes	M	S	Ps	θ	π	D	d_N/d_S
ATP6	232	45	0.067265	0.011244	0.002034	−2.391682	0.29
ATP8	232	17	0.087179	0.014573	0.001415	−2.325058	1.03
COX1	232	75	0.050302	0.008409	0.001305	−2.556932	0.10
COX2	232	34	0.050595	0.008458	0.001166	−2.449534	0.15
COX3	232	40	0.053763	0.008987	0.001757	−2.323988	0.12
CYTB	232	95	0.085818	0.014346	0.002837	−2.457115	0.19
ND1	232	56	0.06041	0.010098	0.002071	−2.361573	0.12
ND2	232	63	0.0625	0.010448	0.00161	−2.533238	0.12
ND3	232	16	0.048048	0.008032	0.003149	−1.548463	0.15
ND4	232	74	0.055306	0.009245	0.002533	−2.195525	0.04
ND4L	232	31	0.050542	0.016855	0.001867	−2.440526	0.06
ND5	232	121	0.068789	0.011499	0.002293	−2.476299	0.14
ND6	232	31	0.060429	0.010102	0.001778	−2.31697	0.08

m: number of sequences. S: Number of segregating sites. Ps: S/*n* (total number of sites). θ: Watterson theta. π: nucleotide diversity. D: Tajima's D. d_N/d_S: ratio of nonsynonymous/synonymous substitution rate.

3.5. Demographic History Reconstruction

BEAST output was used to perform an extended Bayesian skyline plot (EBSP) analysis. We reconstructed the distinct demographic epochs, which highlights a significant and transient contraction in population size some 1400 years before the present (Figure 5). Giving the uncertainty that is associated with the reconstruction method, the plot shows a contraction and reduction for several thousand years and then returned to the level that was before the event. Sd the rate at which lineage coalesces is inversely proportional to population size, our analysis suggests that the bottleneck could affect the different haplogroups disproportionally which is obvious from the branch length distance (Figure 2A).
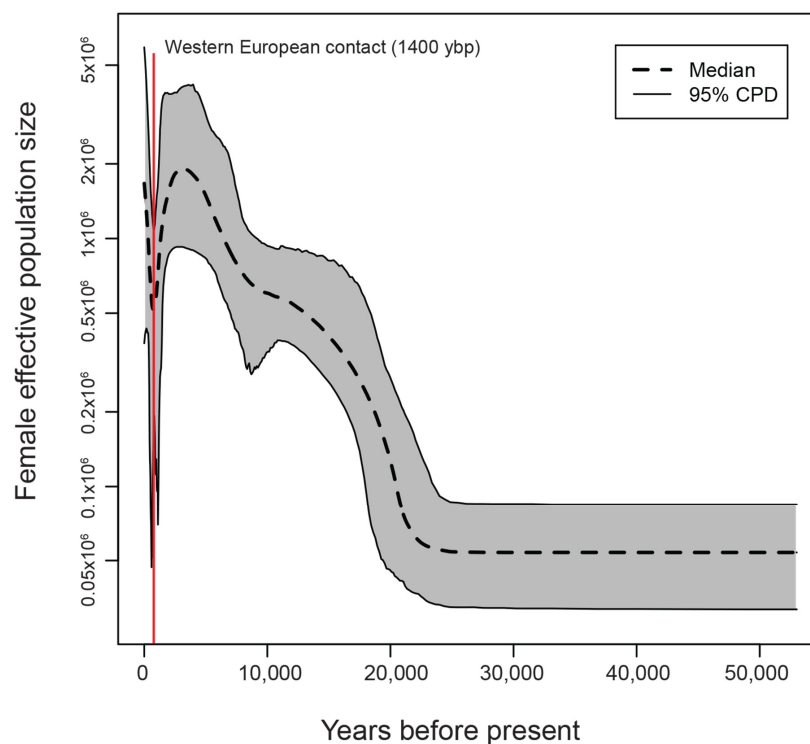
Figure 5. Demographic reconstruction of the UAE mitogenome. Extended Bayesian skyline plot of female effective population size and vertical red line to highlight the bottleneck event that is around 1400 ybp.

3.6. Heteroplasmy and Structural Variation Identification

Heteroplasmy level estimated using mtDNA server shows that D-loop has 305 followed by ND5 276, which is the highest in the genome. These mutations harbored in these genes hold different known pathogenic diseases, which are reported in the human mitogenome. We documented the number as well as the disease associated according to HmtNote (Table S5). There are three mutations in tRNA: two are annotated according to Mitomap with some tumorigenic risk, while one heteroplasmic mutation in tRNA^His gene at position 12172 A->G, with no Mitomap annotation. This mutation is not at higher frequency (1.5%) in the population compared to the others. The RNA fold server was used to predict the change in the secondary structure of the wild type (A) versus the mutant (G). The results highlight a change in the structure in the mutant versus the wild type for only this mutation (Supplementary Figure S2). The eKLIPse results found no significant SV in the 232 different BAMs in the study (Supplementary data zipped file).

4. Discussion

The Arabian Peninsula holds the answers to the out-of-Africa migration and the start of modern human continental genetic diversity and structure. Whether a Levantine terrestrial cross between Africa and Southwest Asia or an East African cross of the Red Sea to the south of Arabian Peninsula and moving eastward [3], archeologists and geneticists are still searching for evidence to favor this route or the other that contributed to human evolution. Although the southern route is the favored option [3], many studies confirm admixture events between the Levant and Arabia, which is likely through the Gulf corridor [61–64].

In our study, the direct descendent (N1, X) of lineage affiliated with L haplogroups suggests an ancient ancestry in the region, which most likely dispersed through the Gulf corridor towards the Levant and Europe 24–55 Ka [65]. Another evidence, in our study, of the Levantine corridor dispersal preference over the horn of Africa is the presence of H, J*, N1b, and T1 haplogroups; other studies, however, confirmed its distribution to be higher in the Levant populations (Iraq, Israeli Druze, Jordan, Palestine, and Syria) compared to Arabian Peninsula groups [11,60,66–70]. In contrast, the M1 haplogroup in our study points to the horn of Africa migration because the frequency of this haplogroup is reported to be high in Ethiopia, low by polymorphic in Yemen and reduced in the Middle East [2,66,68]. Thus, either an Indian or East African origin suggests a favors the horn of Africa route over the Levantine corridor. On the other hand, the distribution of K and HV1 haplogroups of Eurasian origin [67,69] in our study indicate the dispersion by both routes. However, Rowold et al. (2007) showed that the old TMRCA of the star-shaped UAE HV1 network that points to a southern route.

The presence of L1, L2, and L3, in our study, points to sub-Saharan mtDNA lineages. Sub-haplogroup L2a, for example, is associated with the Bantu expansion [71], which is pervasive throughout African and sub-Saharan habitants. However, the presence of the L sub-haplogroup of deep ancestry in the Arabia/Near East reflects its restriction to the horn of African countries (e.g., Ethiopia, Somalia, Sudan, Egypt, etc.). For instance, "ancient" L clades in the Middle East are (L0f2, L0a1d, L0a1c, L1b1a2, L5a1, and L2a1d), and the more prevalent clades are L4, L6, L3i, L3k, L3h, and L3x haplogroups. In contrast, the recently introduced L clades to the Middle East are L0s, L1s, L2s (L2a1a and L2a1b), L3s (L3e, L3b/d's, and L3f L3f1b), excluding the ancient L clades (L0s and L1s). The former appears to be associated with slave trade in the Middle East. We also observed sub-haplogroups L2a1d and L4 as well as L3x1b, which cluster with Ethiopian individuals (Figures 2C and 3A), indicating that it is has been in the Arabia/Near East for a long time. The majority of the sub-Saharan genetic contribution in our population is the product of Arab slave trade, which involved the movement of African slaves through an East African trade route 2500 years ago [72].

The mitochondrial genomic landscape of the UAE mitogenomes has 968 variants and the majority is in the D-loop and ND5 gene (Figure 1D). This is expected as the D-loop accumulates more mutation, as it does not encode any specific protein. As for ND5, the accumulation of synonymous variants is not surprising given the high rate of the heteroplasmic allele in ND5 [73] (Table S5). Consanguinity is

high in the UAE, which increases the incidence of recessive genetic disorder at the genomic level and might affect the mitochondria due to the genetic interaction between nucleus and mitochondria [74]. Heteroplasmy mutations are involved in cancer and aging, but they are also common in healthy humans, and one allele frequency can change over generations (mother-to-child) because of the bottleneck effect. The bottleneck effect can shift the ratio of alleles in a heteroplasmic mitochondrion causing a generational-dependent disease prevalence [75]. In our study, heteroplasmy is more pronounced in D-loop, ND5, ND4, and CYB, and it is associated with known mapped diseases such as cyclic vomiting syndrome; CPEO/Stroke/CM/breast, renal, and prostate cancer risk/altered brain pH/sCJD (Sporadic Creutzfeldt–Jakob disease), and LHON; PD protective factor/longevity/altered cell pH/metabolic syndrome/breast cancer risk/LS risk/ADHD/cognitive decline; and primary open-angle glaucoma (POAG) (Table S5). In addition, the heteroplasmic mutation in tRNA^{His} reported previously [76] is associated with lung cancer. Further studies are required to elucidate the generation-dependent allelic ratio in heteroplasmy cases. This knowledge is useful especially in genetic counseling and diagnostics among UAE citizens especially in cases involving close relative marriages [77,78].

As it is expected, the mitogenome harbors more synonymous mutations than nonsynonymous mutations (Figure 1A), which suggests a strong purifying selection at purging deleterious mutations to maintain fully functioning mitochondria. This is supported by the allele frequency spectrum that is skewed toward low frequency of polymorphism and the diversity estimate theta $(\theta) > $ Pi (π), a negative Tajima's D and dN/dS < 1 of 12 mitochondrial genes. In contrast, ATP8 shows a signature of diversifying selection, with dN/dS = 1.03 (Table 2) and significant departure from neutrality. One explanation could be that the regional distribution of haplogroups in Eurasia and Africa has been shaped by natural selection on the oxidative phosphorylation pathway in response to change in the climate conditions (e.g., from a cool/warm environment to a hot arid environment) [79].

As expected, higher diversity is observed in Africa (0.004 ± 0.00279) than UAE (0.0021 ± 0.0012) to (0.0033 + 0.00310). The lower diversity is consistent with the paucity of population differentiation and structure as shown in the Fst for the UAE population, which is also consistent with the genomic result from the UAE population that was recently published [80]. The haplogroup diversity, especially the African L (0.0037 ± 0.00302) and non-African M and N mtDNA, shows that L3 has higher diversity (0.004 ± 0.00279), whereas M (Tajima's D= −1.024, *p* < 0.01) and N (Tajima's D = −0.908, *p* < 0.01) analyzed separately, show low diversity (0.0016 ± 0.00162; 0.0021 ± 0.0022) and significant deviation from neutrality (Table 3), which is consistent with population expansion out of Africa that distorted the frequency of the mtDNA variants [81,82].

Table 3. Population genetic parameters per cities and haplogroups.

City/Haplogroup	π	D
Abu Dhabi	0.0033 ± 0.00310	−0.9003 ± 0.67193
Al Ain	0.0022 ± 0.00241	−1.0783 ± 0.59466
Dubai, Ajman, UQA	0.0029 ± 0.00264	−0.8337 ± 0.54551
Fujairah	0.0023 ± 0.00258	−1.0999 ± 0.74124
RAK	0.0021 ± 0.00237	−0.8034 ± 0.62341
Sharjah	0.0024 ± 0.00211	−0.9012 ± 0.53421
UAQ	0.0021 ± 0.0012	−0.5231 ± 0.32310
Africa (L3)	0.004 ± 0.00279	−1.7485 ± 0.38859
L	0.0037 ± 0.00302	−0.8079 ± 0.70818
M	0.0016 ± 0.00162	−1.0248 ± 0.56632
N	0.0021 ± 0.00221	−0.9080 ± 0.85152

π: nucleotide diversity. D: Tajima's D.

The reconstructed demographic history of the UAE mitogenome sheds light on a bottleneck event around 1400 years ago that coincides with western European contact (Figure 5). The eastern Mediterranean region witnessed Crusader settlements between 11th and 13th centuries that could create immense genetic drift and bottleneck effect in the introduction of western European lineages

into the Levant, which will affect a big portion of today's gene pool [83]. This western European gene background expanded to the eastern Arabian Peninsula, where the influence of the Portuguese was eminent in major parts for the following 150 years.

5. Conclusions

This study describes the genomic landscape of the UAE mitochondrial genome and the distribution of haplogroups in different geographic regions in the UAE. The analyzed mitogenomes from 232 female students of UAE University, aged 18–24 years, highlights the high resolution of 15 different haplogroups that share ancestry with Africa, East Asia, and the Near East. Furthermore, it elucidates migration routes to the UAE. The low diversity and population differentiation highlight that the low movement between cities. The Demographic history highlights a bottleneck event that coincides with European contact 1400 ybp. In conclusion, this study also provides a matrilineal history of the UAE and will serve as an asset for genetic counseling, forensic science, and anthropology among other fields.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/8/876/s1, Table S1. The haplogroups/sub-haplogroups distribution per city in the UAE population. Table S2. Pairwise Fst among the different cities of the UAE population. Table S3. Summary statistics of Hs and Ht for different cities of the UAE population. Table S4. Summary of the statistic of Gst, Htmax, Gstmax, and G'st. Table S5. Summary of filtered heteroplasmy counts and disease associated annotation. Figure S1. Venn diagram of the different sub-haplogroups among cities of the UAE. Figure S2. Prediction of the MFE structure tRNAHis gene in wild type and mutant (A12172G). Supplementary data zipped. Zipped plots from eKLPse software for the 232 individuals long structural variants.

References

1. Ingman, M.; Kaessmann, H.; Pääbo, S.; Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **2000**, *408*, 708–713. [CrossRef] [PubMed]
2. Maca-Meyer, N.; González, A.M.; Larruga, J.M.; Flores, C.; Cabrera, V.M. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.* **2001**, *2*, 13. [CrossRef] [PubMed]
3. Macaulay, V.; Hill, C.; Achilli, A.; Rengo, C.; Clarke, D.; Meehan, W.; Blackburn, J.; Semino, O.; Scozzari, R.; Cruciani, F. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **2005**, *308*, 1034–1036. [CrossRef]
4. Kivisild, T.; Rootsi, S.; Metspalu, M.; Mastana, S.; Kaldma, K.; Parik, J.; Metspalu, E.; Adojaan, M.; Tolk, H.-V.; Stepanov, V. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* **2003**, *72*, 313–332. [CrossRef]
5. Thangaraj, K.; Chaubey, G.; Singh, V.K.; Vanniarajan, A.; Thanseem, I.; Reddy, A.G.; Singh, L. In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup'M'in India. *BMC Genom.* **2006**, *7*, 1–6. [CrossRef]
6. Merriwether, D.A.; Hodgson, J.A.; Friedlaender, F.R.; Allaby, R.; Cerchio, S.; Koki, G.; Friedlaender, J.S. Ancient mitochondrial M haplogroups identified in the Southwest Pacific. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13034–13039. [CrossRef]
7. Hudjashov, G.; Kivisild, T.; Underhill, P.A.; Endicott, P.; Sanchez, J.J.; Lin, A.A.; Shen, P.; Oefner, P.; Renfrew, C.; Villems, R. Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 8726–8730. [CrossRef]

8.   Forster, P. Ice Ages and the mitochondrial DNA chronology of human dispersals: A review. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **2004**, *359*, 255–264. [CrossRef]

9.   Forster, P.; Torroni, A.; Renfrew, C.; Röhl, A. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol. Biol. Evol.* **2001**, *18*, 1864–1881. [CrossRef] [PubMed]

10.   Oppenheimer, S. *Out of Eden: The Peopling of the World*; Jonathan Ball Publishers: Johannesburg, South Africa, 2012.

11.   Kivisild, T.; Reidla, M.; Metspalu, E.; Rosa, A.; Brehm, A.; Pennarun, E.; Parik, J.; Geberhiwot, T.; Usanga, E.; Villems, R. Ethiopian mitochondrial DNA heritage: Tracking gene flow across and around the gate of tears. *Am. J. Hum. Genet.* **2004**, *75*, 752–770. [CrossRef] [PubMed]

12.   Abu-Amero, K.K.; Gonzalez, A.M.; Larruga, J.M.; Bosley, T.M.; Cabrera, V.M. Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. *BMC Evol. Biol.* **2007**, *7*, 32. [CrossRef] [PubMed]

13.   Rowold, D.; Luis, J.; Terreros, M.; Herrera, R.J. Mitochondrial DNA geneflow indicates preferred usage of the Levant Corridor over the Horn of Africa passageway. *J. Hum. Genet.* **2007**, *52*, 436–447. [CrossRef] [PubMed]

14.   Pilkington, M.M.; Wilder, J.A.; Mendez, F.L.; Cox, M.P.; Woerner, A.; Angui, T.; Kingan, S.; Mobasher, Z.; Batini, C.; Destro-Bisol, G. Contrasting signatures of population growth for mitochondrial DNA and Y chromosomes among human populations in Africa. *Mol. Biol. Evol.* **2008**, *25*, 517–525. [CrossRef]

15.   Quintana-Murci, L.; Quach, H.; Harmant, C.; Luca, F.; Massonnet, B.; Patin, E.; Sica, L.; Mouguiama-Daouda, P.; Comas, D.; Tzur, S. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter–gatherers and Bantu-speaking farmers. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1596–1601. [CrossRef] [PubMed]

16.   Reed, F.A.; Tishkoff, S.A. African human diversity, origins and migrations. *Curr. Opin. Genet. Dev.* **2006**, *16*, 597–605. [CrossRef]

17.   Tishkoff, S.A.; Gonder, M.K.; Henn, B.M.; Mortensen, H.; Knight, A.; Gignoux, C.; Fernandopulle, N.; Lema, G.; Nyambo, T.B.; Ramakrishnan, U. History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol. Biol. Evol.* **2007**, *24*, 2180–2195. [CrossRef] [PubMed]

18.   Wood, E.T.; Stover, D.A.; Ehret, C.; Destro-Bisol, G.; Spedini, G.; McLeod, H.; Louie, L.; Bamshad, M.; Strassmann, B.I.; Soodyall, H. Contrasting patterns of Y chromosome and mtDNA variation in Africa: Evidence for sex-biased demographic processes. *Eur. J. Hum. Genet.* **2005**, *13*, 867–876. [CrossRef]

19.   Campbell, M.C.; Tishkoff, S.A. African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genom. Hum. Genet.* **2008**, *9*, 403–433. [CrossRef]

20.   Shtolz, N.; Mishmar, D. The mitochondrial genome–on selective constraints and signatures at the organism, cell, and single mitochondrion levels. *Front. Ecol. Evol.* **2019**, *7*, 342. [CrossRef]

21.   Elson, J.L.; Andrews, R.M.; Chinnery, P.F.; Lightowlers, R.N.; Turnbull, D.M.; Howell, N. Analysis of European mtDNAs for recombination. *Am. J. Hum. Genet.* **2001**, *68*, 145–153. [CrossRef]

22.   De Fanti, S.; Vicario, S.; Lang, M.; Simone, D.; Magli, C.; Luiselli, D.; Gianaroli, L.; Romeo, G. Intra-individual purifying selection on mitochondrial DNA variants during human oogenesis. *Hum. Reprod.* **2017**, *32*, 1100–1107. [CrossRef] [PubMed]

23.   Stewart, J.B.; Freyer, C.; Elson, J.L.; Wredenberg, A.; Cansu, Z.; Trifunovic, A.; Larsson, N.G. Strong purifying selection in transmission of mammalian mitochondrial DNA. *PLoS Biol.* **2008**, *6*, e10. [CrossRef] [PubMed]

24.   Bannwarth, S.; Procaccio, V.; Lebre, A.S.; Jardel, C.; Chaussenot, A.; Hoarau, C.; Maoulida, H.; Charrier, N.; Gai, X.; Xie, H.M.; et al. Prevalence of rare mitochondrial DNA mutations in mitochondrial disorders. *J. Med. Genet.* **2013**, *50*, 704–714. [CrossRef] [PubMed]

25.   Wallace, D.C.; Chalkia, D. Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a021220. [CrossRef] [PubMed]

26.   Wang, J.; Schmitt, E.S.; Landsverk, M.L.; Zhang, V.W.; Li, F.Y.; Graham, B.H.; Craigen, W.J.; Wong, L.J. An integrated approach for classifying mitochondrial DNA variants: One clinical diagnostic laboratory's experience. *Genet. Med.* **2012**, *14*, 620–626. [CrossRef]

27.   Damas, J.; Samuels, D.C.; Carneiro, J.; Amorim, A.; Pereira, F. Mitochondrial DNA rearrangements in health and disease—A comprehensive study. *Hum. Mutat.* **2014**, *35*, 1–14. [CrossRef]

28.   Neiman, M.; Taylor, D.R. The causes of mutation accumulation in mitochondrial genomes. *Proc. R. Soc. B Biol. Sci.* **2009**, *276*, 1201–1209. [CrossRef] [PubMed]

29.　Holt, I.J.; Harding, A.E.; Morgan-Hughes, J.A. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature* **1988**, *331*, 717–719. [CrossRef]

30.　Wallace, D.C.; Zheng, X.X.; Lott, M.T.; Shoffner, J.M.; Hodge, J.A.; Kelley, R.I.; Epstein, C.M.; Hopkins, L.C. Familial mitochondrial encephalomyopathy (MERRF): Genetic, pathophysiological, and biochemical characterization of a mitochondrial DNA disease. *Cell* **1988**, *55*, 601–610. [CrossRef]

31.　Calloway, C.D.; Reynolds, R.L.; Herrin, G.L., Jr.; Anderson, W.W. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *Am. J. Hum. Genet.* **2000**, *66*, 1384–1397. [CrossRef]

32.　He, Y.; Wu, J.; Dressman, D.C.; Iacobuzio-Donahue, C.; Markowitz, S.D.; Velculescu, V.E.; Diaz, L.A., Jr.; Kinzler, K.W.; Vogelstein, B.; Papadopoulos, N. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature* **2010**, *464*, 610–614. [CrossRef] [PubMed]

33.　Santos, C.; Montiel, R.; Sierra, B.; Bettencourt, C.; Fernandez, E.; Alvarez, L.; Lima, M.; Abade, A.; Aluja, M.P. Understanding differences between phylogenetic and pedigree-derived mtDNA mutation rate: A model using families from the Azores Islands (Portugal). *Mol. Biol. Evol.* **2005**, *22*, 1490–1505. [CrossRef] [PubMed]

34.　Santos, C.; Sierra, B.; Alvarez, L.; Ramos, A.; Fernandez, E.; Nogues, R.; Aluja, M.P. Frequency and pattern of heteroplasmy in the control region of human mitochondrial DNA. *J. Mol. Evol.* **2008**, *67*, 191–200. [CrossRef] [PubMed]

35.　Kirches, E.; Michael, M.; Warich-Kirches, M.; Schneider, T.; Weis, S.; Krause, G.; Mawrin, C.; Dietzmann, K. Heterogeneous tissue distribution of a mitochondrial DNA polymorphism in heteroplasmic subjects without mitochondrial disorders. *J. Med Genet.* **2001**, *38*, 312–317. [CrossRef] [PubMed]

36.　Irwin, J.A.; Saunier, J.L.; Niederstätter, H.; Strouss, K.M.; Sturk, K.A.; Diegoli, T.M.; Brandstätter, A.; Parson, W.; Parsons, T.J. Investigation of heteroplasmy in the human mitochondrial DNA control region: A synthesis of observations from more than 5000 global population samples. *J. Mol. Evol.* **2009**, *68*, 516–527. [CrossRef] [PubMed]

37.　Al-Gazali, L.; Ali, B.R. Mutations of a country: A mutation review of single gene disorders in the United Arab Emirates (UAE). *Hum. Mutat.* **2010**, *31*, 505–520. [CrossRef]

38.　Al-Jasmi, F.A.; Tawfig, N.; Berniah, A.; Ali, B.R.; Taleb, M.; Hertecant, J.L.; Bastaki, F.; Souid, A.K. Prevalence and Novel Mutations of Lysosomal Storage Disorders in United Arab Emirates: LSD in UAE. *JIMD Rep.* **2013**, *10*, 1–9. [CrossRef]

39.　Al-Shamsi, A.; Hertecant, J.L.; Al-Hamad, S.; Souid, A.K.; Al-Jasmi, F. Mutation Spectrum and Birth Prevalence of Inborn Errors of Metabolism among Emiratis: A study from Tawam Hospital Metabolic Center, United Arab Emirates. *Sultan Qaboos Univ. Med. J.* **2014**, *14*, e42–e49. [CrossRef]

40.　Fendt, L.; Zimmermann, B.; Daniaux, M.; Parson, W. Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences. *BMC Genom.* **2009**, *10*, 139. [CrossRef]

41.　Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef]

42.　Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef] [PubMed]

43.　Bragg, L.M.; Stone, G.; Butler, M.K.; Hugenholtz, P.; Tyson, G.W. Shining a light on dark sequencing: Characterising errors in Ion Torrent PGM data. *PLoS Comput. Biol.* **2013**, *9*, e1003031. [CrossRef] [PubMed]

44.　McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef] [PubMed]

45.　Preste, R.; Clima, R.; Attimonelli, M. Human mitochondrial variant annotation with HmtNote. *BioRxiv* **2019**, 600619. [CrossRef]

46.　Team, R.C. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.

47.　Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [CrossRef] [PubMed]

48.　Rambaut, A. 2006–2012. Fig. Tree. Tree Figure Drawing Tool, Version 1.4. 0. University of Edinburgh: Institute of Evolutionary Biology [on-line]. Available online: http://tree.bio.ed.ac.uk/ (accessed on 30 January 2015).

49. Weissensteiner, H.; Pacher, D.; Kloss-Brandstatter, A.; Forer, L.; Specht, G.; Bandelt, H.J.; Kronenberg, F.; Salas, A.; Schonherr, S. HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **2016**, *44*, W58–W63. [CrossRef]

50. Jombart, T. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **2008**, *24*, 1403–1405. [CrossRef]

51. Jombart, T.; Devillard, S.; Balloux, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **2010**, *11*, 94. [CrossRef] [PubMed]

52. Kozlov, A.M.; Darriba, D.; Flouri, T.; Morel, B.; Stamatakis, A. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **2019**, *35*, 4453–4455. [CrossRef]

53. Alexander, D.H.; Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **2011**, *12*, 246. [CrossRef]

54. Jakobsson, M.; Rosenberg, N.A. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **2007**, *23*, 1801–1806. [CrossRef] [PubMed]

55. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **1989**, *123*, 585–595. [PubMed]

56. Drummond, A.J.; Suchard, M.A.; Xie, D.; Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **2012**, *29*, 1969–1973. [CrossRef] [PubMed]

57. Heled, J.; Drummond, A.J. Bayesian inference of population size history from multiple loci. *BMC Evol. Biol.* **2008**, *8*, 289. [CrossRef] [PubMed]

58. Weissensteiner, H.; Forer, L.; Fuchsberger, C.; Schopf, B.; Kloss-Brandstatter, A.; Specht, G.; Kronenberg, F.; Schonherr, S. mtDNA-Server: Next-generation sequencing data analysis of human mitochondrial DNA in the cloud. *Nucleic Acids Res.* **2016**, *44*, W64–W69. [CrossRef]

59. Goudenège, D.; Bris, C.; Hoffmann, V.; Desquiret-Dumas, V.; Jardel, C.; Rucheton, B.; Bannwarth, S.; Paquis-Flucklinger, V.; Lebre, A.S.; Colin, E. eKLIPse: A sensitive tool for the detection and quantification of mitochondrial DNA deletions from next-generation sequencing data. *Genet. Med.* **2019**, *21*, 1407–1416. [CrossRef] [PubMed]

60. Fernandes, V.; Brucato, N.; Ferreira, J.C.; Pedro, N.; Cavadas, B.; Ricaut, F.X.; Alshamali, F.; Pereira, L. Genome-Wide Characterization of Arabian Peninsula Populations: Shedding Light on the History of a Fundamental Bridge between Continents. *Mol. Biol. Evol.* **2019**, *36*, 575–586. [CrossRef] [PubMed]

61. Černý, V.; Mulligan, C.J.; Fernandes, V.; Silva, N.M.; Alshamali, F.; Non, A.; Harich, N.; Cherni, L.; El Gaaied, A.B.A.; Al-Meeri, A. Internal diversification of mitochondrial haplogroup R0a reveals post-last glacial maximum demographic expansions in South Arabia. *Mol. Biol. Evol.* **2011**, *28*, 71–78. [CrossRef] [PubMed]

62. Al-Abri, A.; Podgorná, E.; Rose, J.I.; Pereira, L.; Mulligan, C.J.; Silva, N.M.; Bayoumi, R.; Soares, P.; Černý, V. Pleistocene-Holocene boundary in Southern Arabia from the perspective of human mtDNA variation. *Am. J. Phys. Anthropol.* **2012**, *149*, 291–298. [CrossRef]

63. Fernandes, V.; Triska, P.; Pereira, J.B.; Alshamali, F.; Rito, T.; Machado, A.; Fajkošová, Z.; Cavadas, B.; Černý, V.; Soares, P. Genetic stratigraphy of key demographic events in Arabia. *PLoS ONE* **2015**, *10*, e0118625. [CrossRef]

64. Vyas, D.N.; Kitchen, A.; Miro-Herrans, A.T.; Pearson, L.N.; Al-Meeri, A.; Mulligan, C.J. Bayesian analyses of Yemeni mitochondrial genomes suggest multiple migration events with Africa and Western Eurasia. *Am. J. Phys. Anthropol.* **2016**, *159*, 382–393. [CrossRef]

65. Fernandes, V.; Alshamali, F.; Alves, M.; Costa, M.D.; Pereira, J.B.; Silva, N.M.; Cherni, L.; Harich, N.; Cerny, V.; Soares, P.; et al. The Arabian cradle: Mitochondrial relicts of the first steps along the southern route out of Africa. *Am. J. Hum. Genet.* **2012**, *90*, 347–355. [CrossRef] [PubMed]

66. Macaulay, V.; Richards, M.; Hickey, E.; Vega, E.; Cruciani, F.; Guida, V.; Scozzari, R.; Bonne-Tamir, B.; Sykes, B.; Torroni, A. The emerging tree of West Eurasian mtDNAs: A synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* **1999**, *64*, 232–249. [CrossRef] [PubMed]

67. Richards, M.; Macaulay, V.; Hickey, E.; Vega, E.; Sykes, B.; Guida, V.; Rengo, C.; Sellitto, D.; Cruciani, F.; Kivisild, T.; et al. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* **2000**, *67*, 1251–1276. [CrossRef]

68. Al-Zahery, N.; Semino, O.; Benuzzi, G.; Magri, C.; Passarino, G.; Torroni, A.; Santachiara-Benerecetti, A.S. Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Mol. Phylogenet. Evol.* **2003**, *28*, 458–472. [CrossRef]

69. Achilli, A.; Rengo, C.; Magri, C.; Battaglia, V.; Olivieri, A.; Scozzari, R.; Cruciani, F.; Zeviani, M.; Briem, E.; Carelli, V. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am. J. Hum. Genet.* **2004**, *75*, 910–918. [CrossRef]

70. Luis, J.R.; Rowold, D.J.; Regueiro, M.; Caeiro, B.; Cinnioglu, C.; Roseman, C.; Underhill, P.A.; Cavalli-Sforza, L.L.; Herrera, R.J. The Levant versus the Horn of Africa: Evidence for bidirectional corridors of human migrations. *Am. J. Hum. Genet.* **2004**, *74*, 532–544. [CrossRef]

71. Salas, A.; Richards, M.; De la Fe, T.; Lareu, M.V.; Sobrino, B.; Sanchez-Diz, P.; Macaulay, V.; Carracedo, A. The making of the African mtDNA landscape. *Am. J. Hum. Genet.* **2002**, *71*, 1082–1111. [CrossRef]

72. Richards, M.; Rengo, C.; Cruciani, F.; Gratrix, F.; Wilson, J.F.; Scozzari, R.; Macaulay, V.; Torroni, A. Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. *Am. J. Hum. Genet.* **2003**, *72*, 1058–1064. [CrossRef]

73. Just, R.S.; Irwin, J.A.; Parson, W. Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. *Forensic Sci. Int. Genet.* **2015**, *18*, 131–139. [CrossRef]

74. Rogell, B.; Dean, R.; Lemos, B.; Dowling, D.K. Mito-nuclear interactions as drivers of gene movement on and off the X-chromosome. *BMC Genom.* **2014**, *15*, 330. [CrossRef] [PubMed]

75. Zaidi, A.A.; Wilton, P.R.; Su, M.S.; Paul, I.M.; Arbeithuber, B.; Anthony, K.; Nekrutenko, A.; Nielsen, R.; Makova, K.D. Bottleneck and selection in the germline and maternal age influence transmission of mitochondrial DNA in human pedigrees. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 25172–25178. [CrossRef] [PubMed]

76. Wang, L.; Chen, Z.J.; Zhang, Y.K.; Le, H.B. The role of mitochondrial tRNA mutations in lung cancer. *Int. J. Clin. Exp. Med.* **2015**, *8*, 13341–13346. [PubMed]

77. Al-Gazali, L.I.; Dawodu, A.H.; Sabarinathan, K.; Varghese, M. The profile of major congenital abnormalities in the United Arab Emirates (UAE) population. *J. Med. Genet.* **1995**, *32*, 7–13. [CrossRef]

78. Fahmy, N.; Benson, P.; Al-Garrah, D. Consanguinity in UAE: Prevalence and analysis of some risk factors. *Emirates Med. J.* **1993**, *1*, 39–41.

79. Mishmar, D.; Ruiz-Pesini, E.; Golik, P.; Macaulay, V.; Clark, A.G.; Hosseini, S.; Brandon, M.; Easley, K.; Chen, E.; Brown, M.D.; et al. Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 171–176. [CrossRef]

80. Tay, G.K.; Henschel, A.; Daw Elbait, G.; Al Safar, H.S. Genetic Diversity and Low Stratification of the Population of the United Arab Emirates. *Front. Genet.* **2020**, *11*, 608. [CrossRef]

81. Merriwether, D.A.; Clark, A.G.; Ballinger, S.W.; Schurr, T.G.; Soodyall, H.; Jenkins, T.; Sherry, S.T.; Wallace, D.C. The structure of human mitochondrial DNA variation. *J. Mol. Evol.* **1991**, *33*, 543–555. [CrossRef]

82. Schurr, T.G.; Wallace, D.C. Mitochondrial DNA diversity in Southeast Asian populations. *Hum. Biol.* **2002**, *74*, 431–452. [CrossRef]

83. Zalloua, P.A.; Xue, Y.; Khalife, J.; Makhoul, N.; Debiane, L.; Platt, D.E.; Royyuru, A.K.; Herrera, R.J.; Hernanz, D.F.S.; Blue-Smith, J. Y-chromosomal diversity in Lebanon is structured by recent historical events. *Am. J. Hum. Genet.* **2008**, *82*, 873–882. [CrossRef]