Animal Biotechnology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/labt20

The usefulness of maternally inherited genetic markers for phylogeographic studies in village chicken

Ahmed S. Al-Jumaili & Olivier Hanotte

To cite this article: Ahmed S. Al-Jumaili & Olivier Hanotte (2022): The usefulness of maternally inherited genetic markers for phylogeographic studies in village chicken, Animal Biotechnology, DOI: 10.1080/10495398.2021.2000429

To link to this article: https://doi.org/10.1080/10495398.2021.2000429

ANIMAL BIOTECHNOLOGY
https://doi.org/10.1080/10495398.2021.2000429

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

The usefulness of maternally inherited genetic markers for phylogeographic studies in village chicken

Ahmed S. Al-Jumaili[a,b] and Olivier Hanotte[b,c]

[a]Medical Laboratory Techniques Department, Al-Maarif University College, Anbar, Iraq; [b]School of Life Sciences, The University of Nottingham, University Park, Nottingham, UK; [c]LiveGene–CTLGH, International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia

ABSTRACT

Phylogeography plays a major role in understanding micro and macroevolutionary processes dealing with evolutionary interpretations of geographical distribution. This field integrates information from molecular genetics, population genetics, demography, and phylogeny for the interpretation of the geographical distribution of lineages. The full mtDNA sequence and W chromosome polymorphisms were exploited to assess the usefulness of two maternally-inherited genetic markers for phylogeographic studies of village chickens. We studied 243 full mtDNA sequences from three countries (Iraq, $n = 27$; Ethiopia, $n = 211$; and Saudi Arabia, $n = 5$) and a 13-kb fragment of the W chromosome from 20 Iraqi and 137 Ethiopian female chickens. The results show a high level of genetic diversity for the mtDNA within and among countries as well as within populations. On the other hand, sequence analysis of the W chromosome shows low genetic diversity both within and among populations. Six full mtDNA haplogroups (A, B, C1, C2, D1, and E1) were observed and 25 distinct W haplotypes. The results support the effectiveness of full mtDNA sequences but not the W chromosome in tracing the maternal historical genome background with, however, weak within a country phylogeographic signal.

KEYWORDS

Full mtDNA; W chromosome; phylogeography; village chicken

Introduction

Phylogeography is widely known as the study of the geographic distribution of genetic lineages among populations within species and closely related species.[1] It has grown rapidly in the last three decades since the word 'Phylogeography' was coined by Avise et al.[2] It combines different subject areas to interpret the geographic distribution patterns of genetic markers and by extension populations or species.[3] It can infer the evolutionary background of a population, including the timing of colonization, the geographic origin, speciation events, and population structure and geographical dispersal between regions.[4–8]

Maternally inherited genetic markers are widely used to explore the evolutionary history of animals.[1,6,9] Mitochondrial DNA (mtDNA) is utilized in most evolutionary studies, being nearly exclusively inherited from the maternal side, with no recombination and rapid mutation rates.[9,10] The mtDNA D-loop diversity has been exploited extensively in analyzing the history of domestic chickens, with findings, such as the identification of candidate domestication geographic areas and dispersion routes.[11,12] In chicken, the first study reporting the analysis of the full mtDNA genome is the one from Miao et al.,[13] which attempted to solve, unsuccessfully, the ongoing debate about the center(s) of origin of the domestic chicken. Later, Di Lorenzo et al.[14] confirmed that mtDNA could be helpful in providing information about human-mediated dispersal of chicken from candidate domestication center(s).

The W chromosome has some features in common with mtDNA. They are both maternally inherited (with female chicken being the heterozygous ZW sex), and they represent haploid non-recombining parts of the genome, apart from the pseudo-autosomal regions (PAR) of the W chromosome.[15,16] Smeds et al.[16] mentioned that the W chromosome and mtDNA both offer a stable maternally inherited genome that can be used in phylogenetic studies. Using many resequenced samples

CONTACT Ahmed S. Al-Jumaili ahmedaljumiliy@gmail.com Medical Laboratory Techniques, Al-Maarif University College, Anbar, Iraq; Olivier Hanotte olivier.hanotte@nottingham.ac.uk School of Life Sciences, University of Nottingham, UK.

Supplemental data for this article is available online at https://doi.org/10.1080/10495398.2021.2000429.

of flycatcher birds, they also show that the W chromosome has a reduced level of genetic diversity within species and a higher level of diversity between species. This reduced genetic variation within species in the W chromosome was attributed to the Hill-Robertson effect (i.e., the effect of a long evolutionary process in the absence of recombination) that decreases diversity due to the effects of linked selection. No phylogeographic study in chicken has attempted so far to use W chromosome sequences as genetic markers.

In this study, we assessed if the analysis of the full mitochondrial DNA genome allows us a deeper understanding of the distribution of domestic chicken populations compared to the *D*-loop,[17,18] through the analysis of 243 new full mitochondrial DNA genome sequences from Ethiopia, Iraq, and Saudi Arabia. In addition, we assess the usefulness of W chromosome polymorphisms as a maternally inherited genetic marker for the understanding of the diversity of domestic chicken in two countries (Ethiopia and Iraq).

Materials and methods

Collection of samples and genomic DNA isolation

For the full mtDNA analysis, 243 samples were analyzed from three countries (Fig. 1), Iraq ($n = 27$), Ethiopia ($n = 211$), and Saudi Arabia ($n = 5$). The Saudi Arabia full genome sequences were downloaded from the NCBI database (accession no. SRP142580).[19] For the W chromosome, the same female birds included in the mtDNA work were analyzed. The Iraqi samples ($n = 20$) were collected from Baghdad ($n = 11$) and Karbala ($n = 3$) representing the Central region ($n = 14$) of the country, and Sulimania representing the North-Eastern part ($n = 6$). Ethiopian samples ($n = 137$) were collected from 19 different geographical areas representing populations from both highland and lowland areas of the country. The Saudi samples, which had been collected from a region in the East part of the country, were not included here because there were only males. The extraction of DNA was done accordingly to the protocols described in Al-Jumaili et al.[17]

Sequencing, alignment, and variants calling

After checking the DNA quality for whole-genome sequencing, samples were sent to Edinburgh Genomics (https://www.ed.ac.uk/roslin/facilities-resources/ark-genomics) at the Roslin Institute (Edinburgh University) for Next Generation Sequence (NGS) data generation as part of a bigger project

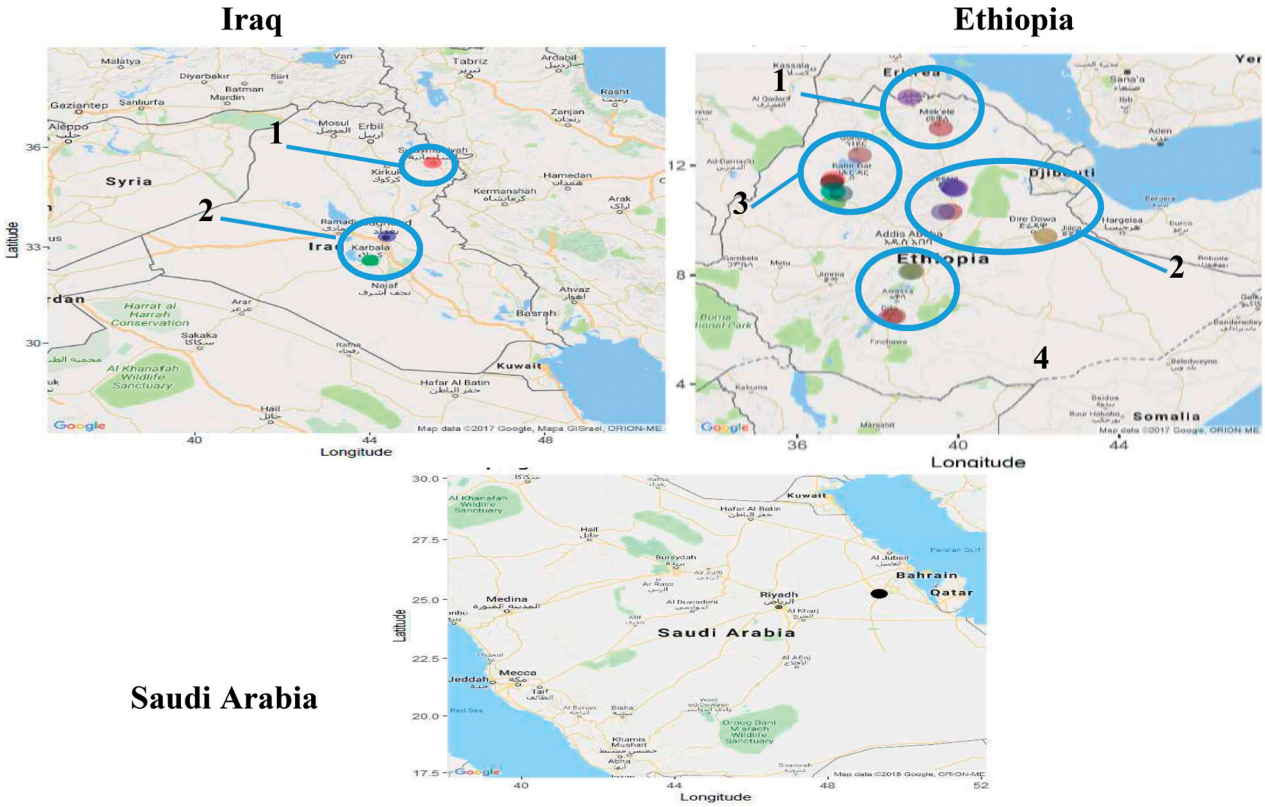

Figure 1. Sampling locations for the countries included in this study. For Iraq, 1 = North-Eastern region, 2 = Central; Ethiopia, 1 = North, 2 = Central-East, 3 = West, and 4 = South.

called 'Analysis of genotypes and phenotypes for backyard poultry production in sub-Saharan Africa'. In this institute, technical staff followed the following protocol: Genomic DNA (gDNA) samples were evaluated for quantity and quality using an AATI Fragment Analyzer and the DNF-487 Standard Sensitivity Genomic DNA Analysis Kit. Genomic DNA samples found to have a quality score <7 and have little or no high molecular weight material failed sample QC and replacement samples were requested. Next-Generation sequencing libraries were prepared using Illumina SeqLab specific TruSeq Nano High Throughput library preparation kits in conjunction with the Hamilton MicroLab STAR and Clarity LIMS X Edition. The gDNA samples were normalized to the concentration and volume required for the Illumina TruSeq Nano library preparation kits, then sheared to a 450 bp mean insert size using a Covaris LE220 Focused-ultrasonicator. The inserts were then ligated with blunt-ended, A-tailed, size-selected, TruSeq adapters, and enriched using eight cycles of PCR amplification. The libraries were normalized, denatured, and pooled in eights for clustering and sequencing using a Hamilton MicroLab STAR with Genologics Clarity LIMS X Edition. Libraries were clustered onto HiSeqX Flow cell v2.5 on cBot2s, and the clustered flow cell is transferred to a HiSeqX for sequencing using a HiSeqX Ten Reagent kit v2.5. After trimming the two adapters and applying the demultiplexing process, two compressed 'fastq.gz' files for each sample have been obtained.

Following the sequencing process and after obtaining the fastq files, we aligned the sample reads against the chicken reference genome Galgal 5.0 using the Burrows-Wheeler Aligner (BWA) v.0.7.5a software.[20] After sorting and indexing the bam files, Picard tools v.2.8 (http://broadinstitute.github.io/picard/) were used for marking duplicate reads and adding read groups. Insertion, deletions (indels), and realignments were performed with the Genome Analysis Toolkit (GATK) v.3.7,[21] and a Base Quality Score Recalibration (BQSR) step was done to correct for sequencing errors. The haplotype caller tool from GATK was used for calling the variants from the bam files, to generate the gVCF file for each sample and to perform SNPs and INDEL discoveries. Then these gVCF files were jointly genotyped to generate a single VCF file. The above steps were followed according to the recommendations of the GATK Best Practices for Variant Quality Score Recalibrations (VQSR).[22,23] The data file was restricted to bi-allelic SNPs and two files were generated: a full mtDNA file and the W chromosome file.

Distribution of SNPs and genetic diversity

mtDNA

The extracted vcf files of mtDNA for samples of all three countries were exploited to unravel SNP distributions. Genomic positions and SNPs were extracted from these files and plotted along with the ~17 kb of mtDNA using R software v. 3.3.2 (https://www.r-project.org/). Subsequently, the consensus sample sequences for the full mtDNA genome were generated from the same vcf files by using bcftools v.1.8.[24] The output was aligned by ClustalX v.2.1 for multiple sequence alignment et al., 1997) for each country separately.

Sequences of the mtDNA genome for each country were analyzed and genetic diversity parameters, such as the number of haplotypes (H), haplotype diversity (Hd), nucleotide diversity (π), and the average number of nucleotide differences were calculated using DnaSP v.5,[25] for populations, geographical regions, and countries.

W Chromosome

After checking the metadata for samples used in the mtDNA analysis, female samples were selected to extract ~6 Mb of W chromosome sequences. We did not include any Saudi samples because these were all males. The same way of plotting SNP positions in mtDNA was applied by extracting them from the vcf files and using histogram plot in R software v. 3.3.2 (https://www.r-project.org/). For a better comparison of SNP distributions between mtDNA and the W chromosome, a search for SNP polymorphism along the chromosome was done to identify W chromosome genomic regions with a high number of SNPs. A W-chromosome genomic region with high SNPs number was found at positions 4101210–4114277. It is 13 kb in length (Fig. S1) was annotated using the Ensemble (http://www.ensembl.org/) BioMart tool for gene identification. Then, the same process for generating sequencing data like that for the mtDNA was followed using bcftools v.1.8.[24] Multiple sequences aligned by ClustalX v.2.1.[26] DnaSP v.5[25] was used to the selected ~13 kb variable region of the W chromosome to calculate genetic diversity parameters.

Phylogeographic analysis

Maximum likelihood unrooted trees were constructed for the mtDNA sequences and the W chromosome

separately for each country. Besides the mtDNA samples from Iraq, Ethiopia, and Saudi Arabia, we included 19 reference sequences from Miao et al.[13] representing the following haplogroups: GU261695—haplogroup A, NC_007235—haplogroup B, GU261701—haplogroup C1, GU261680—haplogroup C2, GU261707—haplogroup C3, NC_007237—haplogroup D1, GU261683—haplogroup D2, GU261697—haplogroup D3, GU261709—haplogroup E1, HQ857209—haplogroup E2, GU261708—haplogroup E3, GU261691—haplogroup F, GU261690—haplogroup G, GU261715—haplogroup H, GU261698—haplogroup I, GU261706—haplogroup W, GU261692—haplogroup X, GU261693—haplogroup Y, and GU261674—haplogroup Z. The relationship among the haplotypes was examined through Median-Joining network using the PopArt software,[27] which uses the same network algorithm as Bandelt et al.[28] but for sequences larger than 5000 bp.

Analysis of molecular variance (AMOVA)

The analysis of molecular variance was implemented using Arlequin v 3.5.2 with 1,000 permutations.[29] This analysis was performed for all the samples and within a country. For the within-country analysis, we followed the grouping of the populations as described in the sampling section. Iraqi samples were divided into two groups [North-East (Sulimania) $n = 9$ and Central (Baghdad and Karbala) $n = 18$], Ethiopia into four groups (North $n = 20$, Central-East $n = 84$, West $n = 67$, and South $n = 40$). Saudi Arabia included five samples.

Neutrality test and demographic dynamics

The demographic profiles for each population were calculated from mismatch distribution patterns.[30] Mean Absolute Error (MAE) was used between the observed and the theoretical (expected) mismatch distribution to assess demographic expansion.[31] Then, Fu's Fs[32] and Tajima's D[33] were estimated by using the infinite site model in DnaSP v.5.[25]

Bayesian Skyline Plots (BSPs)[34] were investigated for a deeper insight into the demographic history of the chicken within countries. This was accomplished using the piecewise constant function in BEAST V 2.4.7.[35] First, the HKY + G nucleotide substitution model was used for the analysis, and then separate Markov Chain Monte Carlo simulation (MCMC) runs were applied for twenty-million generations sampled every one-thousand generation with the first two-million generations used as burn-in. Tracer software V.1.7[36] was used to calculate the convergence of the posterior estimates of *Ne* to the likelihood of stationary distribution. The BSPs were standardized using the molecular mutation rate of evolution for chicken mtDNA, 3.13×10^{-7} mutations/site/year (m/s/y) following Alexander et al.[9] This estimation of BSPs was performed on each country sample separately, and then they were plotted together.

Results

Distribution of SNPs—mtDNA

Figure 2 shows the distribution of SNPs for every 1000 bp along with the ~17 kb of the full mtDNA, compared to reference NCBI accession no. NC_007235.1,[37] for the Iraqi, Ethiopian, and Saudi Arabian samples. Overall, the highest number of polymorphisms are present in the Ethiopian populations ($n = 255$) compared with Iraq ($n = 71$) and Saudi Arabia ($n = 29$). For all three countries, the highest number of SNPs are present within the first 1 kb region, which includes the hypervariable (*D*-loop) or control region.

In Iraqi populations, Baghdad has the highest number of polymorphisms ($n = 40$), then Sulimania ($n = 8$) and Karbala ($n = 5$). Comparing the central area of Iraq with the North-East (represented by Sulimania), a higher number of variants is observed in the former than in the latter. Ethiopian populations show the greater variation in SNP numbers, with the highest in Adane ($n = 72$), Loya ($n = 65$), ShubiGemo ($n = 62$), and Kumato ($n = 60$) and the lowest in Midir ($n = 9$) and Negasi Amba ($n = 9$). The Central-Eastern region of Ethiopia exhibit the highest number of SNPs ($n = 136$), then the South ($n = 117$), North (81), and West ($n = 77$) regions.

MtDNA protein-coding genes were specifically examined for SNPs in the three countries. Iraqi samples show a range of 0–7 SNPs per gene for the 13 mtDNA genes [ND4 ($n = 7$), COXI ($n = 6$) and COXII ($n = 6$), ND2 ($n = 2$), ND3 ($n = 2$), and ND4L ($n = 2$)]. No polymorphism was found in the ATP8 gene. More SNPs are present in the mtDNA of the Ethiopian chicken. Variant numbers range from 2 to 34, with the highest number ($n = 34$) in the COXI gene, then ND5 and CYTB with 19 SNPs in each. The lowest number of variants were found in four genes, ATP8 ($n = 2$), ND4L ($n = 4$), ND3 ($n = 5$), and COXII ($n = 6$). The mtDNA genes of Saudi Arabian samples has a range of 0–7 SNPs, with the highest number in ND4 ($n = 7$) and COXI ($n = 6$), then between 0 and 4 SNPs for the remaining genes, such as COX3 ($n = 4$), ATP6 ($n = 3$), and no variants for ND1 and ATP8 ($n = 0$).

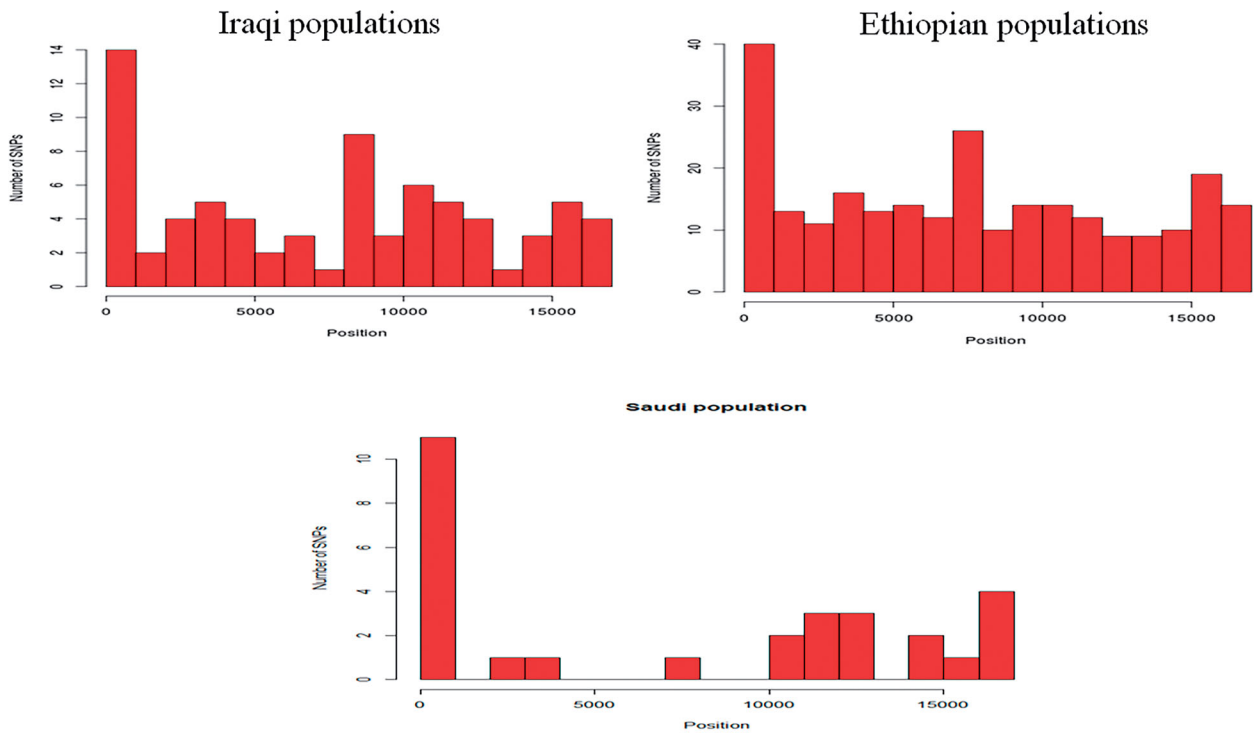

Figure 2. Distribution of SNPs along the full mtDNA sequences for the three countries included in this study.

Distribution of SNPs—W chromosome

Figure 3 shows the distribution of SNPs along with ~6 Mb of the W chromosome for the Iraqi and Ethiopian populations. A similar pattern of SNP diversity along the chromosome is observed, with the average number of SNPs ranging from 59 to 318 for the Iraqi and 86 to 363 SNPs for the Ethiopian populations. We chose a 13-kb region (position 4101210–4114277) to represent the W chromosome (Fig. S1). This 13-kb region includes no protein-coding genes (Ensemble Genes 94 database for *Galgal* 5.0, coordinates 4101210–4114277 and Ensemble Genes 95 *Galgal* 6.0, coordinates 6507911–6520978). Examining this genomic region closely, we identify 4 SNPs in the Iraqi samples; these variants have been reported before, and they have rs ID in the database of chicken variants (db SNP ftp://ftp.ensembl.org/pub/release-94/variation/gvf/gallus_gallus/). All these variants are intergenic variants. This is also the case for the 27 polymorphisms present in Ethiopian populations. These variants include 25 known SNPs from the chicken database (db SNP) with rs ID and two novel ones.

Genetic diversity—mtDNA

A high level of genetic diversity was present in the full mtDNA sequences from the three countries (Table 1). The total number of segregating sites was 292. It ranges between 5 and 72 within individual populations. The highest number of haplotypes ($n = 10$) within populations is found in a few populations from Ethiopia (Adane, Jarso, and Meseret). Haplotype diversity ranges between 0.655 and 1.000; the highest (1.000) being found in Adane, Meseret, and Saudi, and the lowest in Horro (0.655 ± 0.086). The haplotype diversity of the Iraqi populations is higher in the Central area (Baghdad 0.872 ± 0.067 and Karbala 0.900 ± 0.161) compared to the North-Eastern region (Sulimania 0.833 ± 0.127).

Genetic diversity—W chromosome

While much lower than the mtDNA (Table 2), the genetic diversity of the W chromosome indicates that, overall, that there is more diversity in Ethiopian compared to Iraqi populations. In fact, the Iraqi populations do not show any segregating sites, with the exception of those of the Central region, and the same applied to the majority of Ethiopian populations. A few Ethiopian populations have 1–4 polymorphic sites, except for Horro, Jarso, and Tzion (4, 3, 11 segregating sites) with the highest nucleotide and haplotype diversities (Table 2). At the regional level in Ethiopia, the Central-Eastern and South regions have about the same level of haplotype and nucleotide diversity. The West shows the

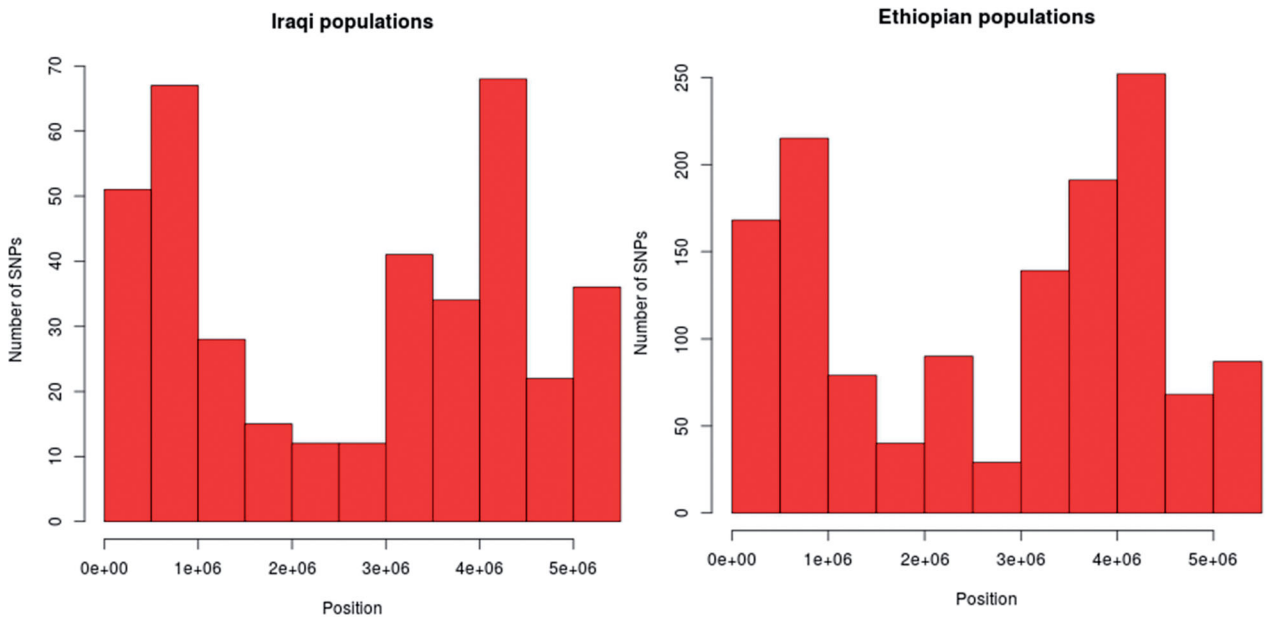

Figure 3. Distribution of SNPs (for every 500 kb) along the full W chromosome sequences of two countries included in this study.

Table 1. Location, sample size, and genetic diversity of full mtDNA sequences.

Country/Population	N	S	H	Hd (SD)	π (SD)	K
Iraqi populationsNorth-East						
Sulimania (1)	9	8	6	0.833 (0.127)	0.0001 (0.0000)	1.778
Central						
Baghdad	13	40	7	0.872 (0.067)	0.0004 (0.0002)	7.462
Karbala	5	5	4	0.900 (0.161)	0.0001 (0.0000)	2.200
Central (2)	18	55	11	0.928 (0.040)	0.0006 (0.0000)	10.660
Total	27	71	17	0.952 (0.023)	0.0010 (0.0001)	17.111
Ethiopian populationsNorth						
Meseret	10	59	10	1.000 (0.045)	0.0008 (0.0003)	14.089
Mihquan	10	55	8	0.956 (0.059)	0.0007 (0.0003)	12.800
North (1)	20	81	18	0.989 (0.019)	0.0008 (0.0002)	13.589
Central-East						
Adane	10	72	10	1.000 (0.045)	0.0012 (0.0004)	21.244
Arabo	10	19	9	0.978 (0.054)	0.0003 (0.0000)	5.956
Horro	30	48	7	0.655 (0.086)	0.0007 (0.0001)	12.184
Jarso	14	19	10	0.945 (0.045)	0.0002 (0.0000)	4.462
Midir	10	9	5	0.844 (0.080)	0.0001 (0.0000)	2.489
Negasi Amba	10	9	5	0.822 (0.097)	0.0001 (0.0000)	3.067
Central-East (2)	84	136	39	0.942 (0.016)	0.0013 (0.0001)	23.330
West						
Amshi	10	23	9	0.978 (0.054)	0.0003 (0.0000)	6.067
Ashuda	10	17	8	0.933 (0.077)	0.0002 (0.0000)	4.556
Batambie	8	13	6	0.929 (0.084)	0.0002 (0.0000)	3.786
Dikuli	10	22	9	0.978 (0.054)	0.0003 (0.0000)	5.822
Gafera	10	14	7	0.933 (0.062)	0.0002 (0.0000)	4.378
Surta	9	14	7	0.917 (0.092)	0.0002 (0.0000)	3.444
Tzion Teguaz	10	26	9	0.978 (0.054)	0.0004 (0.0001)	7.511
West (3)	67	77	50	0.991 (0.004)	0.0004 (0.0000)	7.745
South						
Girissa	10	14	6	0.844 (0.103)	0.0002 (0.0000)	4.200
Kumato	10	60	8	0.933 (0.077)	0.0011 (0.0004)	19.311
Loya	10	65	9	0.978 (0.054)	0.0015 (0.0002)	26.267
Shubi Gemo	10	62	7	0.933 (0.062)	0.0016 (0.0002)	28.178
South (4)	40	117	30	0.982 (0.011)	0.0018 (0.0000)	31.426
Total	211	255	134	0.988 (0.0030)	0.0013 (0.0000)	22.272
Other population						
Saudi Arabia	5	29	5	1.000 (0.126)	0.0007 (0.0001)	11.800
Total						
All the samples included in this study	243	292	156	0.990 (0.0023)	0.0013 (0.0000)	22.777

N: number of samples; S: segregating sites; H: number of haplotypes; Hd (SD): haplotype diversity (standard deviation); π (SD): nucleotide diversity (standard deviation); K: average number of nucleotide differences.

Table 2. Location, sample size, and genetic diversity of W chromosome sequences.

Country/Population	N	S	H	Hd (SD)	π (SD)	K
Iraqi populationsNorth-East						
Sulimania (1)	6	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Central						
Baghdad	11	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Karbala	3	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Central (2)	14	4	2	0.363 (0.130)	0.0001 (0.000)	1.451
Total	20	4	3	0.616 (0.077)	0.0001 (0.000)	1.832
Ethiopian populationsNorth						
Meseret	6	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Mihquan	6	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
North (1)	12	1	2	0.545 (0.062)	0.0000 (0.000)	0.545
Central-East						
Adane	6	1	2	0.333 (0.215)	0.0000 (0.000)	0.333
Arabo	6	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Horro	24	4	2	0.083 (0.075)	0.0000 (0.000)	0.333
Jarso	9	3	3	0.639 (0.126)	0.0001 (0.000)	1.611
Midir	6	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Negasi Amba	6	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Central-East (2)	57	19	10	0.799 (0.042)	0.0003 (0.000)	5.129
West						
Amshi	6	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Ashuda	6	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Batambie	6	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Dikuli	6	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Gafera	6	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Surta	6	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Tzion Teguaz	6	11	3	0.733 (0.155)	0.0003 (0.0001)	4.467
West (3)	42	15	9	0.891 (0.015)	0.0003 (0.000)	4.566
South						
Girissa	7	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Kumato	6	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Loya	6	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
Shubi Gemo	7	0	1	0.000 (0.000)	0.0000 (0.000)	0.000
South (4)	26	9	4	0.778 (0.025)	0.0003 (0.000)	4.197
Total	137	27	24	0.943 (0.008)	0.0003 (0.000)	5.051
All the samples included in this study	157	27	25	0.946 (0.007)	0.0003 (0.000)	4.784

N: number of samples; S: segregating sites; H: number of haplotypes; Hd (SD): haplotype diversity (standard deviation); π (SD): nucleotide diversity (standard deviation); K: average number of nucleotide differences.

highest level of haplotype diversity (0.891 ± 0.015) and the North the lowest one (0.545 ± 0.062).

Within-country phylogenetic relationships

Iraq

The result of the maximum likelihood unrooted tree shows the presence of three haplogroups in Iraqi mtDNA samples (Fig. 4). These haplogroups are A, D1, and E1 based on the classification of Miao et al.[13] Haplogroup A is found in the Sulimania population (North-East region) with six haplotypes in nine samples. The second haplogroup (D1) is represented by one haplotype in the Baghdad population (Central region). The last haplogroup (E1), the most frequent one, is found in the two populations (Baghdad and Karbala). It is represented by 10 haplotypes and 17 samples. The presence of three haplogroups in Iraqi chickens is highly supported by a range from 84 to 100 bootstrap values for all the tree branches.

Seventeen mtDNA haplotypes are found in 27 Iraqi samples (Fig. 5). The Baghdad population has the highest number of haplotypes ($n = 7$), then Sulimania ($n = 6$) and Karbala ($n = 4$). No common haplotypes are identified among the populations, but rather the distribution of the haplotypes reflects the geographical distribution of the sampling locations (see Fig. 1). Most haplotypes are represented by one sample (12 out of 17 haplotypes), while two (H_1 and H_10) are represented each by four samples. The haplotype in the middle of the network (H_17 from Baghdad) belongs to haplogroup D1, and it separates haplogroup A (Sulimania) from haplogroup E1 (Baghdad and Karbala). Haplogroup E1 is separated from haplogroup D1 and haplogroup A by 32 and 20 mutations, respectively, with the presence of 2 median vectors (mv).

The maximum likelihood unrooted tree for Iraqi W chromosome haplotypes is shown in Fig. 6. The tree shows the presence of three haplotypes. Each population is represented by one haplotype.

The median-joining network of the Iraqi W chromosome shows only three haplotypes (Fig. 7). Each population clustered in a single haplotype,

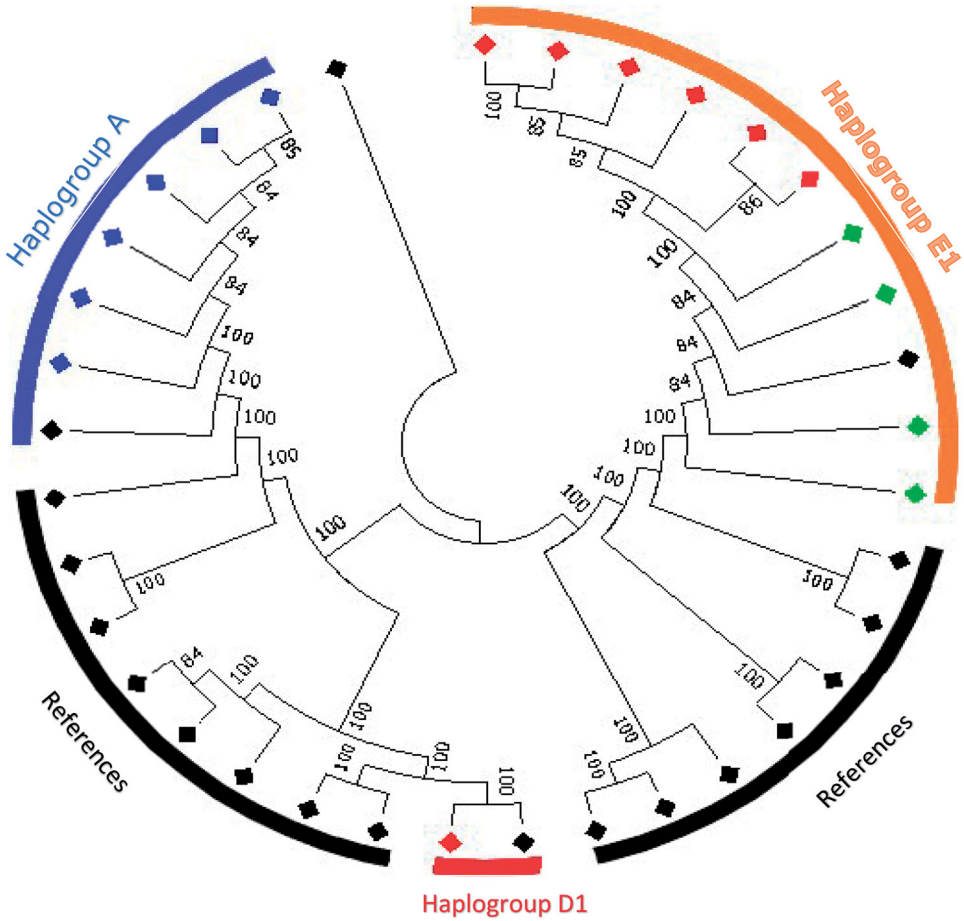

Figure 4. Maximum likelihood unrooted tree for the 17 Iraqi mtDNA haplotypes with haplogroup references from Miao et al.[13] Blue = Sulimania, Red = Baghdad, and Green = Karbala.

Baghdad being represented by eleven samples (red), Karbala by three (green), and Sulimania by six (blue). The North-Eastern population (Sulimania) links the two central populations, which are separated from Baghdad by three mutations and a single mutation from Karbala.

Ethiopia

The phylogenetic tree of Ethiopian mtDNA sequences revealed the presence of five haplogroups and 134 haplotypes (Fig. 8). These haplogroups were A, B, C1, C2, and E1, represented by different numbers of samples: haplogroup A includes six samples [Meseret ($n = 1$) and Horro ($n = 5$)] Haplogroup (B) is found in one bird (Mihquan), and the same applied to C1 represented by one bird from Adane; C2 is present in five populations (Jarso, Adane, Kumato, Shubi, and Loya) and 34 birds. Most birds ($n = 169$) clustered with haplogroup E1. So, the most frequent haplogroups are E1 and C2, with the other three haplogroups observed at low frequencies. Most branches in this maximum likelihood unrooted tree are supported by high bootstrap values.

The relationships among Ethiopian mtDNA haplotypes are illustrated in Fig. 9. This median-joining network shows 134 haplotypes for the 19 populations. As a general observation, the network appears to divide the populations into two groups, the first includes six populations (Dikuli, Horro, Jarso, Kumato, Loya, and Shubi) that do not share haplotypes with the others. The second group consists of the other 13 populations that have some haplotypes in common and are grouped together. Most of the haplogroups other than E1 are found in the first group: for example, haplogroup A consists of six samples represented by two haplotypes in the first group; and haplogroup C2 consists of 34 individuals, also from populations in the first group.

The maximum likelihood tree for the Ethiopian W chromosome haplotypes is shown in Fig. 10. It shows three groups of haplotypes, the first includes 11 haplotypes found in 11 populations (Arabo, Amshi, Batambi, Horro, Loya, Meseret, Midir, Mihquan,

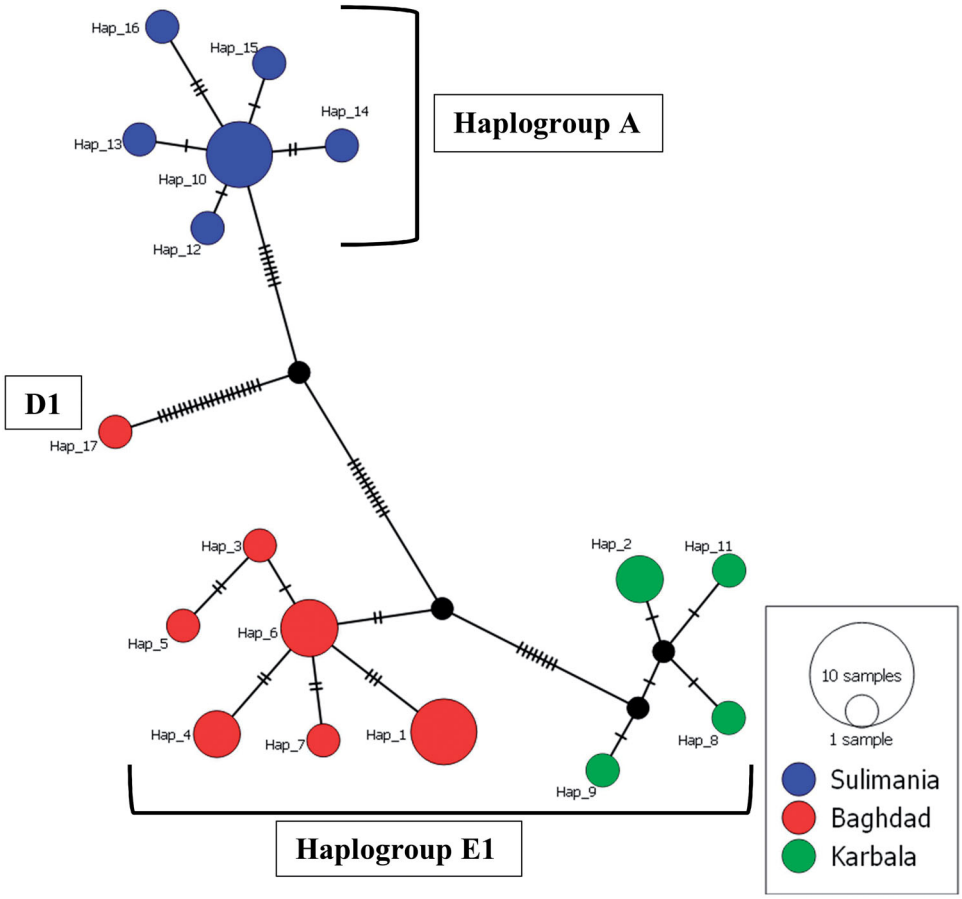

Figure 5. Median-joining haplotype network of Iraqi mtDNA haplotypes and haplogroups.

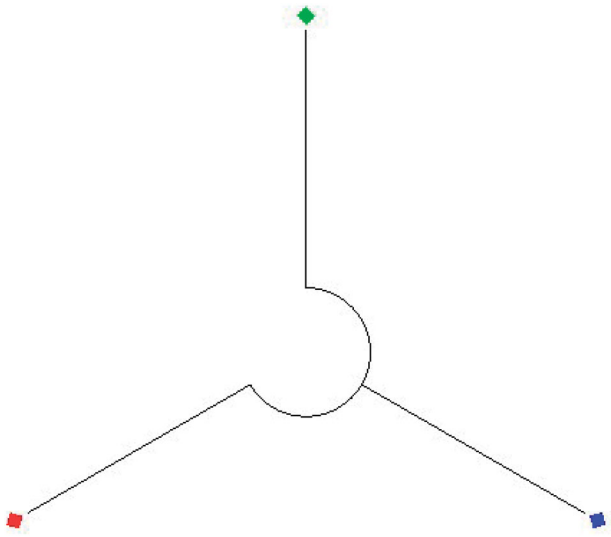

Figure 6. Maximum likelihood unrooted tree for Iraqi W chromosome haplotypes. Blue = Sulimania, Red = Baghdad, and Green = Karbala.

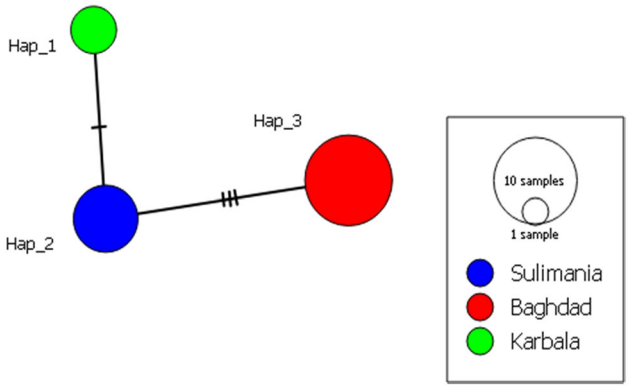

Figure 7. Median-joining haplotype network of Iraqi W chromosome haplotypes.

Shubi, Surta, and Tzion), the second four haplotypes in four populations (Dikuli, Kumato, Horro, and Negasi), and the third nine haplotypes in six populations (Adane, Ashuda, Girrisa, Gafera, Jarso, and Tzion). Two populations (Horro and Tzion) have haplotypes belonging to two groups. The majority of the bootstrap confidence levels of the tree branches range from 70 to 100.

The relationship of 24 Ethiopian W chromosome haplotypes is shown in Fig. 11. Many populations have only one haplotype; the highest number of haplotypes are in Jarso, Tzion, Adane, and Horro (3, 3, 2, and 2, respectively). All but one haplotype is

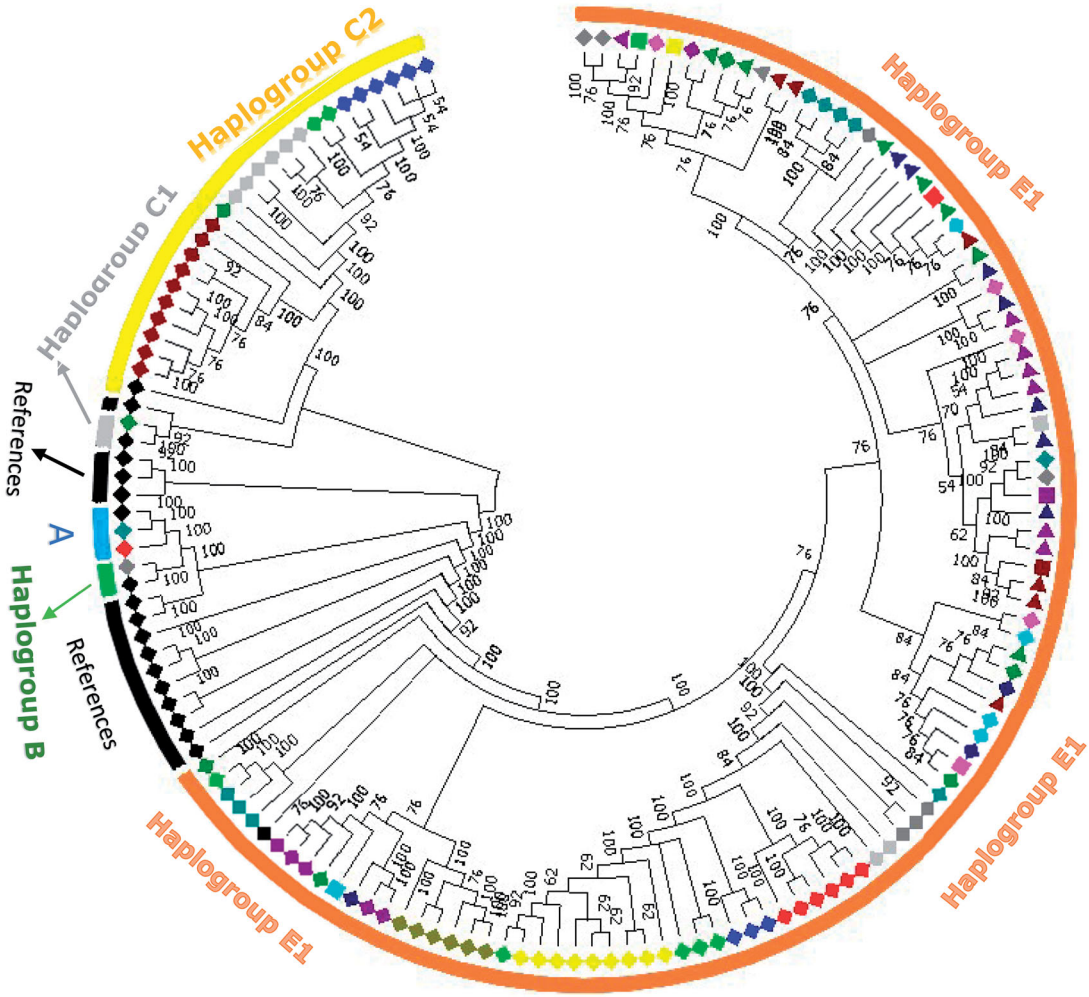

Figure 8. Maximum likelihood unrooted tree for Ethiopian mtDNA haplotypes with haplogroup references from Miao et al.[13] ◆ = Jarso, ◆ = Adane, ◆ = Girissa, ◆ = Midir, ◆ = Arabo, ◆ = Meseret, ◆ = Mihquan, ◆ = Kumato, ◆ = Horro, ◆ = Shubi, ◆ = Dikuli, ◆ = Loya, ◆ = Gafera, ◆ = Batambie, ▲ = Tzion, ▲ = Amshi, ▲ = Ashuda, ▲ = Surta, ▲ = Negasi, ■ = Amshi + Batambie, ■ = Adane + Negasi + Midir + Gafera + Batambie, ■ = Midir + Negasi + Adane + Arabo, ■ = Negasi + Arabo + Adane, ■ = Arabo + Gafera + Tzion, ■ = Negasi + Ashuda, ■ = Amshi + Tzion, ■ = Gafera + Tzion, ◆ = References.

population-specific, the shared haplotype being found in Meseret (six samples) and Tzion (three samples). The number of mutations between haplotypes ranges from 1 to 4.

Saudi Arabia

Figure 12 shows the relationship between the five Saudi Arabia mtDNA haplotypes. All the haplotypes belong to one haplogroup (E1). All but four of the branches of the tree have a high confidence bootstrap value (100).

The relationship among the Saudi Arabia haplotypes is shown at the median-joining haplotype network (Fig. 13). It shows five distinct haplotype sequences and two median vectors. The most distinct haplotype is Hap_4, separated from haplotype 5 by 19 mutations. Hap_4 represents one of the basal sequences in the phylogenetic tree in Fig. 12.

Analysis of molecular variance (AMOVA)

The results of the analysis of molecular variances for the full mtDNA sequences are shown in Table 3. For the Iraqi populations, the highest variance is found among the Iraqi groups (43.1%), then among populations within groups (39.7%), and finally among individuals within Iraqi population (17.2%). For the Ethiopian samples, the highest variance was found among populations within groups (47.2%), with the among-groups variance being 9.6%, and 43.3% for the within-population variation. For the three countries together, the highest percentage of variation was

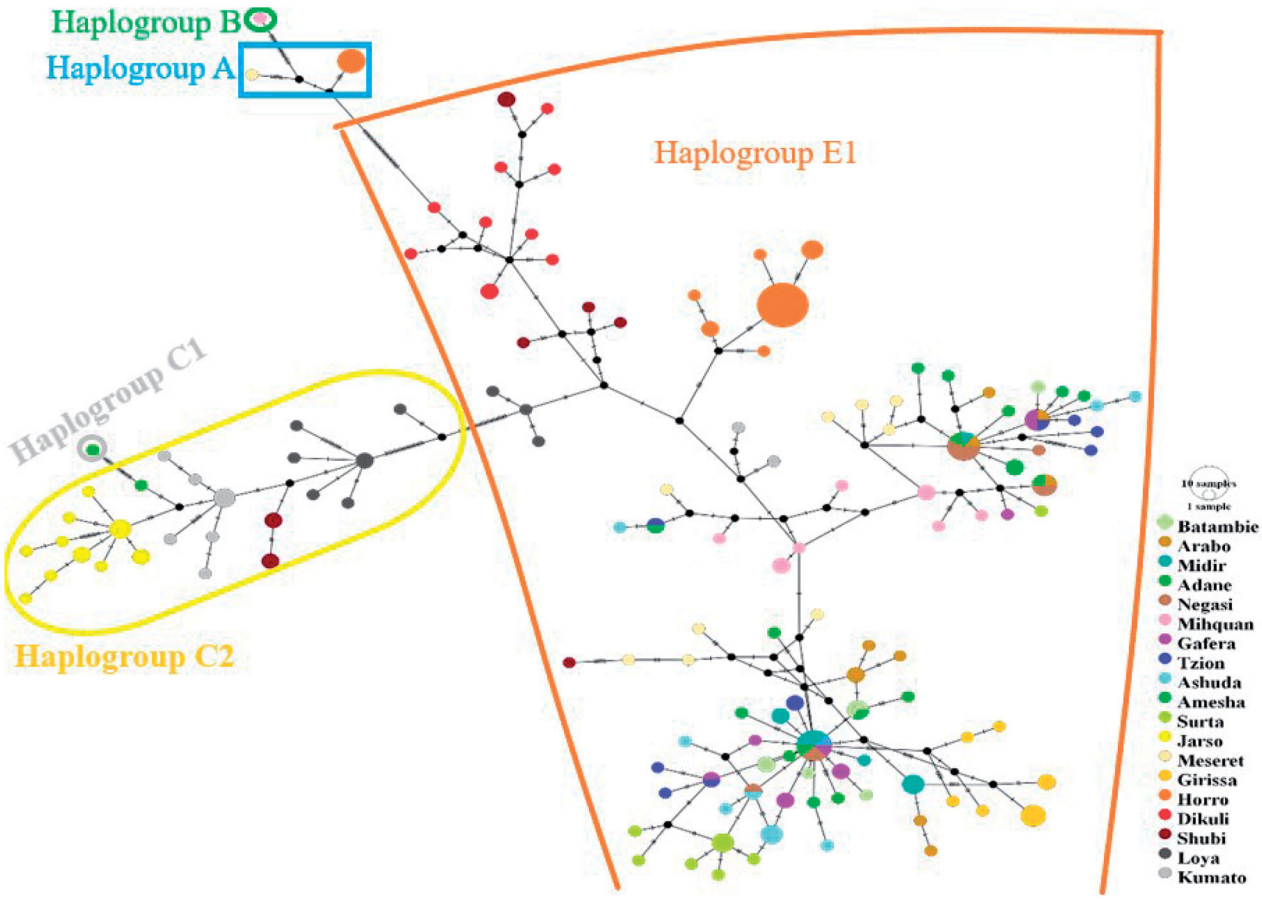

Figure 9. Median-joining network of Ethiopian mtDNA haplotypes.

among populations within countries (53.3%), then among individuals within populations (38.6%), with among countries the lowest (8.1%).

Considering the low level of polymorphisms for the W chromosome fragment, AMOVA analysis was not performed.

Population history and demography

The demographic patterns for each population and country are presented in Table 4 for the mtDNA analysis. Given the low number of W haplotypes, population demography analysis was not performed. The Mean Absolute Error (MAE) values for the difference between the actual and expected variance for Iraqi populations are all <1. For the Ethiopian samples, nine populations have values >1, and the Saudi population has a value >1.

Tajima's *D* results for the populations were mostly negative and non-significant, except for Baghdad and Sulimania for Iraq, and Meseret, Mihquan, and Surta for Ethiopia, all of which show significant negative Tajima's *D* values. Horro, Loya, and Shubi all have positive values, although non-significant. The second test of neutrality is Fu's *Fs*. Again, the majority of populations display negative non-significant values, with the exception of Sulimania (Iraq), Amshi, Arabo, and Dikuli (Ethiopia). At a population level, the Harpending raggedness indices (*r*) were not significant, except for Meseret, Surta, and Tzion from Ethiopia with low significant values.

The pairwise differences or mismatch distributions displayed three patterns (Fig. 14 and Fig. S2), uni-, bi-, and multimodal, the commonest being the bi- and multimodal patterns. The unimodal pattern was present in one Iraqi population only (Karbala) (Fig. S2). These patterns support a complex history for these populations: the results of Table 4 could indicate demographic expansion, bottleneck, or purifying selection. This complexity was endorsed by the mismatch distribution, which revealed three patterns of distribution for the countries included in this study.

Bayesian Skyline Plots (BSPs) (Fig. 15) show that the chicken populations from the three countries witnessed past demographic expansion. For the Ethiopian samples, the expansion was more rapid than for the others, and it also showed a recent decrease. The Iraqi chickens showed a historic expansion, which was less

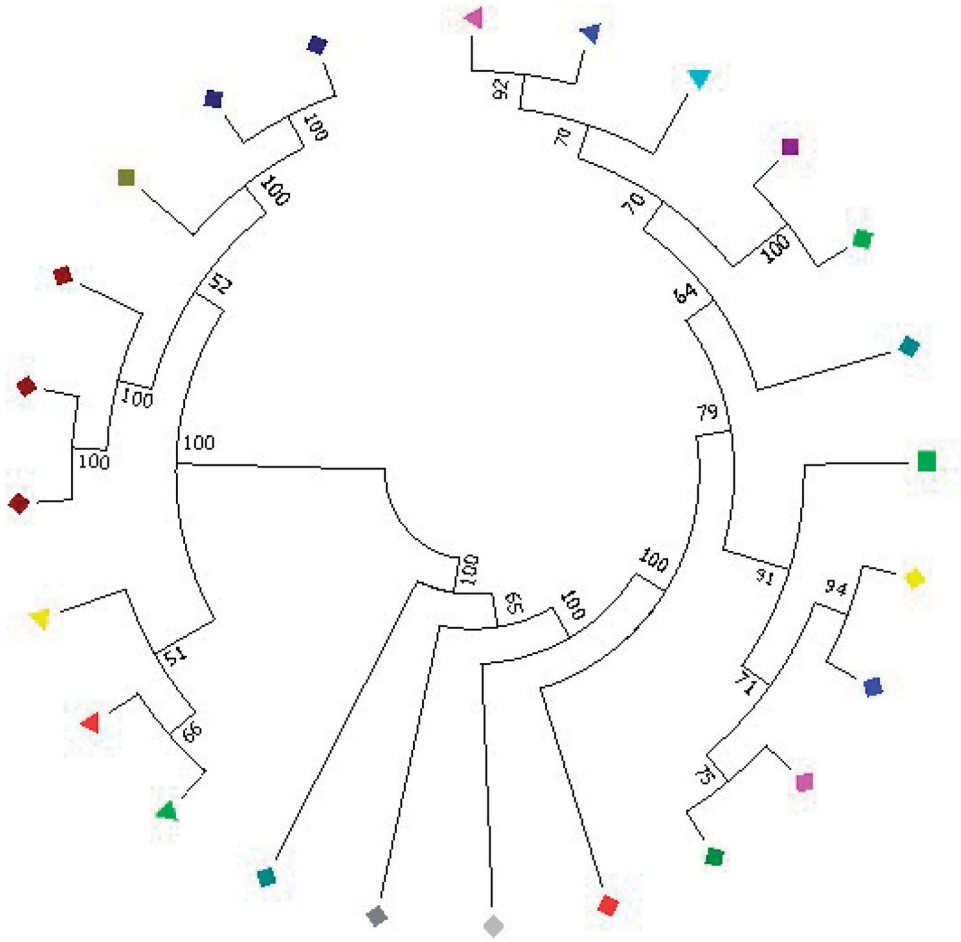

Figure 10. Maximum likelihood unrooted tree for Ethiopian W chromosome samples. = Midir, = Arabo, = Amshi, = Loya, = Shubi, = Horro, = Meseret + Tzion, = Batambie, = Surta, = Mihquan, = Tzion, = Negasi, = Dikuli, = Kumato, = Gafera, = Ashuda, = Tzion, = Jarso, = Girissa, = Adane.

rapid than for the two other countries. They also witnessed a bottleneck like the one observed in Ethiopia. The Saudi chicken was the only one with evidence for an old expansion without showing signs of a recent decrease.

Discussion

In this study, 243 full mtDNA and 157 W chromosome sequences were analyzed. The mtDNA sequences were from three countries [Ethiopia ($n = 211$), Iraq ($n = 27$), and Saudi Arabia ($n = 5$)], while the W sequences were from two countries [Ethiopia ($n = 137$) and Iraq ($n = 20$)]. As far as we know, this is the first study to investigate these two maternally inherited chicken genetic markers from these countries for the purpose of phylogeographic analysis.

While the SNPs were found across the entire mtDNA (Fig. 2), and one genomic region displayed more variation than the others. This region is the first 1 kb of the mtDNA that includes the hypervariable *D-loop* region, known as a non-coding part of the mtDNA. Among the three as expected with a much larger number of samples analyzed, more polymorphisms, over the entire mtDNA, were present in the Ethiopian populations. Importantly, other factors may also have here contributed to the mtDNA diversity of Ethiopian chicken. In particular, more complex history of the arrival of chicken in the Horn of Africa compared to Iraq.[17]

The distribution of SNPs shows large variation along the ~6 Mb W chromosomal length studied here Compared to the chicken autosomes, the W chromosome compared to the average number of polymorphisms observed on the chicken autosomes.[38,39] The W chromosome lower effective population size, the absence of recombination at the non-pseudo-autosomal W region selection may have all contributed to the low W chromosome polymorphism.

The 13-kb of the W chromosome genomic region was chosen on particular criteria. We needed a region with a high SNP number and of similar length to the

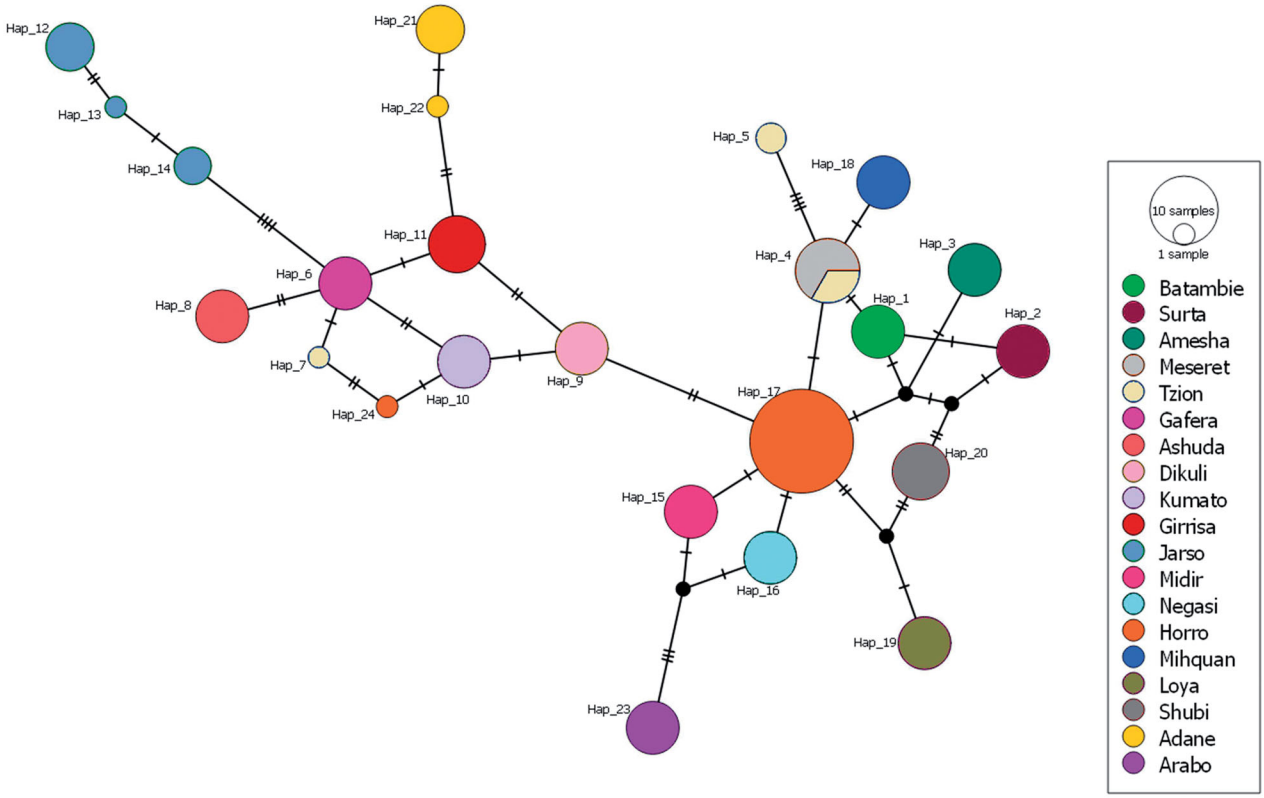

Figure 11. Median-joining haplotype network for the Ethiopian W chromosome sequences.

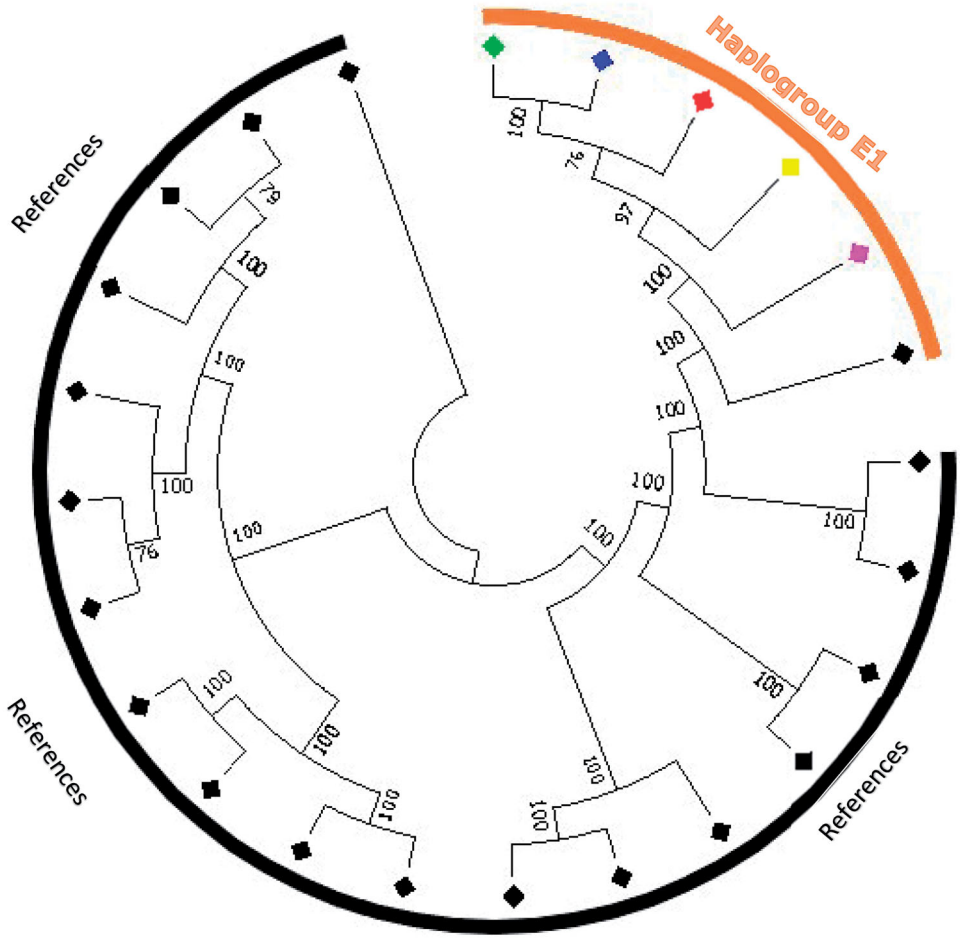

Figure 12. Maximum likelihood unrooted tree for the five Saudi mtDNA haplotypes with haplogroup reference sequences from Miao et al.[13] ◆ = Saudi sample 1, ◆ = Saudi sample 4, ◆ = Saudi sample 5, ◆ = Saudi sample 2, ◆ = Saudi sample 3.

mtDNA one to perform the same analysis on the two sets of maternally inherited markers facilitating the comparison of the results.[9] These criteria were satisfied for the non-coding genomic region we used to represent the W chromosome in this study (Fig. S1). Even so, we observed a large difference in the number of polymorphisms between the two sets of markers, e.g., the total number of SNPs for Iraqi and Ethiopian W chromosome samples were 4 and 27, respectively, whereas they show a much higher number of SNPs for the mtDNA (Iraq = 71 and Ethiopia = 255).

Genetic diversity parameters show results compatible with the distribution of SNPs for both genetic markers. For mtDNA, the samples have a high level of genetic diversity (Table 1) at both population and country levels. The Iraqi central region has more variation than the North-East one. Here, the difference in the number of samples but also distinct past trading networks between regions may explain this difference. Indeed, Baghdad (the capital of Iraq) has been the main center of Islamic civilization and trading for many centuries. Accordingly, it has attracted people and livestock from different geographic regions.

For the W chromosome, all the populations in both countries displayed no or low genetic variation, except for Jarso and Tzion from Ethiopia, with W haplotypes distinct from the ones observed in the other populations and with unique polymorphism compared to the W chromosome sequence of reference (Table 2). This reduced diversity of the W chromosome has been mentioned before. For example, the genetic diversity of the W chromosome of European flycatcher is reduced within species, but higher rates of variation are evident among species. In this respect, the origin of the higher diversity and uniqueness of the W haplotypes observed at Jarso and Tzion populations requires further investigation.

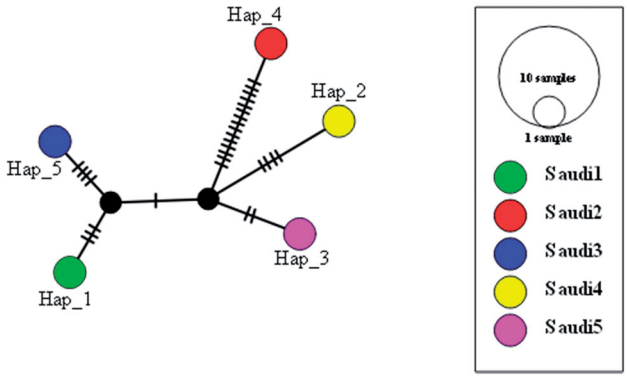

Figure 13. Median-joining haplotype network of Saudi mtDNA haplotypes. Saudi1 = Hap 1, Saudi2 = Hap 4, Saudi3 = Hap 5, Saudi4 = Hap 2, and Saudi5 = Hap 3.

Interestingly, if no or little W chromosome genetic diversity was present within and across populations from a single geographical area, the W diversity increased by combining different populations from several geographical locations. Domestic chicken from different geographic regions may have different origins on the Asian continent and accordingly, different histories of crossbreeding with Red Junglefowl subspecies or even introgression with other *Gallus* species.

The phylogenetic tree of Iraqi mtDNA showed the presence of three haplogroups (A, D1, and E1). The majority of Iraqi samples have a probable Indian subcontinent origin with the clustering of their mtDNA with clade E1. Liu et al.[11] has shown that clades A and E are widely distributed in Eurasia. The reference sequence for clade D1, a cockfighting chicken, comes from Bali (Indonesia).[37] It is present in one Iraqi sample. The presence of haplogroup D1 is unique to Iraq, is not found in Saudi Arabia and Ethiopia. The presence of this haplotype is perhaps the result of a maritime route of chicken introduction to Iraq besides the terrestrial route represented by haplogroup E1. The other haplotypes are unique population haplotypes, with the network reflecting the geographical sampling locations of the Iraqi mtDNA. The mtDNA sequences of Baghdad and Karbala were more closely to each other than to the Sulimania ones. Iraqi W chromosome network shows no phylogeographic structure at the opposite of the Iraqi mtDNA haplotypes.

Table 3. Analysis of molecular variances (AMOVA) for the full mtDNA.

Grouping	Source of variation	Degrees of freedom	Variance components	Percentage of variation
Iraq	Among groups	1	5.87071 Va	43.07
	Populations within groups	1	5.41341 Vb	39.72
	Within populations	24	2.34501 Vc	17.21
	Total	26	13.62914	
Ethiopia	Among groups	3	1.13075 Va	9.59
	Populations within groups	15	5.56228 Vb	47.16
	Within populations	192	5.10101 Vc	43.25
	Total	210	11.79404	
All countries	Among countries	2	1.01521 Va	8.13
	Populations within countries	20	6.65227 Vb	53.29
	Within populations	220	4.81489 Vc	38.57
	Total	242	12.48237	

Table 4. Neutrality and demographic expansion parameters for the mtDNA dataset.

Population/Country	N	S	MAE	Tajima's D (*p*-value)	Fu's Fs (*p*-value)	Harpending *r* (*p*-value)
Iraqi populationsNorth-East						
Sulimania (1)	9	8	0.545	−1.797 (0.004)	−2.495 (0.014)	0.063 (0.085)
Central						
Baghdad	13	40	0.923	−1.871 (0.022)	1.374 (0.736)	0.093 (0.673)
Karbala	5	5	0.854	−0.561 (0.387)	−0.848 (0.182)	0.430 (0.758)
Central (2)	18	55	0.966	−1.377 (0.074)	0.314 (0.582)	0.062 (0.752)
Total	27	71	0.947	−0.273 (0.449)	0.024 (0.544)	0.020 (0.481)
Ethiopian populationsNorth						
Meseret	10	59	1.081	−1.592 (0.052)	−2.576 (0.070)	0.019 (0.008)
Mihquan	10	55	1.057	−1.673 (0.031)	0.227 (0.469)	0.127 (0.824)
North (1)	20	81	1.006	−1.654 (0.025)	−5.271 (0.015)	0.009 (0.011)
Central-East						
Adane	10	72	1.106	−0.814 (0.234)	−1.689 (0.115)	0.062 (0.515)
Arabo	10	19	0.925	−0.530 (0.314)	−3.029 (0.040)	0.034 (0.052)
Horro	30	48	1.296	0.020 (0.586)	9.245 (0.997)	0.247 (0.999)
Jarso	14	19	0.846	−1.054 (0.157)	−2.935 (0.053)	0.061 (0.303)
Midir	10	9	0.623	−0.95 3 (0.191)	−0.105 (0.449)	0.055 (0.065)
Negasi Amba	10	9	0.708	−0.157 (0.486)	0.373 (0.580)	0.168 (0.582)
Central-East (2)	84	136	0.838	−0.479 (0.363)	−0.984 (0.459)	0.010 (0.693)
West						
Amshi	10	23	1.036	−1.202 (0.127)	−2.972 (0.043)	0.035 (0.052)
Ashuda	10	17	0.728	−1.124 (0.128)	−2.248 (0.105)	0.075 (0.283)
Batambie	8	13	0.644	−1.240 (0.115)	−1.080 (0.215)	0.096 (0.274)
Dikuli	10	22	1.082	−1.188 (0.144)	−3.098 (0.036)	0.047 (0.130)
Gafera	10	14	0.759	−0.528 (0.324)	−1.060 (0.221)	0.064 (0.198)
Surta	9	14	0.516	−1.590 (0.046)	−2.165 (0.072)	0.025 (0.010)
Tzion Teguaz	10	26	0.666	−0.871 (0.193)	−2.358 (0.085)	0.023 (0.013)
West (3)	67	77	0.661	−1.767 (0.016)	−41.684 (0.000)	0.005 (0.002)
South						
Girissa	10	14	0.836	−0.692 (0.260)	−0.021 (0.475)	0.084 (0.327)
Kumato	10	60	1.155	−0.439 (0.351)	1.101 (0.629)	0.044 (0.286)
Loya	10	65	1.283	0.704 (0.798)	0.336 (0.468)	0.038 (0.267)
Shubi Gemo	10	62	1.375	1.403 (0.964)	3.560 (0.931)	0.081 (0.727)
South (4)	40	117	0.881	0.523 (0.778)	−2.059 (0.263)	0.004 (0.058)
Total	211	255	0.811	−1.533 (0.034)	−86.131 (0.003)	0.001 (0.000)
Other populations						
Saudi Arabia	5	29	1.386	−1.137 (0.095)	−0.123 (0.274)	0.140 (0.229)
Total						
All the samples included in this study	243	292	0.653	−1.659 (0.20)	−113.833 (0.00)	0.000 (0.000)

Five mtDNA haplogroups were found in Ethiopia: A, B, C1, C2, and E1. The most frequent were C2 and E1 represented by 34 and 169 samples, respectively. The presence of haplogroups B, C1, and C2 was unique to Ethiopian mtDNA relative to Iraqi and Saudi samples. It supports more than one maternal origin for the Ethiopian chicken and several introductions of different origins. As mentioned above, the origin of haplogroup E1 is likely the Indian subcontinent. Haplogroup B has been reported Laos (South-East Asia), C1 in Henan Province came (R.P. of China) and C2 in Southern India. The presence of these different haplogroups explains the high level of mtDNA genetic diversity present in Ethiopia. The Ethiopian mtDNA network supports the results of the phylogenetic tree with the presence of several haplogroups supporting multiple maternal origins for domestic chicken in Ethiopia. Interestingly, Ethiopian population may be divided into two groups, population sharing haplotypes with other populations and population with unique haplotypes only.

The analysis of full mtDNA reveals mtDNA sequences polymorphisms within populations with monomorphic *D-loop* sequences. This is, for example, the case for the Sulimania population from Iraq and the Jarso population from Ethiopia. These two populations exhibited no genetic diversity using the *D-loop* genetic marker.[17] Still, the full mtDNA data showed that these two populations show a high level of mtDNA variation.

The phylogenetic tree of the W chromosome for the Ethiopian samples reveals three groups. There were features in common between the Ethiopian mtDNA and W-chromosome networks. In both networks, Jarso haplotypes were on one of the sides of the network far from other populations. So, in contrast to the Iraqi W chromosome, the low level of diversity for the Ethiopian W chromosome offered some useful information about the haplotype relationship.

The Saudi Arabia mtDNA phylogeny suggests a possible Indian subcontinent origin with a terrestrial

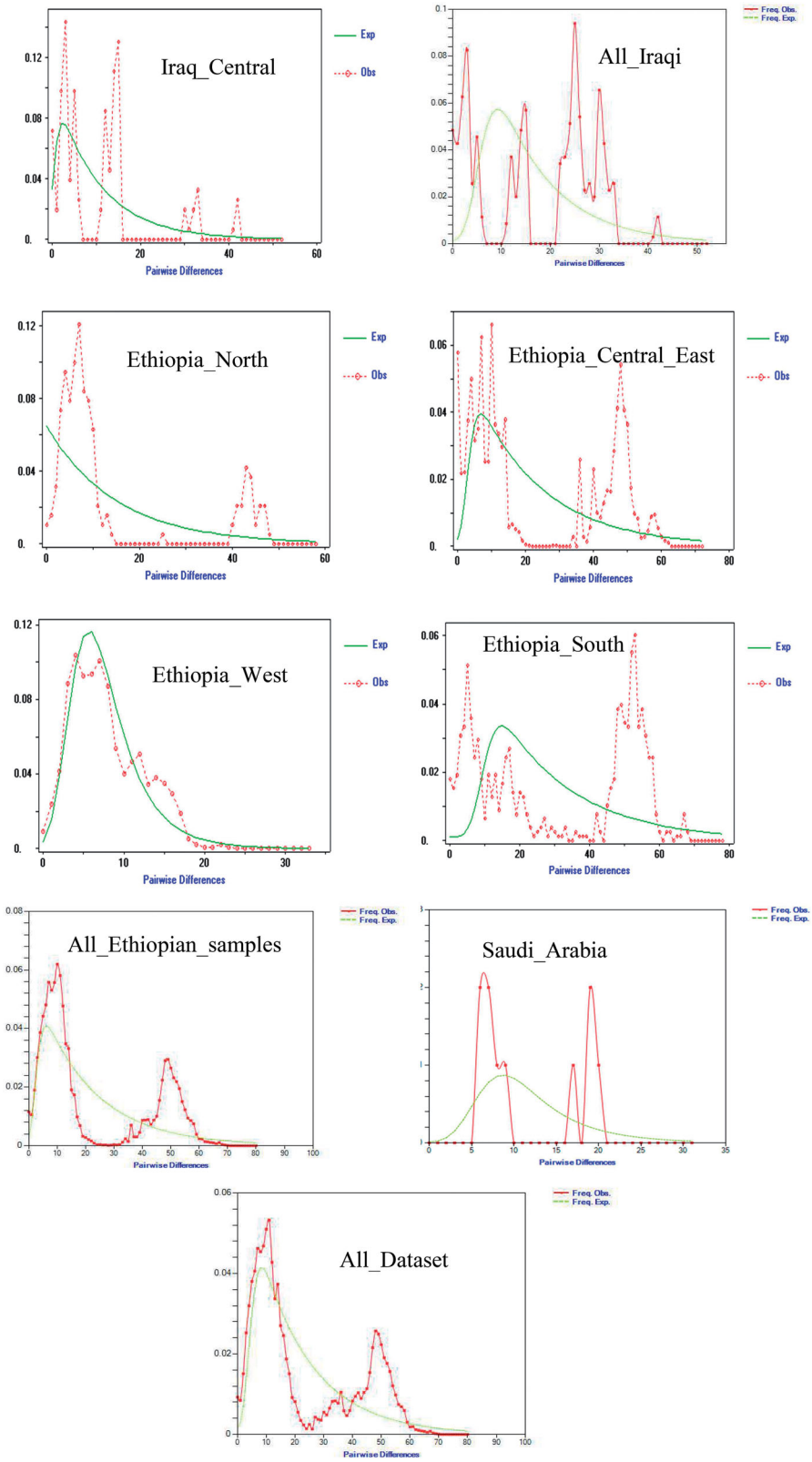

Figure 14. Regions and countries mismatch distribution patterns for the full mtDNA sequences ($n = 243$).

route of introduction. The tree shows five distinct haplotypes for the five samples, indicating the existence of genetic variation among these samples. This variation is more noticeable in the network, which reveals that the five haplotypes representing the five o samples are linked to two median vectors (mv),

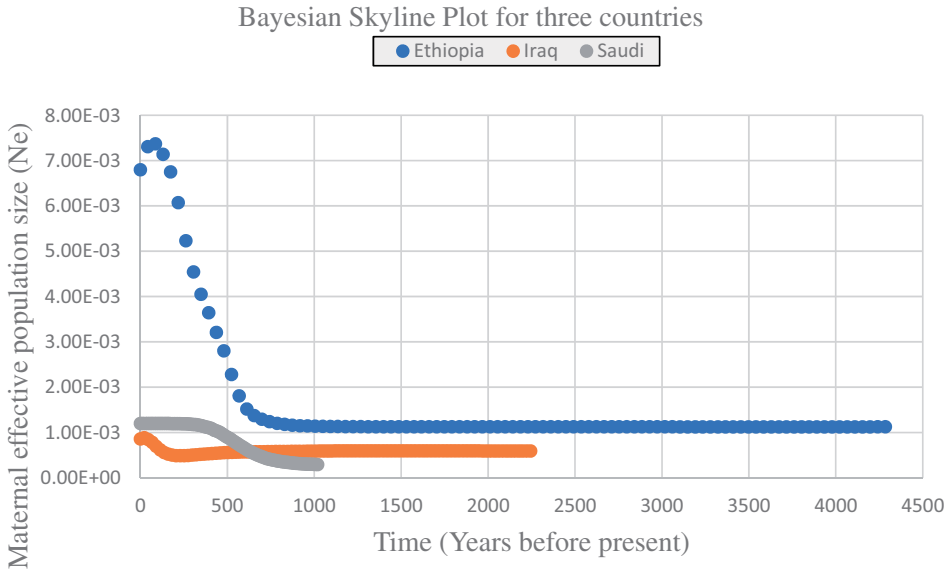

Figure 15. Bayesian Skyline Plot (BSP) for three countries included in this study.

each separated by one mutation. No Saudi Arabia samples were used to analyze the W chromosome because only male samples were available.

The AMOVA results for mtDNA show that the genetic variation is variously distributed among the three levels (among groups, populations within groups, and among individuals within populations). Comparing Iraq with Ethiopia, the former has similar variation among and within groups, whilst Ethiopian variation is slightly different, with most of the variation among populations within groups and among individuals within populations. This is possibly a sampling effect, since the 211 Ethiopian samples may have dominated and affected the overall result.

The BSP plot clarified the picture of demographic expansion, by interpreting the uncertainty of the summary statistics in Table 4. It showed that two countries (Ethiopia and Iraq) had an old expansion with a more recent bottleneck. Arabia chickens were somewhat different, with an old demographic expansion as in the other countries, but then with stability and no recent decline. Depending on these results, action is needed to maintain the genetic material of the countries that have experienced recent decreases to prevent any further loss of diversity.

Conclusion

In conclusion, this study assessed and compared the mtDNA and W chromosome diversity of local populations of chickens from three different countries, which likely witnessed different histories of introductions of domestic chicken. The results confirm the usefulness of mtDNA sequences diversity as a phylogeographic marker. However, very little information was obtained from the W chromosome sequences here, which showed reduced diversity likely explained by strong selection pressure and lack of recombination. These results, therefore, recommend the use of mtDNA as a genetic marker in the phylogeographic study of chicken populations, while the W chromosome could possibly be rather utilized for genetic diversity studies among Gallus species and subspecies.

Acknowledgments

Special thanks to Adebabay Kebede and Ahmed Saadi for collecting Ethiopian and Karbala samples.

Disclosure statement

The authors declare that they have no competing interests.

Funding

This study is part of the first author (Ahmed S. Al-Jumaili) Ph.D., supported and funded by the Iraqi Ministry of Higher Education and Scientific Research (MOHESR).

Ethical approval

This study was approved by the ethics committee of International Livestock Research IAUC guidelines (Reference Number IACUC-RC2017–21) and Ethics committee of Iraqi Ministry of Higher Education and Ministry of Agriculture (Reference Number 17352015/2872). Samples were dispatched in the UK following International Guidelines (Nagoya Protocol).

Author contributions

A.S.A. and O.H. designed the project, and A.S.A. contributed to all data analysis. The manuscript was prepared by A.S.A. and revised by O.H.

ORCID

Ahmed S. Al-Jumaili http://orcid.org/0000-0002-5613-7626

Data availability statement

The datasets generated and/or analyzed during the current study are available in the GenBank repository, [https://www.ncbi.nlm.nih.gov/genbank/] accession numbers OK348287-OK348686.

References

1. Avise JC. Phylogeography: retrospect and prospect. *J Biogeogr*. 2009;36(1):3–15.
2. Avise JC, Arnold J, Ball RM, et al. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst*. 1987;18(1):489–522.
3. Leal BSS, Palma da Silva C, Pinheiro FJ. Phylogeographic studies depict the role of space and time scales of plant speciation in a highly diverse Neotropical region. *Crit Rev Plant Sci*. 2016;35(4):215–230.
4. Avise JC, Bowen BW, Ayala FJ. In the light of evolution X: comparative phylogeography. *Proc Natl Acad Sci USA*. 2016;113(29):7957–7961.
5. Bermingham E, Moritz CJME. Comparative phylogeography: concepts and applications. *Mol Ecol*. 1998;7(4):367–369.
6. Riddle BR. Comparative phylogeography clarifies the complexity and problems of continental distribution that drove A. R. Wallace to favor islands. *Proc Natl Acad Sci USA*. 2016;113(29):7970–7977.
7. Wiens JJ. Perspective: why biogeography matters: historical biogeography vs. phylogeography and community phylogenetics for inferring ecological and evolutionary processes. *Front Biogeogr*. 2012;4(3):128–135.
8. Zhang L-J, Cai W-Z, Luo J-Y, et al. Phylogeographic patterns of *Lygus pratensis* (Hemiptera: Miridae): evidence for weak genetic structure and recent expansion in northwest China. *PLoS One*. 2017;12(4):e0174712.
9. Alexander M, Ho SYW, Molak M, et al. Mitogenomic analysis of a 50-generation chicken pedigree reveals a rapid rate of mitochondrial evolution and evidence for paternal mtDNA inheritance. *Biol Lett*. 2015;11(10):20150561.
10. Lan D, Hu Y, Zhu Q, Liu Y. Mitochondrial DNA study in domestic chicken. *Mitochondrial DNA A*. 2017;28(1):25–29.
11. Liu Y-P, Wu G-S, Yao Y-G, et al. Multiple maternal origins of chickens: out of the Asian jungles. *Mol Phylogenet Evol*. 2006;38(1):12–19.
12. Storey AA, Athens JS, Bryant D, et al. Investigating the global dispersal of chickens in prehistory using ancient mitochondrial DNA signatures. *PLOS One*. 2012;7(7):e39171.
13. Miao Y-W, Peng M-S, Wu G-S, et al. Chicken domestication: an updated perspective based on mitochondrial genomes. *Heredity*. 2013;110(3):277–282.
14. Di Lorenzo P, Ceccobelli S, Panella F, Attard G, Lasagna E. The role of mitochondrial DNA to determine the origin of domestic chicken. *Worlds Poultry Sci J*. 2015;71(2):311–318.
15. Graves JAM. Evolution of vertebrate sex chromosomes and dosage compensation. *Nat Rev Genet*. 2016;17(1):33–46.
16. Smeds L, Warmuth V, Bolivar P, et al. Evolutionary analysis of the female-specific avian W chromosome. *Nat Commun*. 2015;6:7330.
17. Al-Jumaili AS, Boudali SF, Kebede A, et al. The maternal origin of indigenous domestic chicken from the Middle East, the north and the horn of Africa. *BMC Genet*. 2020;21(1):30.
18. Boudali SF, Al-Jumaili AS, Bouandas A, et al. Maternal origin and genetic diversity of Algerian domestic chicken (*Gallus gallus domesticus*) from North-Western Africa based on mitochondrial DNA analysis. *Anim Biotechnol*. 2020;31:1–11.
19. Lawal RA, Al-Atiyat RM, Aljumaah RS, Silva P, Mwacharo JM, Hanotte O. Whole-genome resequencing of red junglefowl and indigenous village chicken reveal new insights on the genome dynamics of the species. *Front Genet*. 2018;9:264.
20. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2010;26(5):589–595.
21. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303.
22. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491–498.
23. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*. 2013;43(1):10–11.
24. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–2079.
25. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009;25(11):1451–1452.
26. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 1997;25(24):4876–4882.
27. Leigh JW, Bryant D. Popart: full-feature software for haplotype network construction. *Methods Ecol Evol*. 2015;6(9):1110–1116.

28. Bandelt H-J, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999;16(1):37–48.
29. 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):564–567.
30. Rogers AR, Harpending H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol*. 1992;9(3):552–569.
31. Rogers AR, Fraley AE, Bamshad MJ, Watkins WS, Jorde LB. Mitochondrial mismatch analysis is insensitive to the mutational process. *Mol Biol Evol*. 1996; 13(7):895–902.
32. Fu Y-X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*. 1997;147(2):915–925.
33. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123(3):585–595.
34. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol*. 2005;22(5):1185–1192.
35. Bouckaert R, Heled J, Kühnert D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Comput Biol*. 2014;10(4):e1003537.
36. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst Biol*. 10:901–904.
37. Nishibori M, Shimogiri T, Hayashi T, Yasue H. Molecular evidence for hybridization of species in the genus Gallus except for *Gallus varius*. *Anim Genet*. 2005;36(5):367–375.
38. Axelsson E, Smith NGC, Sundström H, Berlin S, Ellegren H. Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. *Mol Biol Evol*. 2004;21(8):1538–1547.
39. Wilson Sayres MA. Genetic diversity on the sex chromosomes. *Genome Biol Evol*. 2018;10(4):1064–1078.