The origin and history of genetic diversity of local chickens in the Middle East and Africa

By

Ahmed S. Obaid Al-Jumaili

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

April 2019



University of Nottingham
UK | CHINA | MALAYSIA

Declaration

I, Ahmed Al-Jumaili hereby declare that this thesis has not been previously accepted in substance for any degree to the University of Nottingham or any other one. The work in this thesis is entirely by me except where references to other work have been duly acknowledged.

Ahmed Al-Jumaili

April 2019

Dedication

To *Mesopotamia* and the two rivers of my life (my parents)

Acknowledgement

Firstly, I would like to express my sincere gratitude to my supervisor Professor Olivier Hanotte for the continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study. My sincere appreciation also goes to the Iraqi Ministry of Higher Education and Scientific Research (MOHESR) for funding this PhD work and making my dream a reality.

I am grateful to a number of people, who have provided me through advice and assistance contributed to the accomplishment of this thesis. Special thanks go out to my co-supervisor Professor Francis Gilbert for his valuable comments and Dr Joram Mwacharo for helping in some of the analysis processes. The gratitude goes to the Iraqi Ministry of Agriculture/Animal resources general directorate/Genetic resources department, for collecting samples from different places around Iraq, especially Dr Abdulamir A. Essa, Dr Sahar A. Al-Bayatti and their samples collection team. I wish to acknowledge the support of my colleagues at University of Nottingham/School of Life Sciences in Professor Hanotte's lab, thanks for what you have done to me, and it was my pleasure to work with you.

Finally, this work would not have been possible without the motivation and encouragement of the members of my family. I would like to thank my parents for limitless guidance during my life journey, they are my reservoir of energy. I also wish to thank my sister for her kindness and my three brothers Ehssaan, Mohammed and Abu Baker, you guys make me stronger. Never forget to thank my small family, my wife and my lovely three kids, Rawan, Omar and Ali, thank you all.

Selected Conferences

Midlands Sequencing Consortium Meeting, University of Nottingham, UK. 7-9, September 2015. (Attended).

49th UK population Genetics meeting, University of Edinburgh, UK. 15-18, December 2015. (Attended)

3rd Livestock Genomics meeting, University of Cambridge, UK. 14-16, September 2016 Poster presentation on "Genetic Diversity and History of Iraqi Indigenous Chicken Populations"

36th International Society for Animal Genetics (ISAG) Conference, UCD, Dublin. 16-22, July 2017. Poster presentation on "MtDNA D-loop genetic diversity of Middle East and North African Indigenous Chicken"

School of Life Sciences Postgraduate Research Symposium, University of Nottingham, UK. 11-12, July 2018. Oral presentation on "Outside the centres of origin: maternal diversity of domestic indigenous chicken from the African Horn, the Middle East and the Arabian Peninsula"

37th International Society for Animal Genetics (ISAG) Conference, Lleida, Spain. 7-12, July 2019. Accepted abstract on "Autosomal genome evidence for introgression from other *Gallus* species into African and Middle East indigenous chicken"

Abstract

Among the various livestock animals, chickens and their products (meat and eggs) are becoming increasingly important in sustaining human livelihood. They also represent an important model species in evolutionary studies. Today, four main types of domestic chicken may be recognised, chicken commercial breeds, fancy breeds, inbred research lines and indigenous village chicken breeds or populations. The latter are supplying high-quality protein to poor communities, particularly in Africa and Asia, contributing to food security, poverty alleviation and the management of natural resources.

The importance of chicken arises from many features (e.g short generation time, facility to breed them in captivity) that make them an ideal model species in different areas including embryology, physiology and medicine (e.g. vaccine development). Closely associated with human society, they represent an important proxy for the understanding of the history of human communities.

The two aims of this thesis were (i) to evaluate different genetic markers for the understanding and unravelling of chicken history and (ii) to trace the migration routes of village chickens outside the probable domestication centre on the Indian subcontinent. The study involved analysis of samples of domestic chickens from Iraq, Ethiopia, Algeria, Libya, Pakistan and Saudi Arabia, together with samples of the four junglefowl species (Ceylon, Grey, Green and Red junglefowl). All our domestic chicken populations originate from the West and North-West of the Indian subcontinent, a putative centre of domestication for the species. We used full genome sequence information of high coverage (30 X) with all samples used in this study mapped against the reference genome *Galgal* 5.0.

The first marker investigated (**Chapter 2**) was the mitochondrial (*D*-loop) sequences, which is the hypervariable part of the mtDNA control region. This marker has been used in several studies, and it is today our main source of information to reconstruct the history of the species. In this part, 706 samples from the six countries, including countries where we were lacking any information regarding the diversity of this marker so far were used. The results support an Indian subcontinent origin for the populations investigated as well as both a terrestrial and a more recent maritime, across the Indian Ocean, dispersal route. These two routes of migration have affected today's genetic diversity of the chicken populations examined, although in different proportions. Furthermore, in some countries,

there was evidence of a decline in the maternal population effective size following expansion, while others showed a recent expansion without any decline. However, this genetic marker barely shows any within-country phylogeographic signal.

The next chapter (**Chapter 3**) examined if two other types of markers, the full mtDNA sequence and the W chromosome, may provide us with more information. As for the *D*-loop, they are maternally inherited. In this part, 243 full mtDNA and 157 W chromosome samples were analysed. The mtDNA samples included three countries (Iraq, Ethiopia and Saudi Arabia), while the W chromosome samples included just two (Iraq and Ethiopia), as only data from males were available for Saudi Arabian samples. As expected, the full mtDNA sequence data analysis provided us with more polymorphisms and more evidence for the presence of two routes of migration, terrestrial and maritime. In particular, for Ethiopia, it indicates that different populations may have witnessed different histories. It also shows a greater resolution of genetic diversity at the population level and suggests a decline in effective population size following an expansion in two countries (Iraq and Ethiopia). In contrast, data from the W chromosome had low variation to add any information on the origin of samples and their genetic diversity level. But we did observe some unexpectedly divergent haplotypes in some Ethiopian populations, which require further investigations.

In **Chapter 4**, 315 whole-genome sequences of domestic chickens from three countries (Iraq, Ethiopia and Saudi Arabia) were used together with genome sequence information of four *Gallus* species (Ceylon, Green, Grey and Red junglefowl) and the common pheasant as an outgroup, to detect signals of introgression in domestic chicken. We also aim to assess if markers of introgression may help us to understand the history of the domestic chicken. The results showed no sign of gene flow from Ceylon and Green junglefowl into village chicken populations examined here. In contrast, there were many identified candidate genome regions of gene flow between Grey junglefowl and the domestic chicken populations. Building on previous work, six candidate introgressed genomic regions were further investigated. They all show evidence of Grey junglefowl introgression within domestic chicken as well as signs of gene flow in the opposite direction. These regions have different frequency levels of introgression in each country and also within each country. The highest level of gene flow is present in the north, central-east and western parts of Ethiopia, while the lowest levels of introgression were observed in Iraq, southern Ethiopia and Saudi Arabia. This variation in gene flow

frequencies provided us with new information regarding the past movement of chicken into Africa and, in particular, led us to conclude that the chicken diversity observed in Ethiopia today were brought from the southern part of the Indian subcontinent, while the ones from Iraq and Saudi Arabia originate from the Northern part of the Indian subcontinent.

In conclusion, the results presented here strongly advocate the use of both full mitochondrial DNA sequence information and markers of introgression to unravel the past history of the domestic chicken. However, the information obtained can only provide a window on such history, they allow us to build hypothesis which may be further investigated through the analysis of more populations over large geographic areas and, as much as possible, ancient chicken DNA studies.

Table of Contents

Declaration i
Dedication ii
Acknowledgement iii
Selected Conferences iv
Abstract v
Table of Contents viii
List of Figures xi
List of Tables xiv
Abbreviations xv
Chapter 1 1
The junglefowls 1
The genus *Gallus* 2
Ceylon junglefowl (*Gallus lafayetii*) 3
Green Junglefowl (*Gallus varius*) 3
Grey junglefowl (*Gallus sonneratii*) 4
Red junglefowl (*Gallus gallus*) 5
Subspecies of *G. gallus* 6
Genetic background of domestic chicken 8
Chicken domestication: Archaeological Evidence 9
Chicken domestication: molecular evidence 10
Dispersal of domesticated chicken: archaeological and iconographic evidence 10
Dispersal of domesticated chicken: molecular evidence 13
Genetic markers to study domestic chicken dispersion 14
Chapter 2 16
The maternal origin of indigenous domestic chicken from the Middle East, North and the Horn of Africa 16
Abstract 17
Introduction 17
Materials and Methods 19
Collection of samples and genomic DNA isolation 19
DNA extraction, PCR amplification and sequence alignment 21
Genetic diversity estimation 22
Phylogenetic analysis 22
Analysis of Molecular Variance (AMOVA) 22

Neutrality test and demographic dynamics 23
Results 23
D-loop haplotype variation and genetic diversity 23
Phylogeographic analysis 27
Analysis of Molecular Variance (AMOVA) 31
Population history and demographic dynamics 32
Discussion 38
Chapter 3 43
The usefulness of maternally inherited genetic markers for phylogeographic studies in village chicken 43
Abstract 44
Introduction 44
Materials and Methods 46
Collection of samples and genomic DNA isolation 46
Sequencing, alignment and variants calling 47
Distribution of SNPs and genetic diversity 48
MtDNA 48
W chromosome 48
Phylogeographic analysis 49
Analysis of Molecular Variance (AMOVA) 50
Neutrality test and demographic history 50
Results 50
Distribution of SNPs - MtDNA 50
Distribution of SNPs - W chromosome 51
Genetic diversity - MtDNA 52
Genetic diversity - W chromosome 54
Within-country phylogenetic relationships 55
Iraq 55
Ethiopia 59
Saudi Arabia 63
Analysis of Molecular Variances (AMOVA) 65
Population history and demography 66
Discussion 70
Chapter 4 75

Autosomal genome evidence for introgression from other *Gallus* species in the genome of African and Middle East indigenous chicken 75
Abstract 76
Introduction 77
Materials and Methods 78
Collection of samples and genomic DNA isolation 78
Sequencing, alignment and variants calling 79
Dataset structure and phylogenetic relationships 79
Four-taxon introgression analysis 80
Haplotype phylogenetic tree, introgressed region size and haplotype frequencies 81
Remapping introgressed regions from *Galgal* 5.0 to *Galgal* 6.0 81
Functional annotation of introgressed regions 81
Results 82
Variants and genetic diversity 82
Species structure and variation 85
Phylogenetic relationship among species 85
Introgression Analysis 87
Introgression between Ceylon junglefowl and domestic chicken 87
Introgression between Green junglefowl and domestic chicken 92
Introgression between Grey junglefowl and domestic chicken 96
Discussion 108
Chapter 5 111
General discussion and future prospect 111
References 115
Appendices 124

List of Figures

Figure 1.1: The geographical distribution of the genus *Gallus*…….....2
Figure 1.2: Ceylon junglefowl *Gallus lafayetii*.....3
Figure 1.3: Green junglefowl *Gallus varius*…….....4
Figure 1.4: Grey junglefowl *Gallus sonneratii*.....5
Figure 1.5: Red junglefowl (*Gallus gallus*).....6
Figure 1.6: Possible chicken dispersal route from the putative domestication centres to West Asia and the World…….....11
Figure 1.7: Possible maritime route of chicken dispersion following old trading network on the Indian Ocean.....13
Figure 2.1: Sampling locations and their grouping in regions of the countries included in this study.....20
Figure 2.2: Maximum likelihood tree for the 88 haplotypes and the references.....28
Figure 2.3: Median-Joining Network for Algeria, Ethiopia, Iraq, Libya, Pakistan and Saudi Arabia haplotypes.....29
Figure 2.4: Mismatch distribution patterns for regions and countries.....35
Figure 2.5: Bayesian Skyline Plot (BSP) for the countries (A) and Haplogroups (B)..38
Figure 3.1: Sampling locations for the countries included in this study.....46
Figure 3.2: Distribution of SNPs along the full mtDNA sequences for the three countries included in this study.....51
Figure 3.3: Distribution of SNPs (for every 500 kb) along the full W chromosome sequences of two countries included in this study.....52
Figure 3.4: Maximum likelihood unrooted tree for the 17 Iraqi mtDNA haplotypes……56
Figure 3.5: Median-joining haplotype network of Iraqi mtDNA haplotypes.....57
Figure 3.6: Maximum likelihood unrooted tree for Iraqi W chromosome haplotypes……...58
Figure 3.7: Median-joining haplotype network of Iraqi W chromosome haplotypes…….…..59
Figure 3.8: Maximum likelihood unrooted tree for Ethiopian mtDNA haplotypes.....60
Figure 3.9: Median-joining network of Ethiopian mtDNA haplotypes.....61

Figure 3.10: Maximum likelihood unrooted tree for Ethiopian W chromosome samples.....62
Figure 3.11: Median-joining haplotype network for the Ethiopian W chromosome sequences.....63
Figure 3.12: Maximum likelihood unrooted tree for the five Saudi mtDNA haplotypes.....64
Figure 3.13: Median-joining haplotype network of Saudi mtDNA haplotypes.....65
Figure 3.14: Regions and countries mismatch distribution patterns for the full mtDNA sequences.....68
Figure 3.15: Bayesian Skyline Plot (BSP) for three countries included in this study.....70
Figure 4.1: Sampling locations of the countries included in this study.....79
Figure 4.2: Principal Component Analysis for 315 domestic chicken.....85
Figure 4.3: Maximum likelihood tree for the full mtDNA.....86
Figure 4.4: Autosome maximum likelihood tree using 89,000 SNPs.....87
Figure 4.5: Manhattan introgression plots between domestic chicken and Ceylon junglefowl.....88
Figure 4.6: Haplotype maximum likelihood tree for a Ceylon junglefowl candidate introgressed region.....89
Figure 4.7: Haplotype maximum likelihood tree for a Ceylon junglefowl candidate introgressed region.....90
Figure 4.8: Haplotype maximum likelihood tree for a Ceylon junglefowl candidate introgressed region.....91
Figure 4.9: Manhattan introgression plots between domestic chicken and Green junglefowl.....92
Figure 4.10: Haplotype maximum likelihood tree for a Green junglefowl candidate introgressed region.....93
Figure 4.11: Haplotype maximum likelihood tree for a Green junglefowl candidate introgressed region.....94
Figure 4.12: Haplotype maximum likelihood tree for a Green junglefowl candidate introgressed region.....95
Figure 4.13: Introgression plots between domestic chicken and Grey junglefowl.....96

Figure 4.14: Haplotype maximum likelihood tree for the introgressed region.....98
Figure 4.15: Haplotype maximum likelihood tree for a candidate introgressed region.....99
Figure 4.16: Haplotype maximum likelihood tree for the candidate introgressed region.....100
Figure 4.17: Geographical distribution of introgressed haplotypes.....107

List of Tables

Table 1.1: Details of samples from archaeological sites in China.....9
Table 2.1: Location, sample size and genetic diversity of D-loop.....25
Table 2.2: Distribution of haplotypes across haplogroups and countries.....29
Table 2.3: Analysis of Molecular Variance.....32
Table 2.4: Neutrality and demographic expansion parameters.....33
Table 3.1: Full mtDNA reference sequences included in this study.....49
Table 3.2: Location, sample size and genetic diversity of full mtDNA.....53
Table 3.3: Location, sample size and genetic diversity of W.....54
Table 3.4: Analysis of Molecular Variances.....66
Table 3.5: Neutrality and demographic expansion parameters.....67
Table 4.1: Location, sample size and number of variants.....82
Table 4.2: Locations, number of samples and genetic diversity parameters.....84
Table 4.3: Detected introgressed regions between Grey junglefowl and domestic chicken.....97
Table 4.4: Frequency of introgressed haplotypes from Grey junglefowl to Iraqi domestic chicken.....101
Table 4.5: Frequency of introgressed haplotypes from Grey junglefowl into northern Ethiopian domestic chicken.....102
Table 4.6: Frequency of introgressed haplotypes from Grey junglefowl into central-eastern Ethiopia domestic chicken.....103
Table 4.7: Frequency of introgressed haplotypes from Grey junglefowl into western Ethiopia domestic chicken.....104
Table 4.8: Frequency of introgressed haplotypes from Grey junglefowl into southern Ethiopian domestic chicken.....105
Table 4.9: Frequency of introgressed haplotypes from Grey junglefowl into Saudi Arabian domestic chicken.....106

Abbreviations

MtDNA	Mitochondrial DNA
bp	base pair
PCR	Polymerase Chain Reaction
FTA cards	Flinders Technology Associates cards
GATK	Genome Analysis Toolkit
Hd	Haplotype diversity
π	Nucleotide diversity
MJ	Median Joining
AMOVA	Analysis of Molecular Variance
MAE	Mean Absolute Error
BSPs	Bayesian Skyline Plots
BEAST	Bayesian Evolutionary Analysis by Sampling Trees
MCMC	Markov Chain Monte Carlo
PAR	Pseudo Autosomal Regions
Indel	Insertion/deletion
BQSR	Base Quality Score Recalibration
VQSR	Variant Quality Score Recalibration
F1	First generation
PCA	Principal Component Analysis
AIC	Akaike Information Criterion

Chapter 1

The junglefowls

The genus *Gallus*

This genus belongs to the order Galliformes and family Phasianidae, with a separated branch among the pheasants (Sibley & Monroe, 1990). It shows different phenotypic traits compared to the pheasants. In particular, it has a fleshy large head with a high comb and one or two lappets, an almost naked face and throat, rounded wings and legs longer than the middle toe (Delacour, 1977).

The natural habitat of the wild junglefowls is the warmer parts of Asia at low to moderate altitude, mainly the rainforest of the Indian subcontinent and Southeast Asia (Harrison, 1978). Its large geographic distribution associated with geographic isolation of the species within the genus has led to four species becoming morphologically distinct (**Figure 1.1**). Phenotypic differences include feather colour and pattern, comb shape and colour, voice and behaviour (Crawford, 1990). Depending on these differences, there are four wild living species within the genus *Gallus*: Ceylon junglefowl (*Gallus lafayetii*), Green junglefowl (*Gallus varius*), Grey junglefowl (*Gallus sonneratii*) and Red junglefowl (*Gallus gallus*). The main characteristics for these species are described below.

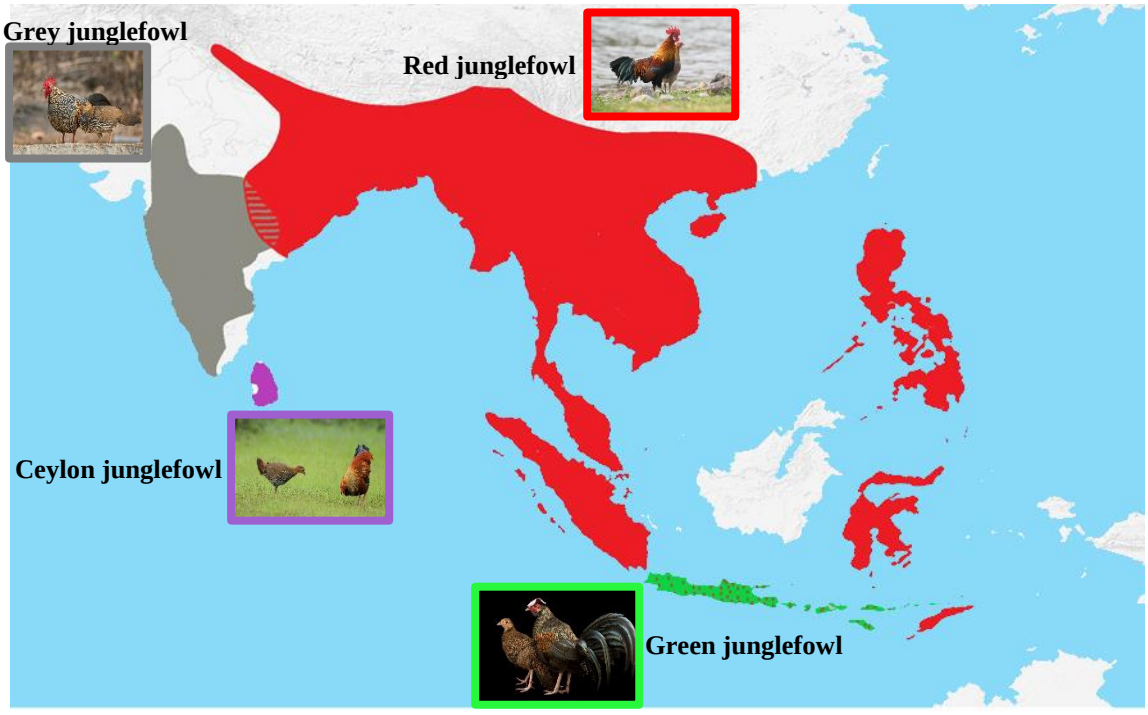


Figure 1.1: The geographical distribution of the genus *Gallus*. (Map from 'Handbook of birds of the world alive' (https://www.hbw.com/).

Ceylon junglefowl (*Gallus lafayetii*)

Gallus lafayetii is only found in Sri Lanka and therefore it is often referred to as the Ceylon or Sri Lanka junglefowl (Tixier-Boichard *et al.*, 2011). Sawai *et al.* (2010)'s study suggests a close similarity between the Sri Lanka and the Red junglefowl, but other studies dispute this (Nishibori *et al.*, 2005).

Delacour (1977) mentioned the morphological characteristics of the male and the female. The male has a border of large purple feathers around the throat with broad feathers at the lower back and rump (**Figure 1.2**). The wings and the tail are metallic dark purplish, and it has a black abdomen with yellow legs; the length of the cock bird ranges from 660 to 725 mm, with wings 216-240 mm long, and the tail 230-400 mm long. The morphology of females was illustrated in Delacour (1977), who indicated that the female has brown crown with dark brown neck, the wings striped black, and the abdomen is white with brownish yellow legs. The length of the female is 350 mm, the wing 170-180 mm, and the tail 110 mm. The author also mentioned that the main habitat of Ceylon junglefowl does not differ from that of the Red junglefowl, i.e. the damp forests of the mountains and the dry coastal scrub.



Figure 1.2: Ceylon junglefowl *Gallus Lafayetii* (credit: www.arkive.org)

Green Junglefowl (*Gallus varius*)

This wild species is found on the Island of Java and nearby Islands. It is considered to be the most distant wild junglefowl from the domestic chicken (Tixier-Boichard *et al.*,

2011). Del Hoyo *et al.* (2017) referred to *Gallus varius* as a bird of seashore, valleys and the lowlands, and to be more common in the western part of Java. The male comb has no indentations and is green near the head with red colour on the outer edge (**Figure 1.3**); the body feathers are glossy greenish black with narrow light-yellow borders. The tail contains sixteen rectrices feathers that are shiny dark blue. This species of chicken has a single three-colour wattle (red, yellow and blue) (Tixier-Boichard *et al.*, 2011). The male is 700 mm long, with the length of the wing ranging from 220 to 245 mm, and the tail from 320 to 330 mm. The female has a brown crown and cheeks, while the upper parts are glossy brownish black; the tail feathers are black with metallic stains on the borders, the breast is faint brown and the legs tend to be greyish white to yellowish pink. Females are approximately 400 mm long, with the length of the wing 195 mm, and the tail 115 mm (Delacour, 1977).



Figure 1.3: Green junglefowl *Gallus varius* (credit: www.joelsartore.com)

Grey junglefowl (*Gallus sonneratii*)

It is found in the south-western part of the Indian subcontinent (Tixier-Boichard *et al.*, 2011). Since the study of Eriksson *et al.* (2008), we know that this species has successfully introgressed with the domestic chicken, and is responsible for the yellow skin observed in some domestic birds. This bird is considered to be the sister species of the Ceylon junglefowl (Fumihito *et al.*, 1996; Nishibori *et al.*, 2005). **Figure 1.4** shows

the phenotypic traits of the male and female *Gallus sonneratii*. According to the description of Delacour (1977), the male has purplish-black back feathers with narrow grey shafts and borders, and the tail is glossy purplish-black. The wing coverts are black with white and long shafts, while the rest of the wing is black. The legs are yellow to pink. The size of the male ranges from 700 to 800 mm, with wing length 220-255 mm, and tail length 330-390 mm. During the summer season the males replace the long tail feathers with shorter ones. Females have a brown crown with reddish shaft lines, a faintly brown face, and white breast feathers with brown edges. The size of the female is about 380 mm, with wing lengths ranging from 200 to 215 mm, and tail lengths 130-170 mm.

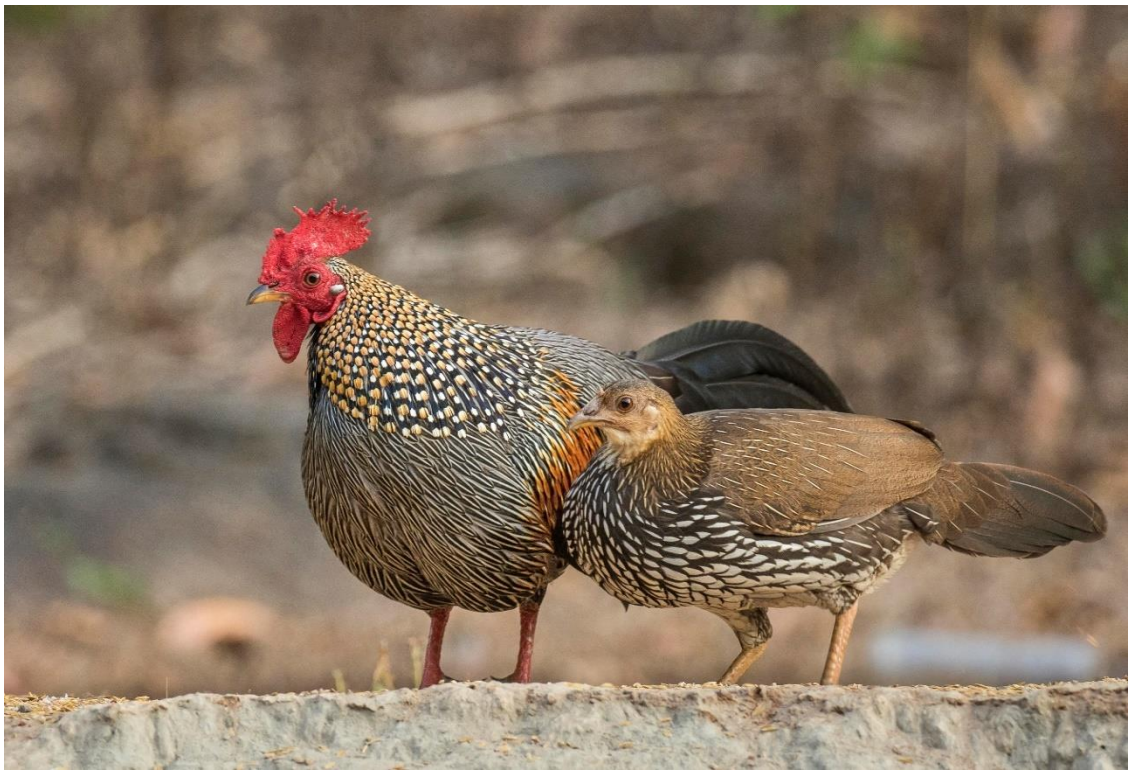


Figure 1.4: Grey junglefowl *Gallus sonneratii* (credit: http://rahulalvares.com)

Red junglefowl (*Gallus gallus*)

Also known previously as *Gallus bankiva* or *Gallus ferrugineus*, this wild species is the main ancestor of domestic chicken *Gallus gallus domesticus* and can have fully fertile offspring when crossed together with the domestic form (**Figure 1.5**) (Darwin, 1868; Morejohn, 1968; Crawford, 1990; Fumihito *et al.*, 1996). Consequently, *Gallus gallus* is considered to be the closest relative of the domestic chicken (Darwin, 1868). The main homeland for the Red junglefowl is humid and sub-humid forests with a large geographic distribution area across tropical Asia (Tixier-Boichard *et al.*, 2011). It can be found in north-west India through Assam, Burma, Siam, Thailand, Malaysia, east to the Indo-

China region, the southern part of China, Hainan, Sumatra and Java (Delacour, 1977). These areas are the main habitat of five Red junglefowl subspecies defined by slight differences in morphology and plumage and their geographical distribution (Delacour, 1977; Johnsgard, 2002). Delacour (1977) offered a comprehensive summary of the phenotypic traits of the different Red junglefowl subspecies, and described them as gathering in larger flocks than the other species and laying more eggs compared to the Ceylon and Grey junglefowls. The cock is large and has a high comb on the head. The tail has sixteen rectrices feathers, and the wings are rounded. The female is smaller than the male. The face is naked and light red, and the colour of the hen bird is dotted brown with different markings of orange, yellow, white and black; it has no spur (Maguelonne, 2009). The Red junglefowl subspecies are described below.



Figure 1.5: Red junglefowl (*Gallus gallus*) (credit: www.thainationalparks.com)

Subspecies of *G. gallus*

Gallus gallus gallus

Known also as the Cochin-Chinese red junglefowl, this subspecies of the Red junglefowl is found in South Vietnam, Cambodia, Thailand and Laos; it has white earlobes, a large dentate comb and has some differences in plumage compared with the other subspecies (Delacour, 1977; Tixier-Boichard *et al.*, 2011). It is widely believed that this Cochin-

Chinese subspecies has contributed significantly to the genetic background of domestic chicken, more so than the other subspecies (Fumihito *et al.*, 1994; Liu *et al.*, 2006). There are clear morphological differences between the male and the female. The male has milk-white earlobes with crown and long neck feathers, a large dentate comb, and males range in size from 650 to 750 mm, with wings 230-250 mm long and a tail 260-275 mm long. The female is characterized by small bluish-white earlobes with the comb reduced to small fold. Female range in size from 420 to 460 mm long, with wings 185-200 mm long and a tail 140-155 mm long (Delacour, 1977).

Gallus gallus spadiceus

Known as the Burmese red junglefowl, the main territory for this bird is in Yunnan province in China, Myanmar, Malaysia, northern Laos, Sumatra and Thailand. The main feature is the red colour of the earlobes and the fancy feathers (Tixier-Boichard *et al.*, 2011). This bird lives sympatrically with *Gallus gallus jabouillei* in Southeast Asia, where the probable hybridization happened. The phenotypic traits of the male are similar to those of *G. g. gallus*, but the fancy feathers are shorter and the ear folds are much smaller, while the female differs in the neck feathers, which are shorter and pale (Delacour, 1977).

Gallus gallus jabouillei

The main homeland of the Tonkinese red junglefowl is in North Vietnam, Hainan and Yunnan in South China. It is characterized by its red earlobes (Tixier-Boichard *et al.*, 2011). The main traits for the male are the darker parts of the red plumage, and the shorter and less pointed ornamental feathers, while the female is darker than *G. g. gallus* and the neck feathers are deep yellow (Delacour, 1977).

Gallus gallus murghi

The habitat of Indian red junglefowl is north-east India, Nepal, Bhutan, Bengal, Ganjam and Bangladesh. The main distinguishing feature of this subspecies is the white earlobes (Tixier-Boichard *et al.*, 2011). The male has yellow neck feathers, more yellow than in *G. spadiceus*, while the female is slightly lighter in colour than the *spadiceus* female. This subspecies can be easily domesticated and it is the most frequently found in collections (Delacour, 1977).

Gallus gallus bankiva

The main geographic region for the Javan red junglefowl is Sumatra, Bali and Java. It is characterized by its red earlobes and a distinctive neck feather (Tixier-Boichard *et al.*, 2011). The male is different in feather shape from other subspecies. Its neck feathers are broader and rounded at the end, and the female also has rounded neck feathers as in the male (Delacour, 1977). No solid evidence has been found to support any contribution of this subspecies to the genetic background of the domestic chicken (Liu *et al.*, 2006; Tixier-Boichard *et al.*, 2011), except a recent study (Lawal, 2018) that mentioned several genomic regions have been introgressed from *G. gallus bankiva* to domestic chicken.

Genetic background of domestic chicken

The origin of the domestic chicken is a debatable subject with both a monophyletic and a polyphyletic species origin suggested. Crawford (1990) mentioned in his book that the archaeological evidence cannot determine the genetic contribution of *Gallus* species to the genetic pool of the domestic chicken due to the similarity of their bones. For Darwin (1868), Hashiguchi (1981) and Fumihito *et al.* (1994) the Red junglefowl was the only ancestor of domestic chicken, an opinion further supported by Punnett (1923), who showed that crossings between the Red junglefowl and the domestic chicken lead to fertile offspring, while crossing with other wild *Gallus* species generally fail to produce any progeny. However, Morejohn (1968) succeeded in producing F2 progeny from the crossing and backcrossing of Red and Grey junglefowl.

This situation prevailed until the study of Nishibori *et al.* (2005), where the analyses of full sequencing of mtDNA and some parts of the nuclear genome showed evidence of introgression among domestic chickens, Red junglefowl and Grey junglefowl. Such introgression have been confirmed by Eriksson *et al.* (2008), when they found that the BCDO2 gene, the gene responsible for the yellow skin phenotype in domestic chickens, originated from *G. sonneratii* rather than from *G. gallus*. More recently, a study by Lawal (2018) used full-genome sequences to trace signs of introgression, and found that while the Red junglefowl is the main ancestor of *Gallus gallus domesticus*, all other *Gallus* species (especially the Grey junglefowl) have contributed to the genetic pool of domestic chicken.

Chicken domestication: Archaeological Evidence

The chicken domestication process remains unclear. From an archaeological viewpoint, in 2000 BC people of the Indus valley kept chicken for cockfighting and entertainment. The bones found in Mohenjo-Daro were longer in size compared to the bones of wild junglefowl, but smaller than commercial chicken bones (Zeuner, 1963). The longer size of chicken bones in the Indus valley supports the idea that these bones may have belonged to domestic chicken. Domestic animal bones were found from all levels of the Mohenjo-Daro site (Sewell & Guha, 1931), but chicken remains were found only in the upper levels. Zeuner (1963) speculated from this that the domestication period in the Indus Valley occurred around 2000 BC. It was widely believed that the chicken spread from the Indus valley into Greece through Mesopotamia. However, West and Zhou (1988)'s study concluded that the domestication happened in South-East Asia, and then the chicken went north to China and may have then dispersed to Europe following a northern route. These authors mentioned that domestic chicken in India are more recent than those of South-East Asia, and that their presence may be explained in two ways: independent domestication or the diffusion of chicken from South-East Asia. Their evidence relies on excavation dates in Asia and Europe. Most of them indicate the presence of domestic chicken before the sixth millennium BC in South-East Asia and in China by 6000 BC (West & Zhou, 1988). It is obvious from them following the study of 29 sites that the domestication happened in South-East Asia and China earlier than in Mohenjo-Daro. Hence, we can no longer accept that the Indus valley was the only place for chicken domestication. The study of Xiang *et al.* (2014) of several archaeological sites (**Table 1.1**) in China used the direct accelerator mass spectrometry radiocarbon dating procedure to determine the approximate age of faunal chicken remains. The outcome of their study supports the suggestion that chicken domestication happened as early as ~10,000 years ago, although the results of this study have now been challenged (Peters *et al.*, 2015).

Table 1.1: Details of samples from archaeological sites in China studied by Xiang *et al.* (2014).

Location	Sample size	Age (years)	Source
Nanzhuangtou site, Hebei	22	10,430-10,465	Xiang *et al*. (2014)
Cishan site, Hebei	7	7,845-7,960	Xiang *et al*. (2014)
Wangyin site, Shandong	6	3,500-4,500	Xiang *et al*. (2014)

Jiuliandun Chu tombs, Hubei	4	2,300-3,000	Xiang *et al*. (2014)

Chicken domestication: molecular evidence

Most studies have used mitochondrial DNA (mtDNA) to unravel the process of chicken domestication. The studies of Fumihito *et al.* (1994) and Fumihito *et al.* (1996) support a chicken domestication in Thailand and adjacent regions. These two studies are supported by a more recent study (Riztyan *et al.*, 2014), which, although using the different approach of SNP markers instead of mtDNA sequencing, reached the same conclusion in favour of Thailand as the centre of chicken domestication. But a common weakness among these research studies is that they did not include samples from the Indian subcontinent and China. In contrast, Liu *et al.* (2006)'s study included 900 mtDNA sequences from domestic chicken and several red junglefowl subspecies from a wide geographic area including China, South East Asia and to some extent South Asia. They found 169 distinct haplotypes that they grouped in 9 clades labelled A-I. Seven out of these nine clades included both samples from both wild and domestic chicken. Clade E was the commonest in Europe, Middle East and India, while the two clades A and B were found in both China and Japan. Of the other clades, clade C was found in Japan and Southeast China, and clades F and G may have originated from the northern part of the Yunnan Province. The most dispersed clade was D, perhaps linked to the dispersal of cockfighting birds. Additionally, clade I was found in Vietnam and clade H was present in Red junglefowl only. The data from this paper and a subsequent paper (Kanginakudru *et al.*, 2008) support independent chicken domestication in South and South-East Asia. More recently, a study (Miao *et al.*, 2013) using full mitochondrial DNA analysis has shown that several local domestication events may have happened in south Asia, South-West China and south-east Asia with distinct haplogroups (W-Z) added to the findings of Liu *et al.* (2006)'s study.

Dispersal of domesticated chicken: archaeological and iconographic evidence

The chicken is largely a terrestrial bird that does not migrate for long distances (Johnsgard, 2002). Furthermore, these birds do not have the ability to swim (Collias & Collias, 1996; Kelly, 2006). Therefore, the dispersal of chicken may be fully attributed to human interventions (Storey *et al.*, 2012; Mwacharo *et al.*, 2013).

The dispersal routes of chicken from its domestication centres to West Asia remain largely unknown, but probably took place around the third-second millennium BC depending on archaeological and iconographic evidence (**Figure 1.6**) (Perry-Gal *et al.*, 2015). Bennecke (1994) showed that the first faunal remains found in Iran, Anatolia and Syria dated back to the third millennium. According to Liu *et al.* (2006), the chicken is the only domesticated species of the Old World that has no origin in the Middle East. Most studies confirm that the first stage of chicken dispersal happened from the domestication centres to west Asia, especially into the Middle East (Iraq and Iran) (Liu *et al.*, 2006; Perry-Gal *et al.*, 2015). Perry-Gal *et al.* (2015) considered the Fertile Crescent as the area of origin for dispersal to Europe through Greece, from archaeological evidence. Carter (1923) mentioned that the Sumerian civilization in Iraq called the chicken "egg-bird", and the words Tarru and Tarlugallu became in Syriac *Gallus*, the cock. This study also showed that the bird was known in Babylon in the early Sumerian period before 2500 B.C., and the Greek named the cock as "Akkadian" (coming from Akkad in Iraq), suggesting an origin from Iraq to Europe through Syria. This date of chicken occurrence in Iraq is also supported by Houlihan and Goodman (1988), who reported textual references to the chicken from the third Dynasty of Ur in about 2000 BC, and also mention the presence of chicken in Jordan by 1200 BC.

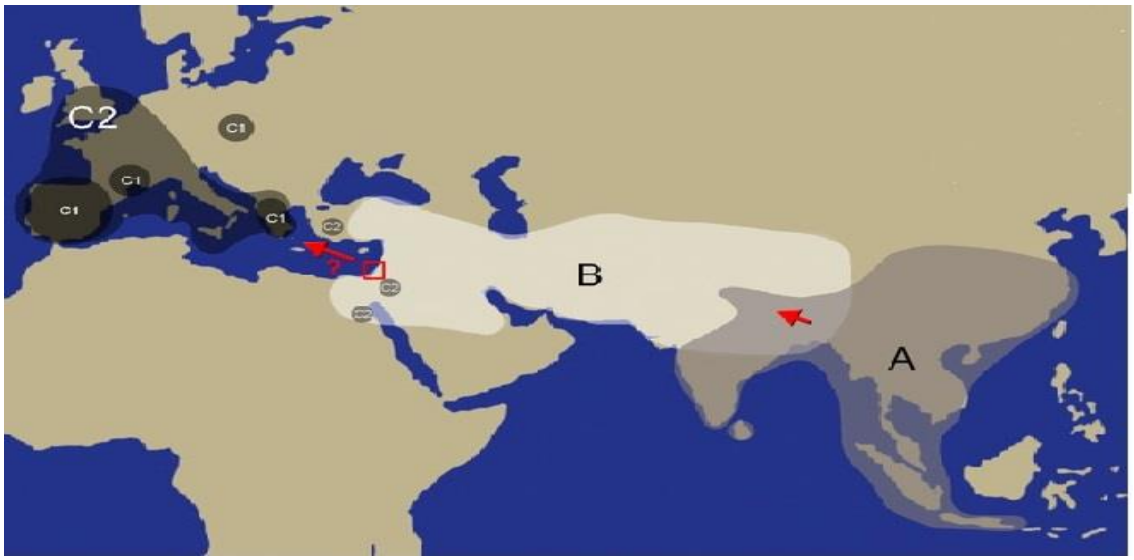


Figure 1.6: Possible chicken dispersal route from the putative domestication centres to West Asia and the World. A = chicken domestication centres, B = chicken dispersal to West Asia and C1, C2 = chicken migratory waves to Europe and Africa (Perry-Gal *et al.*, 2015).

Williamson *et al.* (2000) pointed out that domestic fowl reached Africa through two routes; one through the Sahara from the north, and the second across Central Africa from the east. The northern route was through Egypt down the Nile River into the south. The date of chicken presence in Egypt is debatable (Redding, 2015). Some argued that chicken were introduced but then vanished, only to be re-introduced during the Ptolemaic period (Zeuner, 1963). Others indicate that the discovery of a drawing of a cock on a limestone flake near Tutankhamun's tomb is not a solid evidence of chicken introduction to Egypt (MacDonald & Edwards, 1993), they also speculated that chicken were rare or absent before 300 BC in Egypt. Redding (2015) finally suggested that the stone from Tutankhamun's tomb and the proof of chicken presence in Jordan dated back to 1200 BC could indicate that the time of chicken introduction in Egypt was 1120 BC or earlier.

The second route of chicken arrival to Africa (the maritime route) through the east coast. This suggestion involved chicken moving from the domestication centres using the ancient maritime trading networks of the Indian Ocean after the mid-first millennium CE, to enter Africa through the east (Prendergast *et al.*, 2017). Accordingly, this route has been considered as a more recent diffusion route compared with the northern route (**Figure 1.7**). The suggestion is supported by the findings of another study (Woldekiros & D'Andrea, 2017), which found that the oldest archaeological evidence (from the first millennium BC) in the Horn of Africa were from a pre-Aksumite site dated to around 800–450 BC. The findings of this paper endorse other evidence concerning the maritime route of chicken introduction to Africa, and the maritime exchange networks with the Arabian Peninsula through ports along the Red Sea coast (Galli, 2017).

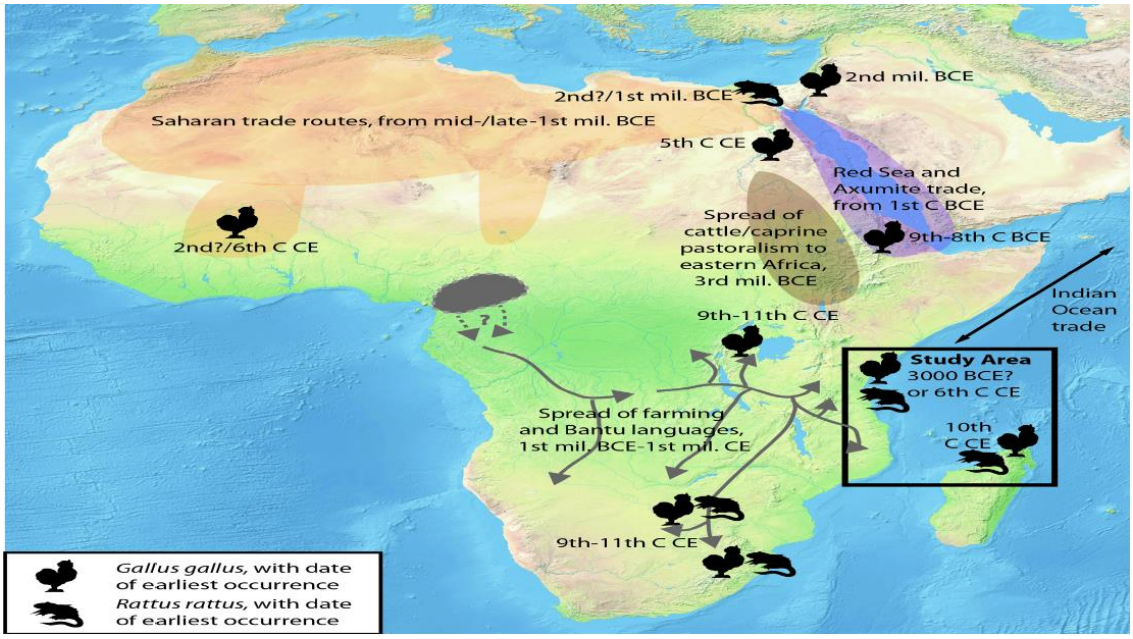


Figure 1.7: Possible maritime route of chicken dispersion following old trading network on the Indian Ocean (Prendergast *et al.*, 2017).

Dispersal of domesticated chicken: molecular evidence

Genetic markers especially mtDNA are used to identify livestock domestication centres and routes of animal dispersal (Shaista *et al.*, 2017). Liu *et al.* (2006) mentioned in his molecular study that clade E is indigenous in India, and the majority of the samples from Europe and Middle East fall in the same clade. The results of this study could indicate that the village chicken in the Middle East and Europe came from the Indian subcontinent through human trade and migration routes. This conclusion was further supported by Perry-Gal *et al.* (2015), who suggested that the expansion of chicken from the domestication centres to Middle East and Europe likely occurred through Mesopotamia (the Iraq of today). After clade E, clade D is the second commonest and is the most widespread in India (Berthouly-Salazar *et al.*, 2010; Tixier-Boichard *et al.*, 2011). The origin of clade D might be south-east Asia and the surrounding regions (Komiyama *et al.*, 2004; Miao *et al.*, 2013). Another molecular study (Storey *et al.*, 2012), also on mtDNA, showed that haplogroup E was present in samples dated back to 3000 BP in the Pacific region, and 1000 BP in Europe. The same haplogroup (E) has also been noticed in Iran and Turkey alongside clade A (Meydan *et al.*, 2016).

In Africa, there are different mtDNA haplogroups; these haplogroups include A, B, C, D and E (Mwacharo *et al.*, 2013). The most frequent haplogroup in Africa is E, reported at high frequency in Madagascar, Zimbabwe, Sudan, Nigeria, Kenya, Ethiopia, Sudan,

Uganda and South Africa (Muchadeyi *et al.*, 2008; Razafindraibe *et al.*, 2008; Adebambo *et al.*, 2010; Mtileni *et al.*, 2011; Mwacharo *et al.*, 2011). Haplogroup D has been observed as the second commonest haplogroup in Madagascar, Zimbabwe, Malawi, Kenya and South Africa. It might have been introduced into the African continent following maritime trading networks between East Africa, the Indian subcontinent and South-East Asia. Clade A has been reported in Kenya and South Africa only. Clade B has been noticed in Ethiopia and South Africa, while clade F is found in Ethiopia, Sudan and South Africa. Depending on the information available from previous studies on the geographical distribution of different haplogroups identified in Africa, it is widely agreed that haplogroup E entered Africa from the North, likely through Egypt (Gifford-Gonzalez & Hanotte, 2011; Mwacharo *et al.*, 2011), with the arrival of chicken from the Indian subcontinent to Egypt through the Fertile Crescent. Haplogroups A, B and F were found in very low frequencies (Mwacharo *et al.*, 2011), and the presence of haplogroup C in East Africa could be a consequence of the maritime route introducing chicken to the area from Yunnan province in south China.

Genetic markers to study domestic chicken dispersion

Animal movement outside the domestication centre(s) has a major impact on the evolutionary background of a domesticated animals with genetic changes attributed to founder effect, gene flow, genetic drift and local adaptation (Saastamoinen *et al.*, 2018).

Different genetic markers have been used to trace the movement of domestic animals. MtDNA is widely used to infer the maternal history of chicken, since its sequence polymorphism can address maternal origin and domestication events (Liu *et al.*, 2006; Silva *et al.*, 2009; Miao *et al.*, 2013; Di Lorenzo *et al.*, 2015). It is also believed that chicken mtDNA can provide information about human-mediated dispersal out of the domestication centre(s) (Di Lorenzo *et al.*, 2015). The majority of studies have used the *D*-loop part of mtDNA to unravel the history of chicken (Komiyama *et al.*, 2003; Liu *et al.*, 2006; Adebambo *et al.*, 2010; Mwacharo *et al.*, 2011; Ramadan *et al.*, 2011; Wani *et al.*, 2014; Eltanany & Hemeda, 2016; Dyomin *et al.*, 2017), including the study of Storey *et al.* (2012) on the origin of South American chicken from ancient bones. Today, a number of studies have now relied on the analysis of the full mtDNA sequence to assess the diversity of the domestic chicken and wild junglefowl (Nishibori *et al.*, 2005; Miao *et al.*, 2013). Although mtDNA has been found to be a target for non-neutral processes and in linkage disequilibrium with some chromosomal regions (e.g. W chromosome)

under high selective pressure, its phylogeographic structure can be higher than nuclear loci and it can exhibit monophyly, simultaneously, that nuclear markers show poly- or paraphyly (Pluzhnikov *et al.*, 2002; Bensch *et al.*, 2006; Brito & Edwards, 2009).

Some scientists (Zhang & Hewitt, 2003) have mentioned that nuclear markers cannot completely replace mtDNA genetic marker in phylogeographic studies, because of difficulty in data analysis from recombination, and technical laboratory issues for resolving haplotypes. Therefore, nuclear and mtDNA markers should be considered as providing complementary information. Of particular interest might be nuclear markers of introgression across the genus *Gallus*, considering the geographic distribution of the different congeners. As indicated, Nishibori *et al.* (2005) used full mtDNA and two segments of nuclear genome hybridization among the genus *Gallus*, between the Red and the Grey junglefowl. Their result was confirmed by Eriksson *et al.* (2008)'s work, showing introgression at the BCDO2 locus from Grey junglefowl into domestic chicken. Lawal (2018) in his PhD thesis on introgression in *Gallus* used full genome sequences.

This thesis aims to contribute to our global understanding of the origin and migration routes of chicken using samples from the Middle East and Africa. Some assessment of the genetic diversity of these samples will be applied, by using different genetic markers (*D*-loop, full mtDNA, W chromosome and Autosomes). We will hence evaluate the most suitable genetic marker for chicken phylogeographic studies.

Chapter 2

The maternal origin of indigenous domestic chicken from the Middle East, North and the Horn of Africa

Abstract

Indigenous domestic chicken represents a major source of protein for agricultural communities around the world. In the Middle East and Africa, they are adapted to hot dry and semi-dry areas, in contrast to their wild ancestor, the Red junglefowl, which lives in humid and sub-humid tropical areas. Indigenous populations are declining following increased demand for poultry meat and eggs, favouring the more productive exotic commercial breeds. In this chapter, using the *D*-loop of mitochondrial DNA as a maternally inherited genetic marker, we address the question of the origin and dispersal routes of domestic chicken of the Middle East (Iraq and Saudi Arabia), the northern part of the African continent (Algeria and Libya) and the Horn of Africa (Ethiopia). In particular, we report the analysis of the mtDNA *D*-loop of 706 chicken samples from Iraq (n = 107), Saudi Arabia (n = 185), Algeria (n = 88), Libya (n = 23), Ethiopia (n = 211) and Pakistan (n = 92). We analysed a 397 bp fragment, which includes the hypervariable region of the *D*-loop. The results show the presence of five haplogroups (A, B, C, D and E; *sensu* Liu *et al*. 2006), suggesting more than one maternal origin for the studied populations. Haplogroup E, which occurred in 625 samples, is the most frequent in all countries. This haplogroup most likely originates from the Indian subcontinent, and probably migrated following a terrestrial route to these different countries. Haplotypes belonging to haplogroup D are present in all countries except Algeria and Libya: it is likely a legacy of the Indian Ocean maritime trading network. Haplogroup A is present in all countries, and may be of commercial origin. Haplogroup B was found only in Ethiopia. Haplogroup C was only detected in the south-west of Saudi Arabia and Ethiopia. The origin of these two haplogroups in the chicken examined here remains unclear. Most of the diversity occurs within rather than between populations within countries. These results support a major influence of the Indian subcontinent on the maternal diversity of today's chicken populations examined here. The lack of phylogeographic signal agrees with both ancient and more recent trading networks having shaped the modern-day diversity of indigenous chicken populations.

Introduction

Undoubtedly, village chickens are a valuable genetic resource for the countries around the world due to their adaptation to the local environment, including their higher resistance against endemic diseases. They supply high-quality protein to the increasing

numbers of people around the world, and represent a major source of income to poor communities. They therefore contribute greatly to food security, poverty alleviation and management of natural resources (Leta & Endalew, 2010). The main production system of indigenous chicken is scavenging or semi-scavenging, which relies on a low level of inputs. This system makes up to 80% of the poultry stocks in the developing countries of Asia and Africa (Gueye, 2009).

The Red junglefowl is the main ancestor of domestic chicken (Darwin, 1868). Its natural habitat is the sub-humid and humid tropical areas in south and south-east Asia. In contrast to the wild ancestor, village chickens have adapted very effectively to a diversity of environments including the arid and semi-arid areas. They show extensive morphological diversity that may be connected to adaptation to such hot and dry environments, including the naked-neck phenotype, small body size and frizzled plumage (Al-Rawi & Al-Athari, 2002; Al-Yousef, 2007; Dalhoum *et al.*, 2016). Recent genome studies have revealed candidate regions under positive selection which may be related to environmental adaptation in this species (Lawal *et al.*, 2018).

MtDNA analysis has been used intensively to unravel the history of domestic chickens (Dyomin *et al.*, 2017). Analysis of the mitochondrial DNA genome allows us to identify the wild ancestor(s) and the maternal lines that have contributed to a breed or population (Miao *et al.*, 2013; Di Lorenzo *et al.*, 2015). Furthermore, such genetic markers can offer valuable information concerning human-mediated dispersal out of the domestication centres of the chicken (Di Lorenzo *et al.*, 2015). Finally, mtDNA characterisation may help the establishment of effective management practices and sustainable strategies for the conservation of diversity.

This Chapter aims to unravel the history and diversity of indigenous chicken from the Middle East, North and the Horn of Africa. It includes chicken from: (i) Pakistan, a putative ancient centre of origin for domestic chicken in the northern part of the Indian subcontinent (Zeuner, 1963), (ii) Iraq, with its ancient Mesopotamian civilizations in contact with those of the Indus Valley, (iii) Saudi Arabia on the route to the African continent, (iv) the Horn of Africa, represented here by Ethiopia, where the oldest osteological evidence of domestic chicken for Africa have been found dated from pre-Aksumite time (Woldekiros & D'Andrea, 2017), and (v) Algeria and Libya, two countries bordering the Mediterranean Sea, which witnessed ancient Phoenician, Greek and Roman terrestrial and maritime trading networks between North Africa, the Near East

and Europe (Galli, 2017). A total of 706 mtDNA *D*-loop sequences and Asian haplotypes of reference from Liu *et al.* (2006) were analysed.

Materials and Methods

Collection of samples and genomic DNA isolation

This study was conducted on 706 samples. It includes chicken from Pakistan (n = 92), Iraq (n = 107), Libya (n = 23), Saudi Arabia (n = 185), Algeria (n = 88) and Ethiopia (n = 211) (**Figure 2.1, Table 2.1**). The Pakistani sequences were obtained from samples collected around Islamabad by Dr Arifa Naqvi. Algerian samples were provided by Professor Semir Gaouar from five places representing five populations; Iraqi samples were collected by the Genetic Resources Department of the Iraqi Agriculture Ministry under the supervision of Dr Abdulamir Essa and Dr Sahar Al-Bayatti; Saudi Arabian samples were provided by Dr Raed Alatiyat from King Saud University; and Libyan samples were collected by Dr Abulgasim Ahbara from Misurata University in Libya.

Iraqi samples were collected from five different areas (**Figure 2.1**) divided into three groups: (i) the north (Sulimania (n = 9) in the north-east part of Iraq; (ii) central, two sampling locations in the central region of Iraq, Baghdad (n = 51) and Karbala (n = 12); and (iii) south, two regions in the south-east and southern part of the country, Misan (n = 24) and Basra (n = 11). In Saudi Arabia, samples were collected from 17 sites divided into three groups, east (Al-Qatef, Hafar Al-Batin and Al-Hessa: n = 45), central (Hail, Al-Aflaj, Al-Kharj, Al-Amari and Unayzah: n = 43) and west (Abha, Al-Baha, Jeddah, Jazan, Mecca, Medina, Najran, Tabuk and Taif: n = 97) (**Figure 2.1**). Algerian chicken samples were divided into two groups, north (Mascara n = 20, Oran n = 17, Tiaret n = 11 and Tlemcen n = 18) and central (Adrar, n = 22). Due to the political situation, sampling in Libya was limited to one population representing the north-west along the Mediterranean Sea (**Figure 2.1**).

In Ethiopia, chicken samples were collected by Dr Adebabay Kebede (from the International Livestock Research Institute) from 19 different sites from both the highland and lowland parts of the country. Ten samples from each site were examined for all sites except four (Batambie n = 8, Horro n = 30, Jarso n = 14 and Surta n = 9). The Ethiopian populations were divided into four groups: north (Meseret and Mihquan), central-east (Adane, Arabo, Horro, Jarso, Midir and Negasi Amba), west (Ashuda, Amshi, Batambie, Dikuli, Gafera, Surta and Tzion Teguaz) and south (Kumato, Loya, Shubi Gemo and

Girissa) (**Figure 2.1**). All the samples included in this study were from free-range scavenging or semi-scavenging village chickens.

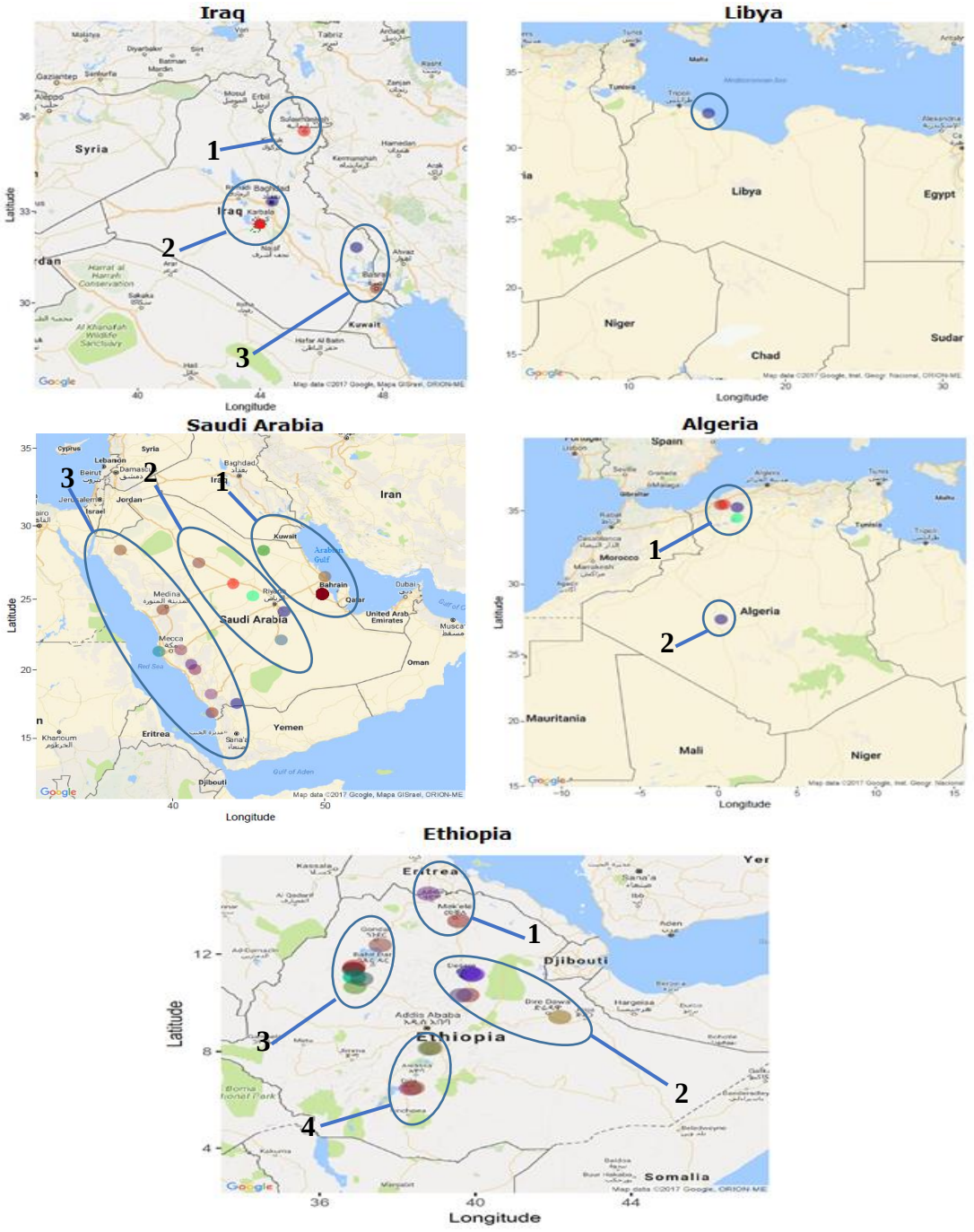


Figure 2.1: Sampling locations and their grouping in regions of the countries included in this study. For Iraq, 1 = North-East region, 2 = Central and 3 = South. Saudi Arabia, 1 = East, 2 = Middle and 3 = West. Algeria, 1 = North-West and 2 = Central. Ethiopia, 1 = North, 2 = Central-East, 3 = West and 4 = South.

DNA extraction, PCR amplification and sequence alignment

Total genomic DNA was extracted from air-dried blood preserved on FTA classic cards (FTA® cards) (Whatman Biosciences) using the MACHEREY-NAGEL DNA extraction kit or from full blood preserved in ethanol using the Qiagen kit following the manufacturers' instructions.

Five hundred and forty-nine base pairs of the mtDNA *D*-loop region were amplified using AV1F2 (5'-AGGACTACGGCTTGAAAAGC-3') (Mwacharo *et al.*, 2011) as the forward primer, and H547 (5'- ATGTGCCTGACCGAGGAACCAG-3') (Komiyama *et al.*, 2003) as reverse primer. PCR amplifications were carried out in a 20-µl reaction volume containing 40 ng genomic DNA, 10 µl PCR ready master mix (Thermo Scientific Ltd), 0.5 µM of each primer, and sterile nuclease-free water to reach the final volume of 20 µl. The PCR was carried out using a Peltier thermocycler with the following conditions: Hot lid (110°C), hot start 98°C (30 sec), denaturation 98°C (5 sec), annealing 63°C (5 sec), elongation 72°C (10 sec), 35 cycles and final extension step at 72°C (1 min) (Chester & Marshak, 1993; Nord *et al.*, 1997; Wikman *et al.*, 2004). The PCR product was electrophoresed on a 1% agarose gel at 100 volts for 45 minutes, and then the gel was stained by 1% ethidium bromide and visualised under ultraviolet light. Amplified DNA fragment size was estimated through size comparison with a 1 kb DNA ladder from New England BioLabs loaded alongside the PCR products. The products were purified using the reSource PCR purification kit from Source Bioscience.

An Applied Biosystems 3730 DNA Analyser was used for Sanger sequencing. For each sample, the primer sequences were trimmed to generate the 549-bp sequence fragment and correct possible base-calling errors using the proseq3 version 3.5 software (Filatov, 2009). Sequences were aligned to the chicken mtDNA reference sequence (GenBank accession no. AB098668) (Komiyama *et al.*, 2003) using Clustal X version 2.1 (Thompson *et al.*, 1997). Analyses were restricted to the first 397 bp of the sequence, which includes the hypervariable region (HV1) of the *D*-loop (Adebambo *et al.*, 2010).

For all the Ethiopian and some of the Iraqi samples (n = 22), the full mtDNA was retrieved from full-genome sequencing data using the Genome Analysis Toolkit (GATK V3.7) (DePristo *et al.*, 2011; Van der Auwera *et al.*, 2013). The sequences were then aligned with the chicken mtDNA reference genome sequence and the 397 bp of the *D*-loop selected for further analysis.

977 **Genetic diversity estimation**

978 DnaSP v5 was used to identify polymorphic sites, the number of haplotypes and to
979 calculate haplotype diversity (Hd), nucleotide diversity (π) and the average number of
980 nucleotide differences (Librado & Rozas, 2009). These parameters were examined both
981 at population and country levels.

982 **Phylogenetic analysis**

983 In order to assess the possible geography origin of the samples included in this study,
984 reference sequences from Liu *et al.* (2006) haplogroups were included (AB114069 –
985 haplogroup A, AB007744 – haplogroup B, AB114070 – haplogroup C, AY588636 –
986 haplogroup D, AB114076 – haplogroup E, AF512285 – haplogroup F, AF512288 –
987 haplogroup G, D82904 – haplogroup H, and AB009434 – haplogroup I). The
988 phylogenetic tree was constructed using jModeltest version 2.1.7 (Darriba *et al.*, 2012)
989 to predict the best-fit model and Phyml version 3.0 for the maximum likelihood tree
990 (Guindon & Gascuel, 2003). The confidence level for each branch in the tree was
991 assessed with 1,000 bootstrap replications. To infer the relationship between the
992 haplotypes, Median-Joining (MJ) network was built using NETWORK 5.0.0 (Bandelt *et*
993 *al.*, 1999).

994 **Analysis of Molecular Variance (AMOVA)**

995 The analysis of molecular variance was implemented using Arlequin v 3.5.2 with 1,000
996 permutations (Excoffier & Lischer, 2010). It was performed across countries and within
997 countries. Across countries, the analysis was performed using all the samples in a
998 country, each country as an individual group. For within-country analyses, we followed
999 the grouping of the populations as described in the sampling section, namely Iraqi
1000 samples divided into three groups (Central n = 63, North n = 9, South n = 35), Algerian
1001 into two groups (North-West, and Central). Ethiopian samples were divided into four
1002 groups ((Meseret and Mihquan), (Adane, Arabo, Horro, Jarso, Midir and Negasi Amba),
1003 (Ashuda, Amshi, Batambie, Dikuli, Gafera, Surta, and Tzion Teguaz), and (Girissa,
1004 Kumato, Loya, Shubi Gemo)), and Saudi Arabian into three groups (East (n = 45),
1005 Central (n = 43) and West (n = 97)). All Libyan and Pakistani samples were grouped as
1006 a single Libyan or Pakistani country population.

Neutrality test and demographic dynamics

The demographic profiles for each population were calculated from mismatch distribution patterns (Rogers & Harpending, 1992). Mean Absolute Error (MAE) was used between the observed and the theoretical (expected) mismatch in order to have more confidence about the decision of demographic expansion (Rogers *et al.*, 1996). Then, Fu's *F*s (Fu, 1997) and Tajima's D (Tajima, 1989) were estimated by using the infinite site model in DnaSP v.5 (Librado & Rozas, 2009).

Bayesian Skyline Plots (BSPs) (Drummond *et al.*, 2005) were investigated in order to have 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 (Bouckaert *et al.*, 2014). 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 generations with the first two-million generations used as burn-in. Tracer software V.1.7 (Rambaut *et al.*, 2018) was used to calculate the convergence of the posterior estimates of *Ne* to the likelihood of stationary distribution. The BSPs were standardised using the molecular mutation rate of evolution for chicken mtDNA, 3.13×10^{-7} mutations/site/year (m/s/y) following Alexander *et al.* (2015). This estimation of BSPs was performed on each countries sample separately and then they were plotted together.

Results

D-loop haplotype variation and genetic diversity

Eighty-eight haplotypes defined by 63 polymorphic sites were identified in the 706 sequences (**Supplementary Figure S1**). Each haplotype was abbreviated with the letter H followed by a number e.g. H_1, H_2, H_3 etc. The most frequent haplotype (41%) was H_3 (291 sequences out of 706), these 291 sequences were represented by 92 (32%), 60 (21%), 58 (20%), 55 (19%), 14 (5%), 12 (4%) of Saudi, Ethiopian, Iraqi, Algerian, Pakistani and Libyan sequences respectively. The next commonest were haplotypes H_23 and H_4 (7.6% and 6% of the sequences). H_23 was present only in Ethiopian samples (n = 54) and H_4 only in Pakistani chicken (n = 43).

The central region of Iraq showed a higher level of genetic diversity compared to the other regions (north-east and south), with haplotype diversities of 0.725 ± 0.053

(Baghdad) and 0.712 ± 0.105 (Karbala), but 0.438 ± 0.121 for Misan in the south-east, and 0.182 ± 0.144 for Basra in the south. The same is applicable for the nucleotide diversity, which shows a higher level of diversity in the central region of the country (**Table 2.1**).

For the Algerian populations, Adrar, Oran and Tlemcen populations show higher haplotype (0.654 ± 0.085, 0.706 ± 0.106 and 0.797 ± 0.088) and nucleotide diversity (0.0035 ± 0.0006, 0.0022 ± 0.0005 and 0.00565 ± 0.0021) compared with the haplotype (0.363 ± 0.131 and 0.182 ± 0.144) and nucleotide (0.001 ± 0.0004 and 0.0009 ± 0.0007) diversities for Mascara and Tiaret, respectively. The north-west area of this country shows a lower level of diversity compared with the central region, which is represented by the Adrar population.

In Ethiopia, different populations show different levels of diversity, with high haplotype diversities (between 0.7-0.9) for Adane, Arabo, Kumato, Loya, Meseret and Shubi, and the nucleotide diversity for the same populations ranging between 0.0033-0.0145. At a regional level, the north, central-east and west regions do not show significant differences in their diversities, but the south region has the highest level of genetic variation compared to the other regions (**Table 2.1**).

The Saudi Arabia regions also show different levels of diversity, with the highest haplotype diversity in the East on the shores of the Arabian Gulf (0.856 ± 0.044), then the west on the shores of the Red Sea (0.702 ± 0.041), while the lowest was in the central region (0.631 ± 0.084). The same trend is observed in Saudi Arabia for the nucleotide variation.

Libya and Pakistan show high diversity levels for both haplotype (0.731 ± 0.099 and 0.756 ± 0.043 respectively) and nucleotide diversity (0.0054 ± 0.0018 and 0.0088 ± 0.001 respectively).

At country level, the highest haplotype diversity was present in Ethiopia and the lowest in Algerian samples. A similar intermediate level of haplotype variation was observed in the Iraqi, Libyan, Pakistani and Saudi Arabian samples.

1066 **Table 2.1**: Location, sample size and genetic diversity of the samples included in this
1067 study.

Country/population	N	S	H	Hd (SD)	π (SD)	K
Iraqi populations						
North-East						
Sulimania (1)	9	0	1	0.000 (0.000)	0.0000 (0.0000)	0.000
Central						
Baghdad	51	18	12	0.725 (0.053)	0.0080 (0.0009)	3.167
Karbala	12	5	4	0.712 (0.105)	0.0043 (0.0010)	1.697
Central (2)	63	20	14	0.736 (0.051)	0.0075 (0.0008)	2.988
South						
Basra	11	1	2	0.182 (0.144)	0.0005 (0.0003)	0.182
Misan	24	11	5	0.438 (0.121)	0.0034 (0.0015)	1.359
South (3)	35	12	6	0.361 (0.103)	0.0025 (0.0011)	1.002
Total	107	22	18	0.686 (0.047)	0.0067 (0.0007)	2.652
Algerian populations						
North-West						
Mascara	20	3	4	0.363 (0.131)	0.0010 (0.0004)	0.389
Oran	17	5	6	0.706 (0.106)	0.0022 (0.0005)	0.882
Tiaret	11	2	2	0.182 (0.144)	0.0009 (0.0007)	0.364
Tlemcen	18	15	8	0.797 (0.088)	0.0056 (0.0021)	2.242
North-West (1)	66	19	12	0.569 (0.072)	0.0025 (0.0007)	1.030
Central						
Adrar (2)	22	5	4	0.654 (0.085)	0.0035 (0.0006)	1.403
Total	88	20	13	0.597 (0.060)	0.0028 (0.0006)	1.145
Ethiopian populations						
North						
Mihquan	10	9	3	0.644 (0.101)	0.0054 (0.0031)	2.156
Meseret	10	11	6	0.867 (0.085)	0.0072 (0.0026)	2.844
North (1)	20	14	7	0.742 (0.071)	0.0061 (0.0021)	2.447
Central-East						
Adane	10	11	5	0.800 (0.100)	0.0082 (0.0029)	3.244
Arabo	10	4	5	0.844 (0.080)	0.0033 (0.0006)	1.289
Horro	30	10	5	0.644 (0.081)	0.0068 (0.0017)	2.699
Jarso	14	0	1	0.000 (0.000)	0.0000 (0.0000)	0.000

Midir	10	2	3	0.600 (0.131)	0.0017 (0.0005)	0.667
Negasi_Amba	10	1	2	0.467 (0.132)	0.0012 (0.0003)	0.467
Central-East (2)	84	20	10	0.774 (0.027)	0.0092 (0.0010)	3.674
West						
Ashuda	10	1	2	0.356 (0.159)	0.0009 (0.0004)	0.356
Amshi	10	1	2	0.533 (0.095)	0.0013 (0.0002)	0.533
Batambie	8	1	2	0.250 (0.180)	0.0006 (0.0004)	0.250
Dikuli	10	2	3	0.600 (0.131)	0.0017 (0.0005)	0.667
Gafera	10	1	2	0.467 (0.132)	0.0012 (0.0003)	0.467
Surta	9	1	2	0.222 (0.166)	0.0006 (0.0004)	0.222
Tzion_Teguaz	10	3	3	0.600 (0.131)	0.0024 (0.0009)	0.933
West (3)	67	5	8	0.707 (0.044)	0.0025 (0.0002)	1.000
South						
Girissa	10	4	5	0.667 (0.163)	0.0027 (0.0008)	1.067
Kumato	10	10	5	0.756 (0.130)	0.0082 (0.0030)	3.244
Loya	10	8	5	0.756 (0.130)	0.0089 (0.0020)	3.533
Shubi_Gemo	10	14	6	0.911 (0.062)	0.0145 (0.0020)	5.756
South (4)	40	23	19	0.929 (0.021)	0.0150 (0.0006)	5.964
Total	211	33	36	0.840 (0.016)	0.0094 (0.0007)	3.732
Saudi Arabia						
East (1)	45	18	15	0.856 (0.044)	0.0042 (0.0008)	1.697
Central (2)	43	20	11	0.631 (0.084)	0.0037 (0.0010)	1.499
West (3)	97	23	16	0.702 (0.041)	0.0041 (0.0006)	1.658
Total	185	34	26	0.727 (0.033)	0.0041 (0.0005)	1.633
Other populations						
Libya	23	14	10	0.731 (0.099)	0.0054 (0.0018)	2.142
Pakistan	92	24	19	0.756 (0.043)	0.0088 (0.001)	3.503
Total						
All the samples included in this study	706	55	88	0.815 (0.014)	0.0077 (0.0004)	3.058

N = number of samples, S = segregating sites, H = Number of haplotypes, Hd (SD) = haplotype diversity (standard deviation), π (SD) = nucleotide diversity (standard deviation) and K = average number of nucleotide differences.

Phylogeographic analysis

Reference haplotypes from Liu *et al.* (2006) were used to name the haplogroups accordingly to their nomenclature (A, B, C, D, E, F, G, H and I). A maximum likelihood tree for the 88 haplotypes was constructed to assess the genetic relationships among them (**Figure 2.2**). Most branches included in this figure show high confidence (bootstrap) relationship values ranging between 70-100, and show the presence of five haplogroups (A, B, C, D, and E) amongst the samples (**Supplementary Figure S2**). The haplotypes were classified in different haplogroups following the results of the tree and network for all the countries together, and an individual networks for each country separately (**Figures 2.2, 2.3 and Supplementary Figures S3, S4, S5, S6, S7 and S8**). The frequency of each haplogroup in each country is indicated in **Table 2.2**. Haplogroup E is the most frequent one (625 individuals out of 706) followed by D (n = 46), A (n = 31), C (n = 3) and B (n = 1). Haplogroup A is represented by 7 haplotypes, B by one, C only two and haplogroup D by 11 haplotypes.

Haplogroup E was represented by different number of individuals from the six countries, they were 87 (14%), 169 (27%), 84 (13.4%), 21 (3.2%), 84 (13.4%) and 180 (29%) samples out of 625 for Algeria, Ethiopia, Iraq, Libya, Pakistan and Saudi Arabia, respectively. Similarly, haplogroup A was represented by 31 samples from all countries. Haplogroup B and C were only found in chicken from Ethiopia and Saudi Arabia: one haplotype from Ethiopia was identical to the reference haplotype B, while C was identified as one Ethiopian haplotype and one Saudi haplotype (represented by two samples from the south-west region). Haplotypes belonging to haplogroup D were present in all countries except Algeria and Libya. This haplogroup was represented by eleven haplotypes beside the reference, which include 34 (74%), 8 (17%), 3 (7%) and 1 (2%) samples out of 46 for Ethiopia, Iraq, Pakistan and Saudi Arabia respectively.

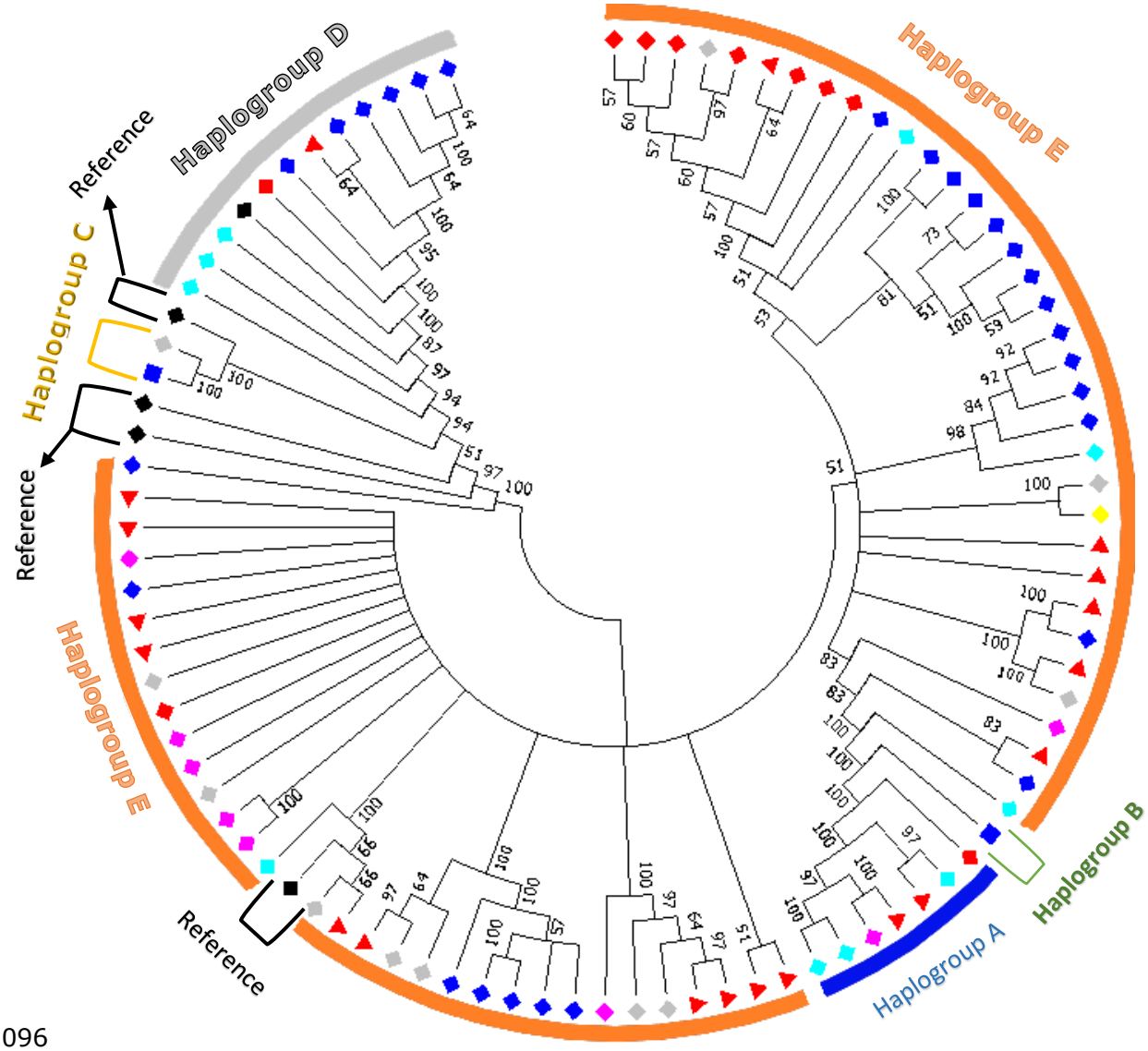


Figure 2.2: Maximum likelihood tree for the 88 haplotypes and the references from Liu *et al.* (2006). ◇ = Iraqi haplotypes, ◆ = Ethiopian haplotypes, ◆ = Algerian haplotypes, ◇ = Saudi haplotypes, ◆ = Pakistani haplotypes, ◇ = Libyan haplotypes, ■ = Ethiopian haplotype with reference, ◆ = References, ▲ = common haplotypes among countries.

Network analysis allows further visualisation of the diversity of the haplotypes and their relationship. The most frequent haplotypes (H_3, H_4 and H_23), belong to haplogroup E (**Figure 2.3**). The first (H_3) is identical to the haplotype reference**,** H_4 is different from the previous one by 4 mutations only, while only a single mutation separated H_23 from H_3. The links between haplogroup E and the other haplogroups are well resolved. Haplogroup E is separated from haplogroup D by several mutations with three median vectors (mv), while C is connected to E by 12 mutations and 5 mv. The same situation is applicable to other clusters except for haplogroup F which was connected to E by 7 mutations without median vectors. The presence of median vectors could be attributed to

un-sampled haplotypes, e.g. haplotypes that have not been introduced to the studied geographical area or haplotypes that became extinct (Mwacharo *et al.*, 2011). Haplotype H_3 (haplogroup E) at the centre of a star-like pattern is likely the ancestral haplotype for this haplogroup.

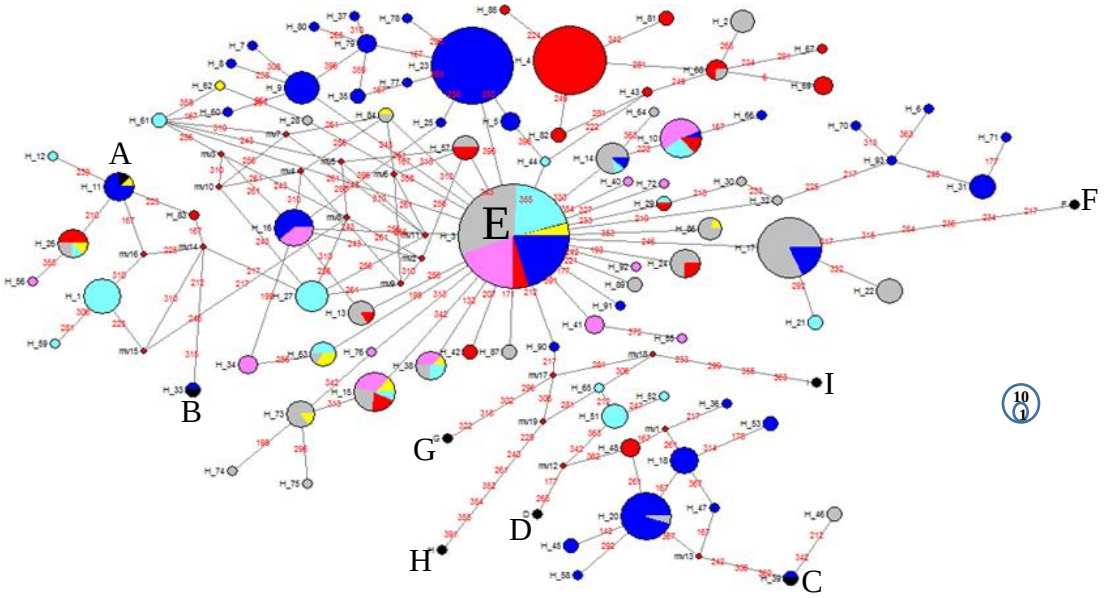


Figure 2.3: Median-Joining Network for Algeria, Ethiopia, Iraq, Libya, Pakistan and Saudi Arabia haplotypes (n = 88). The black circles refer to reference haplotypes from Liu *et al.* (2006), **Pink** = Algerian haplotypes, **Blue** = Ethiopian haplotypes, **Cyan** = Iraqi haplotypes, **Yellow** = Libyan haplotypes, **Red** = Pakistani haplotypes and **Grey** = Saudi Arabian haplotypes. The numbers on the branch indicate the position of the mutations, the circles are proportional to the numbers of haplotypes.

Table 2.2: Distribution of haplotypes across haplogroups and countries

Haplogroup/Haplotypes	**Algeria**		**Ethiopia**				**Iraq**			**Libya**	**Pakistan**	**Saudi Arabia**		
Haplogroup A	**North-West**	**Central**	**North**	**Central_East**	**West**	**South**	**North-East**	**Central**	**South**			**East**	**Central**	**West**
H_1								12						
H_11			1	5						1				
H_12								1						
H_26									1	1	4	1	1	
H_56	1													
H_59								1						
H_83											1			

Haplogroup B														
H_33			1											
Haplogroup C														
H_39				1										
H_46														2
Haplogroup D														
H_18						7								
H_20				15		5								1
H_36						1								
H_45						2								
H_47						1								
H_48											3			
H_51								4	2					
H_52								1						
H_53						2								
H_58						1								
H_65								1						
Haplogroup E														
H_2													2	3
H_3	43	12	7	20	31	2		30	28	12	14	13	31	48
H_4											43			
H_5				3										
H_6						1								
H_7						1								
H_8					1									
H_9					6	4								
H_10	3	5	1					4			2			
H_13											1	4		1
H_14			1					1				2		5
H_15	3	2						1		2	3	3		1
H_16	5				8									
H_17				6									27	
H_21								2						
H_22														6
H_23			8	31	15									
H_24											2	6		
H_25					1									
H_27							9							
H_28												1		
H_29									1		1			
H_30												1		
H_31						6								
H_32														1
H_34		3												
H_35						2								
H_36						1								
H_37					1									
H_38	3							2		1		2		
H_40	1													

H_41	3													
H_42											2			
H_43											1			
H_44								1						
H_55	1													
H_57											3	1	2	
H_60						1								
H_61									2					
H_62										1				
H_63								2	1	2		1		
H_64														1
H_66						1								
H_67											1			
H_68											3			1
H_69											3			
H_70						1								
H_71						1								
H_72	1													
H_73										1		4	1	1
H_74														1
H_75														1
H_76	1													
H_77				1										
H_78				1										
H_79					3									
H_80			1											
H_81											2			
H_82											2			
H_84										1			1	
H_85											1			
H_86										1			4	
H_87														2
H_89													1	1
H_90				1										
H_91						1								
H_92	1													
H_93						1								

Analysis of Molecular Variance (AMOVA)

Maternal genetic differentiation within and among populations was assessed for the samples included in this study (**Table 2.3**). The data were examined at three levels, among groups of populations, populations within groups and within populations. The results showed that the highest percentage of variation for Iraq and Algeria is found among individuals, 66.33% and 94.41% respectively, within populations. The percentage

of differentiation among groups of Iraqi populations was higher than the one in Algeria (28.89% and 5.03% respectively). The Ethiopian situation was more complex than for the other countries: the percentage of variation for both populations within groups (45.84%) and among individuals within populations (42.87%) were relatively large, with lower variation among groups (11.29%). The three regions of Saudi Arabia showed most of the variation (96.66%) among individuals within regions, and hardly any genetic differentiation among regions (3.34%). In general, when we compare all the countries (Algeria, Ethiopia, Iraq, Libya, Pakistan and Saudi Arabia), the highest variation was found among individuals within populations (74.12%), while the next highest was seen among populations within countries (17.01%).

Table 2.3: Analysis of Molecular Variance (AMOVA)

Grouping	Source of variation	Degrees of freedom	Variance components	Percentage of variation
Iraq	Among groups	2	0.44860 Va	28.89
	Populations Within groups	2	0.07417 Vb	4.78
	Within populations	102	1.02984 Vc	66.33
	Total	106	1.55261	
Algeria	Among groups	1	0.02977 Va	5.03
	Populations Within groups	3	0.00331 Vb	0.56
	Within populations	83	0.55855 Vc	94.41
	Total	87	0.59163	
Ethiopia	Among groups	3	0.22407 Va	11.29
	Populations Within groups	15	0.91011 Vb	45.84
	Within populations	192	0.85103 Vc	42.87
	Total	210	1.98520	
Saudi Arabia	Among regions	2	0.02819 Va	3.34
	Within regions	182	0.81532 Vb	96.66
	Total	184	0.84352	
All countries	Among countries	5	0.14091 Va	8.88
	Populations Within countries	8	0.27001 Vb	17.01
	Within populations	692	1.17663 Vc	74.12
	Total	705	1.58755	

Population history and demographic dynamics

The demographic history patterns for each population and each country were evaluated. The Mean Absolute Error (MAE) revealed moderate to low values (**Table 2.4**). Other

results showed that many of the populations have negative non-significant values for Tajima's D. The exceptions here were for Misan, Mascara, Tiaret, Tlemcen, north-west Algeria, south Iraq region, Libya and the east, central, and west regions of Saudi Arabia, where we observed negative and significant values. A negative Tajima's D signifies an excess of low-frequency polymorphism relative to expectation, indicating population size expansion (following a bottleneck or a selective sweep) or purifying selection. Fu's *Fs*, also an index of population expansion, is known to be a more powerful tool than Tajima's D (Ramos-Onsins & Rozas, 2002). This power comes from the ability to differentiate between population growth and genetic hitchhiking (since they have a similar genetic pattern) and rejecting the hypothesis of neutral mutations. No significance was found for the Harpending raggedness index (r) in our dataset except for the Algerian samples when grouped together. These results of MAE, Tajima's D, Fu's *Fs* and raggedness index (r) suggest a complex demographic history for our populations. We also performed mismatch distribution and Bayesian Skyline plot (BSP) analyses to gather more information regarding possible past population expansion.

Table 2.4: Neutrality and demographic expansion parameters

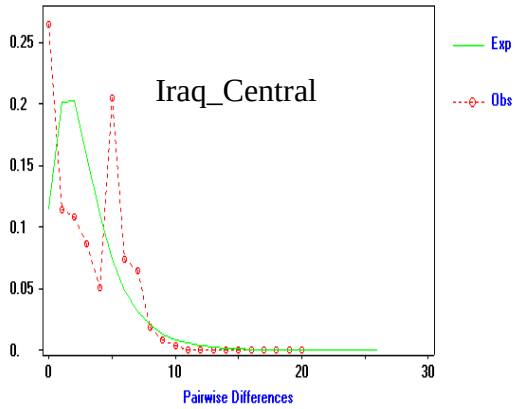
Population/Country	**N**	**S**	**MAE**	**Tajima's D (*P*-value)**	**Fu's *Fs* (*P*-value)**	**Harpending r(*P*-value)**
Iraqi populations						
North-East						
Sulimania (1)	9	0	0.000	0.000 (1.000)	0.000 (0.000)	0.000 (0.000)
Central						
Baghdad	51	18	0.564	-0.653 (0.280)	-1.380 (0.290)	0.123 (0.774)
Karbala	12	5	0.439	0.0918 (0.567)	0.538 (0.616)	0.155 (0.423)
Central (2)	63	20	0.440	-0.911 (0.208)	-2.541 (0.166)	0.067 (0.469)
South						
Basra	11	1	0.055	-1.128 (0.137)	-0.410 (0.350)	0.438 (0.942)
Misan	24	11	0.502	-1.824 (0.016)	-0.090 (0.469)	0.203 (0.632)
South (3)	35	12	0.440	-2.074 (0.001)	-1.444 (0.161)	0.252 (0.688)
Total	107	22	0.432	-1.090 (0.134)	-5.070 (0.050)	0.055 (0.328)
Algerian populations						
North-West						
Mascara	20	3	0.270	-1.440 (0.032)	-2.135 (0.009)	0.187 (0.237)
Oran	17	5	0.560	-1.301 (0.077)	-2.953 (0.001)	0.211 (0.536)
Tiaret	11	2	0.391	-1.429 (0.037)	0.506 (0.622)	0.735 (0.933)
Tlemcen	18	15	0.557	-1.842 (0.015)	-2.089 (0.088)	0.053 (0.128)
North-West (1)	66	19	0.308	-2.255 (0.000)	-7.246 (0.000)	0.066 (0.096)
Central						
Adrar (2)	22	5	0.547	0.0671 (0.586)	0.928 (0.742)	0.280 (0.798)
Total	88	20	0.287	-2.092 (0.001)	-6.980 (0.004)	0.045 (0.048)
Ethiopian populations						

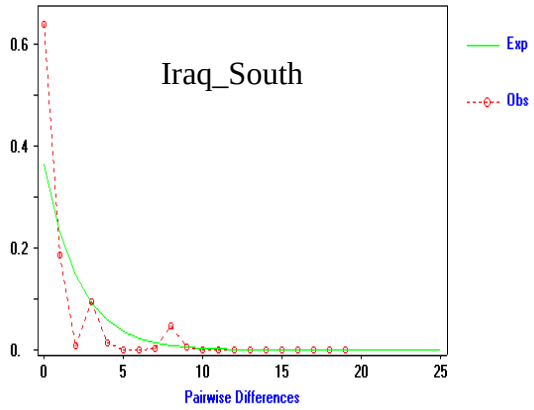
North						
Meseret	10	11	0.609	-1.202 (0.139)	-0.944 (0.237)	0.064 (0.138)
Mihquan	10	9	0.838	-1.411 (0.080)	2.172 (0.898)	0.226 (0.701)
North (1)	20	14	0.503	-1.390 (0.088)	-0.596 (0.419)	0.067 (0.269)
Central-East						
Adane	10	11	0.757	-0.741 (0.272)	0.505 (0.600)	0.176 (0.655)
Arabo	10	4	0.786	-0.339 (0.414)	-1.629 (0.051)	0.185 (0.462)
Horro	30	10	0.782	0.219 (0.638)	2.383 (0.867)	0.125 (0.647)
Jarso	14	0	0.000	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Midir	10	2	0.586	-0.183 (0.322)	-0.272 (0.360)	0.240 (0.509)
Negasi Amba	10	1	0.297	0.819 (0.853)	0.818 (0.672)	0.222 (0.340)
Central-East (2)	84	20	0.781	-0.239 (0.497)	1.547 (0.786)	0.103 (0.748)
West						
Amshi	10	1	0.371	1.302 (0.919)	1.029 (0.789)	0.288 (0.612)
Ashuda	10	1	0.186	0.014 (0.754)	0.417 (0.670)	0.209 (0.326)
Batambie	8	1	0.100	-1.054 (0.187)	-0.182 (0.446)	0.312 (0.792)
Dikuli	10	2	0.586	-0.183 (0.355)	-0.272 (0.374)	0.240 (0.531)
Gafera	10	1	0.297	0.819 (0.827)	0.818 (0.698)	0.222 (0.366)
Surta	9	1	0.080	-1.088 (0.157)	-0.263 (0.389)	0.358 (0.918)
Tzion Teguaz	10	3	0.342	-0.431 (0.331)	0.345 (0.571)	0.133 (0.180)
West (3)	67	5	0.578	-0.104 (0.497)	-2.569 (0.092)	0.124 (0.290)
South						
Girissa	10	4	0.487	-0.942 (0.204)	-2.096 (0.019)	0.073 (0.081)
Kumato	10	10	0.819	-0.364 (0.361)	0.505 (0.583)	0.144 (0.527)
Loya	10	8	0.627	1.076 (0.865)	0.706 (0.646)	0.082 (0.250)
Shubi Gemo	10	14	0.721	0.746 (0.798)	0.723 (0.616)	0.159 (0.757)
South (4)	40	23	0.653	0.345 (0.703)	-4.181 (0.083)	0.020 (0.118)
Total	211	33	0.573	-0.938 (0.176)	-15.722 (0.001)	0.045 (0.364)
Saudi Arabia						
East (1)	45	18	0.729	-1.964 (0.006)	-9.069 (0.000)	0.132 (0.494)
Central (2)	43	20	0.231	-2.207 (0.001)	-4.717 (0.014)	0.048 (0.096)
West (3)	97	23	0.314	-1.862 (0.009)	-7.33 (0.007)	0.062 (0.200)
Total	185	34	0.375	-2.115 (0.000)	-18.713 (0.000)	0.061 (0.186)
Other populations						
Libya	23	14	0.520	-1.539 (0.047)	-3.704 (0.018)	0.058 (0.183)
Pakistan	92	24	0.547	-0.861 (0.210)	-4.096 (0.085)	0.063 (0.482)
Total						
All the samples included in this study	706	55	0.323	-1.668 (0.013)	-97.654 (0.000)	0.025 (0.096)

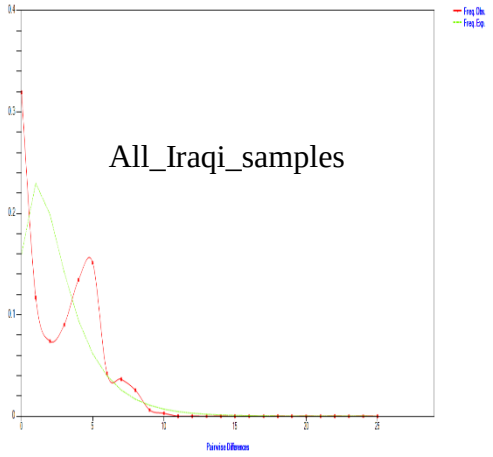
N = Number of samples, S = Segregating sites, MAE = Mean Absolute Error.

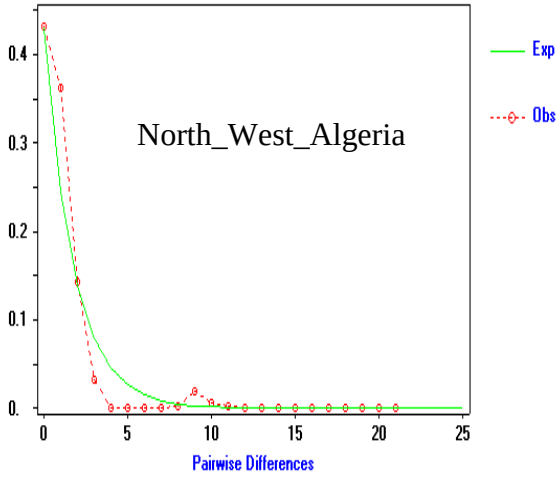
The mismatch distribution graphs show three patterns: uni-, bi- and multi-modal (**Figures 2.4 and Supplementary Figure S9**). The dominant distribution pattern is unimodal, and the multimodal the least frequent among the populations. These results are relatively compatible with the demographic expansion results from **Table 2.4**, which suggested the operation of either demographic expansion, bottleneck or purifying

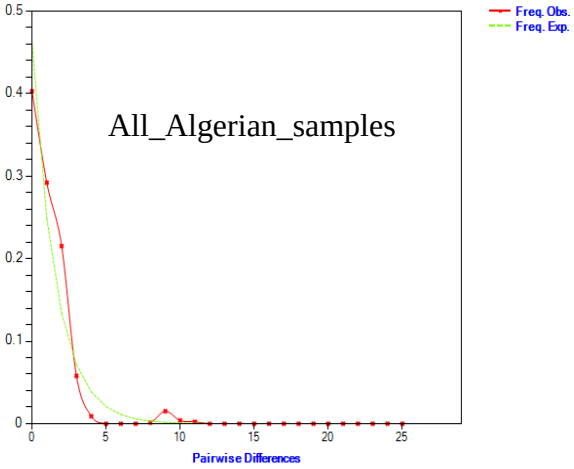
selection and also might be referreing to the presence of more than one haplogroup in populations.

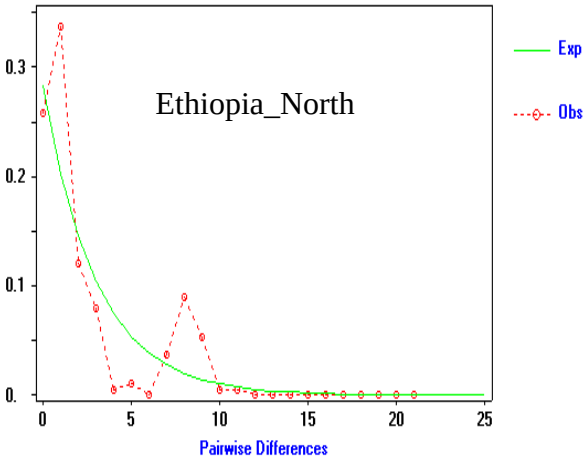


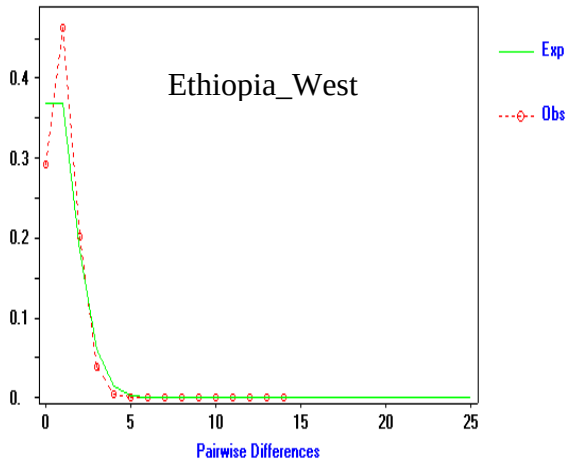


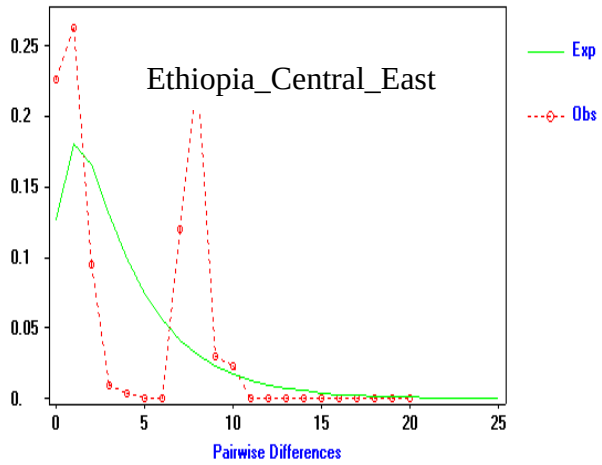


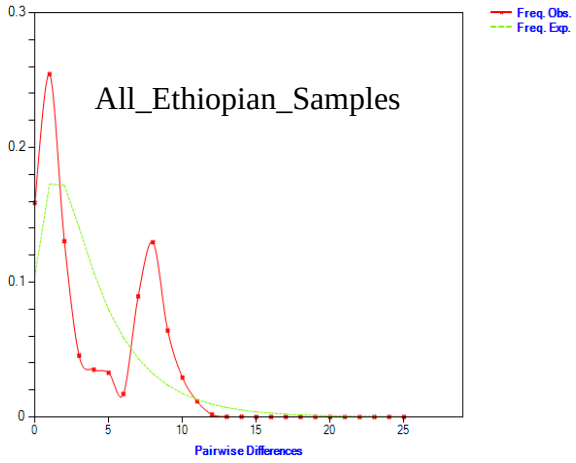


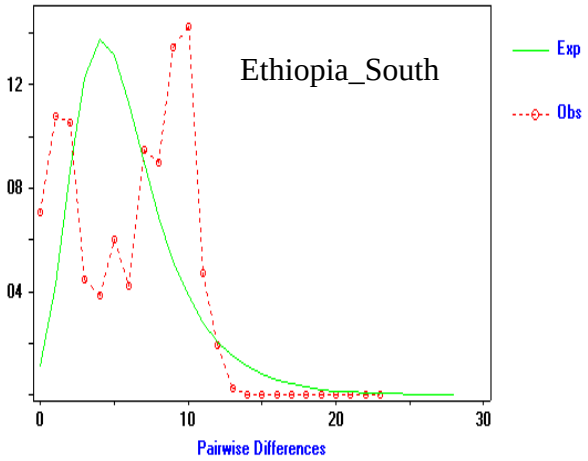


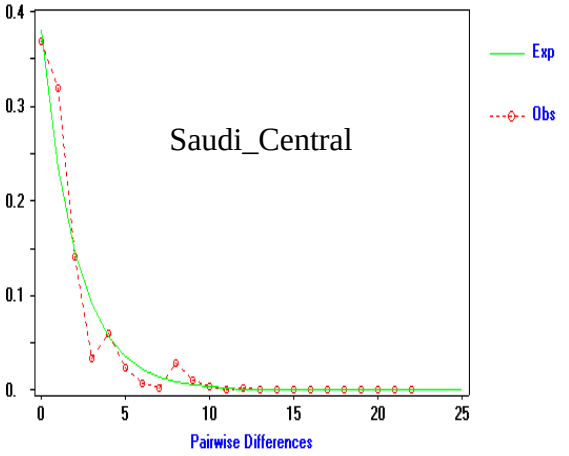


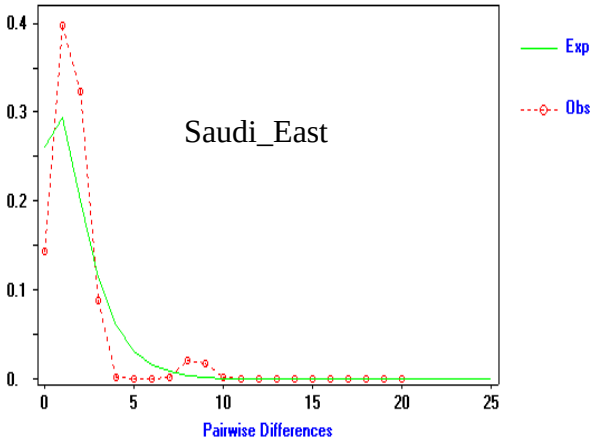

Ethiopia_West
Exp
Obs
Pairwise Differences


Ethiopia_Central_East
Exp
Obs
Pairwise Differences


All_Ethiopian_Samples
Freq. Obs.
Freq. Exp.
Pairwise Differences


Ethiopia_South
Exp
Obs
Pairwise Differences


Saudi_Central
Exp
Obs
Pairwise Differences


Saudi_East
Exp
Obs
Pairwise Differences

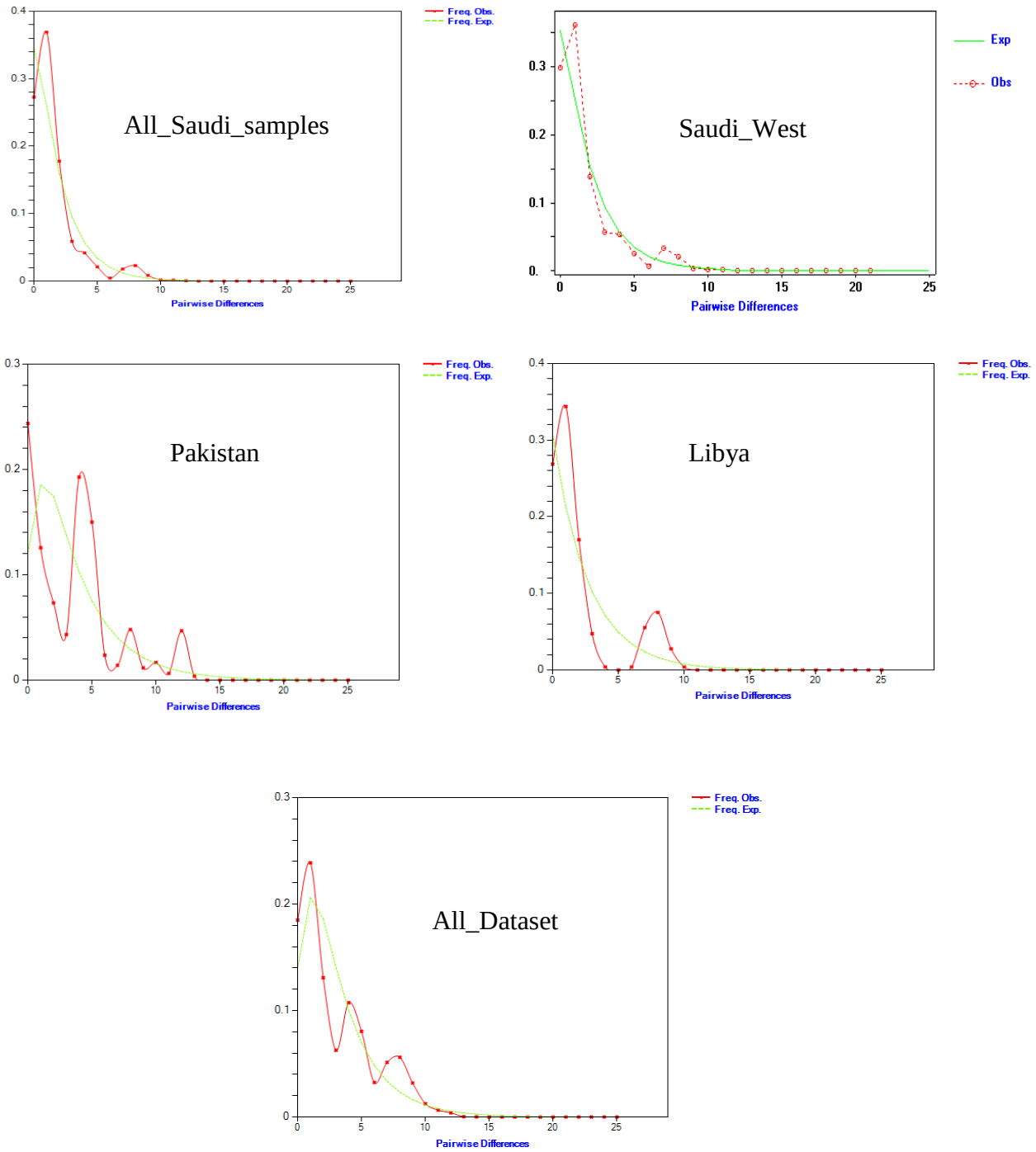
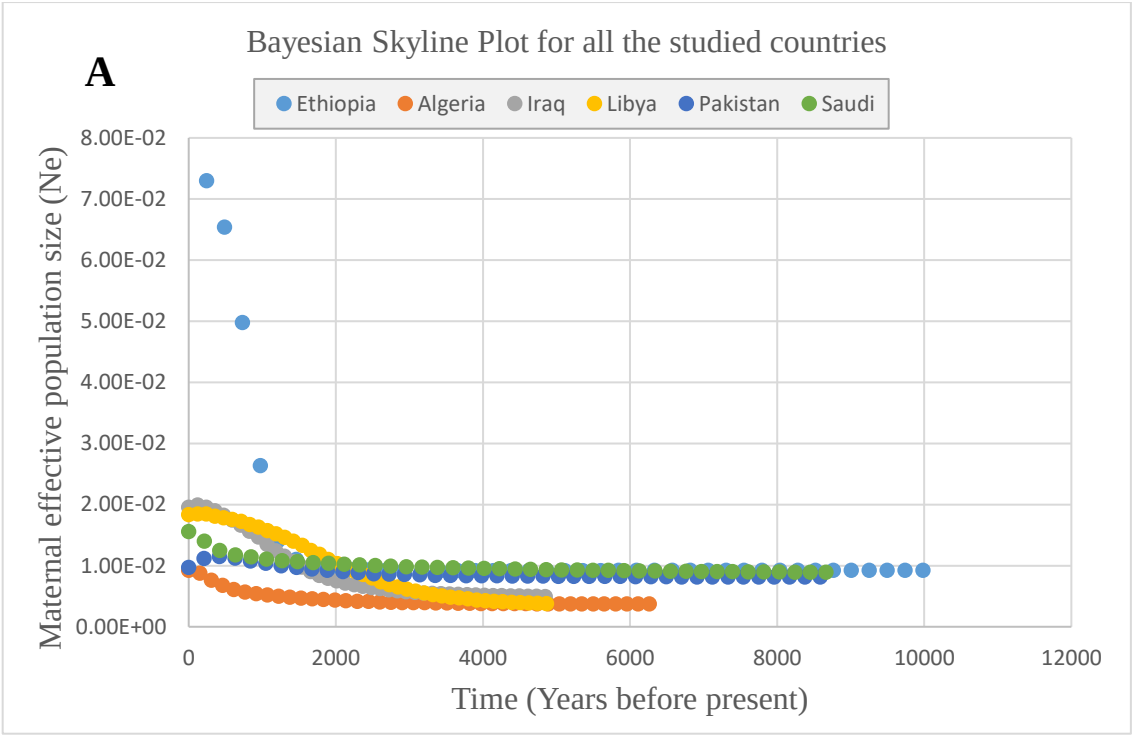


Figure 2.4: Mismatch distribution patterns for regions and countries.

Bayesian Skyline Plot showed evidence of demographic expansion for all the countries in recent times, except for Pakistan (**Figure 2.5 A**). Ethiopia had evidence of a rapid population increase in recent years, much higher than that observed for Iraq, Libya, Saudi Arabia and Algeria. On the other hand, BSP for the three main haplogroups (A, D and E), exhibited more rapid expansion for E than the others, with more recent decline (**Figure 2.5 B**).



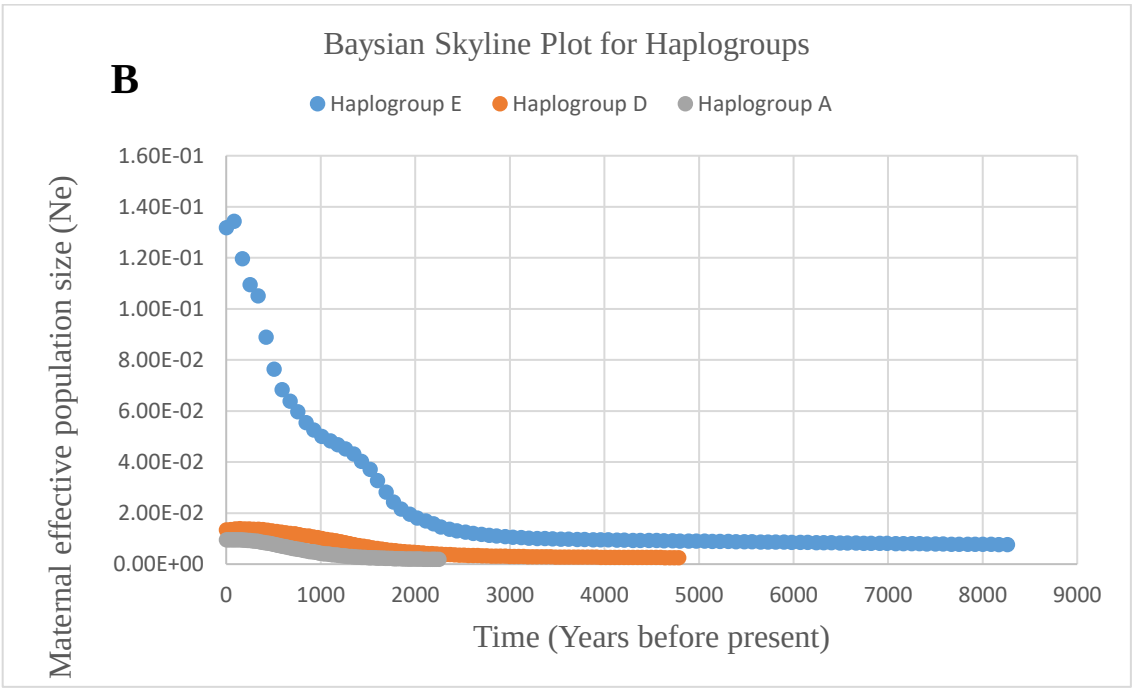


Figure 2.5: Bayesian Skyline Plot (BSP) for the countries (A) and Haplogroups (B).

Discussion

In this study, we analysed the mtDNA genetic diversity of 706 village chickens from Algeria, Ethiopia, Iraq, Libya, Pakistan and Saudi Arabia with the aim of unravelling their history and origin. To the best of our knowledge, it represents the first study on the *D*-loop mtDNA diversity of indigenous chicken from these countries with the exception

of Ethiopia, where different populations (not included here) have been tested in another study (Mwacharo *et al.*, 2011).

The history of chicken domestication remains unsettled. The Indian subcontinent, as well as south-east Asia, has been proposed as major centres of origins of domestic chicken, with (for the former) the Indus Valley as a possible centre of domestication. Accordingly, we would expect that Pakistani chicken would show the highest *D*-loop mtDNA diversity. This is not the case when we consider all haplogroups together for each country (**Table 2.1**). It is also not the case when examining only haplogroup E (**Supplementary Table S1**), believed to have its centre of diversity on the Indian subcontinent, with Liu *et al.* (2006) reporting that it dominated in Europe and India. Considering all haplogroups or only haplogroup E, Ethiopia displays the largest diversity. This may be a consequence of the number of chicken examined here, with more than twice as many Ethiopian as Pakistani samples. Alternatively, or in addition, it may reflect multiple arrivals of chicken in East Africa following different dispersal routes.

Interestingly, previous studies have suggested a possible dual origin for the chicken of East Africa (Mwacharo *et al.*, 2011; Lyimo *et al.*, 2013), with a terrestrial origin from the Indian subcontinent but also a genetic influence from South-East Asian chickens following maritime trading across the Indian Ocean. The second most-frequent haplogroup is haplogroup D, postulated to be a legacy of the past Indian Ocean maritime trading network (Liu *et al.*, 2006). This haplogroup is commonly observed in Ethiopian samples (**Table 2.2**), where it is more frequent in the southern region, completely absent in Libya and Algeria, and also observed but at a low frequency in Iraq, Pakistan and Saudi Arabia. This indeed agrees with maritime dispersal of this haplogroup, with its absence in the Mediterranean Sea not surprising in such context. Also, the absence of haplogroup D in Algeria and Libya, its relatively low frequency in other countries and its geographic distribution in Ethiopia suggest that the arrival and dispersal of the D haplogroup may have occurred before the north and western dispersal of haplogroup E.

Within countries, we observed a different level of diversity. Iraqi chicken populations from the central area show more diversity than the two populations from the south (**Table 2.1**). High genetic diversity in the central region could be attributed to two reasons: first, a terrestrial route of diffusion from the domestication centre could probably have happened through the centre of Iraq rather than the south of the country; and secondly, the central area of the country includes the capital Baghdad, a city built during the 8th

century as a capital for the Islamic world. Accordingly, the region might have witnessed major movements of populations and their livestock from different geographic areas.

In Algeria, with the exception of one haplotype (of haplogroup A), only haplogroup E was observed. Diversity between the northern part of the country and the more central region is nearly the same. The populations studied here are geographically close to each other and in such context, the results might not be surprising. Examination of other populations from other parts of the country is needed to investigate whether Algerian chickens represent a single genetic group or not. In Libya, we only analysed a single population, and as for Algeria, only two haplogroups were observed with haplogroup E being the commonest. Together, these results support a main Indian-continent origin for Mediterranean chickens using a terrestrial route from the northern part of the Indian subcontinent to the shores of the Mediterranean.

Saudi Arabian populations show the presence of four haplogroups (A, C, D and E), with E the commonest. This probably reflects more than one route of introduction, E from the Indian subcontinent following a terrestrial and maritime route, and others through the maritime trading network on the Arabian Gulf and the Red Sea. This is further supported by the genetic diversity values observed for the different regions, where the eastern part on the Arabian Gulf shows higher diversity (0.856 ± 0.044) compared to the central (0.631 ± 0.084) and western (0.702 ± 0.041) regions of the country.

Five haplogroups (A, B, C, D and E) have been identified in Ethiopian populations, with different historical backgrounds of introduction. As mentioned before, the E haplogroup, the commonest, probably originated from the Indian subcontinent via a terrestrial route along the Nile River Basin, although we cannot exclude arrival through the coastal areas of the Horn of Africa. The latter was the most likely route for the second most frequent haplogroup, D. Within the country, the southern region displayed the highest haplotype (0.929 ± 0.021) and nucleotide (0.0150 ± 0.0006) variation compared to the other three Ethiopian regions. Then, Central-East, North and West come next, with slight differences among them.

Compared to other studies, the *D*-loop mtDNA diversity of Ethiopian, Iraqi, Libyan, Pakistani and Saudi chickens are higher than those reported for Iranian, Turkey and Egyptian chickens in the literature (Eltanany & Hemeda, 2016; Meydan *et al.*, 2016). Similarly, among sub-Saharan chicken populations studied previously, we observed

higher genetic diversity for the countries included in this study compared with populations from Nigeria, Chad, Uganda and Sudan (Adebambo *et al.*, 2010; Mwacharo *et al.*, 2011; Hassaballah *et al.*, 2015). These results need to be interpreted with caution as they may reflect sampling artefacts. Globally, however, high mtDNA diversity of Ethiopian chickens could reflect their extensive ancient movements following trading routes linking Ethiopia to the Fertile Crescent civilisations. It also highlights the importance of the Horn of Africa as an entry point of livestock into the continent.

A high frequency of one haplotype (H_3) was found in all countries examined here (**Figure 2.3**). The majority of samples were clustered in haplogroup E, while haplogroup A, B, C and D were observed at low frequencies. Haplogroup E and A were found in all the studied countries. Haplogroup B was found only in Ethiopia, represented by just one individual in the Mihquan population. Haplogroup C is present in Saudi Arabian and Ethiopian samples only, and specifically in the south-western region on the shores of the Red Sea (Abhaa and Jazan) in Saudi Arabia and in the south region (Loya) of Ethiopia.

Haplogroups A and B have been reported before mainly from South China (Liu *et al.*, 2006). Haplogroup C was originally observed in chickens from Japan and south-east China dispersing through the maritime ancient trading network (Liu *et al.*, 2006). The origins of these haplogroups present at a low frequency in our dataset remains speculative; they may be a legacy of ancient dispersal and/or more recent crossbreeding with commercial birds.

Overall, AMOVA showed similarity among countries, with most of the variation observed among individuals within populations (**Table 2.3**). However, more variation within populations within groups are observed for Ethiopia and Iraq in comparison to other countries. This supports our interpretation of haplotype and haplogroup diversity discussed above, suggesting in particular for Ethiopia that multiple routes of chicken origins and dispersal have occurred. It also indicates that once domestic chicken reached a country, genetic exchanges occurred within the country, limiting the usefulness of the *D*-loop mitochondrial DNA as a genetic marker for phylogeographic analysis within the country.

At a country level, summary statistics are consistent with the demographic model of expansion. However, some populations within a country show otherwise e.g. Karbala, Sulimania, Amshi, Ashuda…etc. The mismatch distribution also shows a unimodal

distribution for the majority but not all of the populations. A multimodal distribution might be ascribed to the observation of different haplogroups in these populations.

To sum up, this chapter presents for the first time the genetic diversity of Algerian, Iraqi, Libyan, Pakistani and Saudi Arabian village chickens using mtDNA *D*-loop sequencing information. Combined with findings from other studies, the results presented here add further support for a main Indian subcontinent origin for chickens of the countries examined here, as well as for the importance of the Indian Ocean maritime trading network for the dispersal of the species. However, no or low phylogeographic structure was observed overall across the studied populations, showing the limitation of the *D*-loop chicken mitochondrial DNA diversity for this purpose.

Chapter 3

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

Abstract

Phylogeography plays a major role in understanding micro and macroevolutionary processes, because it deals with evolutionary interpretations of the geographical distribution of lineages among and within closely related species. This field integrates information from molecular genetics, population genetics, demography and phylogeny for the interpretation of the geographical distribution of lineages. In this chapter, we exploit the full mtDNA sequence and W chromosome polymorphisms to assess the usefulness of these 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 mtDNA genome sequences show a high level of genetic diversity within and among countries as well as within populations. On the other hand, sequence analysis of the chosen 13-kb W chromosome fragment shows low genetic diversity both within and among populations. Six full mtDNA haplogroups (A, B, C1, C2, D1 and E1; *sensu* Miao *et al.* (2013)) were observed and 25 distinct W haplotypes. The results support the effectiveness of full mtDNA sequences in tracing the maternal historical background and level of genetic diversity of populations. Even with the low polymorphism level of W chromosome fragment, it shows a number of distinct haplotypes in some populations, which needs further investigation for other genomic regions of this chromosome. However, despite studying polymorphism at the full mtDNA genome level, within country phylogeographic signal is weak.

Introduction

Phylogeography is widely known as the study of the geographic distribution of genetic lineages among populations within species and among closely related species (Avise, 2009). It has grown rapidly in the last three decades since the word “Phylogeography” was coined by Avise *et al.* (1987). It combines different subject areas to interpret the geographic distribution patterns of genetic markers and by extension populations or species (Leal *et al.*, 2016). 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 (Bermingham & Moritz, 1998; Wiens, 2012; Avise *et al.*, 2016; Riddle, 2016; Zhang *et al.*, 2017).

Maternally inherited genetic markers are widely used to explore the evolutionary history of animals (Avise, 2009; Alexander *et al.*, 2015; Riddle, 2016). Mitochondrial DNA (mtDNA) is utilised in the majority of evolutionary studies, being nearly exclusively inherited from the maternal side, with no recombination and rapid mutation rates (Alexander *et al.*, 2015; Lan *et al.*, 2017). The mtDNA *D*-loop diversity has been exploited extensively in analysing the history of domestic chickens, with findings such as the identification of candidate domestication geographic areas and dispersion routes (Liu *et al.*, 2006; Storey *et al.*, 2012). In chicken, the first study reporting the results based on the full mtDNA genome is Miao *et al.* (2013), which attempted to solve, unsuccessfully, the ongoing debate about the centre(s) of origin of the domestic chicken. Later, Di Lorenzo *et al.* (2015) confirmed that mtDNA can be helpful in providing information about human-mediated dispersal of chicken from candidate domestication centre(s).

The W chromosome has some features in common with mtDNA. They are both maternally inherited (with female chicken being the heterozygous sex ZW), and they represent haploid non-recombining parts of the genome, apart from the pseudo-autosomal regions (PAR) of the W chromosome (Smeds *et al.*, 2015; Graves, 2016). Smeds *et al.* (2015) mentioned that the W chromosome and mtDNA both offer a stable maternally inherited genome that can be used in phylogenetic studies. Using a large number of re-sequenced samples 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 a genetic marker.

In this chapter, we assess 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 (Chapter 2), through analysing 243 new full mitochondrial DNA genome sequences from Ethiopia, Iraq and Saudi Arabia. We also assess the usefulness of W chromosome polymorphisms as maternally inherited marker for the understanding of the diversity of domestic chicken. For this, we use samples collected from two countries

(Ethiopia and Iraq), representing the female samples that were included in the mtDNA analysis.

Materials and Methods

Collection of samples and genomic DNA isolation

For the full mtDNA analysis, 243 samples were analysed from three countries (**Figure 3.1**), Iraq (n = 27), Ethiopia (n = 211) and Saudi Arabia (n = 5). The Saudi samples were downloaded from the NCBI database (accession no. SRP142580) (Lawal *et al.*, 2018). For the W chromosome, the same female birds included in the mtDNA work were analysed. 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). The Ethiopian samples (n = 137) were collected from 19 different geographical areas representing populations from both highland and lowland areas of the country (Chapter 2). 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 Chapter 2.

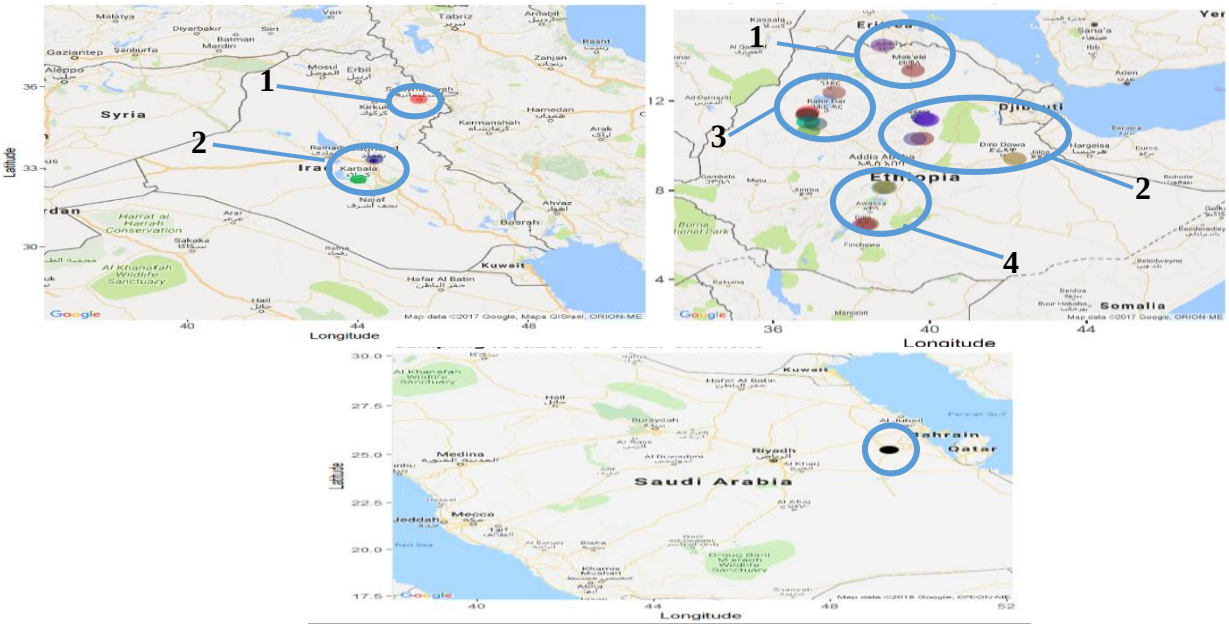


Figure 3.1: Sampling locations for the countries included in this study. For Iraq, 1 = north-eastern region, 2 = central; Ethiopia, 1 = north, 2 = central-east, 3 = west and 4 = south.

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 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 normalised 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 8 cycles of PCR amplification. The libraries were normalised, 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 (Li & Durbin, 2010). 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 group. Insertion, deletions (indels), and realignments were performed with the Genome Analysis Toolkit (GATK) v.3.7 (McKenna *et al.*, 2010), 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) (DePristo *et al.*, 2011; Van der Auwera *et al.*, 2013). The data file was restricted to bi-allelic SNPs and three files were generated: autosomal files (used in Chapter 4), a full mtDNA file and the W chromosome file. The results of the analysis of the latter two are reported here.

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 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 (Li *et al.*, 2009). The output was aligned by ClustalX v.2.1 for multiple sequence alignment (Thompson *et al.*, 1997) for each country separately.

Sequences of the mtDNA genome for each country were analysed and genetic diversity parameters such as the number of haplotypes (H), haplotype diversity (Hd), nucleotide diversity (π) and average number of nucleotide differences were calculated using DnaSP v.5 (Librado & Rozas, 2009), 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 the ~ 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 better comparison of SNP distributions between mtDNA and the W chromosome, an extensive search inside the chromosome was done to identify which W chromosome genomic region had more SNPs than others of a similar length, compared to the full mtDNA. A W-chromosome genomic region with high numbers of SNPs was found at position 4101210 to 4114277. It is 13 kb in length (**Supplementary Figure S1**), and was annotated using the Ensemble (http://www.ensembl.org/) BioMart tool for gene identification.

Also, the same process for generating sequencing data as that for the mtDNA was followed using bcftools v.1.8 (Li *et al.*, 2009). Multiple sequences aligned by ClustalX v.2.1 (Thompson *et al.*, 1997). DnaSP v.5 (Librado & Rozas, 2009) 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 (see Chapter 2). Beside the mtDNA samples from Iraq, Ethiopia and Saudi Arabia, we included 19 reference sequences from Miao *et al.* (2013)'s study (**Table 3.1**). The relationship among the haplotypes was examined through Median-Joining network using the PopArt software (Leigh & Bryant, 2015), which uses the same network algorithm as Bandelt *et al.* (1999) but for sequences larger than 5000 bp.

Table 3.1: Full mtDNA reference sequences included in this study.

No.	NCBI Accession No.	Haplogroup references	Geographic origin	Reference
1	GU261695.1	A	Widely distributed	Miao *et al.* (2013)
2	NC_007235	B	Widely distributed	Nishibori *et al.* (2005)
3	GU261701	C1	Asia	Miao *et al.* (2013)
4	GU261680	C2	Asia	Miao *et al.* (2013)
5	GU261707	C3	Asia	Miao *et al.* (2013)
6	NC_007237	D1	Asia & Africa	Nishibori *et al.* (2005)
7	GU261683	D2	Asia & Africa	Miao *et al.* (2013)
8	GU261697	D3	Asia & Africa	Miao *et al.* (2013)
9	GU261709	E1	Widely distributed	Miao *et al.* (2013)
10	HQ857209	E2	South Asia	Miao *et al.* (2013)
11	GU261708	E3	South Asia	Miao *et al.* (2013)
12	GU261691	F	Asia	Miao *et al.* (2013)
13	GU261690	G	Asia	Miao *et al.* (2013)
14	GU261715	H	East & South Asia	Miao *et al.* (2013)
15	GU261698	I	East & South Asia	Miao *et al.* (2013)
16	GU261706	W	Asia	Miao *et al.* (2013)

17	GU261692	X	Asia	Miao *et al.* (2013)
18	GU261693	Y	Asia	Miao *et al.* (2013)
19	GU261674	Z	Asia	Miao *et al.* (2013)

Analysis of Molecular Variance (AMOVA)

Overall, we followed the same procedure described in Chapter 2, but the Iraqi samples were divided into two groups (central - Baghdad and Karbala, and north-east - Sulimania), while the Saudi Arabian samples were used as one group of five samples all coming from the same area (east).

Neutrality test and demographic history

See Chapter 2 for experimental details.

Results

Distribution of SNPs - MtDNA

Figure 3.2 shows the distribution of SNPs for every 1000 bp along the ~17 kb of the full mtDNA compared to reference NCBI accession no. NC_007235.1 (Nishibori *et al.*, 2005) for the Iraqi, Ethiopian and Saudi Arabian samples. Overall, the highest number of polymorphisms were present in the Ethiopian populations (n = 255) comparing 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 had 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 were observed in the former than the latter. Ethiopian populations showed great variation in SNP numbers, highest in Adane (n = 72), Loya (n = 65), ShubiGemo (n = 62) and Kumato (n = 60) and lowest in Midir (n = 9) and Negasi Amba (n = 9). The central-eastern region of Ethiopia exhibited 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 showed a range of 0-7 SNPs per gene for the 13 mtDNA genes,

highest in the ND4 gene (n = 7), then COXI and COXII (n = 6 in each), while ND2, ND3 and ND4L genes shared the lowest number of SNPs (n = 2); no polymorphism was present in the ATP8 gene. In Ethiopia, mtDNA genes showed more SNPs than the Iraqi samples. Variant numbers ranged 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 showed 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). In general, more variants were detected in the mtDNA protein-coding genes of Ethiopian samples than those from other countries, with Iraqi samples to some extent appearing to be higher than Saudi samples, in accordance with the number of birds and populations examined in each country.

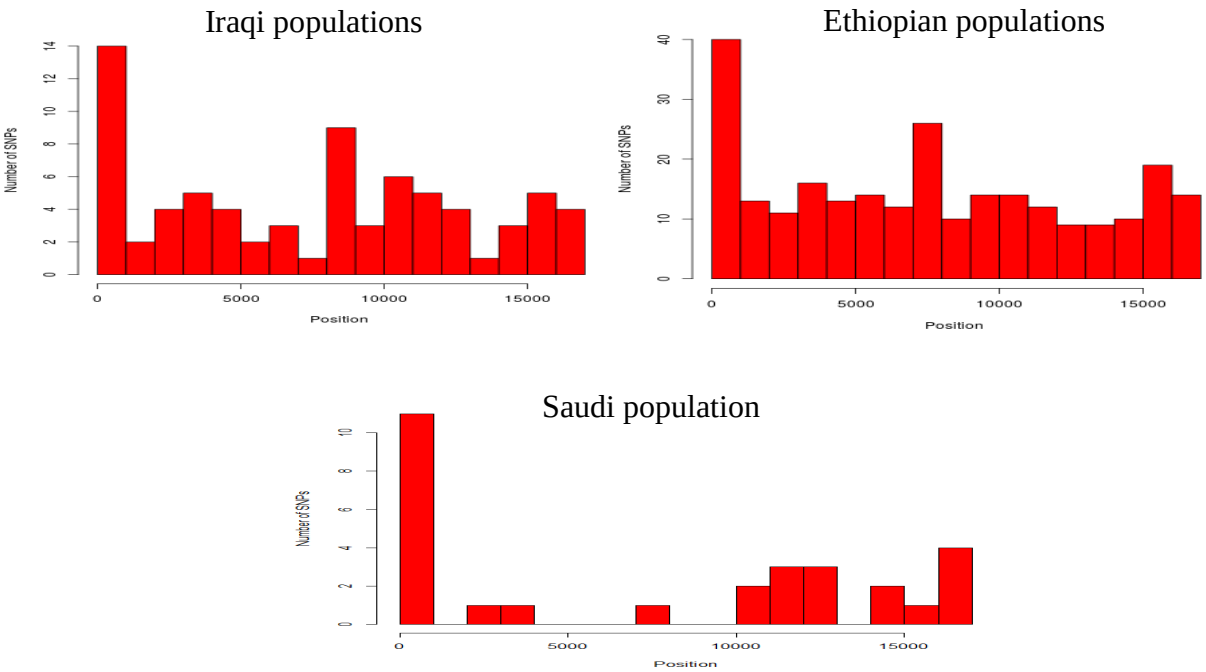


Figure 3.2: Distribution of SNPs along the full mtDNA sequences for the three countries included in this study.

Distribution of SNPs - W chromosome

Figure 3.3 shows the distribution of SNPs along ~ 6 Mb of the W chromosome for the Iraqi and Ethiopian populations. A similar pattern of SNP diversity along the chromosome was observed, with the average number of SNPs ranging from 59 to 318 for the Iraqi populations, and 86 to 363 SNPs for the Ethiopian populations. We chose a

13-kb region (position 4101210 to 4114277) to represent the W chromosome in this study (**Supplementary Figure 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 identified 4 SNPs from Iraqi samples; these variants have been reported before and so 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 was 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 2 novel ones.

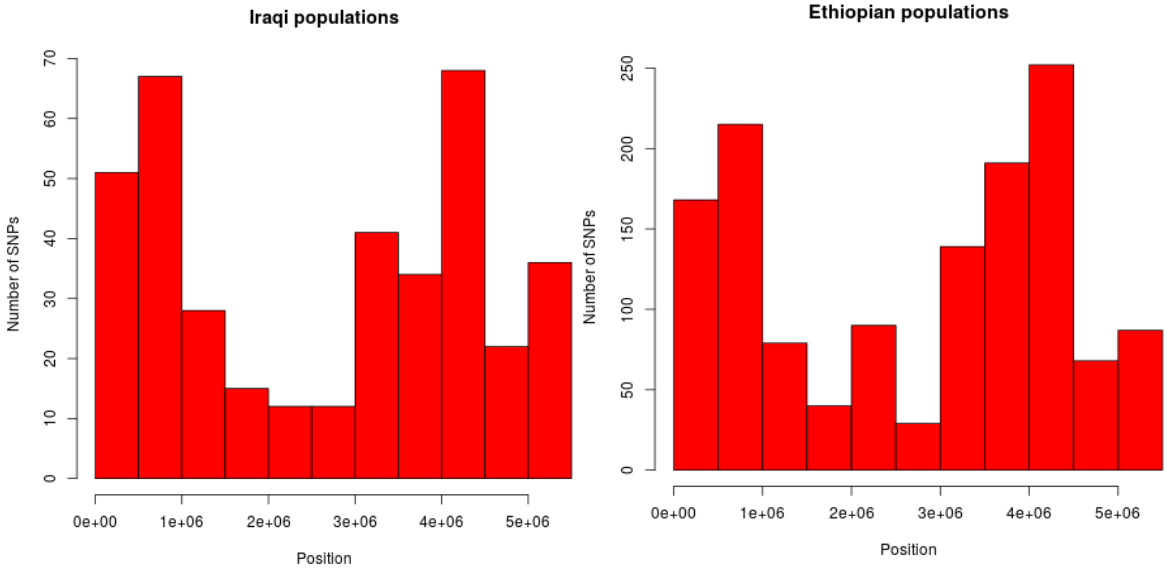


Figure 3.3: Distribution of SNPs (for every 500 kb) along the full W chromosome sequences of two countries included in this study.

Genetic diversity - MtDNA

A high level of genetic diversity was present in the full mtDNA sequences from the three countries (**Table 3.2**). The total number of segregating sites was 292, ranging 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 ranged between 0.655-1.000, the highest (1.000) found in Adane, Meseret and Saudi, while the lowest was in Horro (0.655 ± 0.086). The haplotype diversity of the Iraqi populations was 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).

Table 3.2: Location, sample size and genetic diversity of full mtDNA sequences.

Country/Population	N	S	H	Hd (SD)	π (SD)	K
Iraqi populations						
North-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 populations						
North						
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
ShubiGemo	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) and K = average number of nucleotide differences.

Genetic diversity - W chromosome

While much lower than the mtDNA (**Table 3.3**), the genetic diversity of the W chromosome also indicated that overall that there was more diversity in Ethiopian chickens compared to Iraqi populations. In fact, the Iraqi populations did 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 had 1-4 polymorphic sites, with the exception of Tzion, with 11 segregating sites and the highest nucleotide and haplotype diversities, and Jarso: two other populations (Adane and Horro) had very low differentiation (**Table 3.3**). At the regional level in Ethiopia, the central-eastern and southern regions had about the same level of haplotype and nucleotide diversity, while the west showed the highest level of haplotype diversity (0.891 ± 0.015) and the north has the lowest one (0.545 ± 0.062).

Table 3.3: Location, sample size and genetic diversity of W chromosome sequences.

Country/Population	**N**	**S**	**H**	**Hd (SD)**	**π (SD)**	**K**
Iraqi populations						
North-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 populations						
North						
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
ShubiGemo	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) and K = average number of nucleotide differences.

Within-country phylogenetic relationships

Iraq

The result of the maximum likelihood unrooted tree showed the presence of three haplogroups in Iraqi mtDNA samples (**Figure 3.4**). These haplogroups were A, D1 and E1 based on the classification of Miao *et al.* (2013). Haplogroup A was found in the Sulimania population (north-east region) in 6 haplotypes from 9 samples. The second haplogroup (D1) was represented by one haplotype in the Baghdad population (central region). The last haplogroup (E1), the most frequent, was found in the two populations (Baghdad and Karbala), represented by 10 haplotypes for 17 samples out of the 27 studied here. The presence of three haplogroups in Iraqi chickens were highly supported by a range from 84 to 100 bootstrap values for all the tree branches.

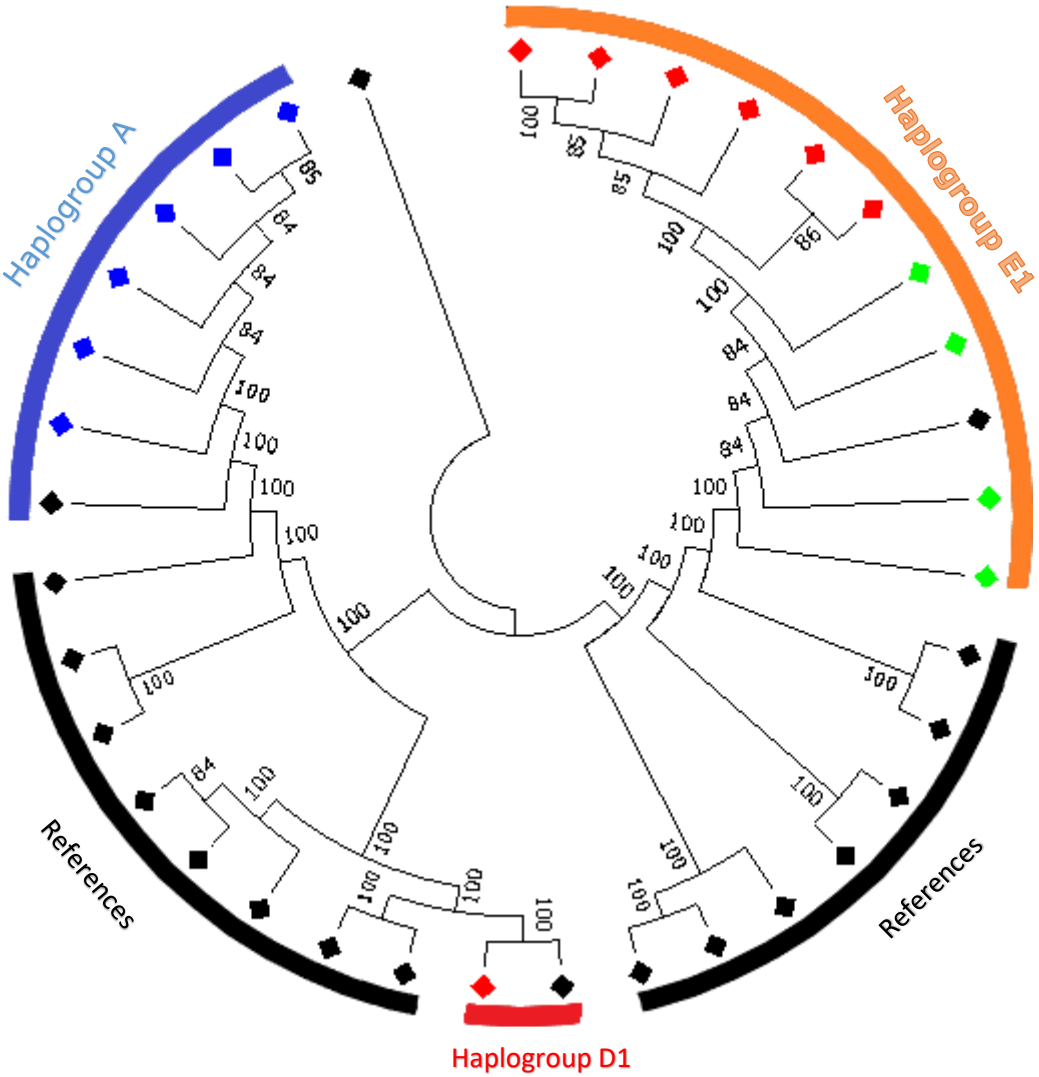


Figure 3.4: Maximum likelihood unrooted tree for the 17 Iraqi mtDNA haplotypes with haplogroup references from Miao *et al.* (2013). Blue = Sulimania, Red = Baghdad and Green = Karbala.

The 27 Iraqi mtDNA samples belonged to 17 distinct haplotypes (**Figure 3.5**). The Baghdad population had the highest number of haplotypes (n = 7), then Sulimania (n = 6) and Karbala (n = 4). No common haplotype was identified among the populations, but rather the distribution of the haplotypes reflected the geographical distribution of the sampling locations (see **Figure 3.1**), with the central region including Baghdad and Karbala linked and closer together than to Sulimania. The majority of haplotypes were represented by one sample only (12 out of 17 haplotypes), while two (H_1 and H_10) were a bit more frequent, represented each by 4 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).

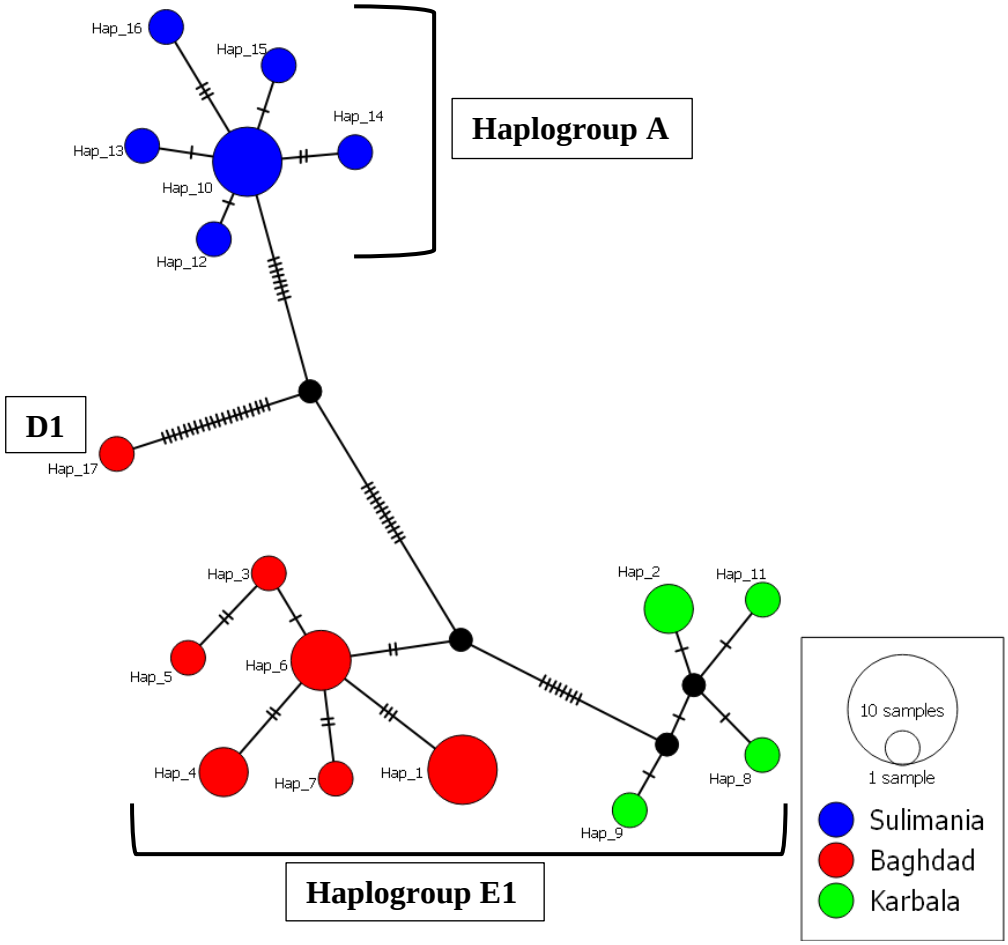


Figure 3.5: Median-joining haplotype network of Iraqi mtDNA haplotypes and haplogroups.

The maximum likelihood unrooted tree for Iraqi W chromosome haplotypes is shown in **Figure 3.6**. The tree shows the presence of three haplotypes, each population represented by one haplotype. Green haplotype colour (upper) representing Karbala population, blue for Sulimania (right) from the North-East part of the country and red (left) for Baghdad.

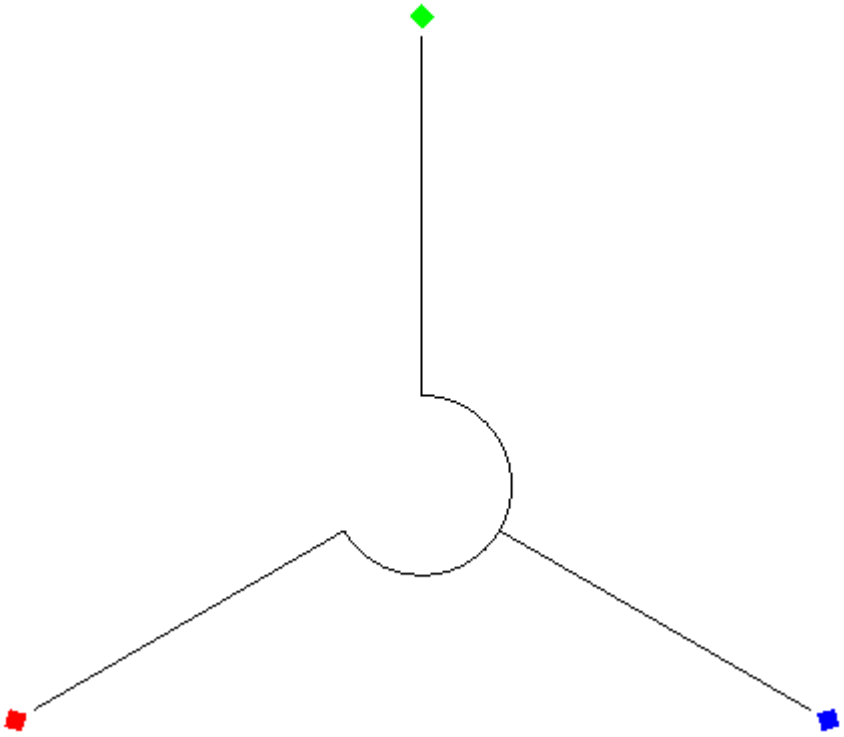


Figure 3.6: Maximum likelihood unrooted tree for Iraqi W chromosome haplotypes. Blue = Sulimania, Red = Baghdad and Green = Karbala.

The median-joining network of Iraqi W chromosome shows only three haplotypes (**Figure 3.7**). Each population clustered in a single haplotype, 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, and is separated from Baghdad by 3 mutations and a single mutation from Karbala.

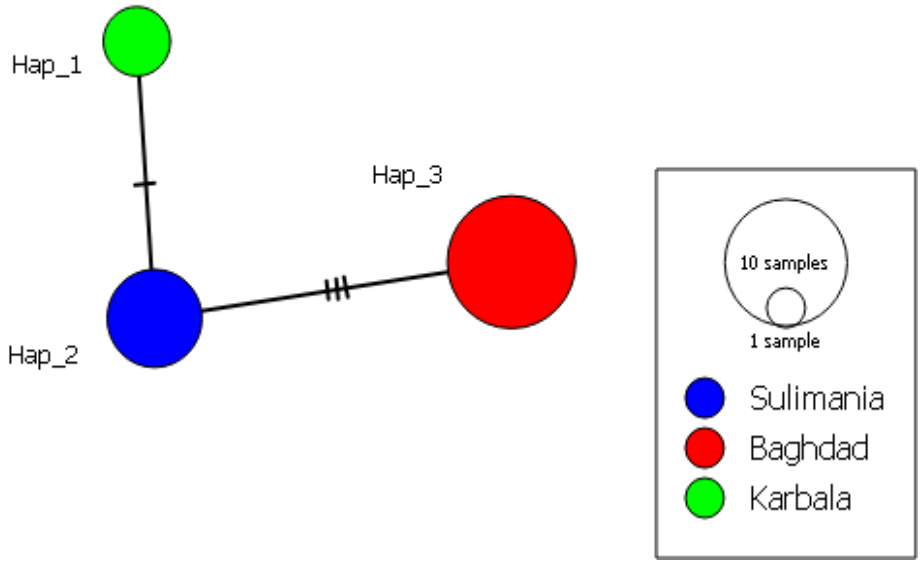


Figure 3.7: Median-joining haplotype network of Iraqi W chromosome haplotypes.

Ethiopia

The phylogenetic tree of Ethiopian mtDNA sequences revealed the presence of five haplogroups and 134 haplotypes (**Figure 3.8**). These haplogroups were A, B, C1, C2 and E1, represented by different numbers of samples: haplogroup A by 6 samples from Meseret (n = 1) and Horro (n = 5); haplogroup (B) was noticed in one bird from the Mihquan population, and the same applied to C1 represented by one bird from Adane; C2 was present in 5 populations (Jarso, Adane, Kumato, Shubi and Loya with 34 birds). The majority of birds (n = 169) clustered with haplogroup E1. It is obvious from these results that the most frequent haplogroups were E1 and C2, with the other three haplogroups at low frequencies. The majority of branches in this maximum likelihood unrooted tree are supported by high bootstrap values.

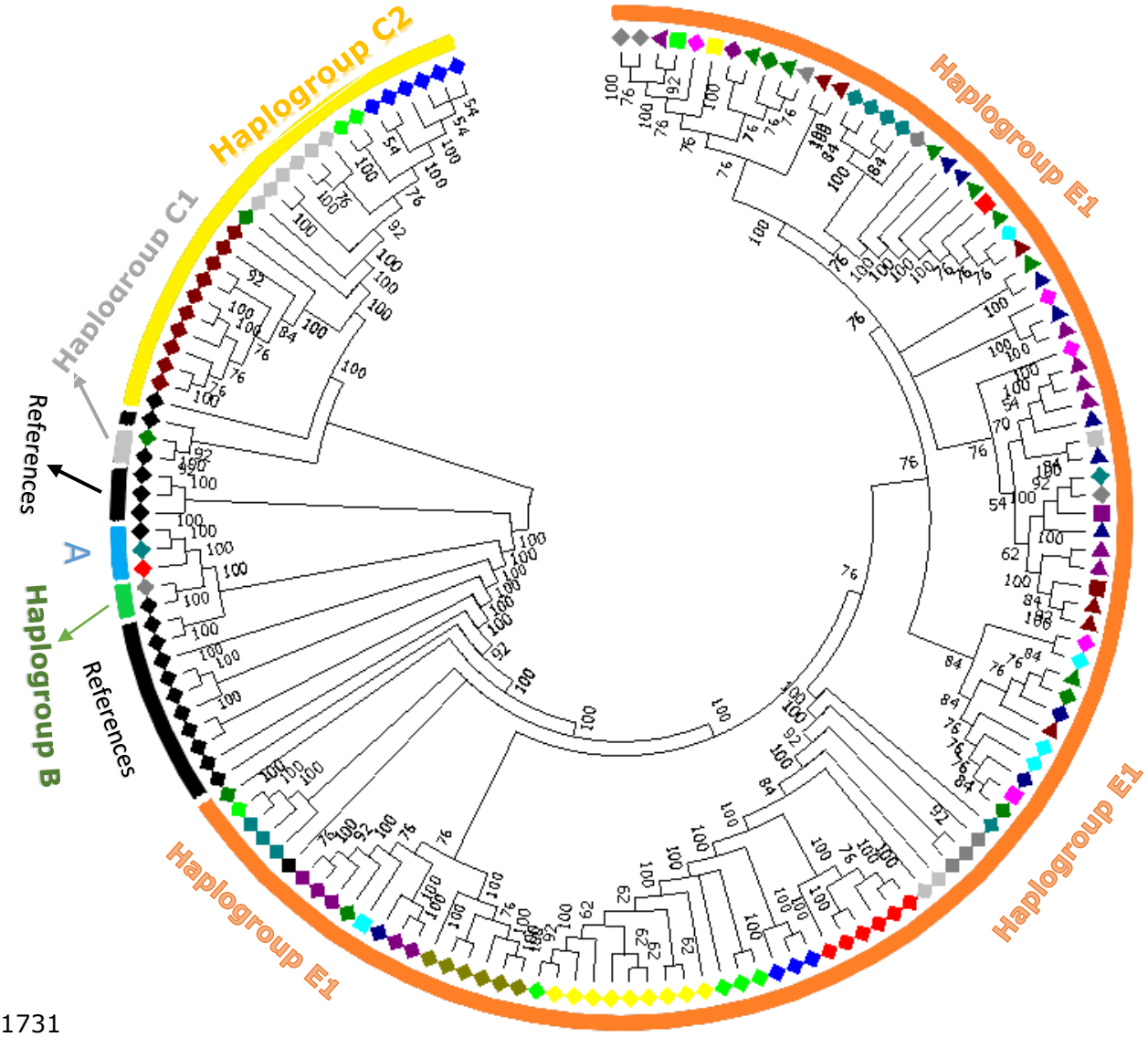


Figure 3.8: Maximum likelihood unrooted tree for Ethiopian mtDNA haplotypes with haplogroup references from Miao *et al.* (2013). ◆ = Jarso, ◆ = Adane, ◆ = Girissa, ◆ = Midir, ◆ = Arabo, ◆ = Meseret, ◆ = Mihquan, ◆ = Kumato, ◆ = Horro, ◆ = Shubi, ◆ = Dikuli, ◆ = Loya, ◆ = Gafera, ◆ = Batambie, ▲ = Tzion, ▲ = 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.

The relationships among Ethiopian mtDNA haplotypes is illustrated in **Figure 3.9**. 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 including six populations (Dikuli, Horro, Jarso, Kumato, Loya and Shubi) that do not

share haplotypes with 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 6 samples represented by two haplotypes in the first group; and haplogroup C2 consists of 34 individuals, also from populations in the first group.

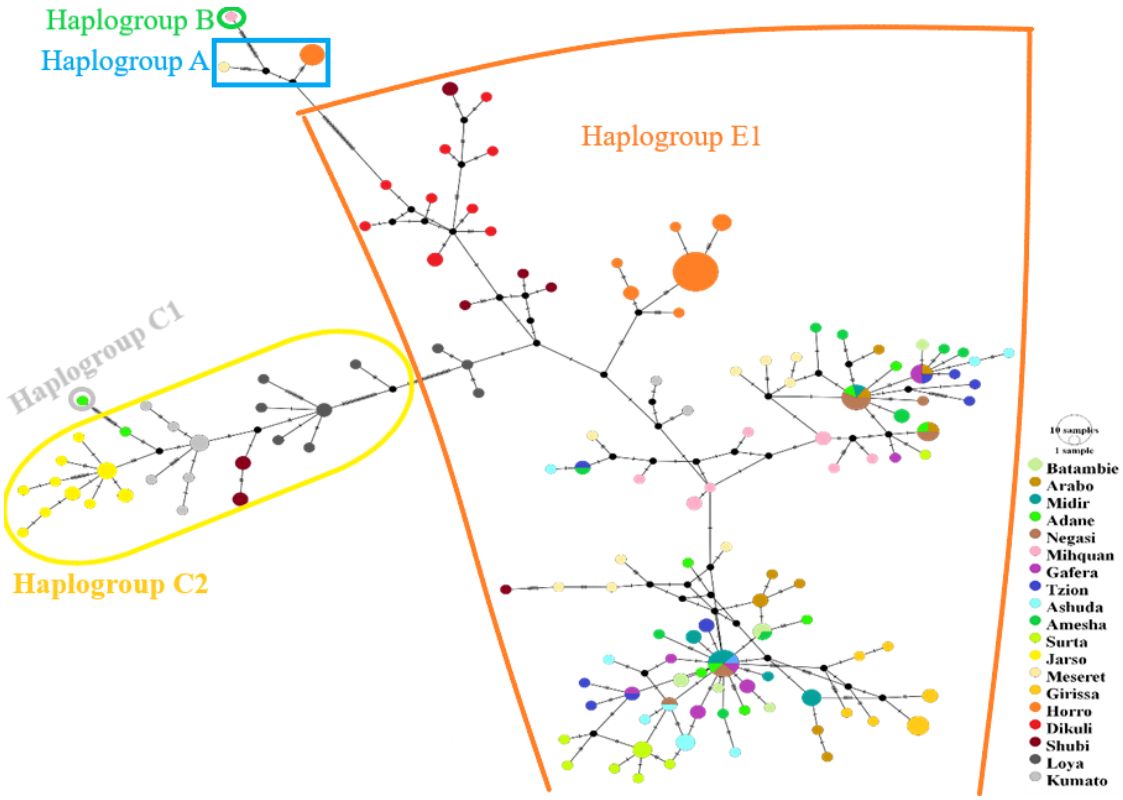


Figure 3.9: Median-joining network of Ethiopian mtDNA haplotypes.

The maximum likelihood tree for the Ethiopian W chromosome haplotypes is shown in **Figure 3.10**. It showed three groups of haplotypes, the first including 11 haplotypes representing 11 populations (Arabo, Amshi, Batambi, Horro, Loya, Meseret, Midir, Mihquan, Shubi, Surta and Tzion), the second of four populations represented by four haplotypes (Dikuli, Kumato, Horro and Negasi), and the third containing six populations (Adane, Ashuda, Girrisa, Gafera, Jarso and Tzion) represented by 9 haplotypes. Two populations (Horro and Tzion) showed haplotypes in both groups. The majority of the bootstrap confidence levels of the tree branches ranged from 70-100.

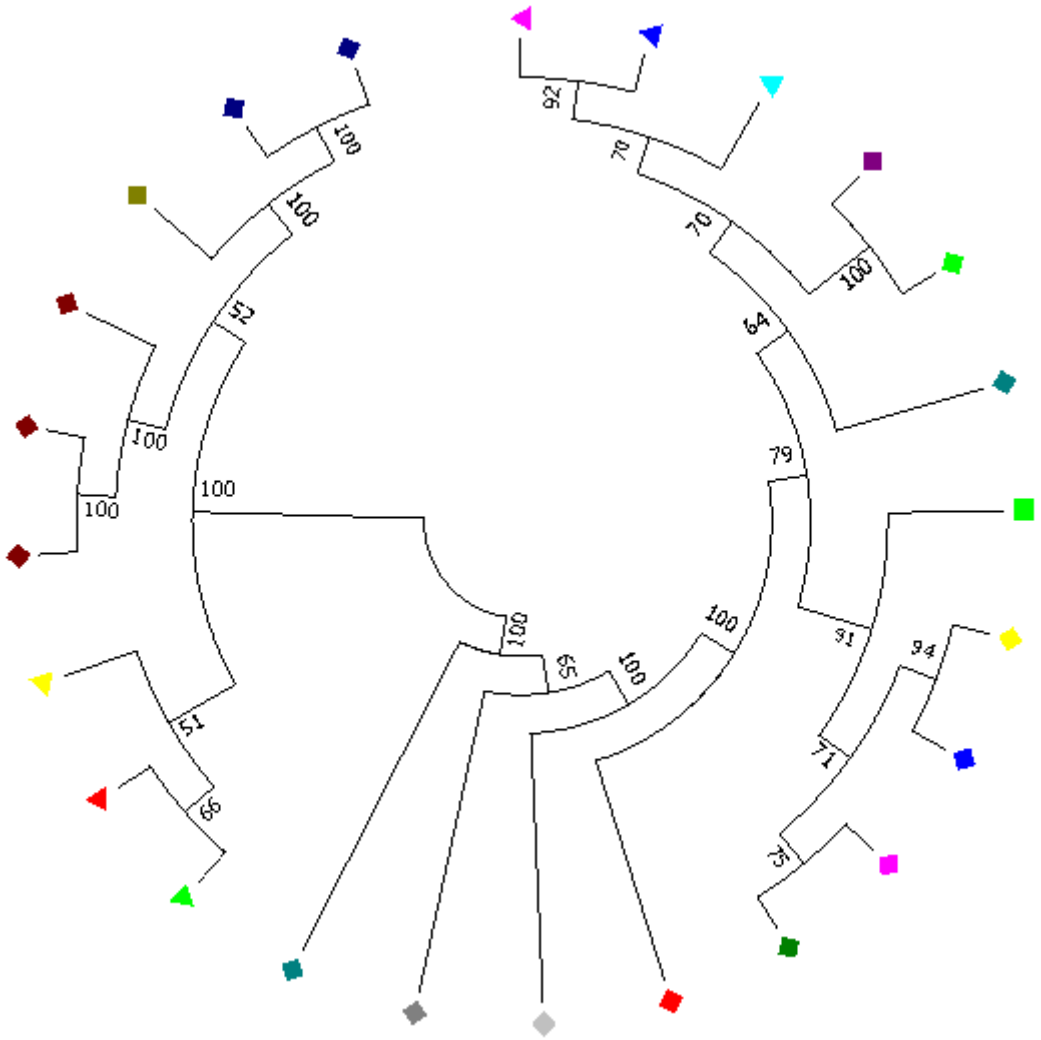


Figure 3.10: Maximum likelihood unrooted tree for Ethiopian W chromosome samples. ▲= Midir, ▲ = Arabo, ▲= Amshi, ◆ = Loya, ◆ = Shubi, ◆ = Horro, ■ = Meseret + Tzion, ◇ = Batambie, ◆ = Surta, ◆ = Mihquan, ◆= Tzion, ◆= Negasi, ◇ = Dikuli, ◆= Kumato, ▲= Gafera, ▲= Ashuda, ▲= Tzion, ◆= Jarso, ◆= Girissa, ◆= Adane.

The relationship of 24 Ethiopian W chromosome haplotypes is shown in **Figure 3.11**. The majority of populations showed only one haplotype containing all their samples; the highest number of haplotypes were found in Jarso, Tzion, Adane and Horro (3, 3, 2 and 2 respectively). All but one haplotype is population-specific, the shared haplotype being found in Meseret (6 samples) and Tzion (3). Mutations among haplotypes range from 1 to 4, where 1 mutation is more frequent than others, then 2 mutations come next and 3 mutations are the lowest.

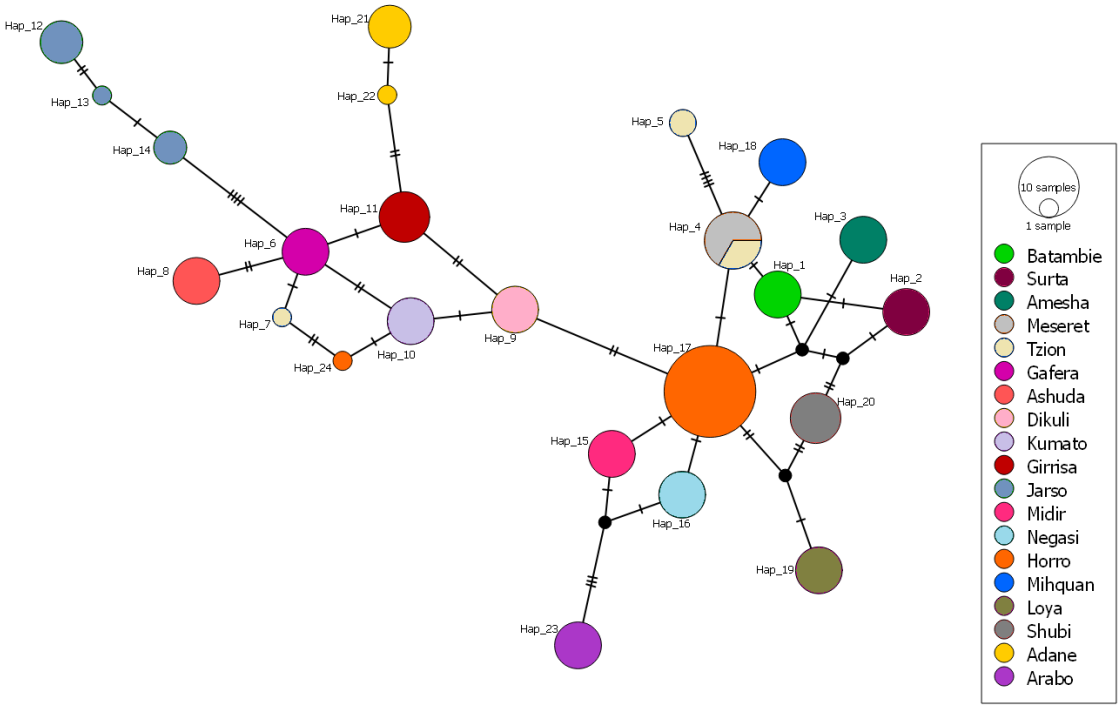


Figure 3.11: Median-joining haplotype network for the Ethiopian W chromosome sequences.

Saudi Arabia

Figure 3.12 shows the relationship between the five Saudi mtDNA haplotypes. All the haplotypes belong to one haplogroup E1. The haplotypes were positioned in different places, two in sister lineages, and the others with a basal position for each one of them comparing to the first two haplotypes. All the branches have a high confidence value (100) except four. Two of these four are linked to two Saudi haplotypes.

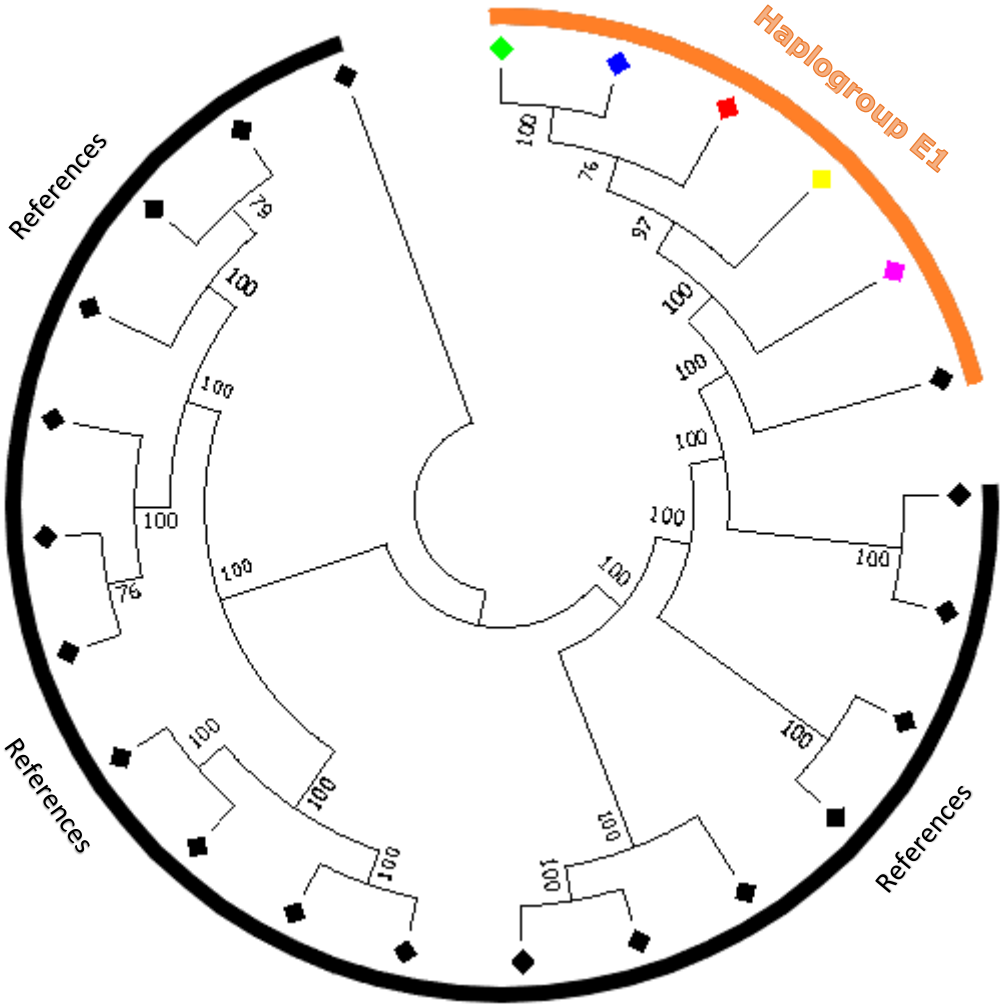


Figure 3.12: Maximum likelihood unrooted tree for the five Saudi mtDNA haplotypes with haplogroup reference sequences from Miao *et al.* (2013). ◆ = Saudi sample 1, ◆ = Saudi sample 4, ◆ = Saudi sample 5, ◆ = Saudi sample 2, ◆ = Saudi sample 3.

The relationship among the Saudi haplotypes is shown at the median-joining haplotype network (**Figure 3.13**). 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 Saudi samples in the phylogenetic tree in **Figure 3.12**.

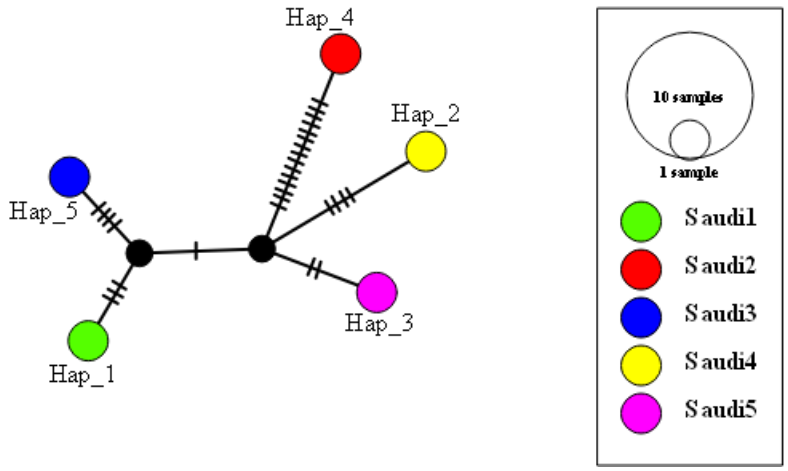


Figure 3.13: Median-joining haplotype network of Saudi mtDNA haplotypes. Saudi1 = Hap 1, Saudi2 = Hap 4, Saudi3 = Hap 5, Saudi4 = Hap 2 and Saudi5 = Hap 3.

Analysis of Molecular Variances (AMOVA)

The results of the analysis of molecular variances for the full mtDNA sequences are shown in **Table 3.4**. 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 among populations within countries (53.3%), then among individuals within populations (38.6%), with among countries the lowest (8.1%).

Table 3.4: 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	

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 3.5** 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, 9 populations have values > 1, and the Saudi population has a value > 1.

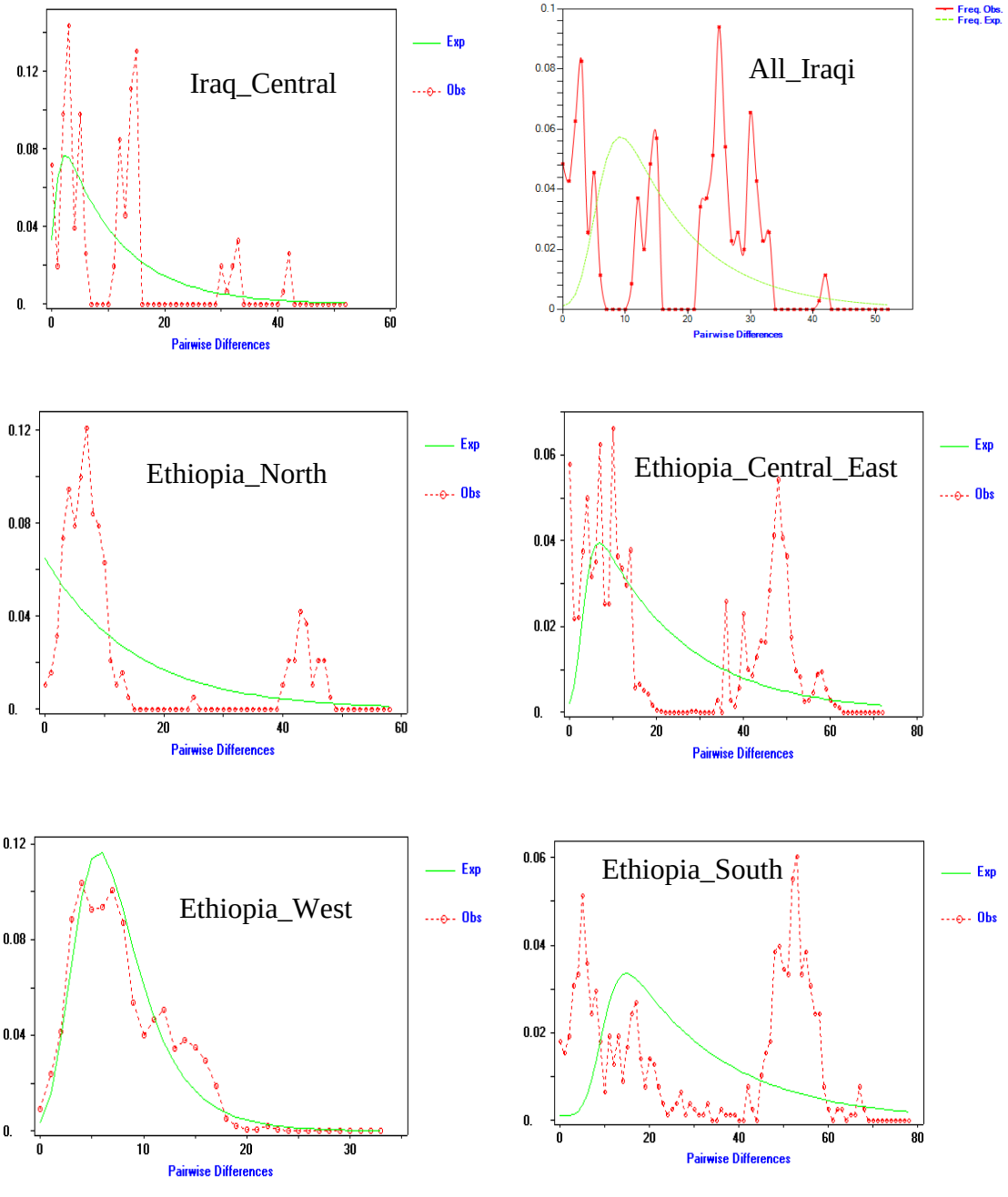
Tajima's D results for the populations were mostly negative and non-significant, with the exception of Baghdad and Sulimania for Iraq, and Meseret, Mihquan, and Surta for Ethiopia, all of which show significant negative Tajima's D values. Horro, Loya and Shubi all have positive Tajima's D values, although 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.

1821 **Table 3.5**: 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 populations						
North-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 populations						
North						
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)

1822

The pairwise differences or mismatch distributions displayed three patterns (**Figure 3.14 and Supplementary Figure S2**), uni-, bi- and multimodal, the commonest being the bi- and multimodal patterns. The unimodal pattern was present in one Iraqi population only (Karbala) (**Supplementary Figure S2**). These patterns support a complex history for these populations: the results of **Table 3.5** 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.



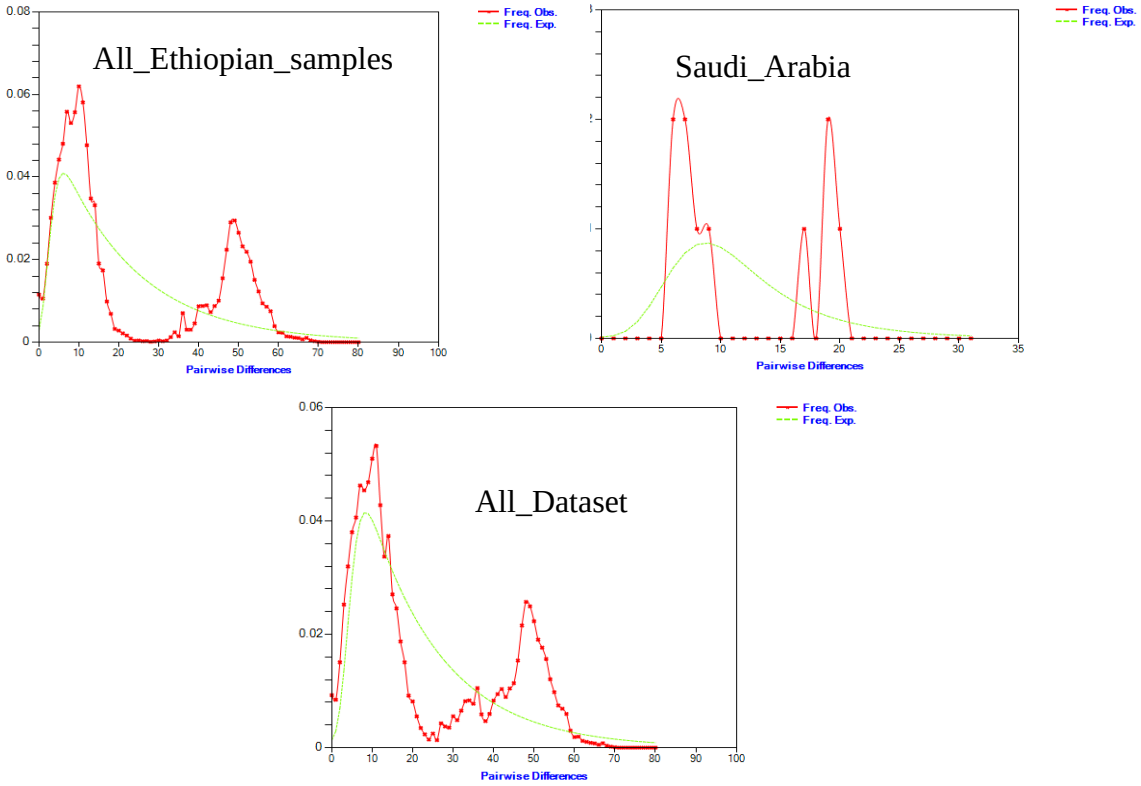


Figure 3.14: Regions and countries mismatch distribution patterns for the full mtDNA sequences (n = 243).

Bayesian Skyline Plots (BSPs) (**Figure 3.15**) 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 rapid than for the two other countries. They also witnessed a bottleneck similar to that of Ethiopia. The Saudi chicken was the only one with evidence for an old expansion without showing signs of recent decrease.

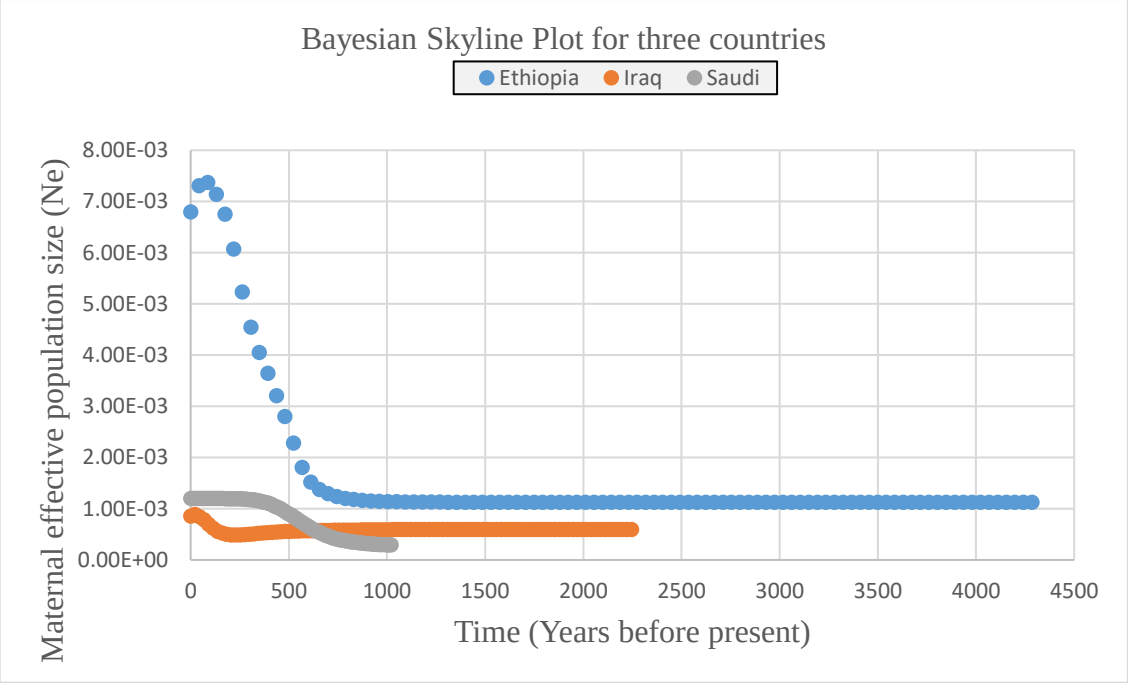


Figure 3.15: Bayesian Skyline Plot (BSP) for three countries included in this study.

Discussion

We analysed 243 full mtDNA sequences and 157 W chromosome samples. The mtDNA was retrieved from three countries (Ethiopia (n = 211), Iraq (n = 27) and Saudi Arabia (n = 5), while W chromosome was sampled from two (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.

SNPs were distributed across ~17 kb of mtDNA (**Figure 3.2**), and one genomic region displayed more variation than the others in samples from all the countries. This region is the first 1 kb of mtDNA that includes the hypervariable *D*-loop region, known as the non-coding part of the mtDNA. More distributed polymorphisms were noticed in Ethiopian populations than other countries, perhaps related to the higher number of samples from Ethiopia than Iraq and Saudi Arabia. This sampling effect can also be seen within protein-coding genes, where SNP numbers are higher in Ethiopian samples compared with the others, and only a small difference between Iraq and Saudi Arabia for the same genes.

At first glance, the distribution of SNPs showed high variation along the W chromosome, but these SNPs were distributed along ~6 Mb length, while the distribution of SNPs in mtDNA is only along ~17 Kb, a more focused area. The level of polymorphism observed

at the W chromosome is lower than the average number of polymorphisms observed on the chicken autosomes (Axelsson *et al.*, 2004; Wilson Sayres, 2018). Together with lack of recombination, it suggests that selection pressure on the W chromosome is higher compared with the average selection pressure on autosomes.

The extracted 13-kb of the W chromosome genomic region was chosen on particular criteria. We needed the most suitable length with the highest number of SNPs. The appropriate length needed to be similar or near the length of the full mtDNA sequence, and to have the highest number of variants because mtDNA is known for its high mutation rate (Alexander *et al.*, 2015). These criteria were all satisfied for the non-coding genomic region that we used to represent the W chromosome in this study. After making the comparison between mtDNA and W chromosome more suitable by selecting 13-Kb of the W chromosome region (**Supplementary Figure S1**) with ~17 Kb mtDNA (**Figure 3.2**), the total number of SNPs for Iraqi and Ethiopian W chromosome samples are 4 and 27 respectively, whereas they show much higher number of SNPs using mtDNA (Iraq = 71 and Ethiopia = 255).

Genetic diversity parameters showed results that are compatible with the distribution of SNPs in both genetic markers. For mtDNA, the studied samples have a high level of genetic diversity (**Table 3.2**) at both population and country levels. The Iraqi central region has more variation than the north-east, either because of the different number of samples or the same interpretations of chapter 2, i.e. either that chicken migration routes went through the central part of the country or that Baghdad (the Capital) being the Centre of Islamic civilisation played a crucial role in attracting people with their animals to live there and contribute to the chicken gene pool.

In the W chromosome, all of the populations of both countries displayed no or low genetic variation, with the exception of Jarso and Tzion Teguaz from Ethiopia, with W haplotypes distinct from the ones observed in the other populations and showing unique polymorphism compared to the W chromosome sequence of reference (**Table 3.3**). This reduced diversity of the W chromosome has been mentioned before (Smeds *et al.*, 2015) from large population samples of flycatcher birds: genetic diversity for the W chromosome tends to be extensively reduced within species, and 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.

An important point was noticed from the results for the genetic diversity of the W chromosome: no genetic diversity was evident for within population from a geographical area, but combining different populations from various geographical locations can show a respectable level of diversity. This was seen in the central region of Iraq and the four geographical regions of Ethiopia. Thus adding different W chromosome samples from different populations can show some level of diversity; this may agree with Smeds *et al.* (2015) about the efficiency of using W chromosome among species rather than within 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 due to clustering with clade E1. Liu *et al.* (2006) suggested in his paper that clade A and E are widely distributed in Eurasia, exactly as we found for Iraq. The reference material for clade D1 (known as cockfighting chicken) came from Bali Island in Indonesia (Nishibori *et al.*, 2005), but has been recognized in one Iraqi sample. The presence of haplogroup D1 is unique to Iraq compared with the other countries. The clustering of this sample with this reference could possibly suggest a maritime route of chicken introduction to this country besides the terrestrial route represented by haplogroup E1. The other haplotypes presented unique haplotypes for each population, and the network reflected the geographical sampling locations of the Iraqi mtDNA. Baghdad and Karbala were closer together relative to Sulimania which is more distantly related. The phylogenetic results for Iraqi W chromosome have not added any new information, because the reduced level of genetic diversity showed no clear structure of the phylogenetic tree among the populations. The network with three haplotypes representing three Iraqi populations did not reflect geographical position, as noticed in the Iraqi mtDNA data.

Five haplogroups were noted in the Ethiopian mtDNA: 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. This could be ascribed to more than one maternal origin of the Ethiopian chicken following more than one route of introduction. As mentioned before, haplogroup E1 signifies the Indian subcontinent origin with a terrestrial route of introduction, haplogroup B originated from Laos in South-East Asia, C1 came from China (Henan province) and C2 from Southern India following a probable maritime route of introduction to Ethiopia (Miao *et al.*, 2013). These different haplogroups could explain

the high level of genetic diversity for Ethiopian samples in **Table 3.2**. The Ethiopian mtDNA network confirmed the results of the phylogenetic tree concerning more than one maternal origin of Ethiopian samples. It showed two groups of populations, one with distinct haplotypes not shared with other populations, and a second with more frequent haplotypes shared with other populations. Therefore, analysing the full mtDNA genome is adding further resolution to the mtDNA *D*-loop analysis. The levels of resolution can be seen in the genetic diversity of Sulimania populations from Iraq and Jarso from Ethiopia. These two populations exhibited no genetic diversity using the *D*-loop genetic marker, but the full mtDNA data showed that these two populations have a high level of variation. Another example of the difference between full mtDNA and the *D*-loop is the confirmation of the maritime route in Ethiopian samples, derived from the presence of both haplogroups C1 and C2 and greater frequency of the latter, and the absence of haplogroup D that was present in *D*-loop data with 7 haplotypes in Ethiopia. The only presence of haplogroup D in the full mtDNA data was in Iraq with a low frequency (of one haplotype). Furthermore, all the samples from Sulimania from Iraq belonged to haplogroup E from *D*-loop data, while the same samples represent haplogroup A in the full mtDNA data

Unlike the other countries, the W chromosome for Ethiopian samples showed some structure in the phylogenetic tree, and revealed 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. In contrast to the Iraqi W chromosome, the low level of diversity for the Ethiopian W chromosome offered some useful information about haplotype relationships.

The Saudi mtDNA phylogeny displayed a possible Indian subcontinent origin with a terrestrial route of introduction. The tree showing groups for the five samples, indicating the existence of genetic variation among these samples. This variation was more noticeable in the network, which revealed that five haplotypes representing five of the samples were linked to two median vectors (mv), each separated from the other by one mutation. No Saudi samples were used to analyse the W chromosome, because only male samples were available.

The AMOVA results for mtDNA showed that the genetic variation was 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 was slightly different, with most of the variation among populations within groups and among individuals within populations. Over all three countries, the pattern was close to the Ethiopian condition, with higher variation for populations within countries and within populations, rather than among countries. 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 3.5**. It showed that two countries (Ethiopia and Iraq) had an old expansion with a more recent bottleneck. Saudi 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.

In conclusion, this study has exploited the mtDNA and W chromosome of local populations of chickens from different countries to trace their genetic diversity and historical background. The results suggest that mtDNA was the most useful genetic marker to fulfil the wanted purposes. Very little information was obtained from the W chromosome because of the reduced diversity of this genetic marker caused by strong selection pressure and lack of recombination. These results recommend the use of mtDNA as a beneficial genetic marker in the phylogeographic study of chickens, while the W chromosome could possibly be utilized in a study among species rather than within species.

Chapter 4

Autosomal genome evidence for introgression from other *Gallus* species in the genome of African and Middle East indigenous chicken

Abstract

Across-species gene flow or introgression is increasingly being recognised as playing an important role in the evolution and historical genetic background of species. In particular, adaptive introgression may be related to environmental adaptation. Considering the geographic distribution of wild species in a genus could help to trace the history of dispersion of livestock species. Here, we investigate if introgression can be used to trace the dispersal routes of indigenous breeds of chicken outside domestication centres and/or primary dispersal centres.

Domestic chicken genomes from three countries (Iraq n = 27, Ethiopia n = 283 and Saudi Arabia n = 5) were analysed together with the four *Gallus* species (Ceylon = 8, Green n = 12, Grey n = 3 and Red n = 6) and the common pheasant genome to detect the genetic background and frequency of introgressed haplotypes in the domestic chicken gene pool. All samples were sequenced using the Illumina HiSeqX sequencing platform and sequencing data were analysed following the GATK best-practices protocol restricted to bi-allelic sites. The four-taxon method, ABBA-BABA test was followed to detect gene flow from Ceylon, Green and Grey junglefowl into the village chicken genome. Haplotype maximum likelihood trees were constructed to assess the number of distinct introgressed haplotypes, and to count haplotype frequency in the different geographic regions. No gene flow from the Ceylon and Green junglefowl was detected into the gene pool of Iraqi, Ethiopian or Saudi village chicken, but in contrast Grey junglefowl introgression was present in all geographic regions. In particular, we studied six introgressed genomic regions into the domestic chicken derived from Grey junglefowl. Overall their frequencies were low in Iraqi, Saudi Arabian and South Ethiopian populations relative to northern, central-eastern and western Ethiopian populations. Combined with information from the literature and previous chapters, the introgression results provide further support for an Indian subcontinent centre of terrestrial dispersal for the Iraqi and Saudi village chickens. For Ethiopian chickens the situation is more complex, with the country likely to have witnessed both a terrestrial arrival of domestic chicken, originating from the northern part of Indian continent and a maritime one possibly from the southern part of Indian subcontinent, considering the current geographic distribution of the Grey junglefowl.

Introduction

Introgression is gene flow from one species into the gene pool of another, with subsequent backcrossing to the introgressed parental population (Arnold, 2006). It is increasingly thought to play a major role in the evolutionary history of species and the speciation process (Abbott *et al.*, 2013), allowing gene flow and recombination among species (Martin *et al.*, 2014). It may also be responsible for adaptive introgression (Abbott *et al.*, 2013). Across species, introgression is believed to be particularly common in birds (Ottenburghs *et al.*, 2015), where morphological differentiation (e.g. plumage colouration and pattern) is often driven by sexual selection, while F1 hybrids from recognised species are still fertile (Fitzpatrick, 2004).

The genus *Gallus* has four wild species in South and South-East Asia. Red junglefowl (*G. gallus*) is widely distributed in the forests and the warm countries of South and South-East Asia (Morejohn, 1968). There are five recognized subspecies, and the species is considered to be the closest relative to the domestic chicken (Darwin, 1868). The distribution of the Ceylon junglefowl (*Gallus lafayettii*) is in South Asia (Sri Lanka), while the Grey junglefowl (*Gallus sonnerattii*) is found in South-East India (Delacour, 1977). The natural distribution of the Green junglefowl (*Gallus varius*) is Java and the adjacent islands, mainly in the seashore and valleys. Darwin (1868) argued that the Red junglefowl is the only ancestor of the domestic chicken, a view supported by several other studies (Punnett, 1923; Fumihito *et al.*, 1994).

However, Hutt (1949) was the first to suggest that the domestic chicken may have multiple ancestors. Subsequently, Morejohn (1968) successfully produced F2 progeny from the crossing and backcrossing of Red and Grey junglefowl. Nishibori *et al.* (2005) showed the first evidence of introgression between domestic/Red junglefowl with Grey junglefowl from their study of nuclear and mitochondrial DNA. This was followed by Eriksson *et al*. (2008) study, which showed that the yellow skin phenotype in domestic chicken is likely to have originated from an introgression event from the Grey junglefowl. With the technological advancement from genes to genomes, more exploration to chicken genome was made. This process includes, for example, identification of candidate genes and chromosomal regions associated with body weight and growth using 60 K SNP panel and NGS data, revealing loci under selection during chicken domestication and assessing the best way of predicting chicken traits from a breeding perspective by using whole genome sequencing data (Rubin *et al.*, 2010; Gu *et al.*, 2011; Ahsan *et al.*, 2013;

Abdollahi-Arpanahi *et al.*, 2015) Last but not least, Lawal (2018) was the first to investigate at a genome-wide level the three congeners of the Red junglefowl for evidence of gene flow into the genetic background of modern domestic chicken, obtaining strong indications that all three have contributed in varying amounts to the genome of the domestic chicken.

In Chapter 2, we established that the main likely geographic origin of the domestic chickens of Iraq, Saudi Arabia and Ethiopia is likely to have been the Indian subcontinent, but that also maritime trading routes may have brought domestic chicken in to Ethiopia and Saudi Arabia. Accordingly, we examine here if the current autosomal genomic background of these domestic chickens may have been influenced by the genes of the Grey, Ceylon and Green junglefowl. Through the analysis of a large number of domestic chicken samples, we examine if the geographical distribution of this introgression provides us with further evidence of the possible routes of dispersal of domestic chicken to the Middle East and the Horn of Africa.

Materials and Methods

Collection of samples and genomic DNA isolation

Three hundred and fifteen domestic chicken were collected from three countries (**Figure 4.1**) Iraq (n = 27), Ethiopia (n = 283) and Saudi Arabia (n = 5). The Iraqi samples were collected from three different places (see Chapter 3 for more details). The Ethiopian samples include 27 populations, each population being represented by 10 samples with a few exceptions, representing the different agro-ecological zones of the country.

Genomic DNA was extracted from air-dried blood on FTA paper and blood preserved in ethanol (see Chapter 2 for details). Saudi Arabian chicken genome sequences were obtained from Lawal *et al.* (2018)'s study (NCBI accession no. SRP142580). Red junglefowl (Red n = 6), and other junglefowl species (Grey n = 3, Ceylon n = 8, Green n = 12) and common Pheasant *Phasianus colchicus* (n = 2) genome sequences were also downloaded from NCBI (accession no. PRJNA432200) (Lawal *et al.*, 2019). See **Figure 4.1** for sampling locations of the populations and regions.

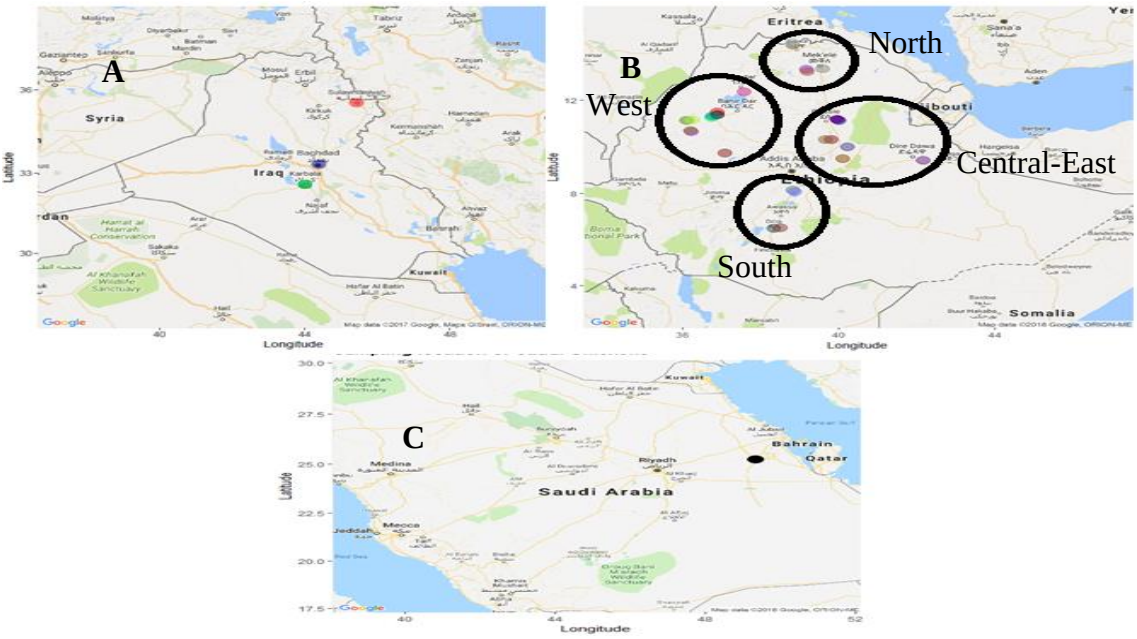


Figure 4.1: Sampling locations of the countries included in this study. **A** = Iraq, **B** = Ethiopia and **C** = Saudi Arabia.

Sequencing, alignment and variants calling

All sequenced reads were aligned against the *Galgal* 5.0 reference genome sequence. Vcf files restricted to autosomes and bi-allelic sites were produced using the GATK tools. (Further details are in Chapter 3.)

Dataset structure and phylogenetic relationships

Genetic diversity parameters were calculated using vcftools v. 0.1.15 (Danecek *et al.*, 2011), where the observed and expected heterozygosity estimated by the --het flag; genome mean nucleotide diversity was estimated for a 20-kb window size with 5-kb sliding window.

Principal Component Analysis (PCA) of the data (346 NGS samples) was explored using plink v. 1.9 (www.cog-genomics.org/plink/1.9/) (Chang *et al.*, 2015). The minor allele frequency flag (--MAF) was applied on a dataset of autosomal SNPs of ~89 million. Genesis PCA (http://www.bioinf.wits.ac.za/software/genesis/) and the Admixture plot Viewer was used for visualising the PCA.

The phylogenetic relationships across samples and species were performed at two levels, mtDNA and all autosomes. In particular, the mtDNA sequences were extracted from the vcf files and the bcftools v. 1.8 consensus option was used to generate the sequence files (Li *et al.*, 2009). These sequence files were aligned using ClustalX v. 2.1 (Thompson *et*

al., 1997) for multiple sequence alignment, removing all gaps and missing sites. Then the best-fit model for the data was predicted by jModeltest v. 2.1.7 (Darriba *et al.*, 2012) and phyml v 3.0 software for tree construction (Guindon & Gascuel, 2003). The autosome tree was constructed in a similar way after randomly selecting 89,000 SNPs, representing 0.1% of the data.

Four-taxon introgression analysis

For the introgression detection process between the domestic chicken of the three countries and the other *Gallus* species (Ceylon, Green and Grey junglefowl), we followed Martin's *fd* protocol (Martin *et al.*, 2014) and the ABBA-BABA (Green *et al.*, 2010) tests.

We selected a 20 kb window size with a 5 kb sliding window, with a minimum of 100 SNPs for each window. *Fd* was calculated following the equation:

$$\hat{f}_d = \frac{S(P_1, P_2, P_3, O)}{S(P_1, P_D, P_D, O)}$$ (Martin *et al.*, 2014)

S = the difference between the sums of ABBAs and BABAs.

P_D = the donor population for independent sites.

This method detects candidate introgressed regions that have an excess of shared derived alleles between two of the four taxa, assuming the other two have the ancestral allele. We denoted P1 as the Red junglefowl, P2 the domestic chicken (for each country separately), P3 as one of *Gallus* species (Ceylon/Green/Grey) and O as the outgroup (the common pheasant).

The two SNP patterns (ABBA-BABA) were defined as follows: In ABBA, the derived allele status (B) will be shared between domestic (P2) and one of the other junglefowl species (P3) (Ceylon/Green/Grey), while the ancestral allele will be shared between (P1) Red junglefowl and the common pheasant (O). In BABA, the gene flow will be between Red junglefowl (P1) and the other *Gallus* species (P3), while domestic and outgroup will be the ancestral condition. The ABBA-BABA tend to have the same value when there is no gene flow or introgression.

Haplotype phylogenetic tree, introgressed region size and haplotype frequencies

In order to examine each significant window from the ABBA-BABA test, haplotype phylogenetic trees were constructed to further support introgression and its frequency, as well as to detect the direction of gene flow.

The significant windows were phased using beagle v. 5.0 (Browning *et al.*, 2018). Then plink v. 1.9 (www.cog-genomics.org/plink/1.9/) (Chang *et al.*, 2015) and customized script were used to generate the sequence (fasta) files. These outputs were then aligned with ClustalX v. 2.1 (Thompson *et al.*, 1997) and the best-fit model for the data was predicted by jModeltest v. 2.1.7 (Darriba *et al.*, 2012). The final stage was to construct a maximum likelihood tree using phyml 3.0.

Although building the haplotype phylogenetic tree for the significant windows helped in recognizing introgression, another step was necessary to obtain the introgressed genomic region length. Therefore, when a candidate introgressed genomic region was spotted, adjacent windows above the threshold were also tested to assess the minimum length of the introgressed fragment.

After the identification of the candidate introgressed haplotypes, the frequency of introgression for each region in each population was calculated. This calculation was done by dividing the number of introgressed haplotypes with the total number present in the population. For visualization of the geographical distribution of introgressed haplotypes, R software v. 3.3.3 (https://www.r-project.org/) was used to draw pie charts of the haplotype frequencies against the map of each country.

Remapping introgressed regions from *Galgal* 5.0 to *Galgal* 6.0

All the samples included in this chapter were mapped against *Galgal* 5.0. After the release of the new reference (*Galgal* 6.0), the coordinates of the introgressed regions were identified corresponding to this new reference. This process was done by using the NCBI remapping service. Default options were used during this remapping process.

Functional annotation of introgressed regions

Genes overlapping with the introgressed regions were identified from their annotation in the Ensembl (Ensembl Genes 95) BioMart online tool (https://www.ensembl.org/biomart/) (accessed 5th February 2019). Gene pathways inferring the biological process were obtained using DAVID v. 6.8 (Database for Annotation, Visualization and

Integrated Discovery) (https://david.ncifcrf.gov/) and PANTHER v. 14.0 (Mi *et al.*, 2016).

Results

Variants and genetic diversity

In Iraq, the highest number of SNPs was found in the north-east population of Sulimania (n = 12796020), followed by Baghdad (n = 12563009) and Karbala (n = 10313701) in the central part of the country. For the Ethiopian populations, the numbers range between 10 and 11 million variants per population. The highest numbers of variants were found in Kumato (n = 11994423), Mihquan (n = 11843800), Loya (n = 11711874), Meseret (n = 11648013), Jarso (n = 11629266), Metkilimat (n = 11628802) and Adane (n = 11625931). The lowest numbers were seen in two populations Local Horro (n = 10070409) and Hugub (n = 10347894). Combining all the samples, Ethiopia shows the highest number of SNPs (n = 19466507) followed by Iraq (n = 15425307) and Saudi Arabia (n = 9884247), in accordance with the number of samples analysed in these three countries (**Table 4.1**).

Table 4.1: Location, sample size and number of variants for 315 domestic chicken.

Country/Population	Number of samples	SNPs count
Iraqi Populations		
Baghdad	13	12563009
Karbala	5	10313701
Sulimania	9	12796020
Total	27	15425307
Ethiopian Populations		
Adane	10	11625931
Amshi	10	11459812
Arabo	10	11419897
Ashuda	10	10867089
Batambie	8	10752196
Dikuli	10	11431418
Gafera	10	11359201
Gesses	10	10759852
Gijet	10	11193400
Girissa	10	11866243
HadushAdi	10	11030421
Improved Horro	30	10632187
Hugub	10	10347894
Jarso	14	11629266
Kefis	10	11048297
Kido	9	10667184

Kumato	10	11994423
Local_Horro	6	10070409
Loya	10	11711874
Meseret	10	11648013
Metikilimat	10	11628802
Midir	10	10750487
Mihquan	10	11843800
Negasi	10	10803666
Shubi	10	11463764
Surta	9	11210073
Tzion	10	11070941
Total	283	19466507
Other population		
Saudi Arabia	5	9884247

Three indicators of genetic diversity were estimated for all the populations and countries (**Table 4.2**). Nucleotide diversity for Iraqi populations demonstrated that Sulimania showed slightly higher diversity (0.0042 ± 0.0017) in relation to Baghdad and Karbala (0.0039 ± 0.0017). For the observed (*Ho*) and expected heterozygosity (*He*), Karbala exhibited the highest values (*Ho* = 0.379 and *He* = 0.376). Ethiopian populations show no major differences among the mean nucleotide diversity values of populations. The highest values were found in Kumato (0.0039 ± 0.0017), Girissa (0.0038 ± 0.0017), Loya (0.0038 ± 0.0017), Meseret (0.0038 ± 0.0016) and Mihquan (0.0038 ± 0.0016), while the lowest were in Jarso (0.0033 ± 0.0015), Midir (0.0032 ± 0.0016) and Negasi (0.0033 ± 0.0016). For the observed and expected heterozygosity, the largest values were found in Batambie (*Ho* = 0.323 and *He* = 0.334), Improved Horro (*Ho* = 0.325 and *He* = 0.318), Hugub (*Ho* = 0.321 and *He* = 0.330) and Local Horro (*Ho* = 0.333 and *He* = 0.352), whereas the lowest value was found in the Jarso population (*Ho* = 0.258 and *He* = 0.278).

Comparing the three countries, they show similar mean nucleotide diversity values. Saudi has the highest observed (0.371) and expected (0.382) heterozygosity, then Ethiopia (*Ho* = 0.290 and *He* = 0.314), and finally Iraq (*Ho* = 0.236 and *He* = 0.250).

Table 4.2: Locations, number of samples and genetic diversity parameters.

Country/Population	Number of samples	Genome mean nucleotide diversity	Observed Heterozygosity (*Ho*)	Expected Heterozygosity (*He*)
Iraqi populations				
Baghdad	13	0.0039 ± 0.0017	0.287	0.295
Karbala	5	0.0039 ± 0.0017	0.379	0.376
Sulimania	9	0.0042 ± 0.0017	0.290	0.309
Total	27	0.004 ± 0.0017	0.236	0.250
Ethiopian populations				
Adane	10	0.0036 ± 0.0016	0.295	0.303
Amshi	10	0.0037 ± 0.0016	0.290	0.314
Arabo	10	0.0034 ± 0.0016	0.291	0.298
Ashuda	10	0.0035 ± 0.0016	0.314	0.319
Batambie	8	0.0036 ± 0.0016	0.323	0.334
Dikuli	10	0.0036 ± 0.0016	0.280	0.311
Gafera	10	0.0036 ± 0.0016	0.295	0.313
Gesses	10	0.0034 ± 0.0016	0.303	0.314
Gijet	10	0.0037 ± 0.0016	0.291	0.324
Girissa	10	0.0038 ± 0.0017	0.312	0.314
HadushAdi	10	0.0037 ± 0.0016	0.297	0.330
Improved Horro	30	0.0034 ± 0.0015	0.325	0.318
Hugub	10	0.0034 ± 0.0016	0.321	0.330
Jarso	14	0.0033 ± 0.0015	0.258	0.278
Kefis	10	0.0036 ± 0.0016	0.289	0.319
Kido	9	0.0035 ± 0.0016	0.298	0.322
Kumato	10	0.0039 ± 0.0017	0.313	0.316
Local_Horro	6	0.0035 ± 0.0016	0.333	0.352
Loya	10	0.0038 ± 0.0017	0.317	0.317
Meseret	10	0.0038 ± 0.0016	0.302	0.315
Metikilimat	10	0.0037 ± 0.0016	0.280	0.314
Midir	10	0.0032 ± 0.0016	0.293	0.298
Mihquan	10	0.0038 ± 0.0016	0.303	0.313
Negasi	10	0.0033 ± 0.0016	0.302	0.306
Shubi	10	0.0037 ± 0.0016	0.313	0.319
Surta	9	0.0037 ± 0.0016	0.291	0.325
Tzion	10	0.0034 ± 0.0016	0.271	0.312
Total	283	0.0039 ± 0.0017	0.290	0.314
Other population				
Saudi	5	0.0037 ± 0.0017	0.371	0.382

Species structure and variation

Within and across species, variation was assessed after quality control on a ~80 million SNPs vcf file. **Figure 4.2** shows PC1 (23% of the variation) and PC2 (16%) of the Principal Component Analysis. In terms of species position, domestic chicken samples (blue) were close to those of the Red junglefowl (red), and separated from other *Gallus* species on PCA1. Ceylon (pink) and Grey junglefowl (grey) were positioned close to each other and distantly separated from Green junglefowl (green) on PC2.

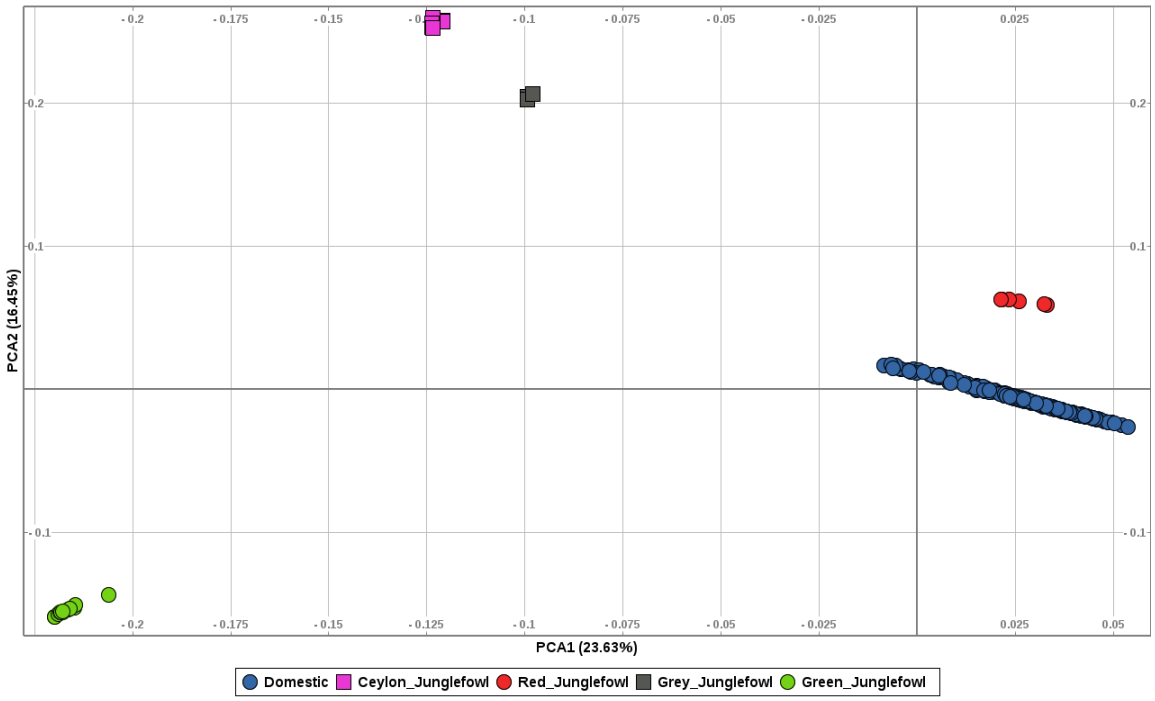


Figure 4.2: Principal Component Analysis (PCA) for 315 domestic chicken and 29 wild Junglefowls.

Phylogenetic relationship among species

This relationship was evaluated using two kinds of data, mtDNA and autosomes. The selected best-fit model for both levels was based on Akaike Information Criterion (AIC) prediction was SYM + I + G. For mtDNA (**Figure 4.3**), the maximum likelihood tree showed high confidence values (100) for all the branches. The tree is rooted on the pheasant, the outgroup. Within the genus *Gallus*, the Green junglefowl lineage separated first. The three samples of Grey junglefowl clustered with the domestic chicken and Red junglefowl lineages. This clustering of Grey junglefowl suggests a maternally inherited introgression from Red/domestic into Grey junglefowl.

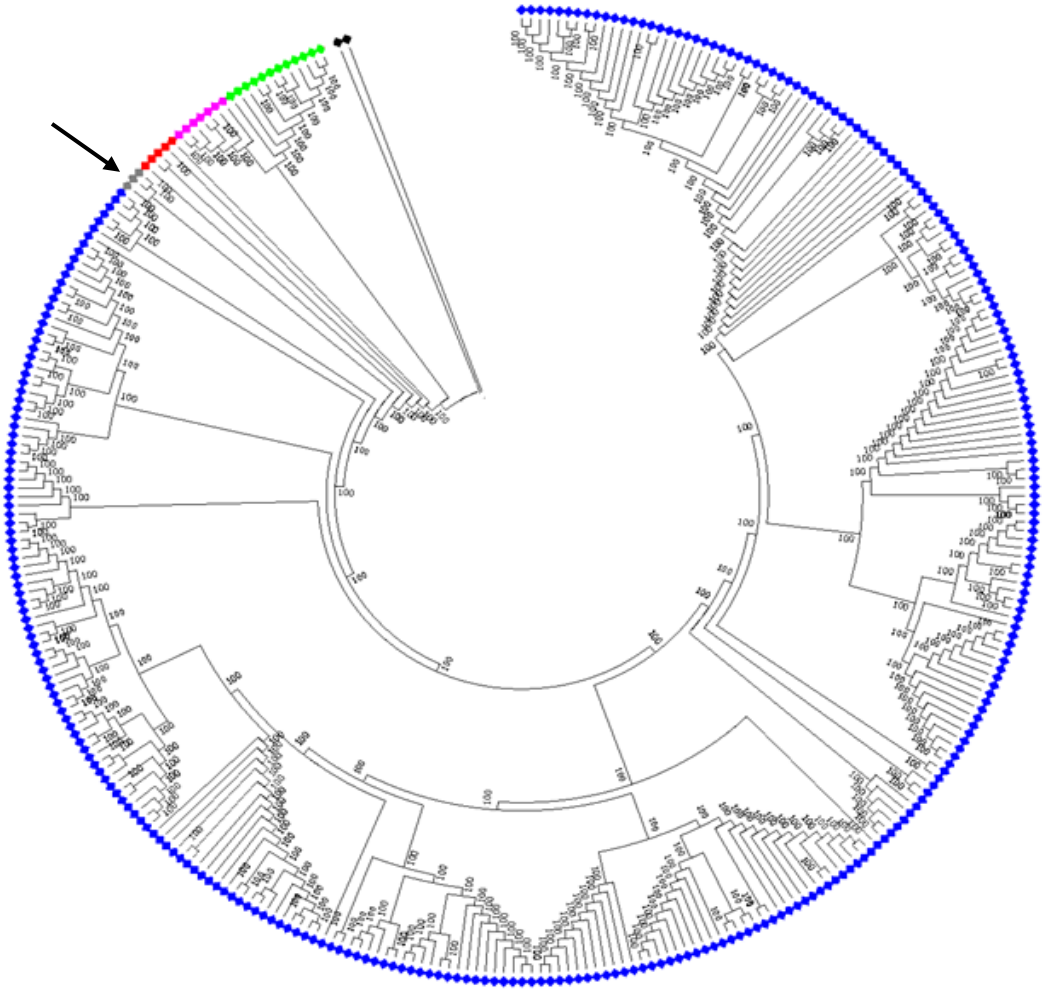


Figure 4.3: Maximum likelihood tree for the full mtDNA. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Red = Red junglefowl, Grey = Grey junglefowl and Blue = Domestic chicken. The arrow refers to the position of the Grey junglefowl within domestic chicken and Red junglefowl clade.

An 89,000 set of SNPs randomly selected from across the entire genome (0.1% of all SNPs) was used for the autosomal phylogenetic dataset. The results also show a high level of bootstrap confidence (100) for every single branch (**Figure 4.4**). As before, the common pheasant lineage formed the outgroup. The Green junglefowl appears to be the basal lineage of all *Gallus* genus. The Ceylon and Grey junglefowl came out as sister species, while the Red junglefowl is positioned at the root of the domestic chicken lineages.

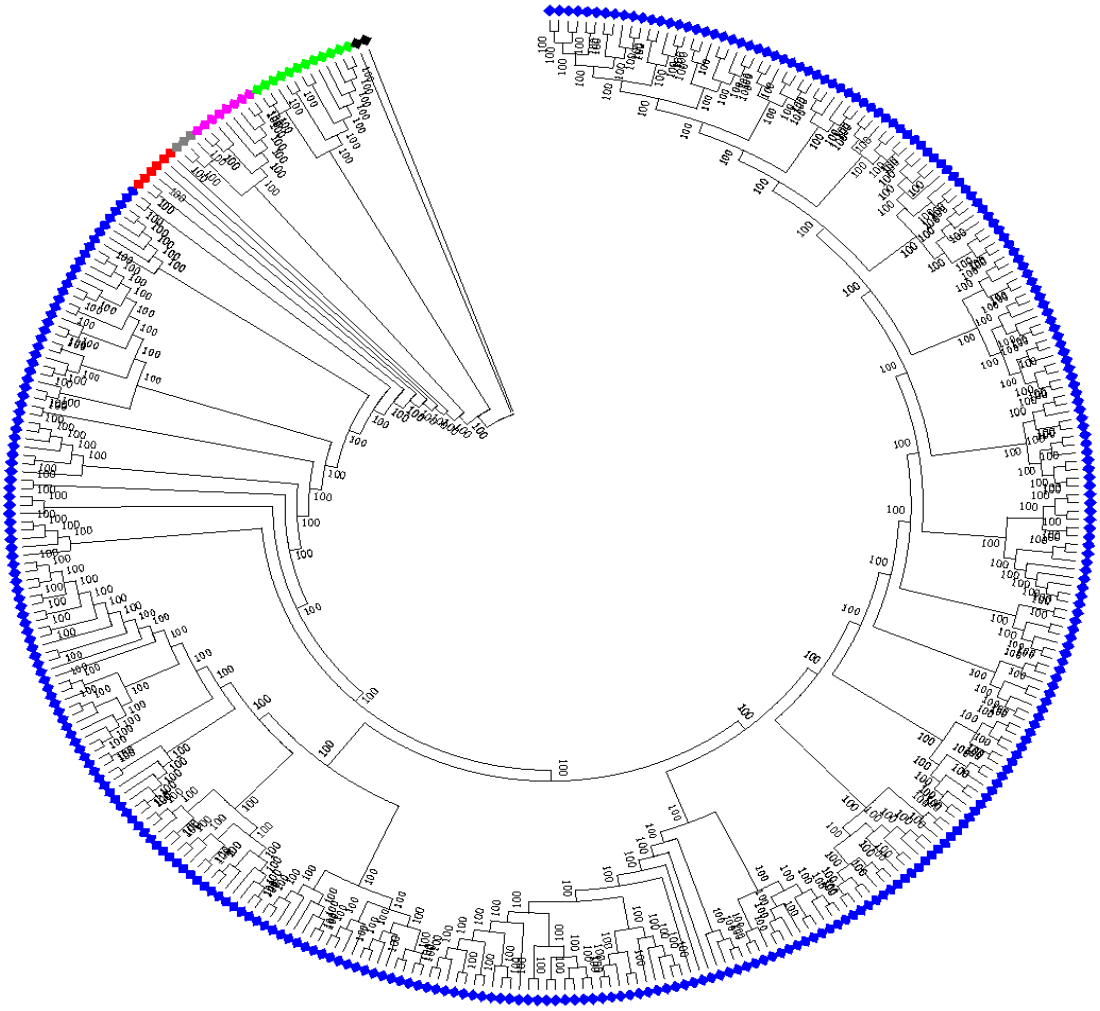


Figure 4.4: Autosome maximum likelihood tree using 89,000 SNPs. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Red = Red junglefowl, Grey = Grey junglefowl and Blue = Domestic chicken.

Introgression Analysis

Introgression between Ceylon junglefowl and domestic chicken

The introgression analysis with domestic chicken was done for each country separately for Ceylon, Green, and Grey. **Figure 4.5** shows the *fd* introgression plots between Ceylon junglefowl and Iraq (A), Ethiopia (B) and Saudi Arabia (C). Windows with at least 100 SNPs per window were only considered. It includes 16823 windows for Iraqi, 12930 for Ethiopian and 9685 for Saudi Arabian samples. In Iraq, significant window regions (*fd* > 0.3) are located in chromosomes 1, 2, 3, 4, 11 and 14; for Ethiopia, these windows are located in chromosomes 1, 2, 4, 6, 7 and 8; and for Saudi Arabia candidate regions are located in chromosomes 1, 2 and 5 only.

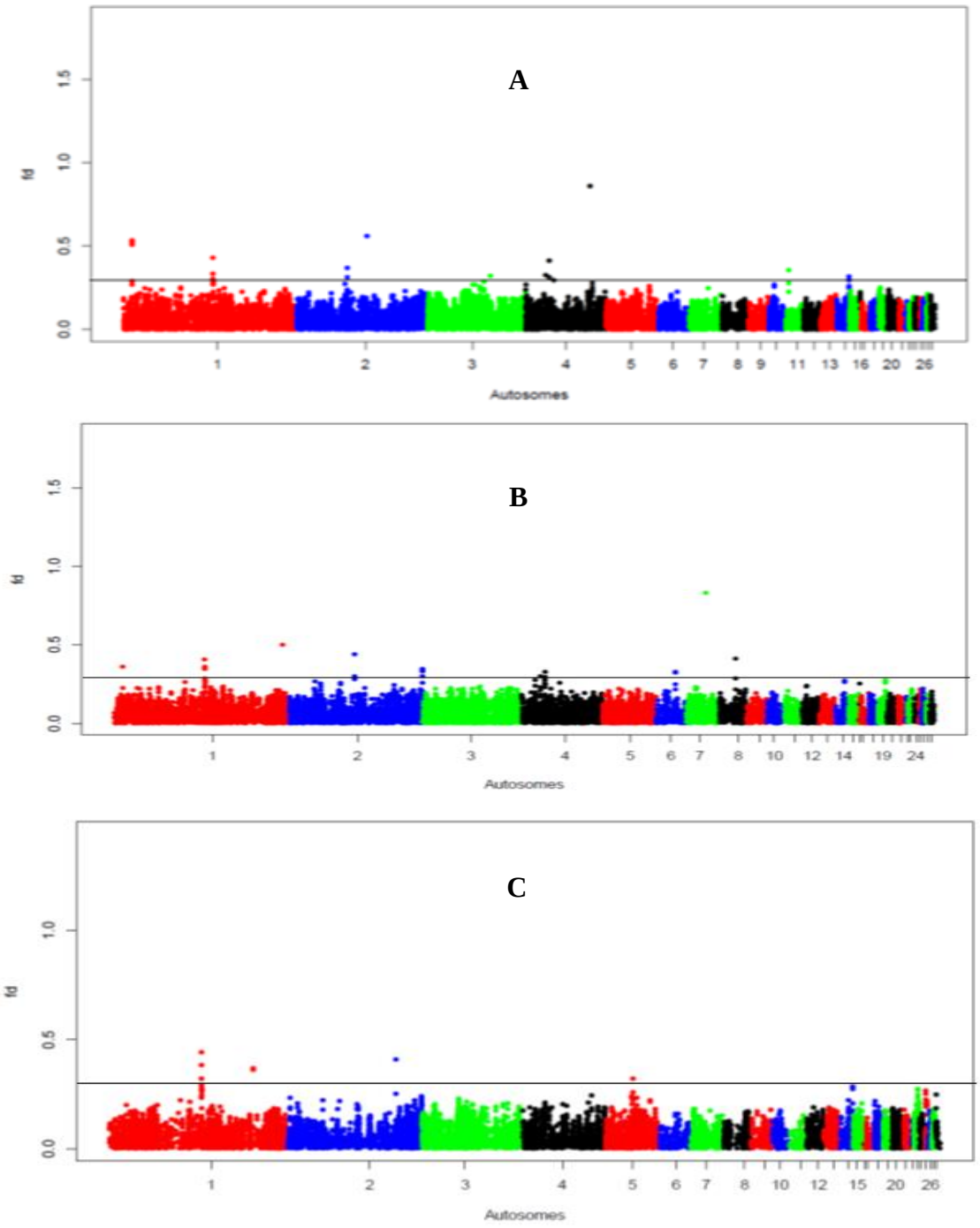


Figure 4.5: Manhattan introgression plots between domestic chicken, Iraq (**A**), Ethiopia (**B**), Saudi Arabia (**C**) and Ceylon junglefowl. The horizontal line represents the chosen threshold ($fd \geq 0.3$).

Windows that have reached the threshold for the three countries were further explored by constructing phylogenetic haplotype trees. These phylogenies do not show any sign or evidence of introgression between domestic chicken and Ceylon junglefowl, but they

mostly exhibit gene flow from Grey junglefowl to domestic chicken or *vice versa* (**Figures 4.6, 4.7, 4.8, and supplementary S1 to S33**).

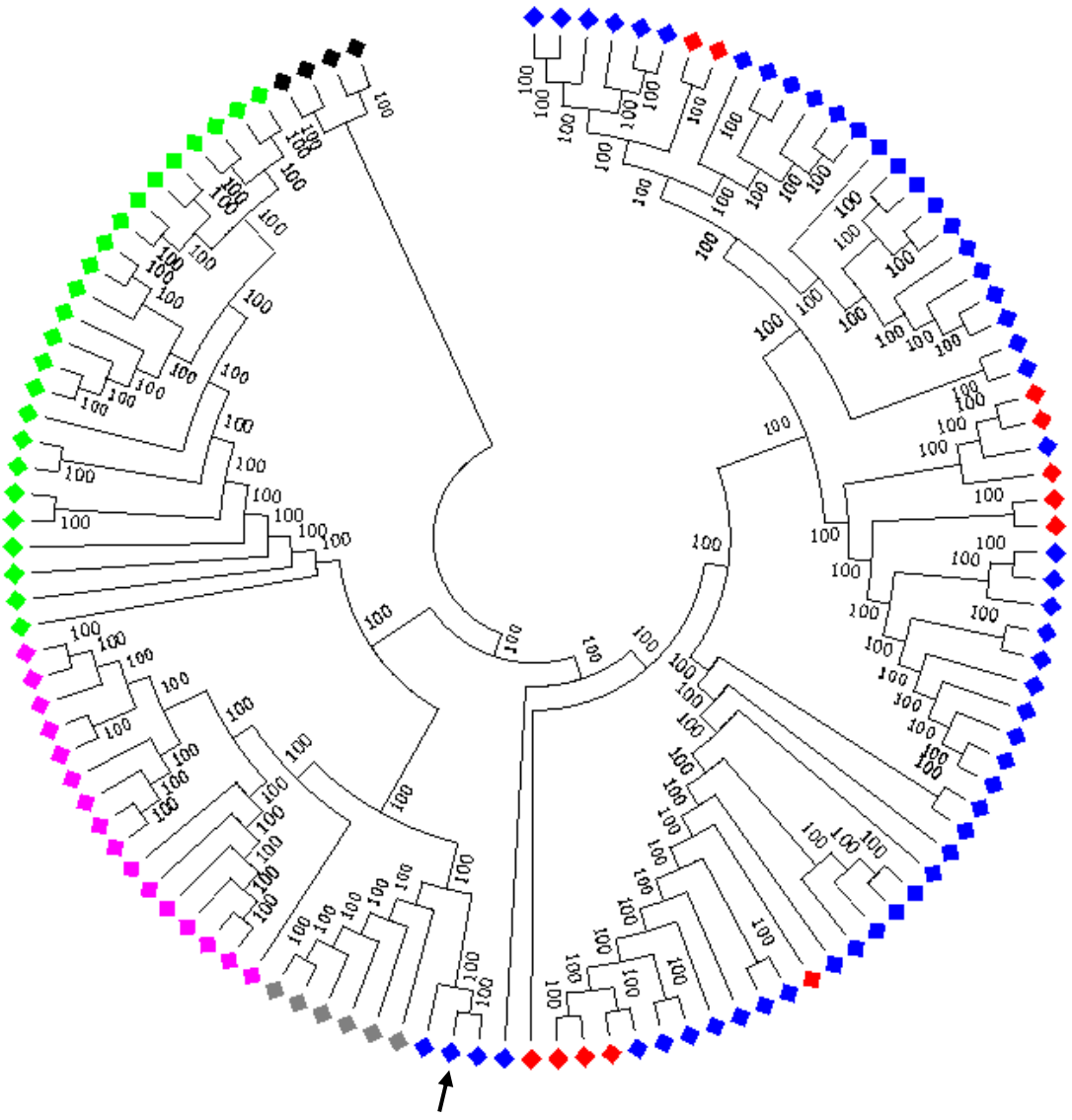


Figure 4.6: Haplotype maximum likelihood tree for a Ceylon junglefowl candidate introgressed region on chromosome 1 (position 9480001-9500000) of Iraqi domestic chicken. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl. The arrow indicates introgression from Grey junglefowl to Domestic chicken for a significant window in Ceylon junglefowl analysis

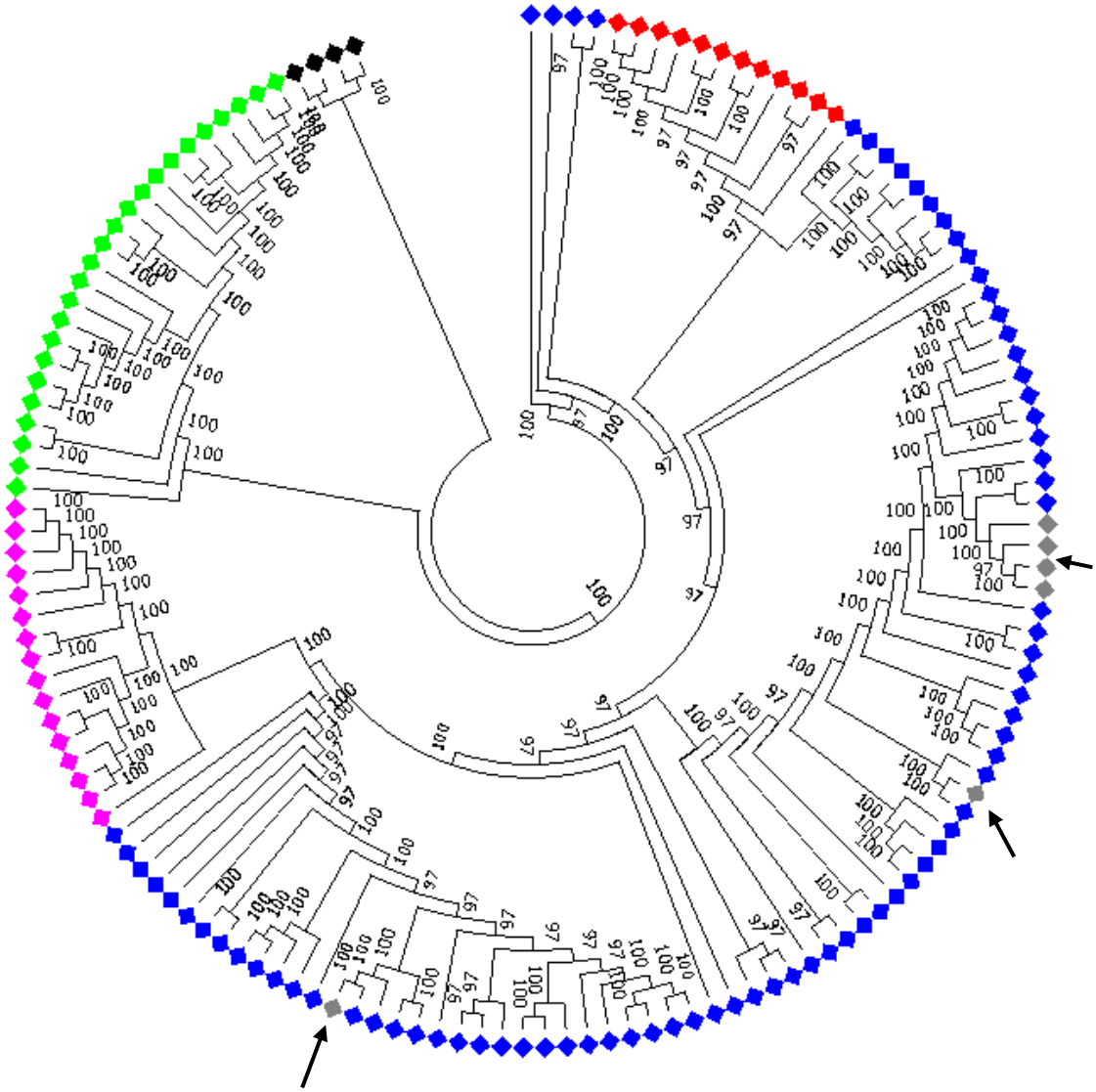


Figure 4.7: Haplotype maximum likelihood tree for a Ceylon junglefowl candidate introgressed region on chromosome 1 (position 101080001-101100000) of Ethiopian domestic chicken. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Arrows indicate introgressed haplotypes between domestic chicken and Grey junglefowl.

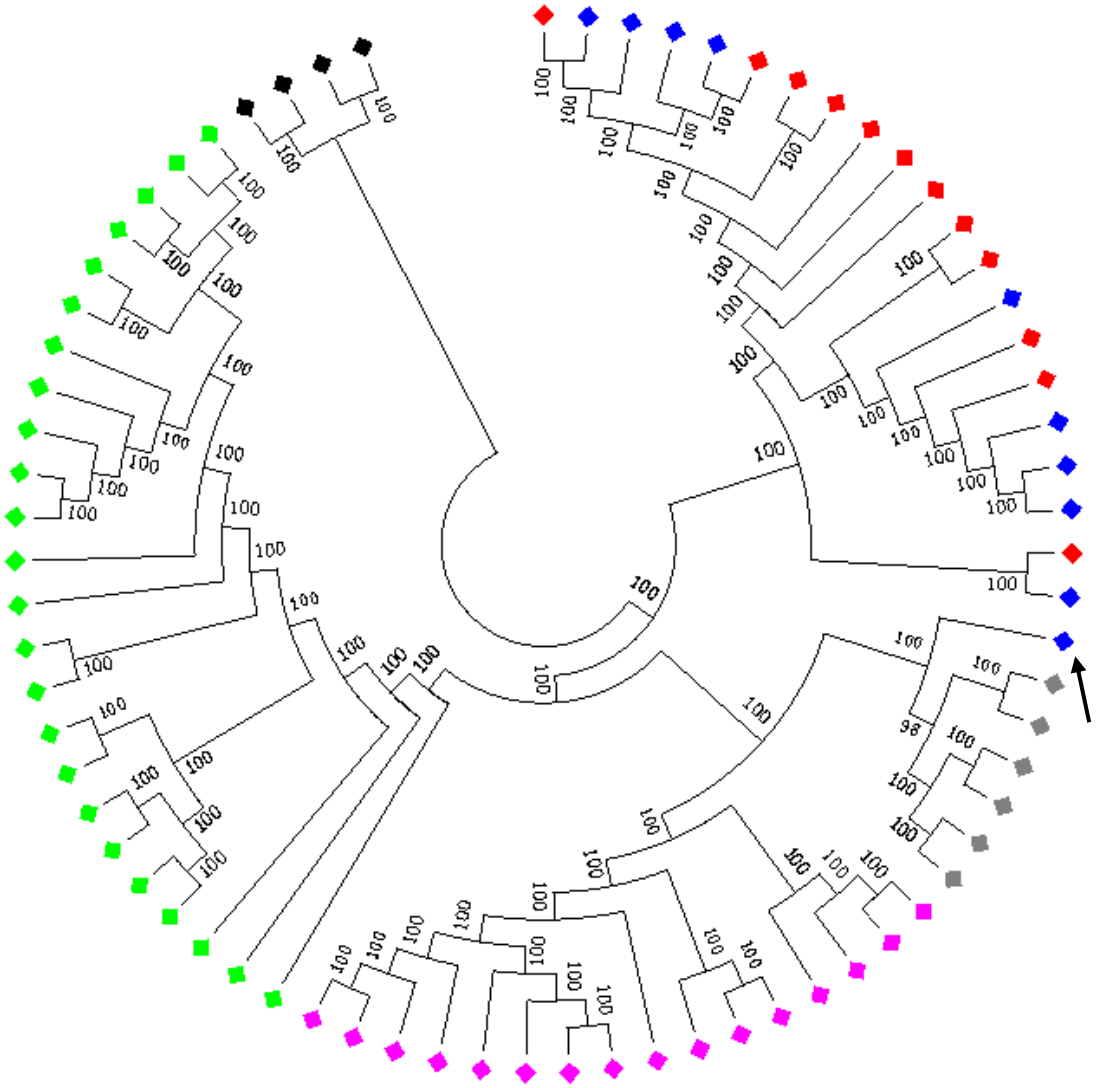


Figure 4.8: Haplotype maximum likelihood tree for a Ceylon junglefowl candidate introgressed region on chromosome 1 (position 101200001-101220000) of Saudi Arabian domestic chicken. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl. The arrow indicates the introgressed haplotype between Grey junglefowl and domestic chicken.

Introgression between Green junglefowl and domestic chicken

Introgression between domestic chickens from Iraq, Ethiopia and Saudi Arabia with Green junglefowl was tested in the same way (**Figure 4.9**). We analysed 16516 significant introgressed windows for Iraq (A), 12075 for Ethiopia (B) and 9372 windows for Saudi Arabia (C). After the threshold of $fd \geq 0.3$ had been applied, the candidate introgressed regions in Iraqi chicken were located on chromosomes 1, 2, 3, 4, 5 and 11; in Ethiopian on chromosomes 1, 2, 11 and 12; and in Saudi Arabian on chromosomes 1 and 3 only.

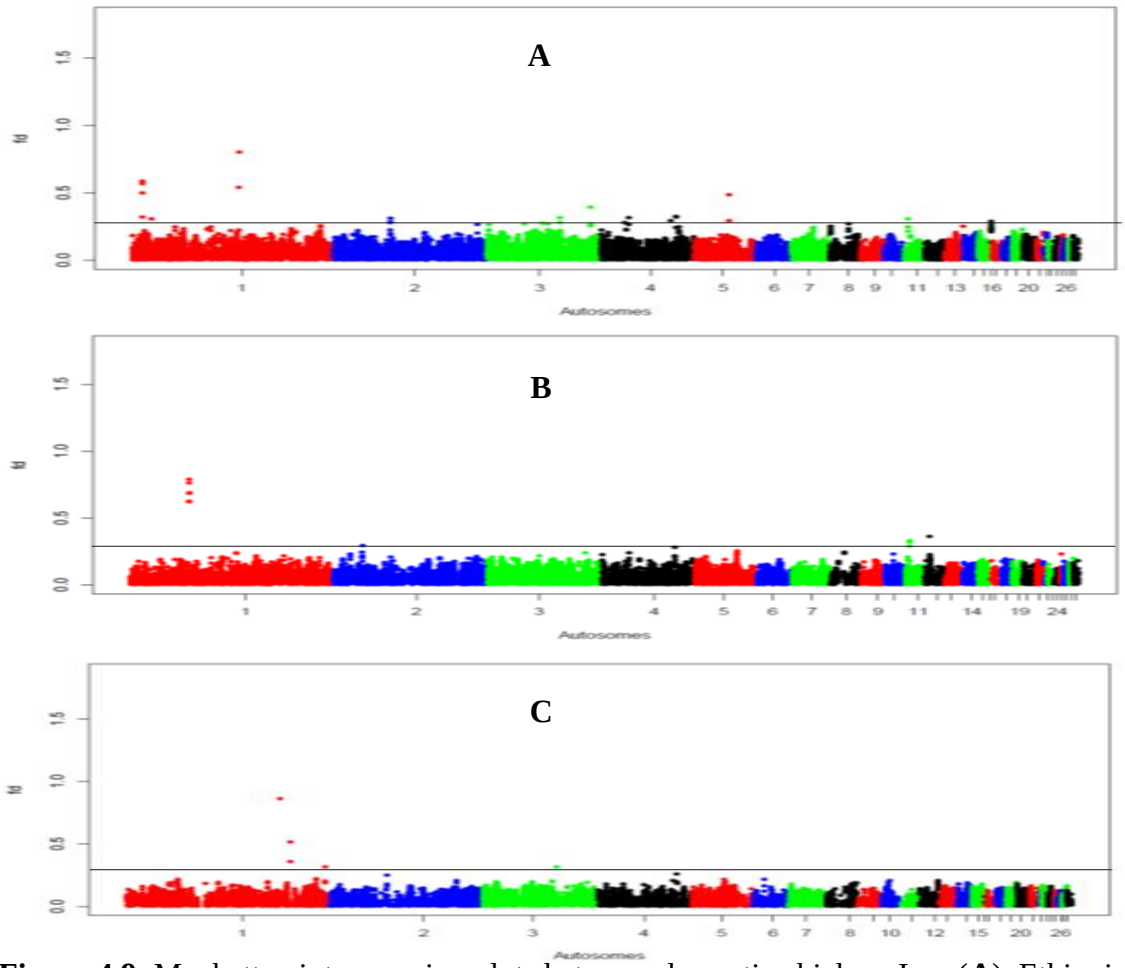


Figure 4.9: Manhattan introgression plots between domestic chicken, Iraq (**A**), Ethiopia (**B**), Saudi Arabia (**C**) and Green junglefowl. The horizontal line represents the chosen threshold ($fd \geq 0.3$).

Following haplotype tree analysis as for the test for Ceylon junglefowl, there was no evidence of haplotype introgression from Green junglefowl into domestic chicken, but there was noticeable gene flow between Grey and Red junglefowl/domestic chicken (**Figures 4.10, 4.11, 4.12 and supplementary S34 to S57**).

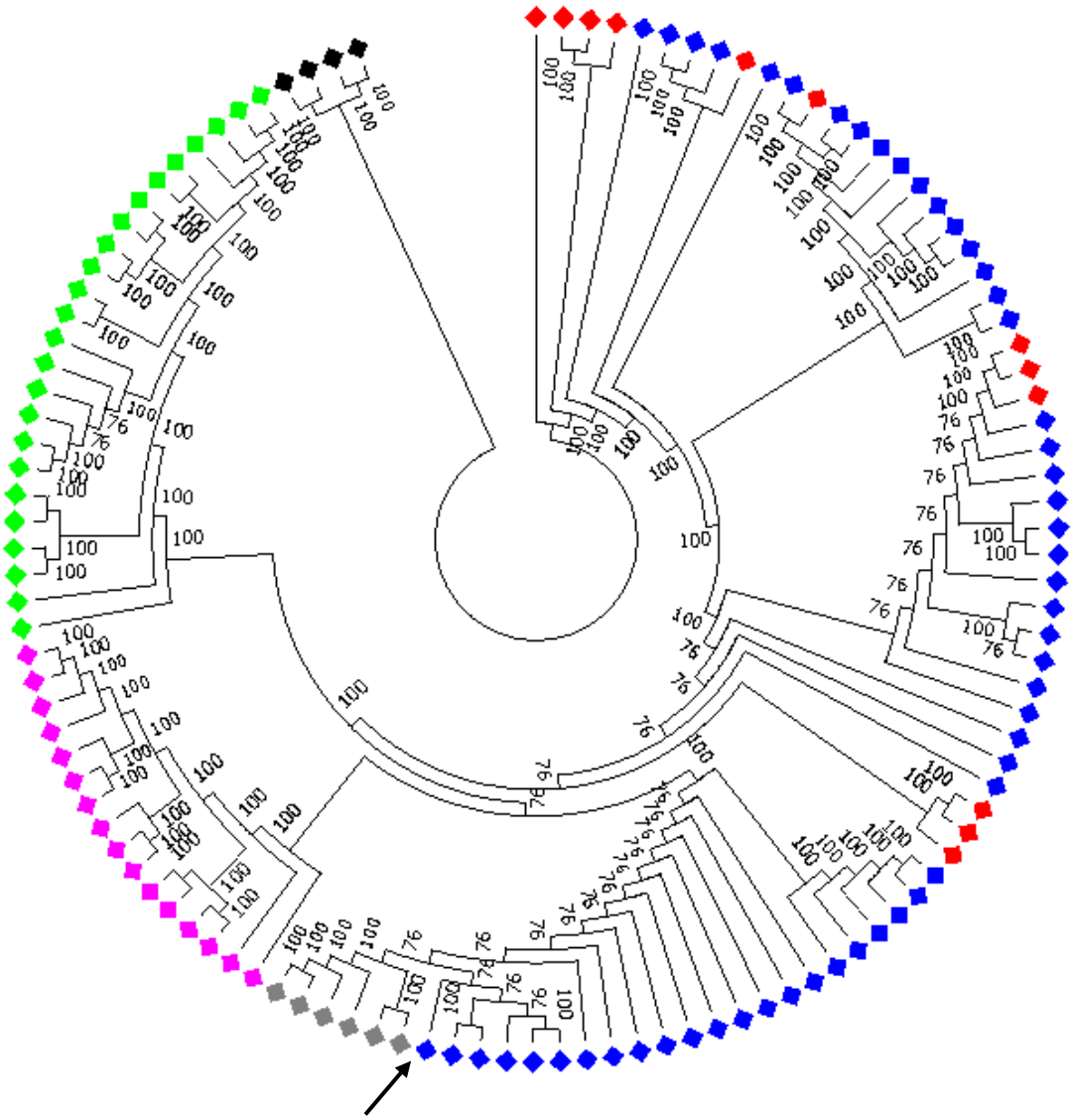


Figure 4.10: Haplotype maximum likelihood tree for a Green junglefowl candidate introgressed region on chromosome 1 (position 9460001-9480000) of Iraqi domestic chicken. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl. The arrow indicates introgressed haplotypes between Grey junglefowl and domestic chicken.

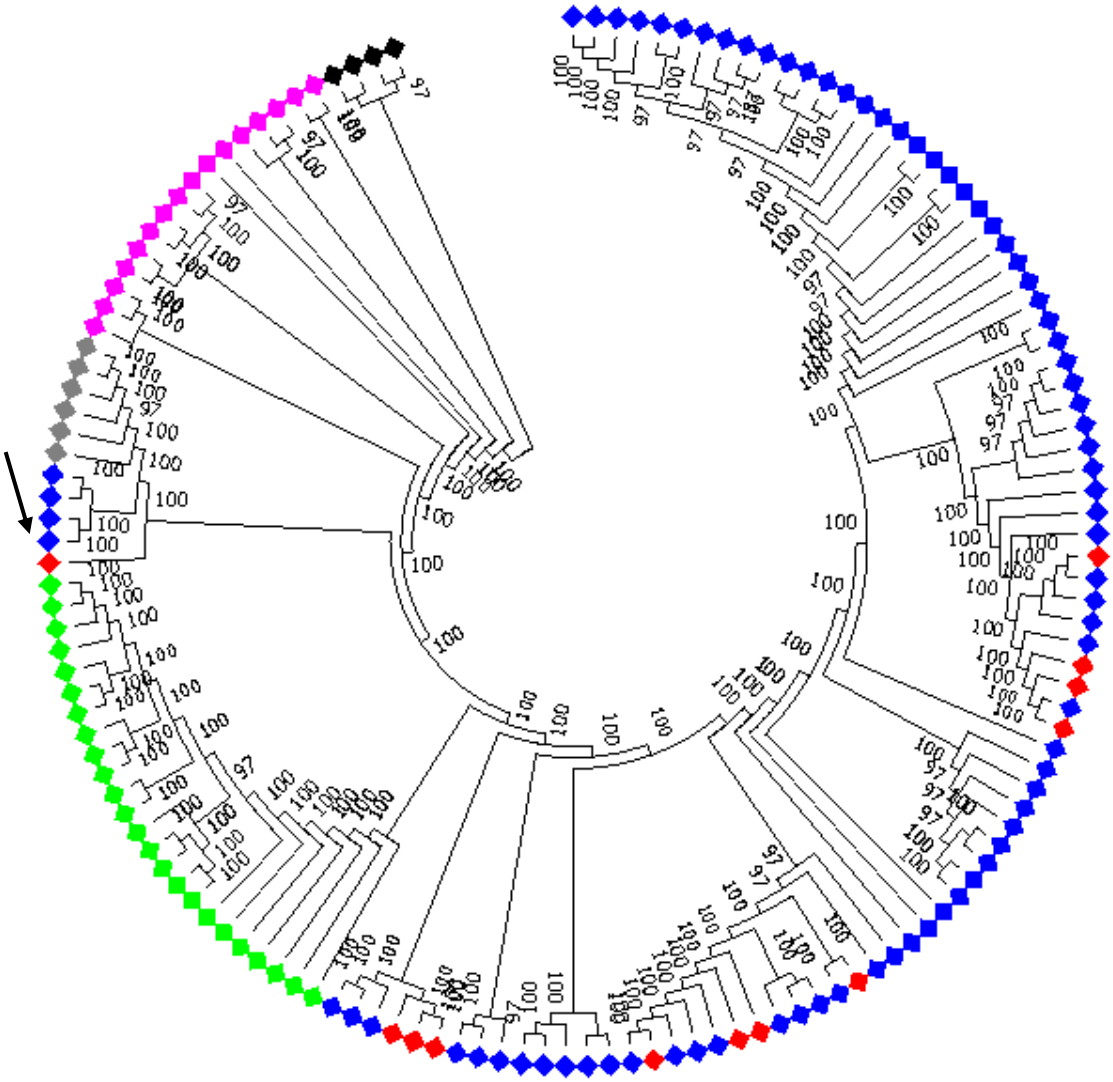


Figure 4.11: Haplotype maximum likelihood tree for a Green junglefowl candidate introgressed region on chromosome 1 (position 56640001-56660000) of Ethiopian domestic chicken. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. The Arrow indicates introgressed haplotypes between Grey junglefowl and domestic chicken.

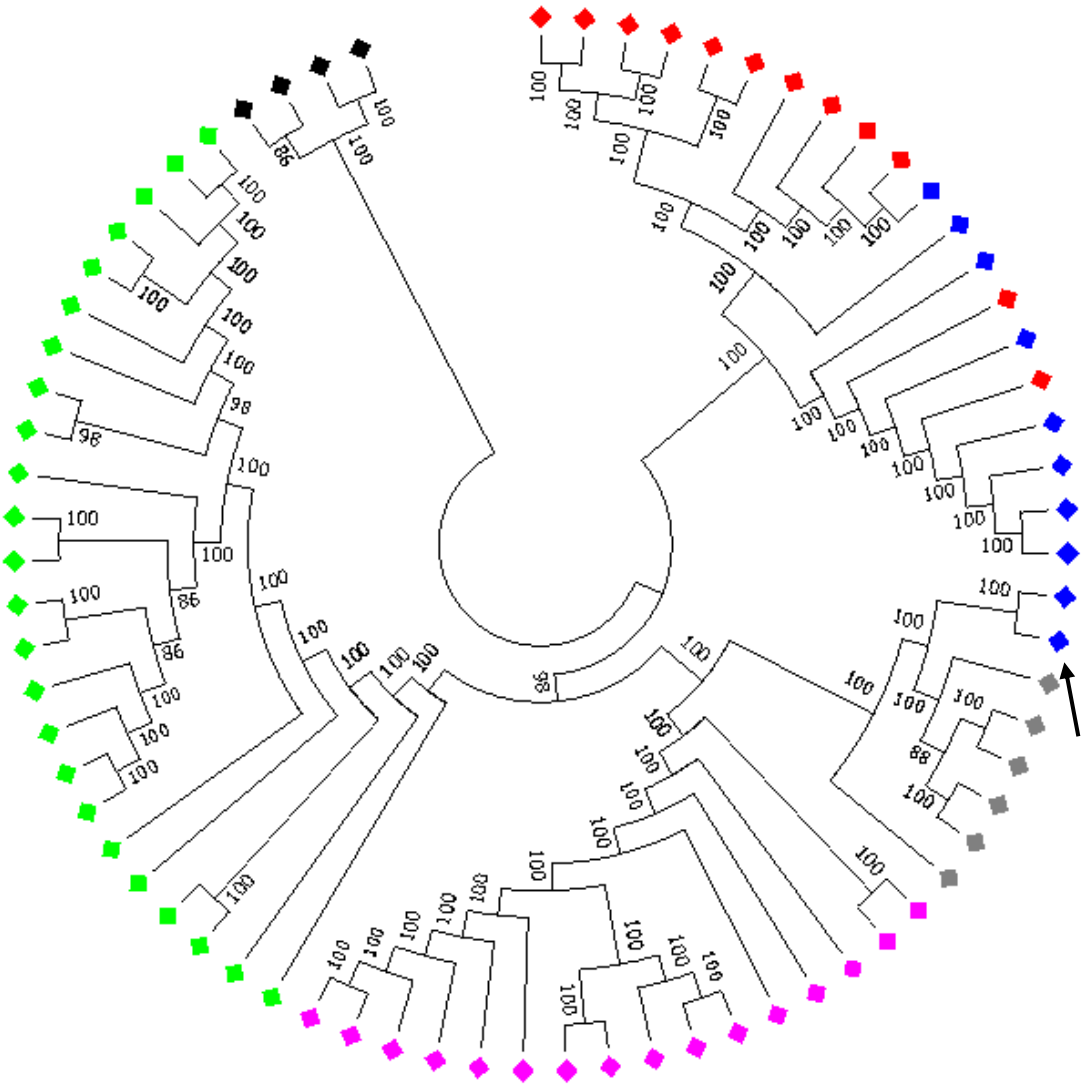


Figure 4.12: Haplotype maximum likelihood tree for a Green junglefowl candidate introgressed region on chromosome 1 (position 158020001-158040000) of Saudi domestic chicken. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl. The arrow indicates introgressed haplotypes between Grey junglefowl and domestic chicken.

Introgression between Grey junglefowl and domestic chicken

In the third part of the introgression analysis, we assessed the gene flow between Grey junglefowl and village chicken (**Figure 4.13**). Applying the same criteria of a minimum of 100 SNPs per window, 19363 windows for Iraq, 14629 for Ethiopia and 11949 for Saudi Arabia were obtained. A significant threshold of *fd* score ≥ 0.5 was applied here. Candidate introgressed regions were present here in most autosomes.

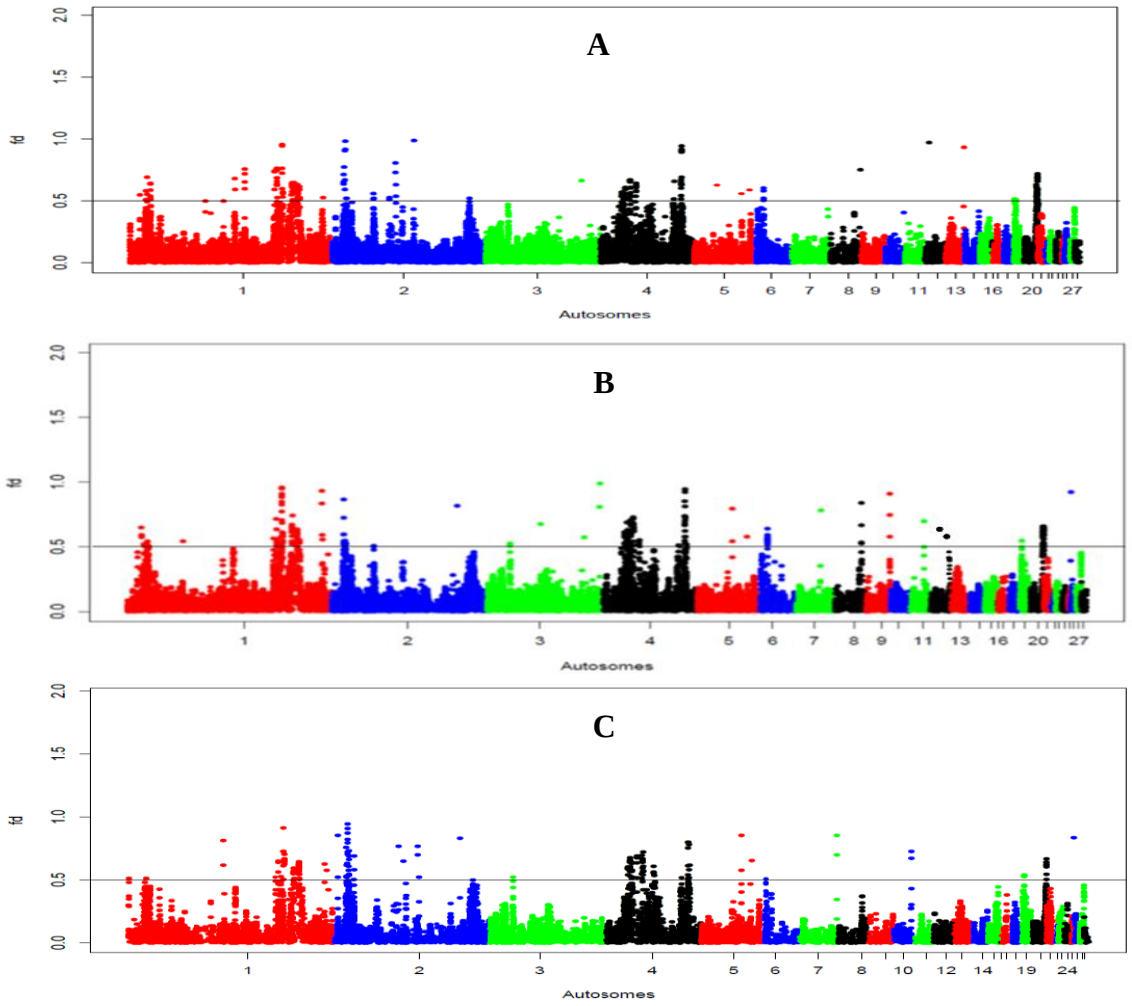


Figure 4.13: Introgression plots between domestic chicken, Iraq (**A**), Ethiopia (**B**), Saudi Arabia (**C**) and Grey junglefowl. The horizontal line represents the chosen threshold of *fd* ≥ 0.5

The results of haplotype trees confirmed the presence of gene flow between Grey junglefowl and domestic chicken from the three countries in six common regions (**Table 4.3**, **Figures 4.14, 4.15, 4.16 and Supplementary S58 to S90**). These regions were on chromosomes 2, 3, 4, 5, 6 and 12. The chromosomal coordinates of these introgressed regions on *Galgal* 5.0 and *Galgal* 6.0, as well as the genes overlapping with these regions, are presented in **Table 4.3**.

Table 4.3: Detected introgressed regions between Grey junglefowl and domestic chicken.

Chr	*Galgal5*_coordinates ($x10^{-6}$)	Length	*Galgal6*_coordinates ($x10^{6}$)	Genes overlapping with the introgressed regions
2	120.12 – 120.40	280 kb	119.67 -119.90	lincRNA
3	50.82 – 50.92	100 kb	50.75-50.85	NOX3
4	62.50 – 62.70	200 kb	62.09-62.29	RF00003
5	45.56 – 45.84	280 kb	45.67-45.95	PPP4R4, SERPINA10, SPIA4, SPIA1
6	20.76 – 20.88	120 kb	21.72-21.84	IDE, MARCH5, CPEB3
12	12.54 – 12.64	100 kb	12.91-13.01	No genes detected

The region on chromosome 2 has a non-protein coding lincRNA and an uncharacterized protein-coding gene related to spermatogenesis. The NOX3 gene located in the chromosome 3 candidate introgressed region is part of the NADPH oxidant family; its function is related to otoconia formation in the inner ear and is believed to increase the cellular oxygen reaction during heat stress (Habashy *et al.*, 2018). RF00003 is a small RNA molecule found in the cell nucleus involved in the processing of pre-messenger RNAs. The genes on chromosome 5 belong to the serpin family, which is involved in the synthesis of ovalbumin in chicken (Remold-O'Donnell, 1993). The IDE gene encodes metallopeptidase that degrades intracellular insulin, terminating insulin activity. MARCH5 is widely known in mammals for controlling cell surface immune-receptors and may be involved in immune regulation in trout (Rebl *et al.*, 2011). CPEB3 appears to play a significant role in chicken embryo skeletal muscle development (Jebessa *et al.*, 2018).

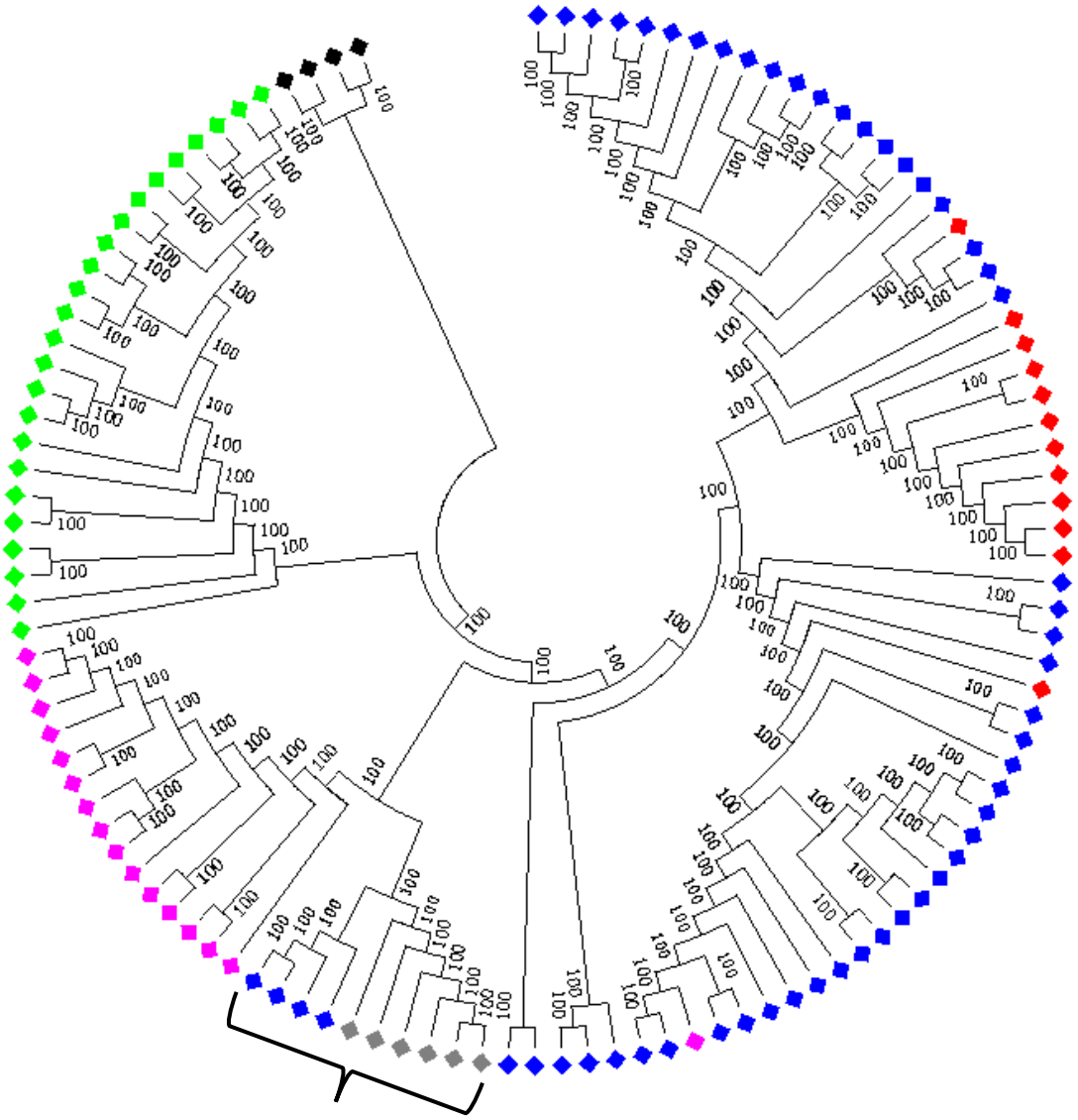


Figure 4.14: Haplotype maximum likelihood tree for the introgressed region on chromosome 5 (position 45560000-45840000) from Grey junglefowl into Iraqi domestic chicken. Black = pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl. Four introgressed haplotypes are detected in Iraqi chicken. Interestingly, we observe also introgression of a candidate domestic chicken introgressed haplotype in Ceylon junglefowl.

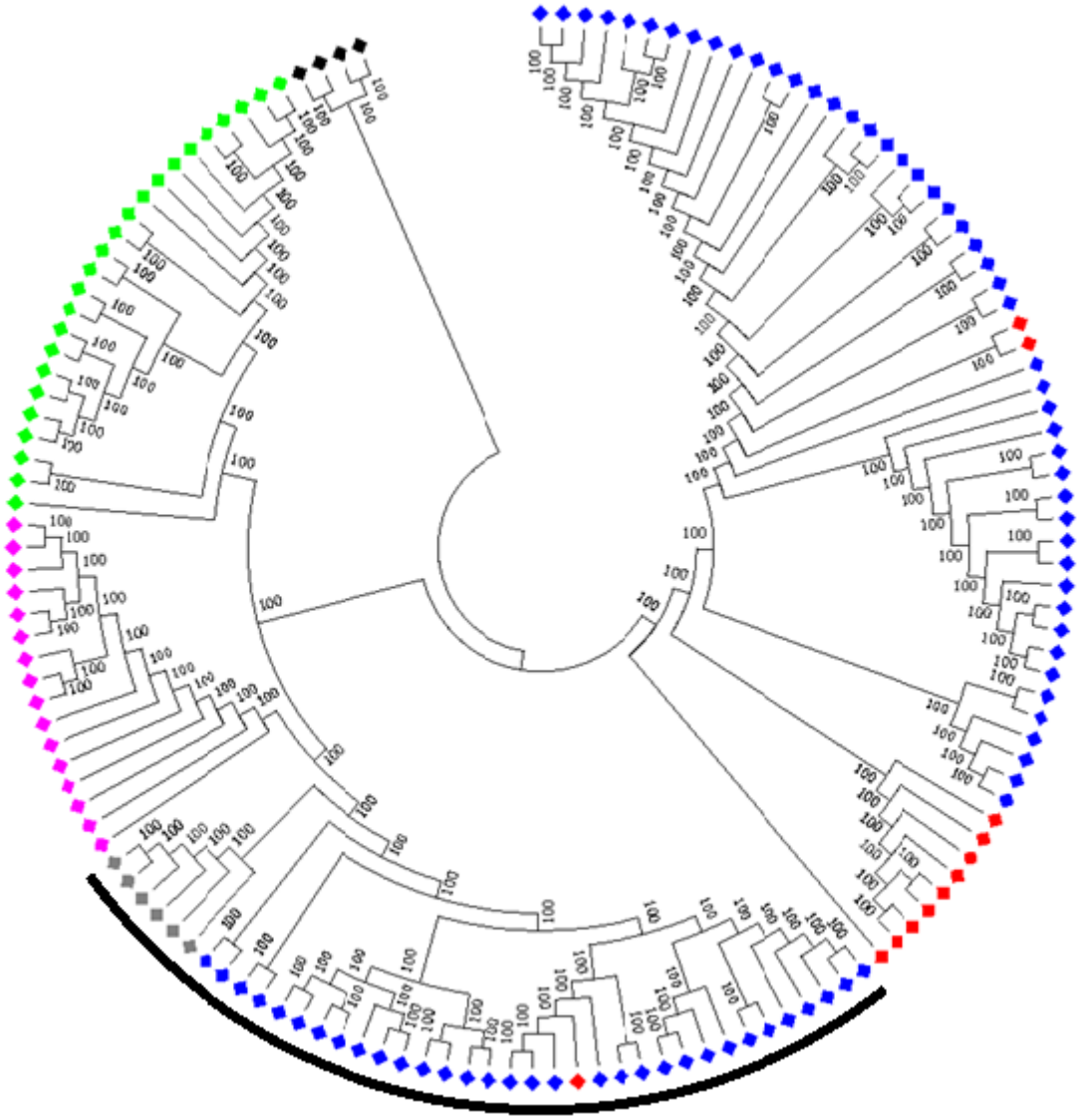


Figure 4.15: Haplotype maximum likelihood tree for a candidate introgressed region on chromosome 2 (position 120120000-120400000) from Grey junglefowl into Ethiopian domestic chicken from the central-east region. Black = pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

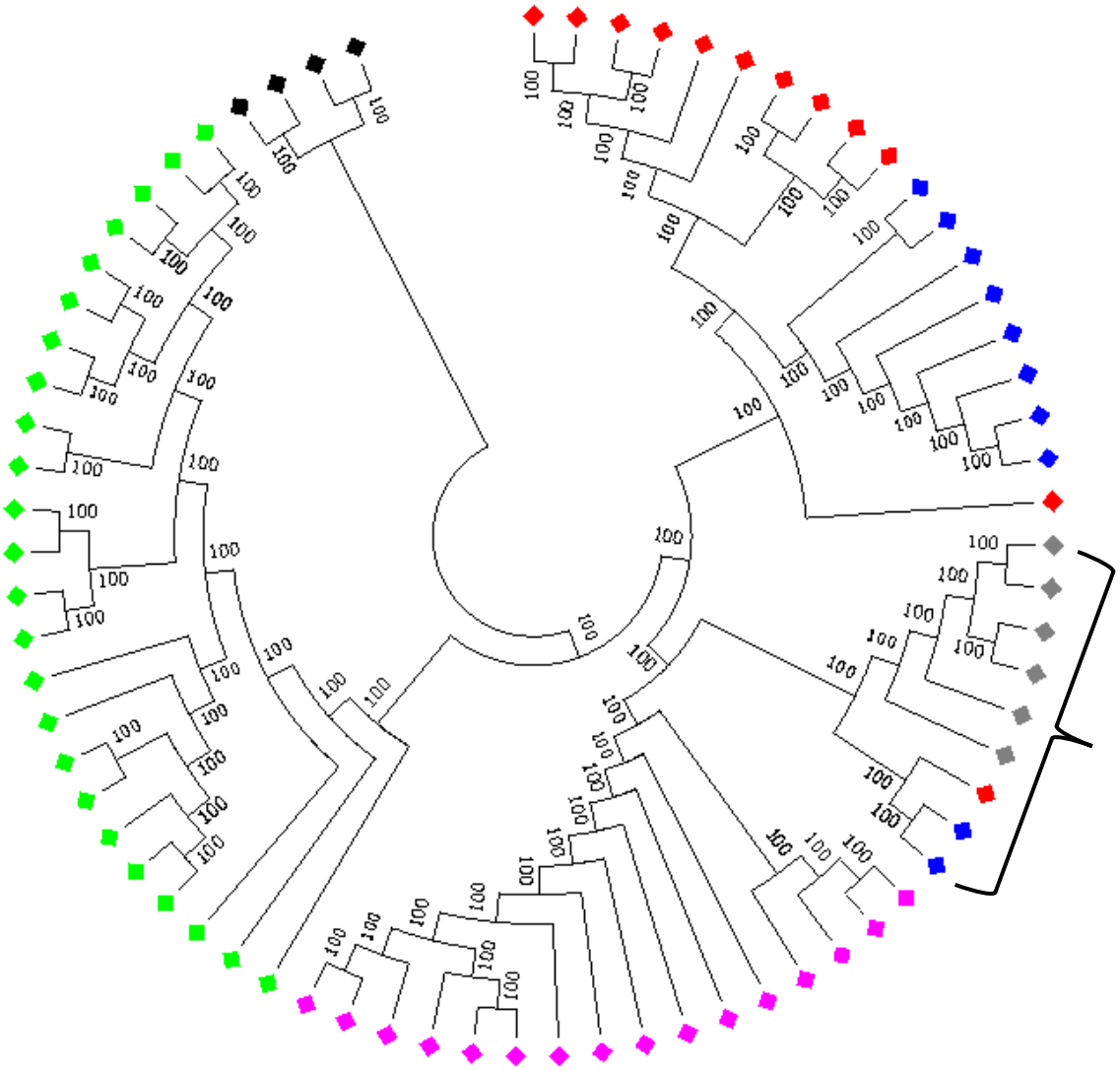


Figure 4.16: Haplotype maximum likelihood tree for the candidate introgressed region on chromosome 2 (position 120120000-120400000) from Grey junglefowl into Saudi Arabia domestic chicken. Black = pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl. Besides two domestic chicken introgressed haplotypes, we observe here one Red junglefowl haplotype of likely Grey junglefowl origin.

The frequencies of introgressed haplotypes from Grey junglefowl into Iraqi domestic chicken are shown in **Table 4.4**. The results showed that Iraqi samples had two introgressed genomic regions out of the total of six examined here. These two regions were on chromosomes 4 and 5, 200 kb for the first and 280 kb for the second. Their frequencies were 7.4% and 5.5%, respectively. The introgressed regions were observed in Baghdad and Sulimania populations, but not in Karbala samples. Chromosome 12 also showed a signal of introgression from domestic chicken into Grey junglefowl.

Table 4.4: Frequency of introgressed haplotypes from Grey junglefowl to Iraqi domestic chicken

Chr	Start	stop	Baghdad n = 26	Karbala n = 10	Sulimania n = 18	Freq %
2	120120000	120400000	0	0	0	0
3	50820000	50920000	0	0	0	0
4	62500000	62700000	2	0	1	5.5
5	45560000	45840000	2	0	2	7.4
6	20760000	20880000	0	0	0	0
12	12540000	12640000	0	0	0	0
Number of candidate introgressed regions			2	0	2	
Total number of introgressed haplotypes for each population			4	0	3	
% of introgressed haplotypes			3	0	3	

Freq = frequency.

The frequency of the introgressed genomic regions into populations of northern Ethiopia can be seen in **Table 4.5**. These northern Ethiopia populations had five introgressed genomic regions (chromosomes 2, 4, 5, 6 and 12). The highest frequency of introgressed haplotypes occurred in a region on chromosome 2 (31.25%) and a region on chromosome 6 (30%); the lowest frequencies were on chromosome 12 (7.5%) and chromosome 4 (6.25%). There was variation in the number of introgressed haplotypes per population, subsequently reflected in the percentage of introgression. The same numbers of introgressed haplotypes were noted in Metkilimat and Meseret, which showed the highest percentage of introgression (23%). The other two populations (HadushAdi and Mihquan) had 14 and 12 introgressed haplotypes respectively. In agreement with previous results,

chromosome 12 also showed a signature of introgression from domestic chicken into Grey junglefowl.

Table 4.5: Frequency of introgressed haplotypes from Grey junglefowl into northern Ethiopian domestic chicken.

Chr	start	stop	HadushAdi n = 20	Metkilimat n = 20	Meseret n = 20	Mihquan n = 20	Freq%
2	120120000	120400000	3	10	9	3	31.25
3	50820000	50920000	0	0	0	0	0
4	62500000	62700000	0	1	2	2	6.25
5	45560000	45840000	5	7	5	3	25
6	20760000	20880000	2	9	11	2	30
12	12540000	12640000	4	0	0	2	7.5
Numbers of candidate introgressed regions			4	4	4	5	
Total number of introgressed haplotypes for each population			14	27	27	12	
% of introgressed haplotypes			12	23	23	10	

Freq = frequency.

The results for central-eastern Ethiopian samples are presented in **Table 4.6**. Here, four introgressed genomic regions out of the six examined were present. These regions were on chromosomes 2, 3, 4 and 12. The most frequent one is on chromosome 2 with a frequency of 40%. The other three genomic regions were observed at low frequencies, 2.5%, 7.5% and 1.25% for the regions on chromosomes 2, 3 and 12 respectively. At a population level, the highest frequencies of introgressed haplotypes were observed at Midir (13%), Negasi (9%) and Arabo (8%), whereas the lowest was observed in Adane (4%). Central-eastern Ethiopia also showed introgression from domestic chicken into Grey junglefowl at three chromosomes (5, 6 and 12).

Table 4.6: Frequency of introgressed haplotypes from Grey junglefowl into central-eastern Ethiopia domestic chicken.

Chr	start	stop	Adane n = 20	Arabo n = 20	Midir n = 20	Negasi n = 20	Freq %
2	120120000	120400000	5	8	11	8	40
3	50820000	50920000	0	0	1	1	2.5
4	62500000	62700000	0	0	4	2	7.5
5	45560000	45840000	0	0	0	0	0
6	20760000	20880000	0	0	0	0	0
12	12540000	12640000	0	1	0	0	1.25
Number of candidate introgressed regions			1	2	3	3	
Total number of introgressed haplotypes for each population			5	9	16	11	
% of introgressed haplotypes			4	8	13	9	

Freq = frequency.

West Ethiopian populations had five introgressed genomic regions out of the six examined in this study (**Table 4.7**). These genomic regions were on chromosomes 2, 3, 4, 6 and 12. A high frequency of introgressed haplotypes was observed in chromosome 6 (40%), lower in chromosome 2 (27.5%) and 3 (12.5%), and the lowest frequency in chromosome 12 (6.25%). Gene flow from domestic into Grey was observed on chromosomes 5 and 12. Variation in the population frequencies of haplotype introgression was present, with the highest frequencies in Gafera (20%), Tzion (18%), Dikuli (14%) and Ashuda (12%).

Table 4.7: Frequency of introgressed haplotypes from Grey junglefowl into western Ethiopia domestic chicken.

Chr	start	stop	Ashuda n = 20	Dikuli n = 20	Gafera n = 20	Tzion n = 20	Freq %
2	120120000	120400000	3	7	5	7	27.5
3	50820000	50920000	3	2	3	2	12.5
4	62500000	62700000	1	0	5	1	8.75
5	45560000	45840000	0	0	0	0	0
6	20760000	20880000	7	7	10	8	40
12	12540000	12640000	0	1	1	3	6.25
Number of candidate introgressed regions			4	4	5	5	
Total number of introgressed haplotypes for each population			14	17	24	21	
% of introgressed haplotypes			12	14	20	18	

Freq = frequency.

The last Ethiopian geographical region examined (south) showed only two introgressed genomic regions (**Table 4.8**) on chromosomes 4 and 12 at frequencies of 8.75% and 1.25% respectively. The introgressed haplotypes were found in Girissa, Kumato, Loya and Shubi in the first region, and Girissa in the second. This region of Ethiopia showed a lower frequency of introgression compared to the other parts. As previously, gene flow from domestic chicken into Grey junglefowl was observed on chromosomes 5, 6 and 12.

Table 4.8: Frequency of introgressed haplotypes from Grey junglefowl into southern Ethiopian domestic chicken.

Chr	start	stop	Girissa n = 20	Kumato n = 20	Loya n = 20	Shubi n = 20	Freq %
2	120120000	120400000	0	0	0	0	0
3	50820000	50920000	0	0	0	0	0
4	62500000	62700000	1	1	2	3	8.75
5	45560000	45840000	0	0	0	0	0
6	20760000	20880000	0	0	0	0	0
12	12540000	12640000	1	0	0	0	1.25
Number of candidate introgressed regions			2	1	1	1	
Total number of introgressed haplotypes for each population			2	1	2	3	
% of introgressed haplotypes			2	1	2	3	

Freq = frequency.

Only one candidate introgressed region (out of six) was found in Saudi Arabian samples (**Table 4.9**). This genomic region was 280 kb long, and was observed at 20% frequency, where 2 haplotypes out of 10 are introgressed. As before, we noticed gene flow from domestic chicken into Grey junglefowl on chromosome 12.

Table 4.9: Frequency of introgressed haplotypes from Grey junglefowl into Saudi Arabian domestic chicken.

Chr	start	stop	Saudi n = 10	Freq%
2	120120000	120400000	2	20
3	50820000	50920000	0	0
4	62500000	62700000	0	0
5	45560000	45840000	0	0
6	20760000	20880000	0	0
12	12540000	12640000	0	0
Number of candidate introgressed regions			1	
Total number of introgressed haplotypes			2	
% of introgressed haplotypes			3	

Freq = frequency.

To address the main objective of this study, i.e. to trace the migration or dispersal routes of local chicken, the geographic distribution of the percentage of introgressed haplotypes was presented in each country (**Figure 4.17**). The Iraqi haplotype distribution (**A**) showed evidence of introgression in north-east (Sulimania) and central (Baghdad) regions only. Both introgressed populations had the same low percentage (3%) of introgression, while it is absent in Karbala. In contrast, Ethiopian populations (**B**) showed a higher percentage of introgression, ranging in the northern part between 10-23%, in the central-eastern between 4-13%, in the western part between 12-20% and in the south the lowest percentage between 1-3%. Saudi samples (**C**) also showed a low frequency of introgression (3%) similar to that for Iraq and southern Ethiopia.

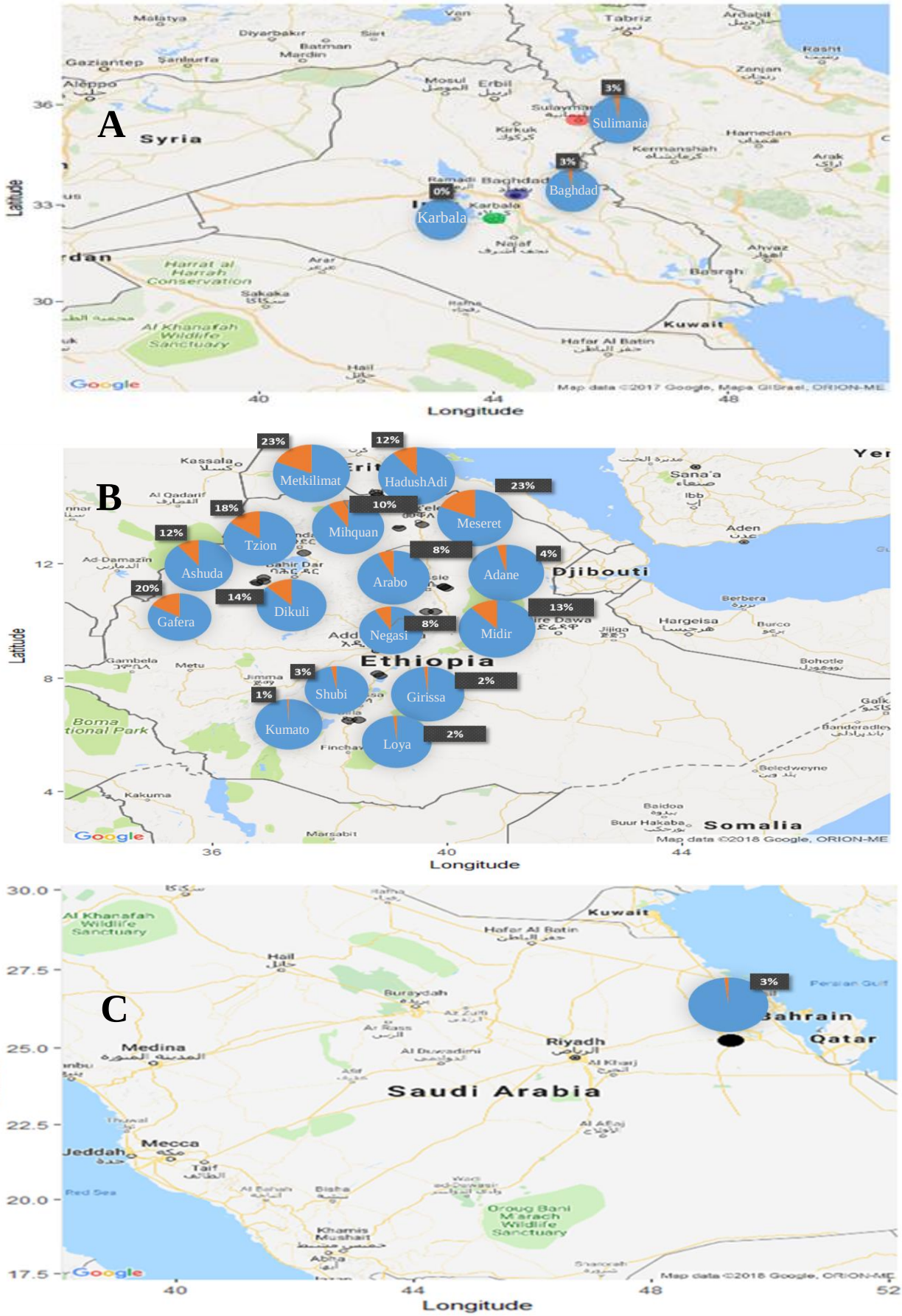


Figure 4.17: Geographical distribution of introgressed haplotypes for (**A**) Iraq, (**B**) Ethiopia and (**C**) Saudi Arabia.

Discussion

The main three objectives of this Chapter were (i) to assess the level of introgression in domestic populations from Iraq, Ethiopia and Saudi Arabia from other *Gallus* species, (ii) to examine the phylogeographic structure of these introgression, (iii) accordingly, if such may provide information on the dispersal routes of domestic chicken from their centres of origin in south and south-east Asia.

Populations in this chapter showed an adequate number of variants, some (such as Karbala from Iraq and Local Horro from Ethiopia) with lower numbers of SNPs than the other populations. This situation could be attributed to the low number of samples for both of them compared with other populations, since Karbala had only 5 samples and Local Horro only 6. The same effect can be applied to the general condition of the three countries. Ethiopia had the highest number of samples and the highest number of SNPs; Iraq had a lower number of samples and SNPs; and finally the lowest in terms of both samples and SNPs was Saudi Arabia. Within countries, in Iraq Sulimania (north-east) had the highest number of SNPs then Baghdad and Karbala from the central region respectively. In Ethiopia, there was more variation among populations, probably due to the high number of populations (27) compared to other countries. The highest number of SNPs were spotted in Kumato (south), Mihquan (north) and Girissa (south) and the lowest was in Local Horro (central-east).

The observed and expected heterozygosity values for the populations included in this study were higher than those found by Zhang *et al*. (2018), where *Ho* ranged between 0.209 to 237 and 0.207 to 0.228 for *He*. Zhang *et al.* (2018) study included three Chinese breeds with nine subpopulations as part of a conservation program that could have affected their genetic diversity level (e.g. inbreeding). Another reason may be attributed to the reference version and mapping protocol: our study used *Galgal* 5.0 through the GATK best-practices process, whereas samples from the other study were mapped against *Galgal* 4.0 using SAMtools for the discovery of variants. Heterozygosity values for Pakistani Aseel and British chicken were higher than the values that have been found by this thesis, these two studies used microsatellite markers to estimate the level of diversities which could explain the reason behind the differences of expected and observed values (Babar *et al.*, 2012; Wilkinson *et al.*, 2012).

Both PCA and autosomal phylogenetic tree support the species relationship as reported in the literature, with the pheasant as a distantly related outgroup, Green junglefowl as the basal lineage of genus *Gallus*, sister lineages for Ceylon and Grey junglefowl and the closest connection between domestic and Red junglefowl (Sawai *et al.*, 2010; Lawal, 2018). The mtDNA maximum likelihood tree showed introgression between Grey Red/domestic chicken, more specifically from Red/domestic chicken into Grey junglefowl. Such introgression in Grey junglefowl has been reported before by Nishibori *et al.* (2005). Petit and Excoffier (2009) mentioned in their study that mtDNA introgression can happen more easily than autosomes due to smaller effective size and maternal inheritance.

Introgression analysis reveals a large number of candidate regions for the Grey junglefowl - Red/domestic chicken and much less so for the Red/domestic chicken - Ceylon/Green junglefowl introgression. Haplotype tree analysis for the latest two comparisons do not support any of these candidate regions. In case of the Red/domestic chicken - Ceylon/Green junglefowl all these regions correspond to Grey junglefowl - Red/domestic introgression events. The sister relationship between the Grey and Ceylon junglefowl likely explain these false positive signals. In the case of the few candidate introgressed regions between Green junglefowl and Red/domestic chicken, they are more difficult to explain, they may correspond to haplotypes under the same selection pressures in both species, and this however required further investigation.

It should be noted that our analysis could fully exclude the possibility that some of these introgression events may reflect the presence of Incomplete Lineage Sorting (ILS), with many studies mentioned that introgression and ILS display similar signal and can be hardly differentiated (McGuire *et al.*, 2007; Choleva *et al.*, 2014; Martin *et al.*, 2014). Also, Martin *et al.* (2014) reported in his paper that a significant excess of ABBA over BABA can be an indication of introgression or non-random mating or sub-structuring in the ancestral population.

In contrast, evidence for gene flow from Grey junglefowl to village chicken was commonly observed in both directions. We examined in further detail six of the regions building-up on the results reported by Lawal *et al*. (2019). These introgression events are present at a different frequency levels within and across countries.

Several genes were located in these introgressed regions, some related to environmental adaptation, spermatogenesis and embryonic development, which may indicate the role that Grey junglefowl played in village chicken acclimation in different weather conditions.

Unexpectedly, the results exposed higher Grey junglefowl frequency of introgression into the gene pool of Ethiopian local chicken than into Iraq or Saudi Arabia. The prediction was that, because Iraq and Saudi Arabia are the closest to the geographic distribution of the Grey junglefowl, then, they would have more frequent gene flow than Ethiopia. This maybe a reflection of different sample size for the countries and/or rather translated a different history of arrival of domestic chicken in these regions. Indeed, it may be reasonably assumed that village chicken from Iraq and Saudi Arabia would have reached these two countries following a terrestrial route and/or short maritime crossing from the northern part of Indian subcontinent, a geographic area where the Grey junglefowl are not present. In case of Ethiopia, it is possible that this country has witnessed a direct arrival of chicken from the southern part of Indian subcontinent where gene flow have happened at a high frequency between domestic chicken and Grey junglefowl, considering the geographic distribution of the latter (**Figure S91**).

To summarize, going back to our three main objectives of this chapter, we may conclude that Grey junglefowl introgression is presented in all three countries though with different frequency. A phylogeographic signal of introgression might be present across these populations and such introgression may indeed provide new markers to assess the migration of domestic chicken from their centres of origin. Analysis of more populations from different geographic areas are however required for further supporting these observations and conclusions.

Chapter 5

General discussion and future prospect

This work provides detailed genomic evidence for tracing the dispersal routes of local chicken from the Horn of Africa, the Arabian Peninsula, the Middle East and North Africa outside the putative centre(s) of domestication of the species. In parallel to archaeological evidence, molecular genetics analysis of modern chicken populations is providing important insight into the history of this domesticate. This is particularly important in a context where chicken bones are rarely unambiguously identified in archaeological remains (Perry-Gal *et al.*, 2015; Woldekiros & D'Andrea, 2017).

The evaluation of a different class of genetic markers supports the use of the full mitochondrial DNA sequence information rather than the *D*-loop for phylogeographic analysis of maternally inherited diversity. At the contrary, the study of a W DNA fragment showed some distinct haplotypes which may need further investigation. Last but not least, the geographic distribution and frequency of candidate autosomal introgressed regions seem promising avenues to investigate further the history of the domestic chicken. By extension, chickens can be exploited as a biological marker to trace ancient human migration routes and trading networks (Mwacharo *et al.*, 2013; Di Lorenzo *et al.*, 2015).

This thesis also offers new and greater detail about the geographical distribution of genetic diversity in local breeds of chicken across different countries and their historical demographic patterns such as expansion and structure. Genetic diversity within local chicken is extensively higher in comparison with commercial breeds. Therefore, it can be considered as a valuable resource that can be used for genetic improvement and consumer demands (Delany, 2004). Genetic variation has a major role to play in constructing chicken demographic history and in tracing their origins. It may also be involved in designing a suitable breeding program for the improvement of economically important traits (Desta, 2015). In addition, the decision to prioritise a population or species conservation is very reliant on genetic diversity indicators.

Based on largely archaeological evidence, chickens have likely been transported from their domestication centre(s) to southern Europe and North Africa through the Middle East following ancient terrestrial trading routes (Perry-Gal *et al.*, 2015). Another more recent route of migration likely utilised the maritime trading network through the Indian Ocean and the Red Sea (Prendergast *et al.*, 2017). However, very few have traced these migration routes of chicken from domestication centre(s) and none until this thesis to Saudi Arabia, Iraq and the shores of the Mediterranean sea. The evidence from the *D*-

loop (**Chapter 2**) indicates that most of the chicken samples in this study have a heritage attributable to a terrestrial route of migration to the Middle East and Africa, with the possibility of a maritime dispersal route through the Indian ocean. This signal of a maritime route can be seen to be stronger when closer to East Africa (Ethiopia), and absent in North Africa (Libya and Algeria). The presence of these routes may have had an impact on the genetic diversity, clearly noticeable in Ethiopian populations in comparison with those from other countries including Pakistan. Our conclusion on different routes of introduction of domestic chicken into East Africa is supported by previous studies (Mwacharo *et al.*, 2011; Desta, 2015).

More resolution has been obtained from the full mtDNA results compared to the *D*-loop (**Chapter 3**). Genetic diversity indicators showed higher diversity levels for the full mtDNA data compared to the *D*-loop. The levels of resolution and genetic variation can be seen in the genetic diversity of Sulimania population from Iraq and Jarso from Ethiopia. These two populations exhibited no genetic diversity using the *D*-loop genetic marker, but the full mtDNA data showed that these two populations have a high level of genetic variation. Other examples of the difference in resolution between the full mtDNA and the *D*-loop are the presence of haplogroups C1 and C2, with a greater frequency of the latter, and the absence of haplogroup D in Ethiopian full mtDNA, instead of seven haplotypes for the *D*-loop analysis. Haplogroup D for the full mtDNA dataset was only found in Iraq where it was observed once. Furthermore, all the samples of Sulimania population from Iraq belonged to haplogroup E following the *D*-loop data, while the same samples represent haplogroup A in the full mtDNA dataset. However, one feature of both the *D*-loop and full mtDNA analysis is the relatively low level of phylogeographic structuring within a country. The third studied genetic marker (the W chromosome) provided little information following its reduced diversity probably due to strong selection pressure and lack of recombination. Smeds *et al.* (2015) reached the same conclusion, although for a wild bird species, and recommended using W chromosome data for study across species rather than within species.

Analysis of across *Gallus* species introgression supports the Red junglefowl as the main ancestor of the domestic chicken, with a genetic contribution of the Grey junglefowl (**Chapter 4**). Gene functions that have introgressed from Grey junglefowl to domestic chickens are mostly related to adaptation and productivity traits. Their introgression frequency could be used to trace livestock dispersal routes. Evidence from such gene

flow suggests that some chickens reached Ethiopia possibly through a maritime route directly from the Southern part of the Indian subcontinent. To the best of our knowledge, this is the first time that such a route has been postulated. Besides the introgression pattern observed here, the presence of haplogroup C2, known to come from the southern part of the Indian subcontinent in Ethiopian chicken (Miao *et al.*, 2013) is another piece of evidence. Another endorsement comes from the incidence of haplogroup B, possibly originating from Laos, and haplogroup C1 previously described from Henan province in the southern part of China, both now also observed in Ethiopia. Autosomal introgression analysis also offers new ideas about the history of Iraqi and Saudi chickens. The frequency of introgressed haplotypes suggests that the chickens of these two countries originally derived from the northern part of the Indian subcontinent, where introgression is rare between Red and Grey junglefowl.

The accumulated information from both the full mtDNA data and introgression analysis probably represents the best combination to trace the dispersal routes of the domestic chicken. It also comes with a negative side, the cost of full-genome sequencing and the number of samples required for such a study.

Future work will need to add samples from other countries such as Oman, places with an intermediate geographical position between East Africa and the Indian subcontinent, to clarify the maritime route(s) through the Indian Ocean to East Africa. With archaeological evidence indicating that Ethiopian chicken originates from two migration routes, an old terrestrial from the North through the Nile River basin and a more recent maritime route through trading networks (Perry-Gal *et al.*, 2015; Prendergast *et al.*, 2017), including chicken samples from Egypt and Sudan may be particularly informative.

References

Abbott R., Albach D., Ansell S., Arntzen J.W., Baird S.J., Bierne N., Boughman J., Brelsford A., Buerkle C.A. & Buggs R. (2013) Hybridization and speciation. *Journal of evolutionary biology* **26**, 229-46.

Abdollahi-Arpanahi R., Morota G., Valente B., Kranis A., Rosa G. & Gianola D. (2015) Assessment of bagging GBLUP for whole-genome prediction of broiler chicken traits. *Journal of Animal Breeding and Genetics* **132**, 218-28.

Adebambo A., Mobegi V., Mwacharo J., Oladejo B., Adewale R., Ilori L., Makanjuola B., Afolayan O., Bjørnstad G. & Jianlin H. (2010) Lack of Phylogeographic Structure in Nigerian Village Chickens Revealed by Mitochondrial DNA D-loop Sequence Analysis. *International Journal of Poultry Science* **9**, 503-7.

Ahsan M., Li X., Lundberg A.E., Kierczak M., Siegel P.B., Carlborg Ö. & Marklund S. (2013) Identification of candidate genes and mutations in QTL regions for chicken growth using bioinformatic analysis of NGS and SNP-chip data. *Frontiers in genetics* **4**, 226.

Al-Rawi A. & Al-Athari A. (2002) Characteristics of indigenous chicken in Iraq. *Animal Genetic Resources/Resources génétiques animales/Recursos genéticos animales* **32**, 87-93.

Al-Yousef Y.M. (2007) A survey study on the distribution of Saudi Baladi chickens and their characteristics. *International Journal of Poultry Science* **6**.

Alexander M., Ho S.Y., Molak M., Barnett R., Carlborg Ö., Dorshorst B., Honaker C., Besnier F., Wahlberg P. & Dobney K. (2015) Mitogenomic analysis of a 50-generation chicken pedigree reveals a rapid rate of mitochondrial evolution and evidence for paternal mtDNA inheritance. *Biology letters* **11**, 20150561.

Arnold M.L. (2006) *Evolution through genetic exchange*. Oxford University Press.

Avise J.C. (2009) Phylogeography: retrospect and prospect. *Journal of biogeography* **36**, 3-15.

Avise J.C., Arnold J., Ball R.M., Bermingham E., Lamb T., Neigel J.E., Reeb C.A. & Saunders N.C. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual review of ecology and systematics* **18**, 489-522.

Avise J.C., Bowen B.W. & Ayala F.J. (2016) In the light of evolution X: comparative phylogeography. *Proceedings of the National Academy of Sciences* **113**, 7957-61.

Axelsson E., Smith N.G., Sundström H., Berlin S. & Ellegren H. (2004) Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. *Molecular biology and evolution* **21**, 1538-47.

Babar M.E., Nadeem A., Hussain T., Wajid A., Shah S.A., Iqbal A., Sarfraz Z. & Akram M. (2012) Microsatellite marker based genetic diversity among four varieties of Pakistani Aseel Chicken. *Pak. Vet. J* **32**, 237-41.

Bandelt H.-J., Forster P. & Röhl A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular biology and evolution* **16**, 37-48.

Bennecke N. (1994) Der Mensch und seine Haustiere. Die Geschichte einer jahrtausendealten Beziehung. *Stuttgart, Theiss*.

Bensch S., Irwin D.E., Irwin J.H., Kvist L. & Åkesson S. (2006) Conflicting patterns of mitochondrial and nuclear DNA diversity in Phylloscopus warblers. *Molecular Ecology* **15**, 161-71.

Bermingham E. & Moritz C. (1998) Comparative phylogeography: concepts and applications. *Molecular Ecology* **7**, 367-9.

Berthouly-Salazar C., Rognon X., Van T.N., Gély M., Chi C.V., Tixier-Boichard M., Bed'Hom B., Bruneau N., Verrier E. & Maillard J.-C. (2010) Vietnamese chickens: a gate towards Asian genetic diversity. *BMC genetics* **11**, 53.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.-H., Xie D., Suchard M.A., Rambaut A. & Drummond A.J. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology* **10**, e1003537.

Brito P.H. & Edwards S.V. (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* **135**, 439-55.

Browning B.L., Zhou Y. & Browning S.R. (2018) A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics* **103**, 338-48.

Carter H. (1923) An ostracon depicting a red jungle-fowl. *The Journal of Egyptian Archaeology* **9**, 1-4.

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M. & Lee J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7.

Chester N. & Marshak D.R. (1993) Dimethyl sulfoxide-mediated primer Tm reduction: a method for analyzing the role of renaturation temperature in the polymerase chain reaction. *Analytical biochemistry* **209**, 284-90.

Choleva L., Musilova Z., Kohoutova-Sediva A., Paces J., Rab P. & Janko K. (2014) Distinguishing between incomplete lineage sorting and genomic introgressions: complete fixation of allospecific mitochondrial DNA in a sexually reproducing fish (Cobitis; Teleostei), despite clonal reproduction of hybrids. *PloS one* **9**, e80641.

Collias N. & Collias E. (1996) Social organization of a red junglefowl, Gallus gallus, population related to evolution theory. *Animal behaviour* **51**, 1337-54.

Crawford R.D. (1990) *Poultry breeding and genetics*. Elsevier Amsterdam.

Dalhoum L., Moula N., Halbouche M. & Mignon-Grasteau S. (2016) Phenotypic characterization of the indigenous chickens (Gallus gallus) in the northwest of Algeria. *Archives Animal Breeding* **59**, 79-90.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T. & Sherry S.T. (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8.

Darriba D., Taboada G.L., Doallo R. & Posada D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature methods* **9**, 772.

Darwin C. (1868) *The variation of animals and plants under domestication*. John Murray, United Kingdom.

Del Hoyo J., Elliott A., Sargatal J., Christie D. & de Juana E. (2017) Handbook of the birds of the world alive. Lynx Edicions, Barcelona.

Delacour J. (1977) *Pheasants of the world*. Spur Publications. England.

Delany M.E. (2004) Genetic variants for chick biology research: from breeds to mutants. *Mechanisms of development* **121**, 1169-77.

DePristo M.A., Banks E., Poplin R., Garimella K.V., Maguire J.R., Hartl C., Philippakis A.A., Del Angel G., Rivas M.A. & Hanna M. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**, 491.

Desta T.T. (2015) Phenomic and genomic landscape of Ethiopian village chickens. University of Nottingham.

Di Lorenzo P., Ceccobelli S., Panella F., Attard G. & Lasagna E. (2015) The role of mitochondrial DNA to determine the origin of domestic chicken. *World's Poultry Science Journal* **71**, 311-8.

Drummond A.J., Rambaut A., Shapiro B. & Pybus O.G. (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular biology and evolution* **22**, 1185-92.

Dyomin A., Danilova M., Mwacharo J., Masharsky A., Panteleev A., Druzhkova A., Trifonov V. & Galkina S. (2017) Mitochondrial DNA D-loop haplogroup contributions to the genetic diversity of East European domestic chickens from Russia. *Journal of Animal Breeding and Genetics* **134**, 98-108.

Eltanany M.A. & Hemeda S.A. (2016) Deeper insight into maternal genetic assessments and demographic history for Egyptian indigenous chicken populations using mtDNA analysis. *Journal of advanced research* **7**, 615-23.

Eriksson J., Larson G., Gunnarsson U., Bed'hom B., Tixier-Boichard M., Strömstedt L., Wright D., Jungerius A., Vereijken A. & Randi E. (2008) Identification of the Yellow Skin Gene Reveals a Hybrid Origin of the Domestic Chicken. *PLoS Genetics* **4**.

Excoffier L. & Lischer H.E. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources* **10**, 564-7.

Filatov D.A. (2009) Processing and population genetic analysis of multigenic datasets with ProSeq3 software. *Bioinformatics* **25**, 3189-90.

Fitzpatrick B.M. (2004) Rates of evolution of hybrid inviability in birds and mammals. *Evolution* **58**, 1865-70.

Fu Y.-X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915-25.

Fumihito A., Miyake T., Sumi S., Takada M., Ohno S. & Kondo N. (1994) One subspecies of the red junglefowl (Gallus gallus gallus) suffices as the matriarchic ancestor of all domestic breeds. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 12505-9.

2927 Fumihito A., Miyake T., Takada M., Shingu R., Endo T., Gojobori T., Kondo N. & Ohno
2928 S. (1996) Monophyletic origin and unique dispersal patterns of domestic fowls.
2929 *Proceedings of the National Academy of Sciences* **93**, 6792-5.

2930 Galli M. (2017) Beyond frontiers: Ancient Rome and the Eurasian trade networks.
2931 *Journal of Eurasian studies* **8**, 3-9.

2932 Gifford-Gonzalez D. & Hanotte O. (2011) Domesticating animals in Africa: implications
2933 of genetic and archaeological findings. *Journal of World Prehistory* **24**, 1-23.

2934 Graves J.A.M. (2016) Evolution of vertebrate sex chromosomes and dosage
2935 compensation. *Nature Reviews Genetics* **17**, 33.

2936 Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., Li
2937 H., Zhai W. & Fritz M.H.-Y. (2010) A draft sequence of the Neandertal genome.
2938 *science* **328**, 710-22.

2939 Gu X., Feng C., Ma L., Song C., Wang Y., Da Y., Li H., Chen K., Ye S. & Ge C. (2011)
2940 Genome-Wide Association Study of Body Weight in Chicken F2 Resource
2941 Population. *PloS one* **6**.

2942 Gueye E. (2009) The role of networks in information dissemination to family poultry
2943 farmers. *World's Poultry Science Journal* **65**, 115-24.

2944 Guindon S. & Gascuel O. (2003) A simple, fast, and accurate algorithm to estimate large
2945 phylogenies by maximum likelihood. *Systematic biology* **52**, 696-704.

2946 Habashy W.S., Milfort M.C., Rekaya R. & Aggrey S.E. (2018) Expression of genes that
2947 encode cellular oxidant/antioxidant systems are affected by heat stress. *Molecular*
2948 *biology reports* **45**, 389-94.

2949 Harrison C. (1978) A new jungle-fowl from the Pleistocene of Europe. *Journal of*
2950 *Archaeological Science* **5**, 373-6.

2951 Hashiguchi T. (1981) Phylogenetic relationships determined by the blood protein types
2952 of fowls. *Japanese Journal of Zootechnical Science* **52**, 713-29.

2953 Hassaballah K., Zeuh V., Lawal R.A., Hanotte O. & Sembene M. (2015) Diversity and
2954 origin of indigenous village chickens (Gallus gallus) from Chad, Central Africa.
2955 *Advances in Bioscience and Biotechnology* **6**, 592.

2956 Houlihan P.F. & Goodman S.M. (1988) *The birds of ancient Egypt*. American University
2957 in Cairo Press.

2958 Hutt F. (1949) Genetics of the Fowl, xi+ 590 pp. New York, McGraw-Hill.

2959 Jebessa E., Ouyang H., Abdalla B.A., Li Z., Abdullahi A.Y., Liu Q., Nie Q. & Zhang X.
2960 (2018) Characterization of miRNA and their target gene during chicken embryo
2961 skeletal muscle development. *Oncotarget* **9**, 17309.

2962 Johnsgard P.A. (2002) The Pheasants of the World: Biology and Natural History. *History*
2963 *and Philosophy of the Life Sciences* **23**, 547-.

2964 Kanginakudru S., Metta M., Jakati R. & Nagaraju J. (2008) Genetic evidence from Indian
2965 red jungle fowl corroborates multiple domestication of modern day chicken. *BMC*
2966 *evolutionary biology* **8**, 174.

2967 Kelly L. (2006) The little chicken book. *Caledon, Ontario: Blue Barn Publishing*.

Komiyama T., Ikeo K. & Gojobori T. (2003) Where is the origin of the Japanese gamecocks? *Gene* **317**, 195-202.

Komiyama T., Ikeo K., Tateno Y. & Gojobori T. (2004) Japanese domesticated chickens have been derived from Shamo traditional fighting cocks. *Molecular phylogenetics and evolution* **33**, 16-21.

Lan D., Hu Y., Zhu Q. & Liu Y. (2017) Mitochondrial DNA study in domestic chicken. *Mitochondrial DNA Part A* **28**, 25-9.

Lawal R.A. (2018) Signatures of selection and introgression in the genus Gallus. University of Nottingham.

Lawal R.A., Al-Atiyat R.M., Aljumaah R.S., Silva P., Mwacharo J.M. & Hanotte O. (2018) Whole-Genome Resequencing of Red Junglefowl and Indigenous Village Chicken Reveal New Insights on the Genome Dynamics of the Species. *Frontiers in genetics* **9**.

Leal B.S.S., Palma da Silva C. & Pinheiro F. (2016) Phylogeographic studies depict the role of space and time scales of plant speciation in a highly diverse neotropical region. *Critical reviews in plant sciences* **35**, 215-30.

Leigh J.W. & Bryant D. (2015) popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6**, 1110-6.

Leta S. & Endalew B. (2010) Survey on village based chicken production and utilization system in Mid Rift Valley of Oromia, Ethiopia. *Global Veterinaria* **5**, 198-203.

Li H. & Durbin R. (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589-95.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G. & Durbin R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-9.

Librado P. & Rozas J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-2.

Liu Y.-P., Wu G.-S., Yao Y.-G., Miao Y.-W., Luikart G., Baig M., Beja-Pereira A., Ding Z.-L., Palanichamy M.G. & Zhang Y.-P. (2006) Multiple maternal origins of chickens: out of the Asian jungles. *Molecular phylogenetics and evolution* **38**, 12-9.

Lyimo C.M., Weigend A., Janßen-Tapken U., Msoffe P.L., Simianer H. & Weigend S. (2013) Assessing the genetic diversity of five Tanzanian chicken ecotypes using molecular tools. *South African Journal of animal science* **43**, 499-510.

MacDonald K.C. & Edwards D.N. (1993) Chickens in Africa: the importance of Qasr Ibrim. *Antiquity* **67**, 584-90.

Maguelonne T.-S. (2009) A history of food. *Chichester, West Sussex, UK*.

Martin S.H., Davey J.W. & Jiggins C.D. (2014) Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Molecular biology and evolution* **32**, 244-57.

McGuire J.A., Linkem C.W., Koo M.S., Hutchison D.W., Lappin A.K., Orange D.I., Lemos-Espinal J., Riddle B.R. & Jaeger J.R. (2007) Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. *Evolution: International Journal of Organic Evolution* **61**, 2879-97.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S. & Daly M. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*.

Meydan H., Jang C.P., Yıldız M.A. & Weigend S. (2016) Maternal Origin of Turkish and Iranian Native Chickens Inferred from Mitochondrial DNA D-loop Sequences. *Asian-Australasian journal of animal sciences* **29**, 1547.

Mi H., Huang X., Muruganujan A., Tang H., Mills C., Kang D. & Thomas P.D. (2016) PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic acids research* **45**, D183-D9.

Miao Y., Peng M., Wu G., Ouyang Y., Yang Z., Yu N., Liang J., Pianchou G., Beja-Pereira A. & Mitra B. (2013) Chicken domestication: an updated perspective based on mitochondrial genomes. *Heredity* **110**, 277.

Morejohn G.V. (1968) Breakdown of isolation mechanisms in two species of captive junglefowl (Gallus gallus and Gallus sonneratii). *Evolution*, 576-82.

Mtileni B., Muchadeyi F., Maiwashe A., Chimonyo M., Groeneveld E., Weigend S. & Dzama K. (2011) Diversity and origin of South African chickens. *Poultry science* **90**, 2189-94.

Muchadeyi F., Eding H., Simianer H., Wollny C., Groeneveld E. & Weigend S. (2008) Mitochondrial DNA D-loop sequences suggest a Southeast Asian and Indian origin of Zimbabwean village chickens. *Animal genetics* **39**, 615-22.

Mwacharo J., Bjørnstad G., Han J. & Hanotte O. (2013) The history of African village chickens: an archaeological and molecular perspective. *African Archaeological Review* **30**, 97-114.

Mwacharo J.M., Bjørnstad G., Mobegi V., Nomura K., Hanada H., Amano T., Jianlin H. & Hanotte O. (2011) Mitochondrial DNA reveals multiple introductions of domestic chicken in East Africa. *Molecular phylogenetics and evolution* **58**, 374-82.

Nishibori M., Shimogiri T., Hayashi T. & Yasue H. (2005) Molecular evidence for hybridization of species in the genus Gallus except for Gallus varius. *Animal genetics* **36**, 367-75.

Nord K., Gunneriusson E., Ringdahl J., Ståhl S., Uhlén M. & Nygren P.-Å. (1997) Binding proteins selected from combinatorial libraries of an α-helical bacterial receptor domain. *Nature biotechnology* **15**, 772.

Ottenburghs J., Ydenberg R.C., Van Hooft P., Van Wieren S.E. & Prins H.H. (2015) The Avian Hybrids Project: gathering the scientific literature on avian hybridization. *Ibis* **157**, 892-4.

Perry-Gal L., Erlich A., Gilboa A. & Bar-Oz G. (2015) Earliest economic exploitation of chicken outside East Asia: Evidence from the Hellenistic Southern Levant. *Proceedings of the National Academy of Sciences* **112**, 9849-54.

Peters J., Lebrasseur O., Best J., Miller H., Fothergill T., Dobney K., Thomas R.M., Maltby M., Sykes N. & Hanotte O. (2015) Questioning new answers regarding Holocene chicken domestication in China. *Proceedings of the National Academy of Sciences* **112**, E2415-E.

Petit R.J. & Excoffier L. (2009) Gene flow and species delimitation. *Trends in Ecology & evolution* **24**, 386-93.

Pluzhnikov A., Di Rienzo A. & Hudson R. (2002) Inferences about human demography based on multilocus analyses of noncoding sequences. *Genetics* **161**, 1209.

Prendergast M.E., Buckley M., Crowther A., Frantz L., Eager H., Lebrasseur O., Hutterer R., Hulme-Beaman A., Van Neer W. & Douka K. (2017) Reconstructing Asian faunal introductions to eastern Africa from multi-proxy biomolecular and archaeological datasets. *PloS one* **12**.

Punnett R.C. (1923) *Heredity in poultry*. Macmillan and Company, limited.

Ramadan H., Galal A., Fathi M., El F.S. & Yakoub H. (2011) Characterization of two Egyptian native chicken breeds using genetic and immunological parameters. *Biotechnology in Animal Husbandry* **27**, 1-16.

Rambaut A., Drummond A.J., Xie D., Baele G. & Suchard M.A. (2018) Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol* **10**.

Ramos-Onsins S.E. & Rozas J. (2002) Statistical properties of new neutrality tests against population growth. *Molecular biology and evolution* **19**, 2092-100.

Razafindraibe H., Mobegi V.A., Ommeh S.C., Bjørnstad G., Hanotte O. & Jianlin H. (2008) Mitochondrial DNA origin of indigenous malagasy chicken. *Annals of the New York Academy of Sciences* **1149**, 77-9.

Rebl A., Köbis J.M., Fischer U., Takizawa F., Verleih M., Wimmers K. & Goldammer T. (2011) MARCH5 gene is duplicated in rainbow trout, but only fish-specific gene copy is up-regulated after VHSV infection. *Fish & shellfish immunology* **31**, 1041-50.

Redding R.W. (2015) The pig and the chicken in the Middle East: Modeling human subsistence behavior in the archaeological record using historical and animal husbandry data. *Journal of Archaeological Research* **23**, 325-68.

Remold-O'Donnell E. (1993) The ovalbumin family of serpin proteins. *FEBS letters* **315**, 105-8.

Riddle B.R. (2016) Comparative phylogeography clarifies the complexity and problems of continental distribution that drove AR Wallace to favor islands. *Proceedings of the National Academy of Sciences* **113**, 7970-7.

Riztyan K.K., Shimogiri T., Kawamoto Y., Rerkamnuaychoke W., Nishida T. & Okamoto S. (2014) Genetic Diversity and Ancestral Relationships of Red Junglefowls and Domestic Chickens in Southeast Asia. *Journal of Poultry Science* **51**, 370.

Rogers A.R., Fraley A.E., Bamshad M.J., Watkins W.S. & Jorde L.B. (1996) Mitochondrial mismatch analysis is insensitive to the mutational process. *Molecular biology and evolution* **13**, 895-902.

Rogers A.R. & Harpending H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular biology and evolution* **9**, 552-69.

Rubin C.-J., Zody M.C., Eriksson J., Meadows J.R., Sherwood E., Webster M.T., Jiang L., Ingman M., Sharpe T. & Ka S. (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587.

Saastamoinen M., Bocedi G., Cote J., Legrand D., Guillaume F., Wheat C.W., Fronhofer E.A., Garcia C., Henry R. & Husby A. (2018) Genetics of dispersal. *Biological Reviews* **93**, 574-99.

Sawai H., Kim H.L., Kuno K., Suzuki S., Gotoh H., Takada M., Takahata N., Satta Y. & Akishinonomiya F. (2010) The origin and genetic variation of domestic chickens with special reference to junglefowls Gallus g. gallus and G. varius. *PloS one* **5**, e10639.

Sewell R.S. & Guha B. (1931) Zoological remains. *Mohenjo-Daro and the Indus Civilisation, London: A. Probsthain*, 649-73.

Shaista R., Bano S. & Afzal S. (2017) Genetic diversity in farm and wild animals: a Review. *J. Inf. Mol. Biol* **5**, 7-26.

Sibley C.G. & Monroe B.L. (1990) *Distribution and Taxonomy of Birds of the World*. Yale University Press.

Silva P., Guan X., Ho-Shing O., Jones J., Xu J., Hui D., Notter D. & Smith E. (2009) Mitochondrial DNA-based analysis of genetic variation and relatedness among Sri Lankan indigenous chickens and the Ceylon junglefowl (Gallus lafayetti). *Animal genetics* **40**, 1-9.

Smeds L., Warmuth V., Bolivar P., Uebbing S., Burri R., Suh A., Nater A., Bureš S., Garamszegi L.Z. & Hogner S. (2015) Evolutionary analysis of the female-specific avian W chromosome. *Nature communications* **6**, 7330.

Storey A.A., Athens J.S., Bryant D., Carson M., Emery K., Higham C., Huynen L., Intoh M., Jones S. & Kirch P.V. (2012) Investigating the global dispersal of chickens in prehistory using ancient mitochondrial DNA signatures. *PloS one* **7**, e39171.

Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-95.

Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F. & Higgins D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic acids research* **25**, 4876-82.

Tixier-Boichard M., Bed'hom B. & Rognon X. (2011) Chicken domestication: From archeology to genomics. *Comptes rendus biologies* **334**, 197-204.

Van der Auwera G.A., Carneiro M.O., Hartl C., Poplin R., Del Angel G., Levy-Moonshine A., Jordan T., Shakir K., Roazen D. & Thibault J. (2013) From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics* **43**, 11.0. 1-.0. 33.

Wani C.E., Yousif I.A., Ibrahim M.E. & Musa H.H. (2014) Molecular characterization of Sudanese and Southern Sudanese chicken breeds using mtDNA D-Loop. *Genetics research international* **2014**.

West B. & Zhou B.-X. (1988) Did chickens go north? New evidence for domestication. *Journal of Archaeological Science* **15**, 515-33.

Wiens J.J. (2012) Perspective: why biogeography matters: historical biogeography vs. phylogeography and community phylogenetics for inferring ecological and evolutionary processes. *Frontiers of Biogeography* **4**.

Wikman M., Steffen A.-C., Gunneriusson E., Tolmachev V., Adams G.P., Carlsson J. & Ståhl S. (2004) Selection and characterization of HER2/neu-binding affibody ligands. *Protein Engineering Design and Selection* **17**, 455-62.

Wilkinson S., Wiener P., Teverson D., Haley C. & Hocking P. (2012) Characterization of the genetic diversity, structure and admixture of British chicken breeds. *Animal genetics* **43**, 552-63.

Williamson K., Blench R. & Macdonald K. (2000) Did chickens go west. *The origins and development of African livestock: archaeology, genetics, linguistics and ethnography*, 368-448.

Wilson Sayres M.A. (2018) Genetic diversity on the sex chromosomes. *Genome Biology and Evolution* **10**, 1064-78.

Woldekiros H. & D'Andrea A. (2017) Early evidence for domestic chickens (Gallus gallus domesticus) in the Horn of Africa. *International Journal of Osteoarchaeology* **27**, 329-41.

Xiang H., Gao J., Yu B., Zhou H., Cai D., Zhang Y., Chen X., Wang X., Hofreiter M. & Zhao X. (2014) Early Holocene chicken domestication in northern China. *Proceedings of the National Academy of Sciences* **111**, 17564-9.

Zeuner F.E. (1963) *A history of domesticated animals:* Hutchinson a Co. (Publishers) Ltd. London.

Zhang D.X. & Hewitt G.M. (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* **12**, 563-84.

Zhang L.-J., Cai W.-Z., Luo J.-Y., Zhang S., Wang C.-Y., Lv L.-M., Zhu X.-Z., Wang L. & Cui J.-J. (2017) Phylogeographic patterns of Lygus pratensis (Hemiptera: Miridae): Evidence for weak genetic structure and recent expansion in northwest China. *PloS one* **12**, e0174712.

Zhang M., Han W., Tang H., Li G., Zhang M., Xu R., Liu Y., Yang T., Li W. & Zou J. (2018) Genomic diversity dynamics in conserved chicken populations are revealed by genome-wide SNPs. *BMC genomics* **19**, 598.

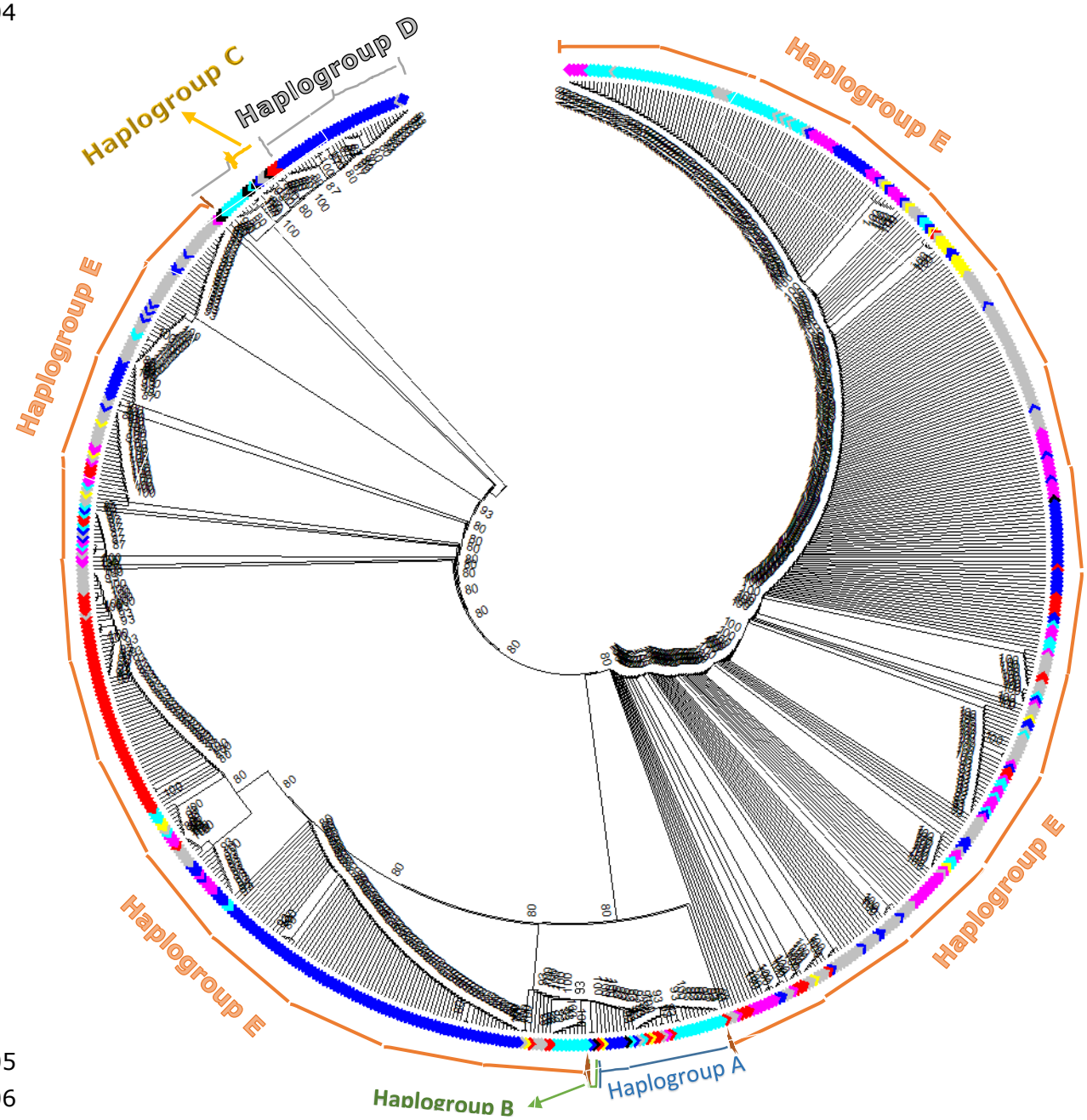
Appendices

```
    01111111112222222222222222222222222222333333333333333333333333
    03467779990111222223333344444455668999900011111223455555566679 9
    63271783897027124593468902367946151126926803457270223458923721 6 N
Ref TTCTTACCCTACATCATCCCCTGACGTTAATTCCAAATTCTCCCCTATTCACATTACCCTACT
H_1 ............G....T.........C.........C....T..C.................. 12
H_2 ............C.G.C........C..G.CTT...................C........ 5
H_3 ............C...C........C....CT.............................291
H_4 ............C.G.C........C..G.CT.G..................C........ 43
H_5 ............C...C........C....CT.....................C.......C 3
H_6 ...........T.............C....CT.........................T.... 1
H_7 ...C.........C...C........C....CT........A..................... 1
H_8 ...C.........C...C....A...C....CT.............................. 1
H_9 ...C.........C...C........C....CT.............................. 10
H_10............C.G.C........C....CT...............T............. 15
H_11...C.....................................C..................... 7
H_12...C...................G..................C..................... 1
H_13............C...C........C....AT.............................. 6
H_14............C...C........C....CT...............T............. 9
H_15............C...C........C....CT.................G............ 15
H_16............C...C........C.....T.............................. 13
H_17............C...C........CT...CT.............................. 33
H_18...C............C........C....C..G.....C.........G......TT.... 7
H_20................C........C....C..G.....C.........G......TT.... 21
H_21............C...C........CT...CT...G........................... 2
H_22............C...C........CT...CT.............C............... 6
H_23............C...C........C....CT..............................C 54
H_24.......A.....C...C........C....CT.............................. 8
H_25............C...C........C.....T............................C 1
H_26...C.......T..............................C..................... 8
H_27............C...C........................C..................... 9
H_28............C...C........C....CT.....................G....... 1
H_29............C...C.T......C....CT.............................. 2
H_30...........T.C...C.T......C....CT.............................. 1
H_31...........T.............CT...CT.............................. 6
H_32...........T.C...C........C....CT.............................. 1
H_33............A....C.........T.............C..T................. 1
H_34.........C...C...C........C.....T............................. 3
H_35...C.........C...C........C....CT.....................T.....C 2
```

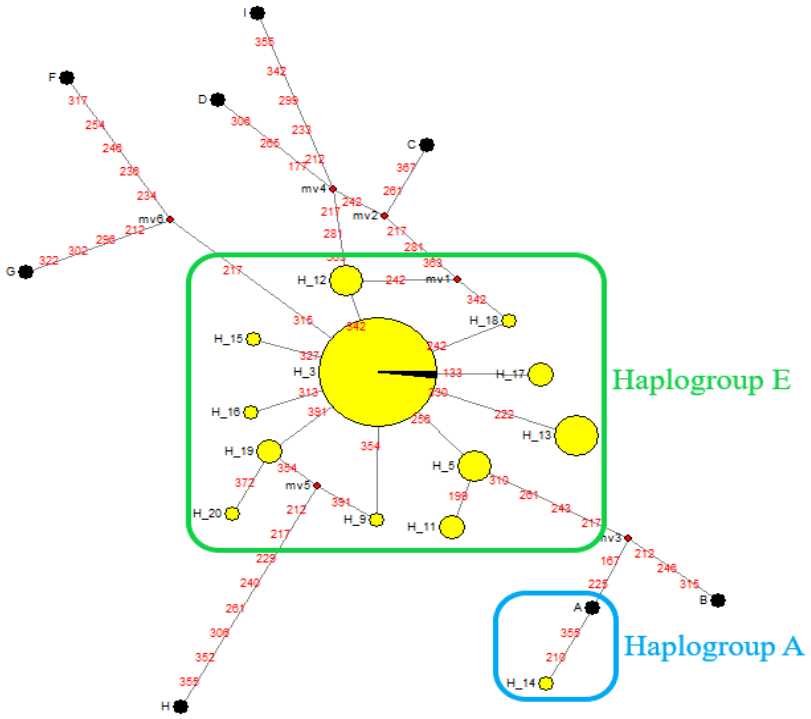
```
H_36...C.........C...C........C....CT.G.....C.........G......TT.... 1
H_37...C.........C...C........C....CT.........C..................C 1
H_38.C...........C...C........C....CT.............................. 8
H_39.................C.......AC....C..G..............G.......TC... 1
H_40.............C...C........C....CT...................C......... 1
H_41.............C...C........C....CT............................T. 3
H_42..........G..C...C........C....CT.............................. 2
H_43.............C.G.C........C....CT.....................C........ 1
H_44.............C...C........C....CT.....................C........ 1
H_45..T..............C........C....C..G.....C.........G......TT.... 2
H_46.............A....C.......AC....C..G.....................TC... 2
H_47...C.............C........C....C..G.....C.........G......TTC... 1
H_48.................C........C....CT.G.....C.........G......TT.... 3
H_51.................C........C....CT.G.....C..................... 6
H_52.................C........C.G..CT.G.....C..................... 1
H_53...C..T..........C........C....C..G.....C...T.....G......TT.... 2
H_55.............C...C........C....CT............................GT. 1
H_56...C.......T.............................C...........C........ 1
H_57.............C...C..............CT.............................. 6
H_58.................C........C....C..G.G...C.........G......TT.... 1
H_59.....................................G.....C..................... 1
H_60...C.........C...C........C....C.............................. 1
H_61.............C...C........C....C.............................. 2
H_62.............C...C........C....C......................G....... 1
H_63.........C...C...C........C....CT.............................. 6
H_64.............C...C........C....CT................T........T.... 1
H_65.............A....C........C....CT.G.....C..................... 1
H_66...C.........C.G.C........C....CT................T............. 1
H_67.............C.G.C..T.....C..G.CT..C..................C........ 1
H_68.............C.G.C........C..G.CT.....................C........ 4
H_69C............C.G.C........C..G.CT.....................C........ 3
H_70............T..............C....CT..........T................... 1
H_71.....G.....T..............CT...CT.............................. 1
H_72.............C...C........C....CT................C.............. 1
H_73.............C...C........C....CT..........T......G............ 7
H_74........T....C...C........C....CT..........T......G............ 1
H_75.............C...C........C....CT....T.....T......G............ 1
H_76.............C...C........C....CT..........T................... 1
```

```
H_77.............C...C........C....CT.......................T.....C 1
H_78.............C...C........C....CT....T.....................C 1
H_79...C.........C...C........C....CT..........................C 3
H_80...C.........C...C........C....CTT.........................C 1
H_81.............C.G.C........C..G.CT.G..............G...C..... 2
H_82.............C.G.C........C....CT.G..................C..... 2
H_83...C.............C.....................C................... 1
H_84.............C...C........C....CT.........C................... 2
H_85.............C.GCC........C..G.CT.G..................C..... 1
H_86.............C...C........C....CT...................G..... 5
H_87....C........C...C........C....CT.......................... 2
H_89.............CT..C........C....CT.......................... 2
H_90.............AC...C........C....CT.......................... 1
H_91.....G.......C...C........C....CT.......................... 1
H_92.............C...C.......AC....CT.......................... 1
H_93...........T.............C....CT.......................... 1
```

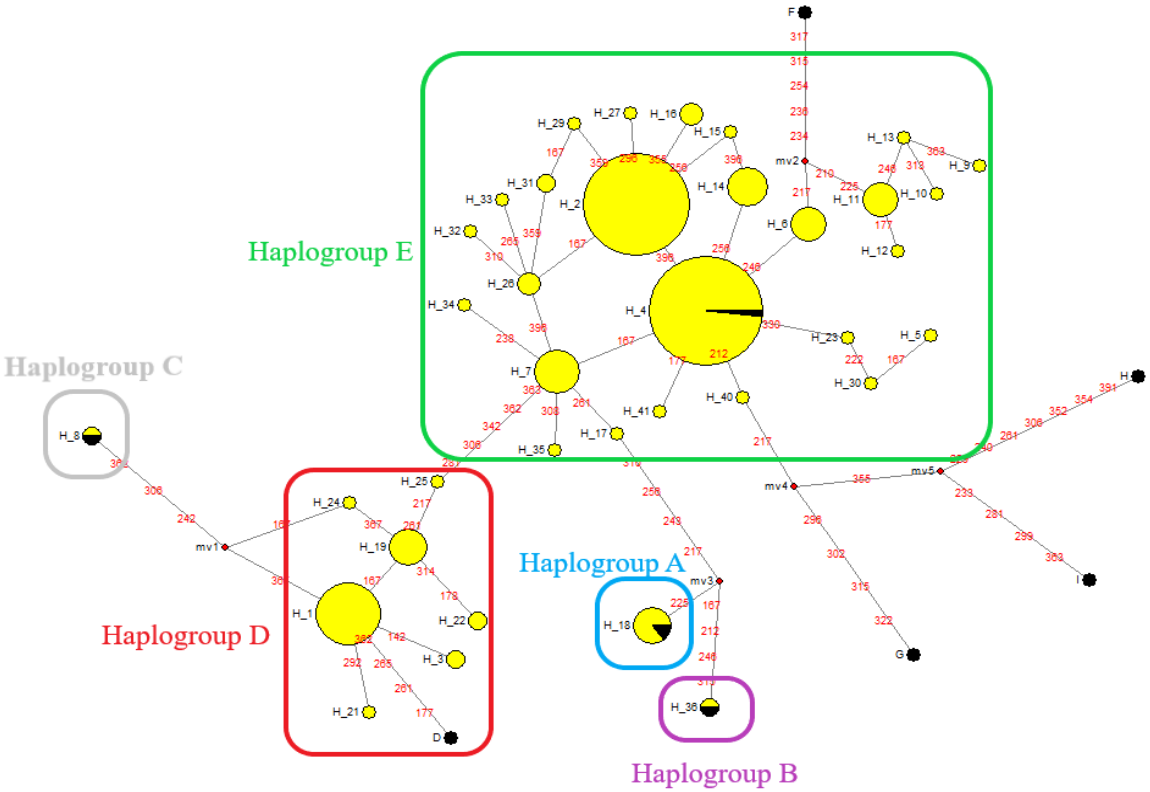
Supplementary Figure S1: mtDNA D-loop variation of 88 haplotypes found in the 706 village chickens. The vertical numbers indicate to the position of the polymorphism compared to the mDNA sequence of reference (Liu *et al*., 2006). The right column under 'N' the number of time a haplotype was observed. Dots (.) indicate that the same nucleotide compared to the reference sequences was found, the letters show the polymorphism.



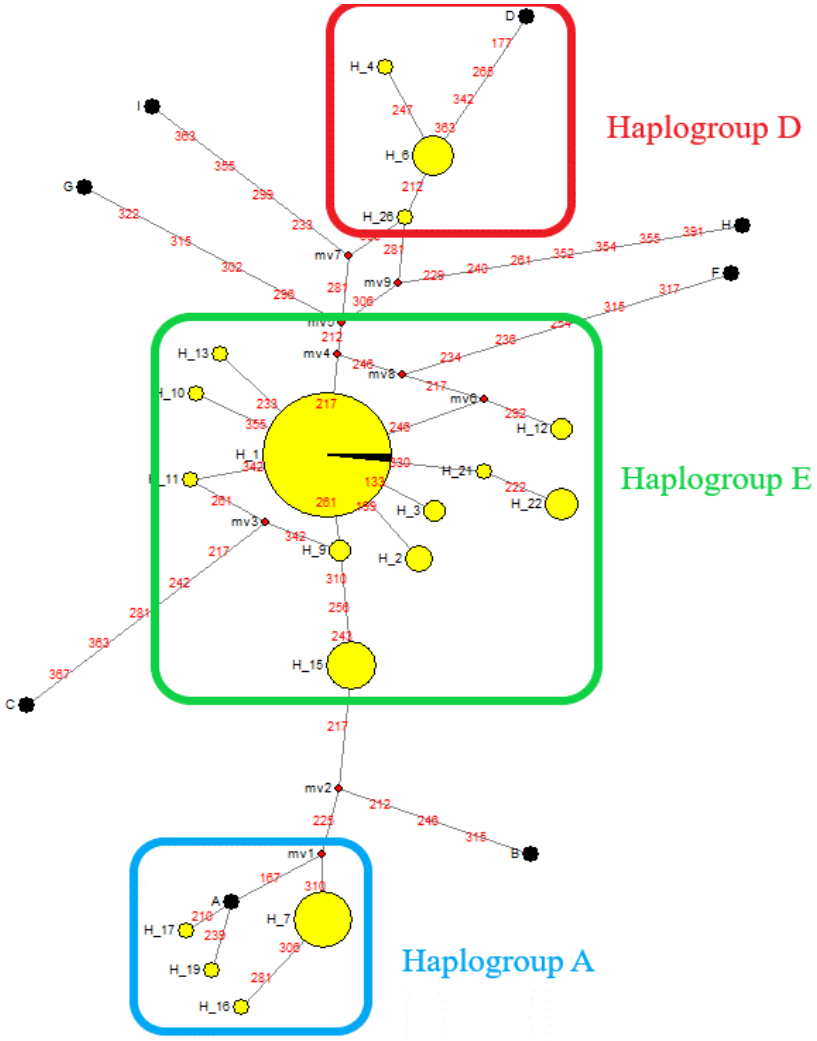
Supplementary Figure S2: Maximum likelihood tree for 706 samples and reference sequences from Liu *et al*. (2006). Blue = Ethiopia, Red = Pakistan, Grey = Saudi, pink = Algeria, yellow = Libya, cyan = Iraq and black = haplogroup references (A, B, C, D, E, F, G, H and I).



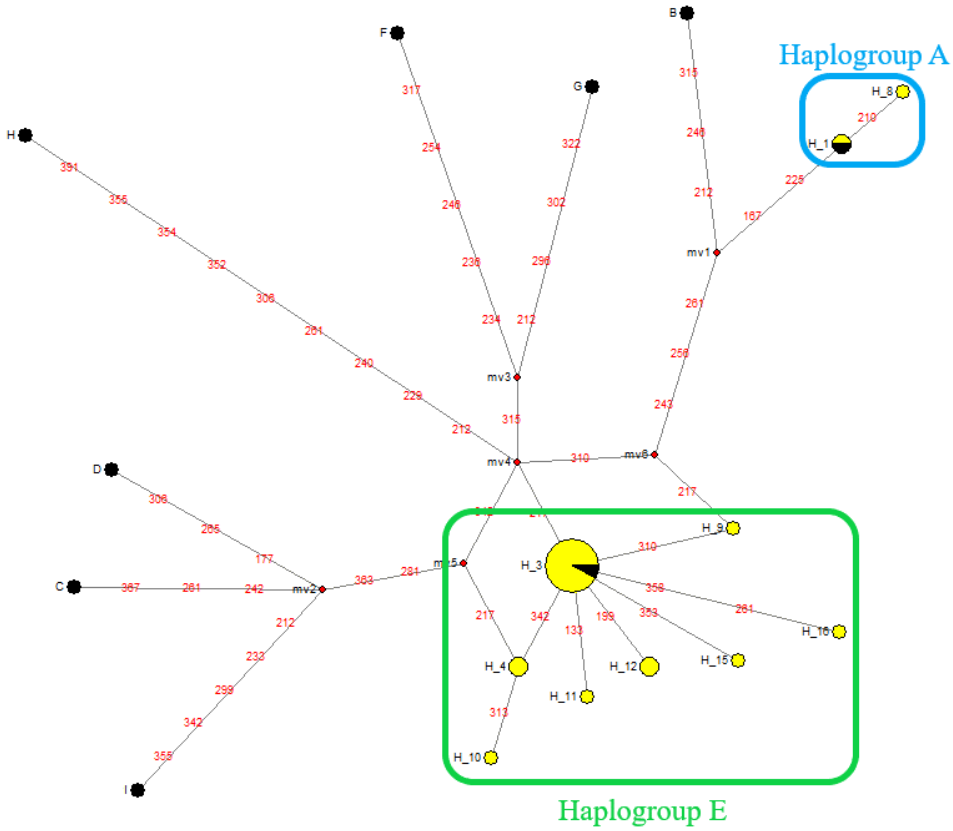
Supplementary Figure S3: Median-Joining network for Algerian haplotypes (n = 13). The black circles refer to reference haplotypes from Liu *et al*. (2006), yellow = Algerian haplotypes. The numbers on the branch indicate the position of the mutations, the circles are proportional to the numbers of haplotypes.



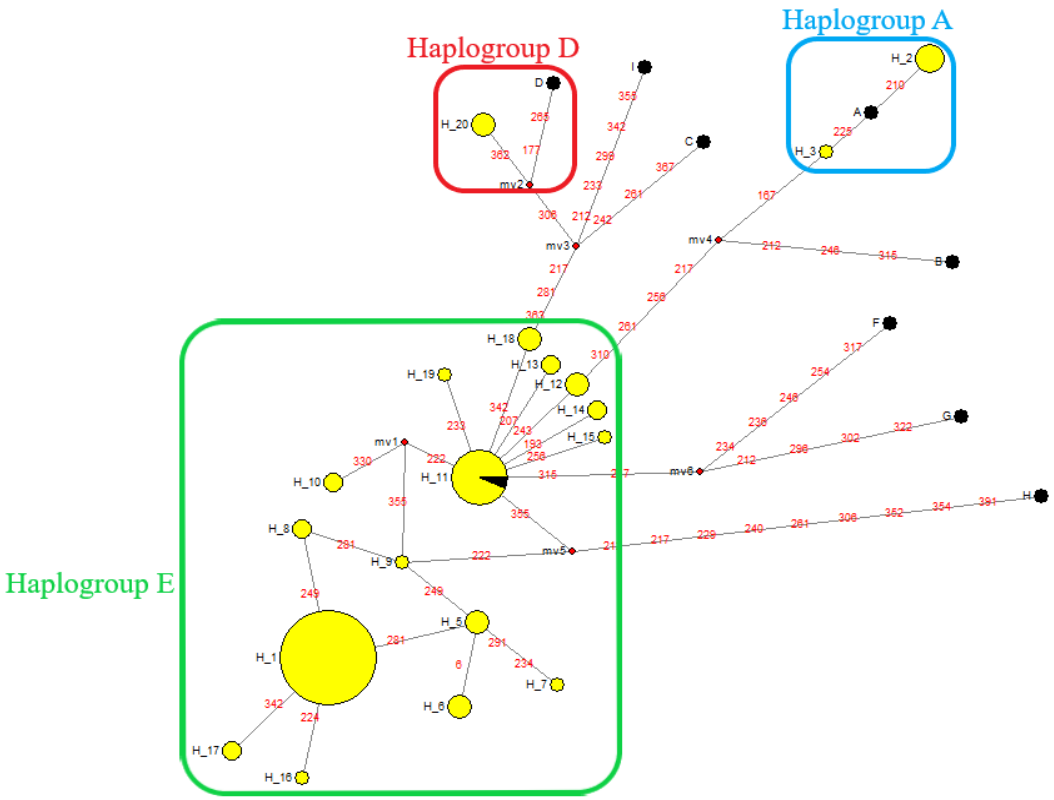
Supplementary Figure S4: Median-Joining network for Ethiopian haplotypes (n = 36). The black circles refer to reference haplotypes from Liu *et al*. (2006), yellow = Ethiopian haplotypes. The numbers on the branch indicate the position of the mutations, the circles are proportional to the numbers of haplotypes.



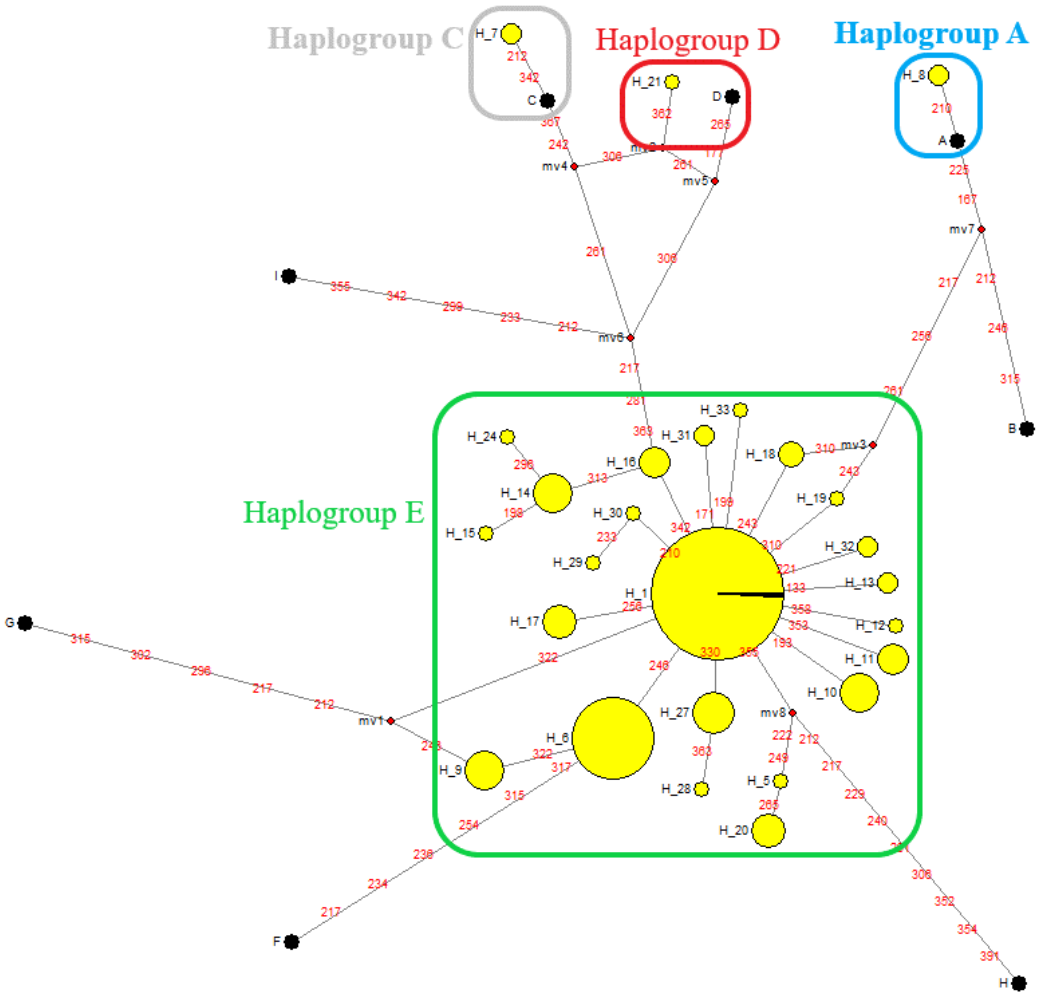
Supplementary Figure S5: Median-Joining network for Iraqi haplotypes (n = 18). The black circles refer to reference haplotypes from Liu *et al*. (2006), yellow = Iraqi haplotypes. The numbers on the branch indicate the position of the mutations, the circles are proportional to the numbers of haplotypes.



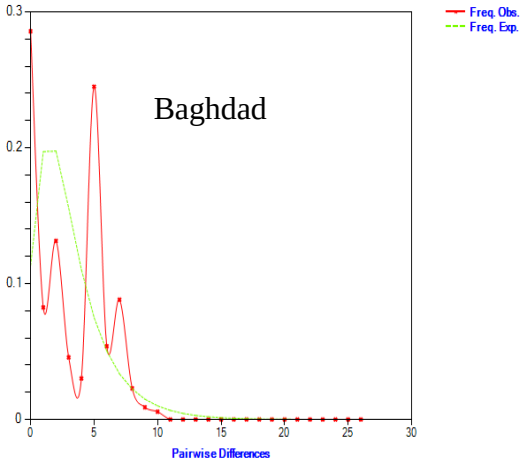
Supplementary Figure S6: Median-Joining network for Libyan haplotypes (n = 10). The black circles refer to reference haplotypes from Liu *et al*. (2006), yellow = Libyan haplotypes. The numbers on the branch indicate the position of the mutations, the circles are proportional to the numbers of haplotypes.

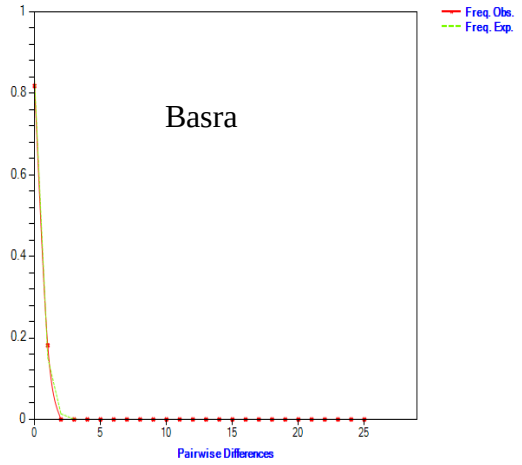


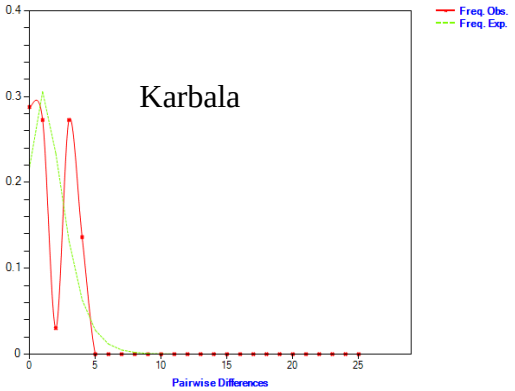
Supplementary Figure S7: Median-Joining network for Pakistani haplotypes (n = 19). The black circles refer to reference haplotypes from Liu *et al*. (2006), yellow = Pakistani haplotypes. The numbers on the branch indicate the position of the mutations, the circles are proportional to the numbers of haplotypes.

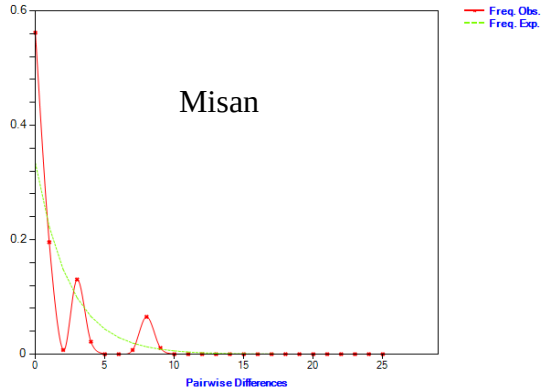


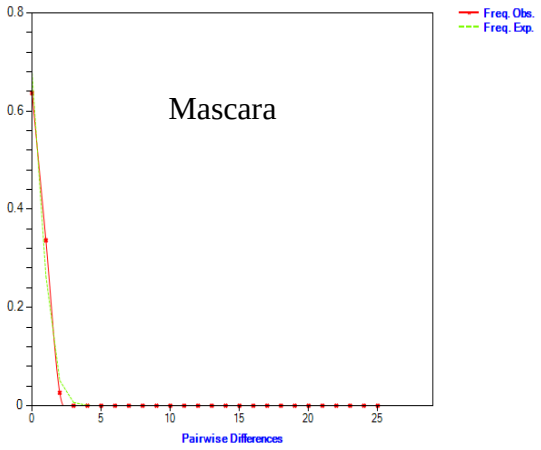
Supplementary Figure S8: Median-Joining network for Saudi Arabia haplotypes (n = 26). The black circles refer to reference haplotypes from Liu *et al*. (2006), yellow = Saudi haplotypes. The numbers on the branch indicate the position of the mutations, the circles are proportional to the numbers of haplotypes.

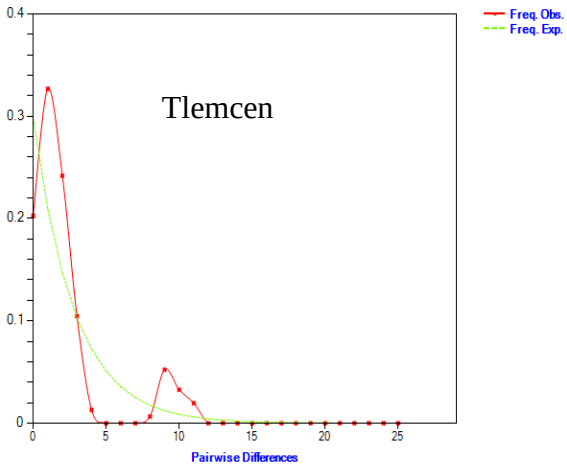

Baghdad
Freq. Obs.
Freq. Exp.
Pairwise Differences

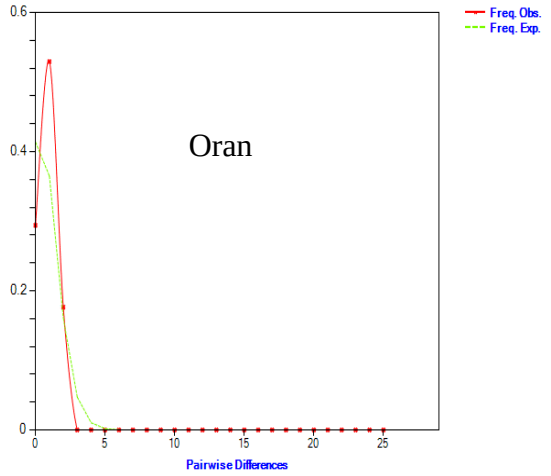

Basra
Freq. Obs.
Freq. Exp.
Pairwise Differences

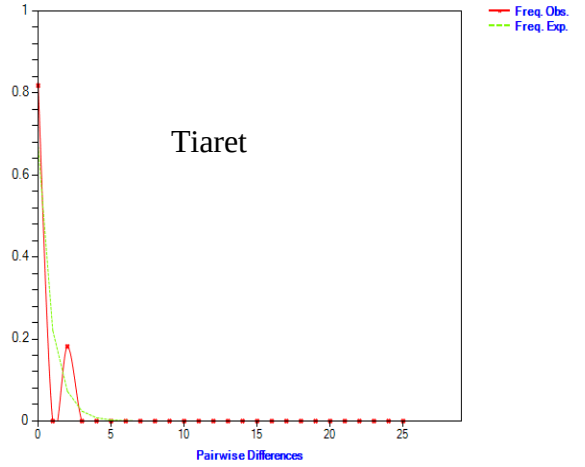

Karbala
Freq. Obs.
Freq. Exp.
Pairwise Differences

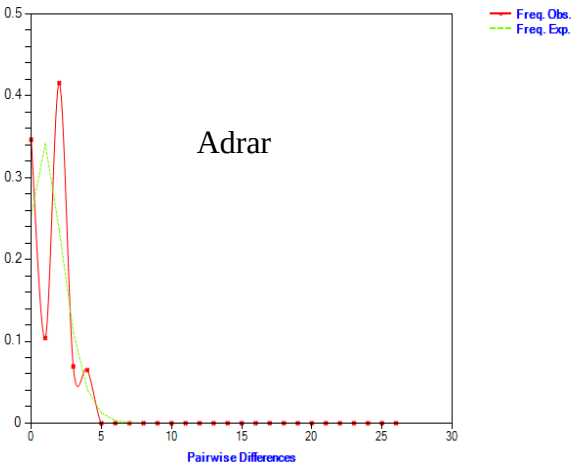

Misan
Freq. Obs.
Freq. Exp.
Pairwise Differences

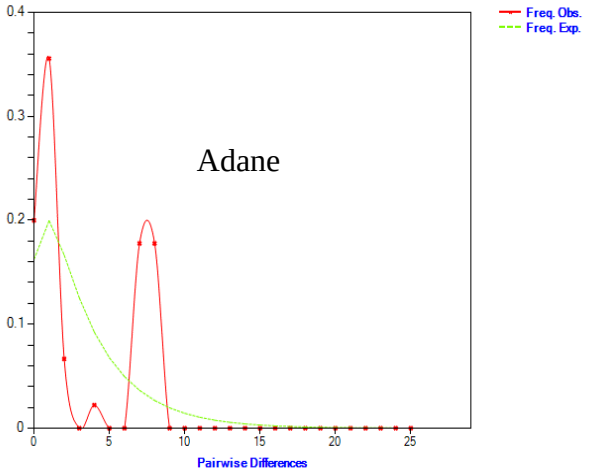

Mascara
Freq. Obs.
Freq. Exp.
Pairwise Differences

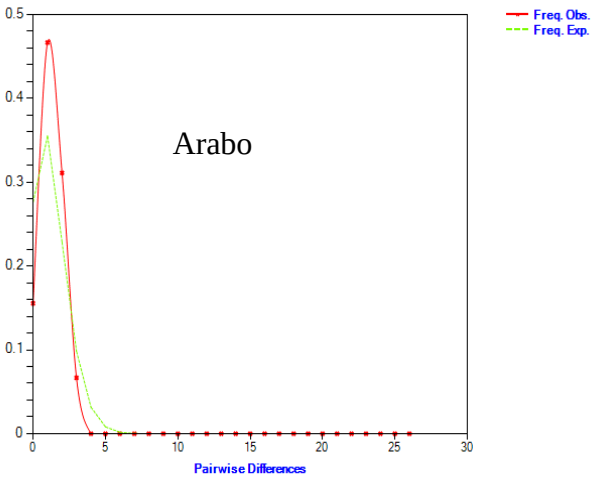

Tlemcen
Freq. Obs.
Freq. Exp.
Pairwise Differences

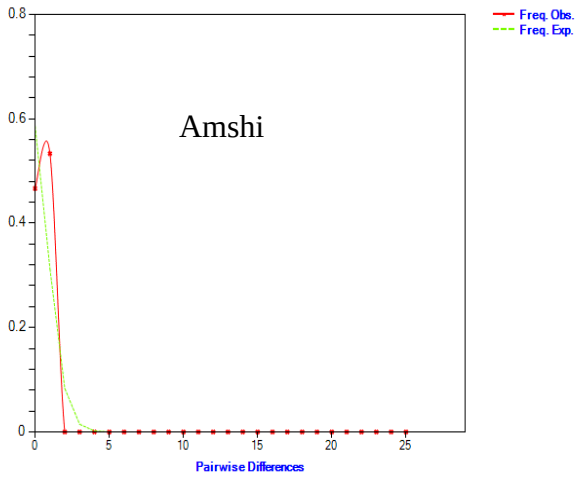

Oran
Freq. Obs.
Freq. Exp.
Pairwise Differences

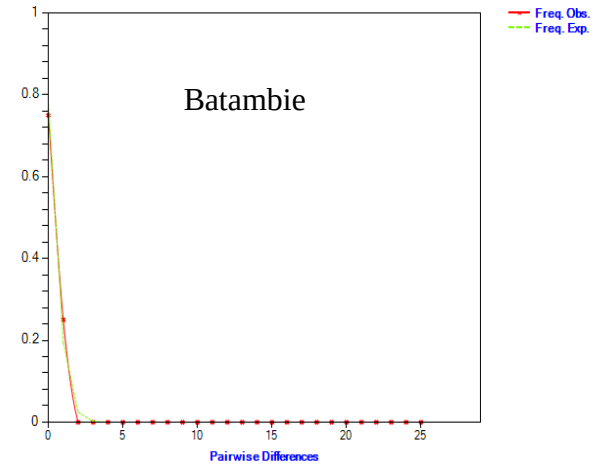

Tiaret
Freq. Obs.
Freq. Exp.
Pairwise Differences

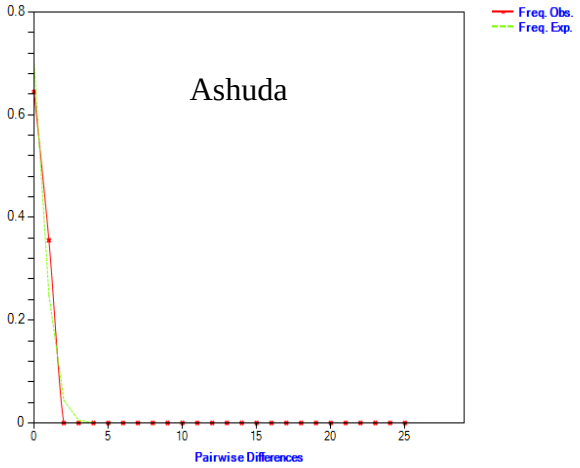

Adrar
Freq. Obs.
Freq. Exp.
Pairwise Differences

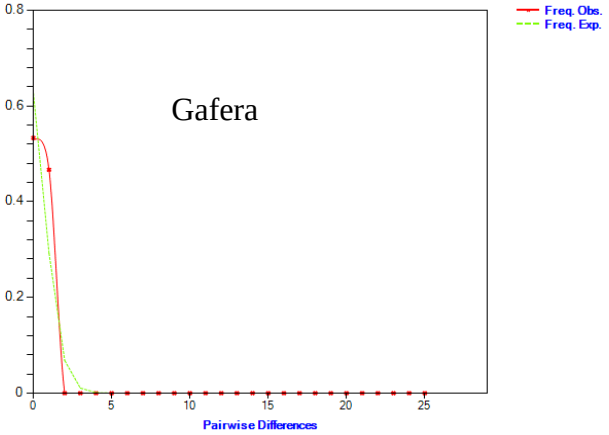

Adane
Freq. Obs.
Freq. Exp.
Pairwise Differences

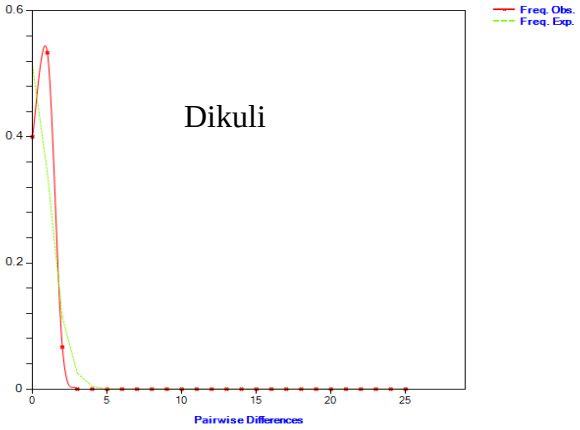

Arabo
Freq. Obs.
Freq. Exp.
Pairwise Differences

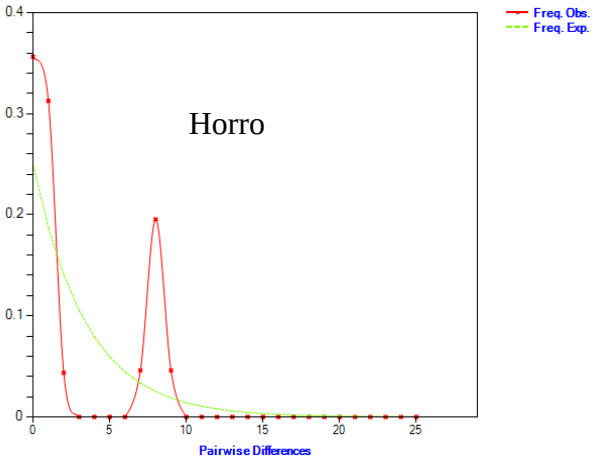

Amshi
Freq. Obs.
Freq. Exp.
Pairwise Differences

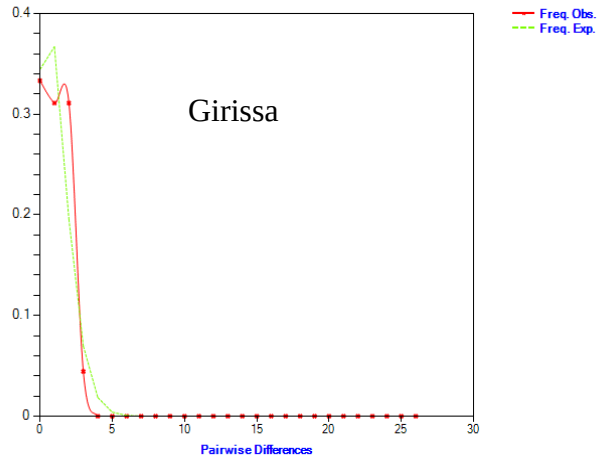

Batambie
Freq. Obs.
Freq. Exp.
Pairwise Differences

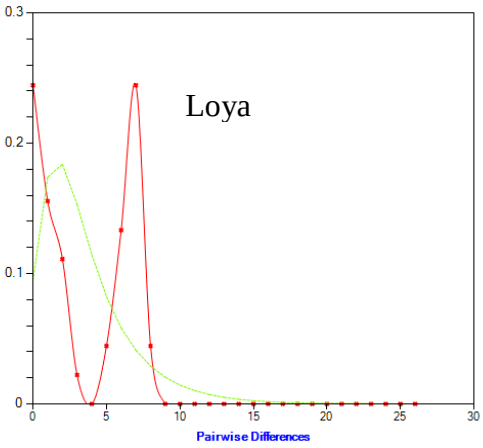

Ashuda
Freq. Obs.
Freq. Exp.
Pairwise Differences

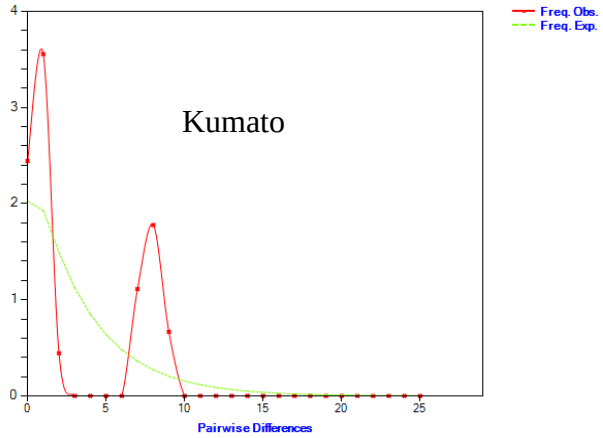

Gafera
Freq. Obs.
Freq. Exp.
Pairwise Differences

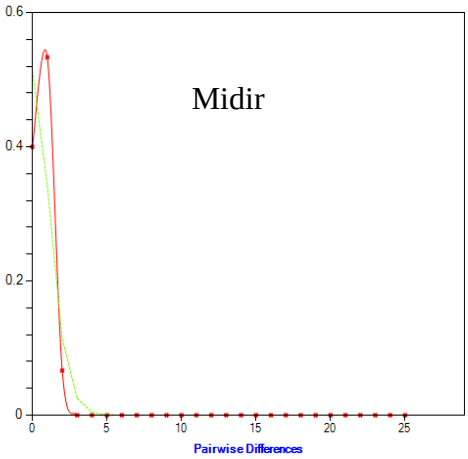

Dikuli
Freq. Obs.
Freq. Exp.
Pairwise Differences

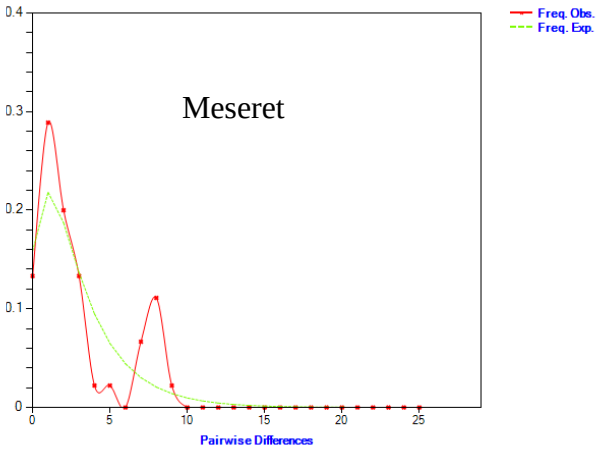

Horro
Freq. Obs.
Freq. Exp.
Pairwise Differences

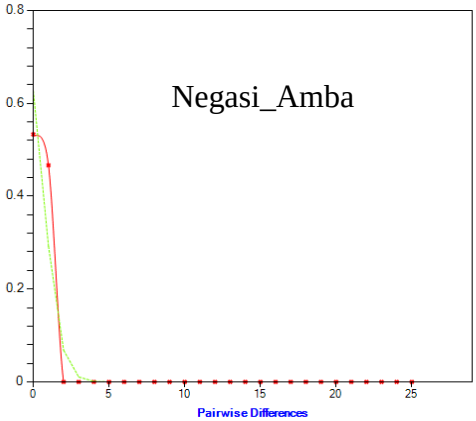

Girissa
Freq. Obs.
Freq. Exp.
Pairwise Differences

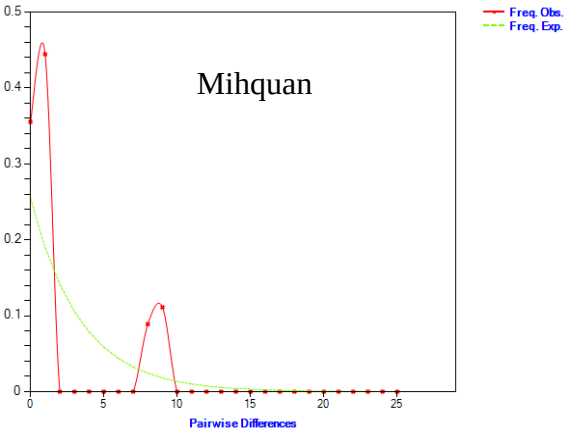

Loya
Freq. Obs.
Freq. Exp.
Pairwise Differences

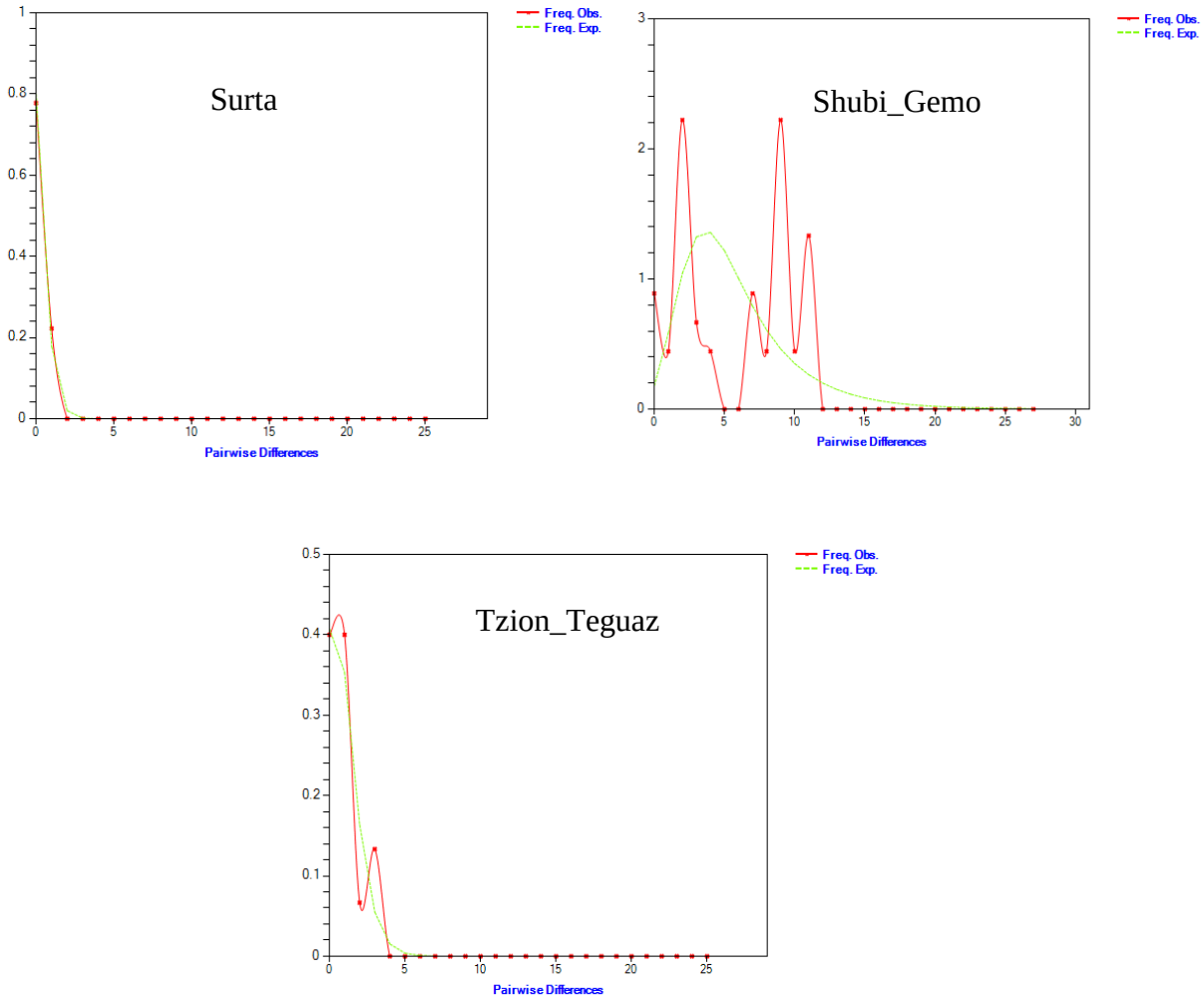

Kumato
Freq. Obs.
Freq. Exp.
Pairwise Differences


Midir
Freq. Obs.
Freq. Exp.
Pairwise Differences


Meseret
Freq. Obs.
Freq. Exp.
Pairwise Differences


Negasi_Amba
Freq. Obs.
Freq. Exp.
Pairwise Differences


Mihquan
Freq. Obs.
Freq. Exp.
Pairwise Differences

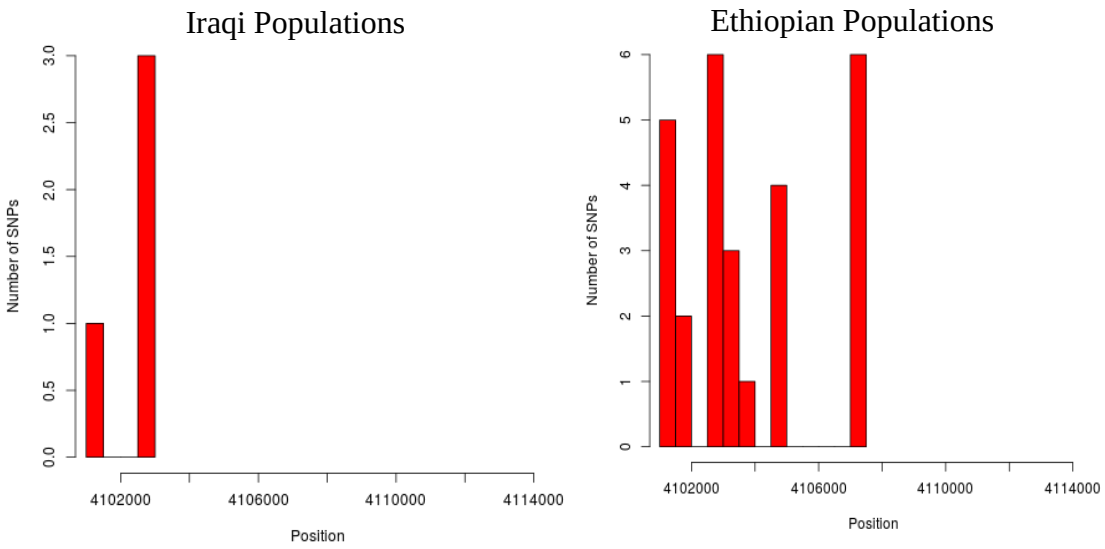


Supplementary Figure S9: Mismatch distribution patterns of populations included in this study.

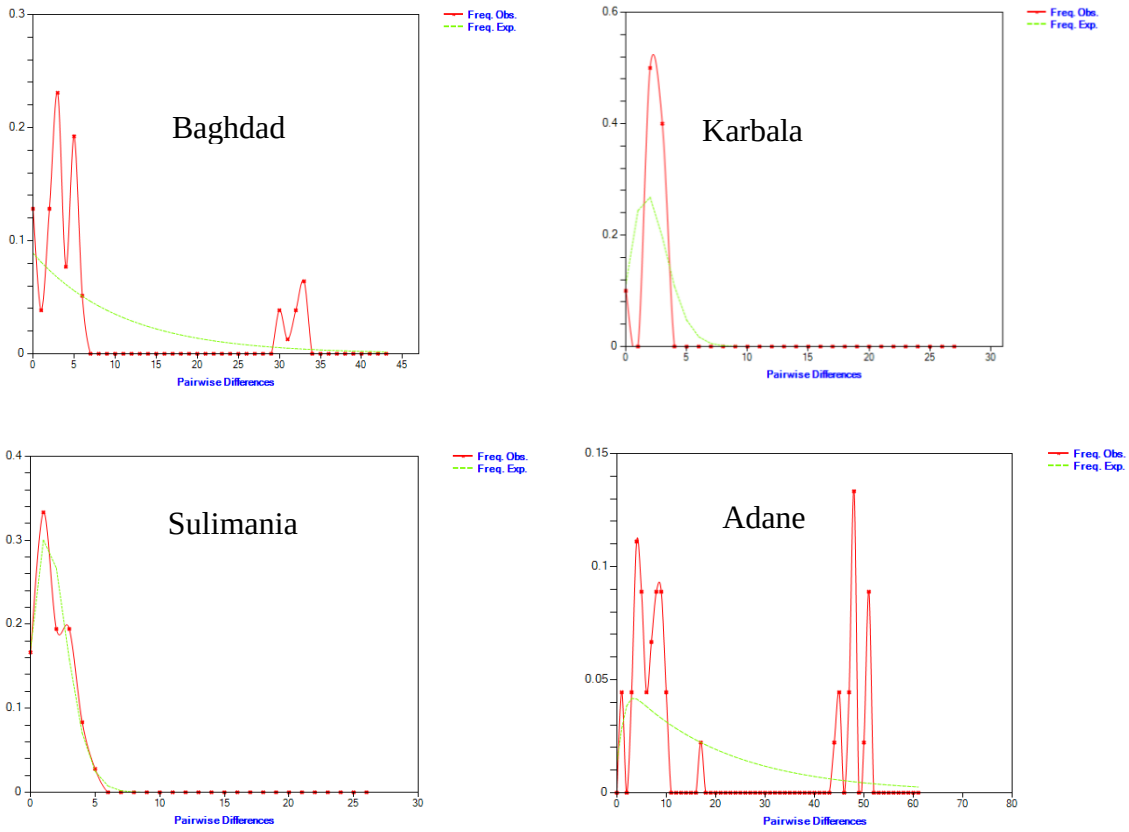
Supplementary Table S1: Sample size and genetic diversity of Haplogroup E

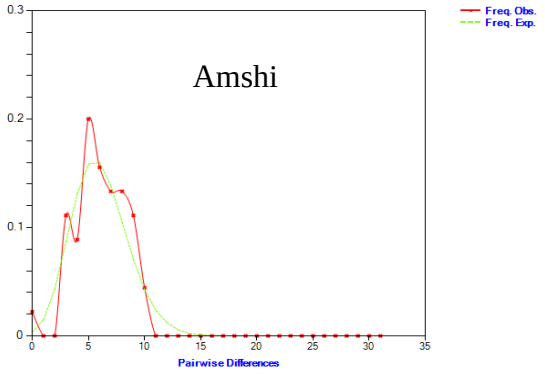
Country/population	N	S	H	Hd (SD)	π (SD)	K
Iraqi populations						
Algeria_Haplogroup E	87	12	12	0.588(0.060)	0.0024(0.0003)	0.955
Ethiopia_Haplogroup E	169	25	27	0.770(0.023)	0.0040(0.0003)	1.613
Iraq_Haplogroup E	84	13	11	0.512(0.064)	0.0032(0.0005)	1.299
Libya_Haplogroup E	21	8	8	0.676(0.111)	0.0025(0.0006)	1.010
Pakistan_Haplogroup E	84	19	17	0.717(0.048)	0.0062(0.0005)	2.485
Saudi_Haplogroup E	180	24	23	0.712(0.034)	0.0032(0.0003)	1.273

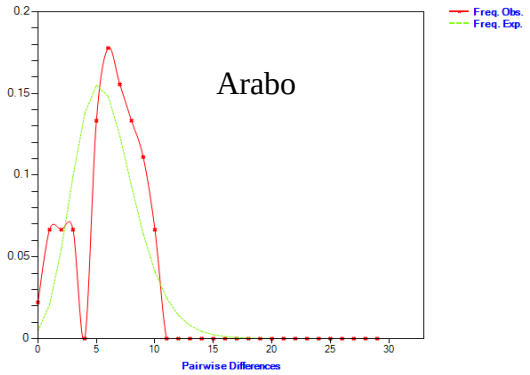
Chapter 3

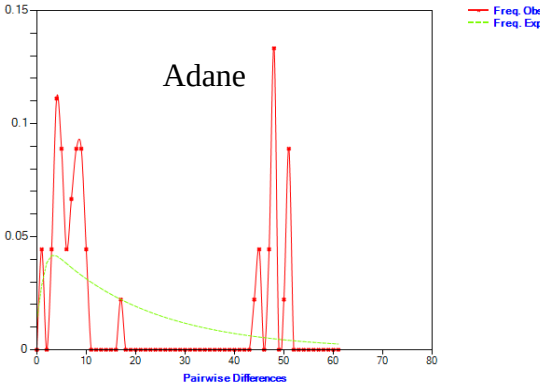


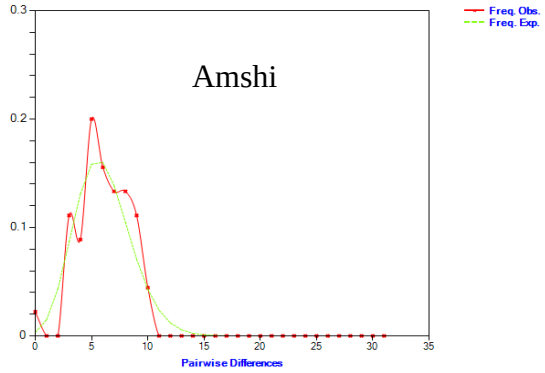
Supplementary Figure S1: 13 kb of W chromosome for both Iraqi and Ethiopian samples.

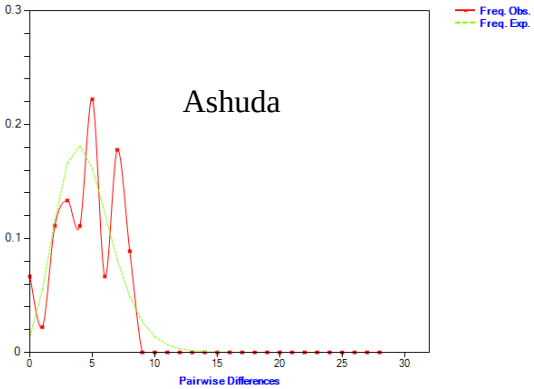


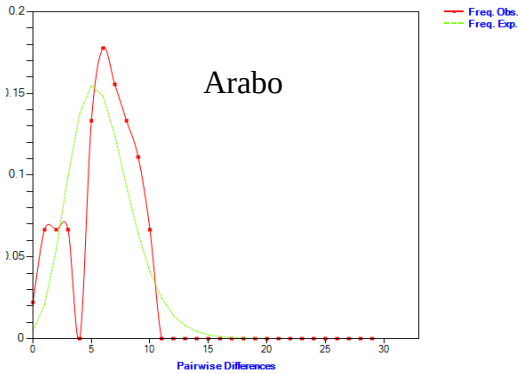

Amshi
Freq. Obs.
Freq. Exp.
Pairwise Differences

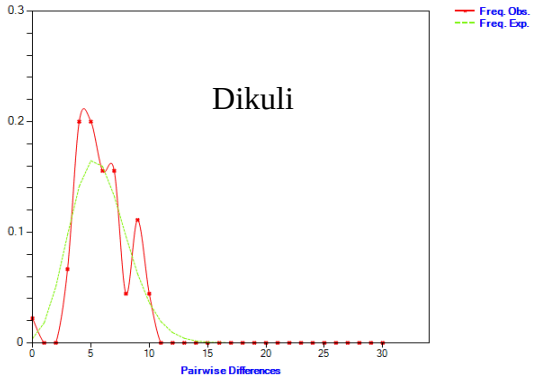

Arabo
Freq. Obs.
Freq. Exp.
Pairwise Differences

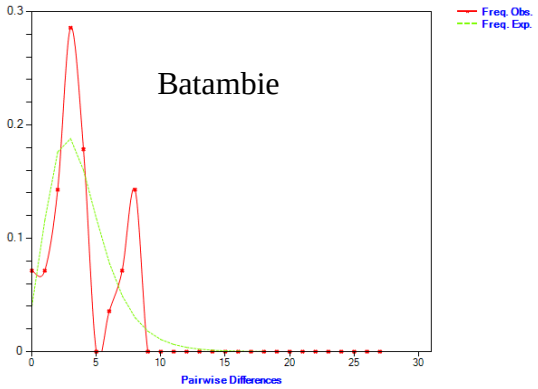

Adane
Freq. Obs.
Freq. Exp.
Pairwise Differences

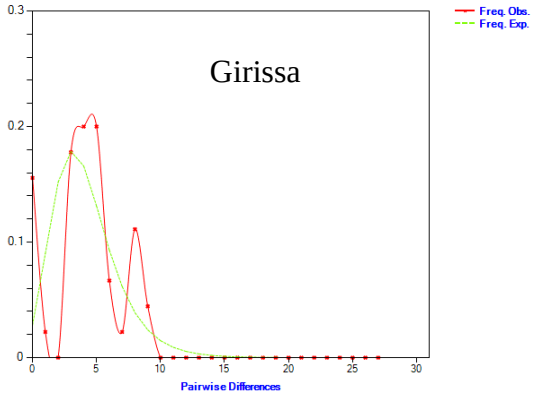

Amshi
Freq. Obs.
Freq. Exp.
Pairwise Differences

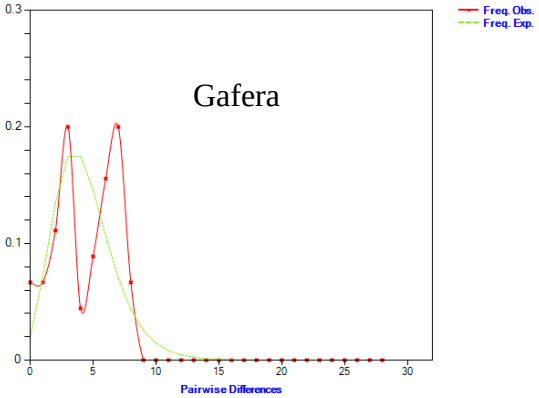

Ashuda
Freq. Obs.
Freq. Exp.
Pairwise Differences

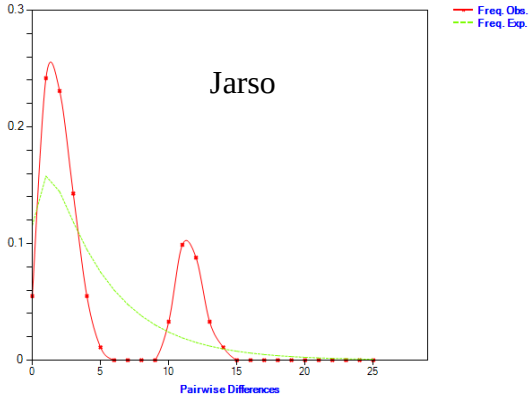

Arabo
Freq. Obs.
Freq. Exp.
Pairwise Differences

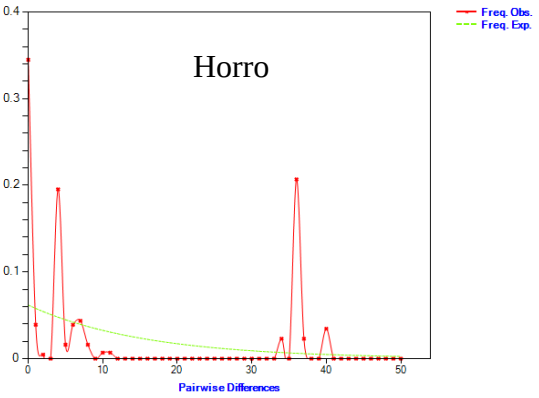

Dikuli
Freq. Obs.
Freq. Exp.
Pairwise Differences

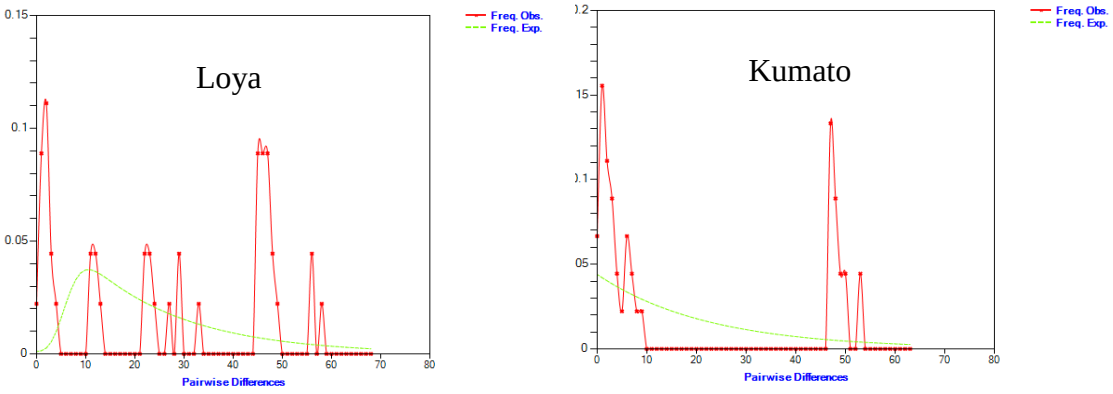

Batambie
Freq. Obs.
Freq. Exp.
Pairwise Differences


Girissa
Freq. Obs.
Freq. Exp.
Pairwise Differences

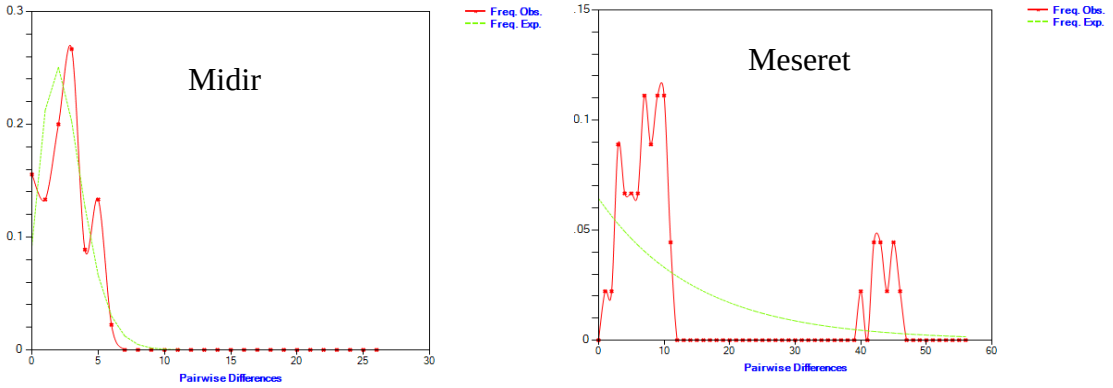

Gafera
Freq. Obs.
Freq. Exp.
Pairwise Differences


Jarso
Freq. Obs.
Freq. Exp.
Pairwise Differences

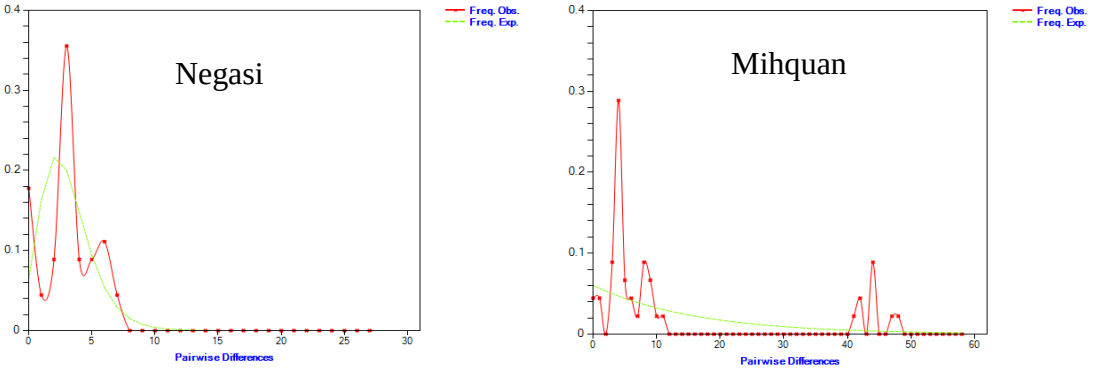

Horro
Freq. Obs.
Freq. Exp.
Pairwise Differences


Loya
Freq. Obs.
Freq. Exp.
Pairwise Differences

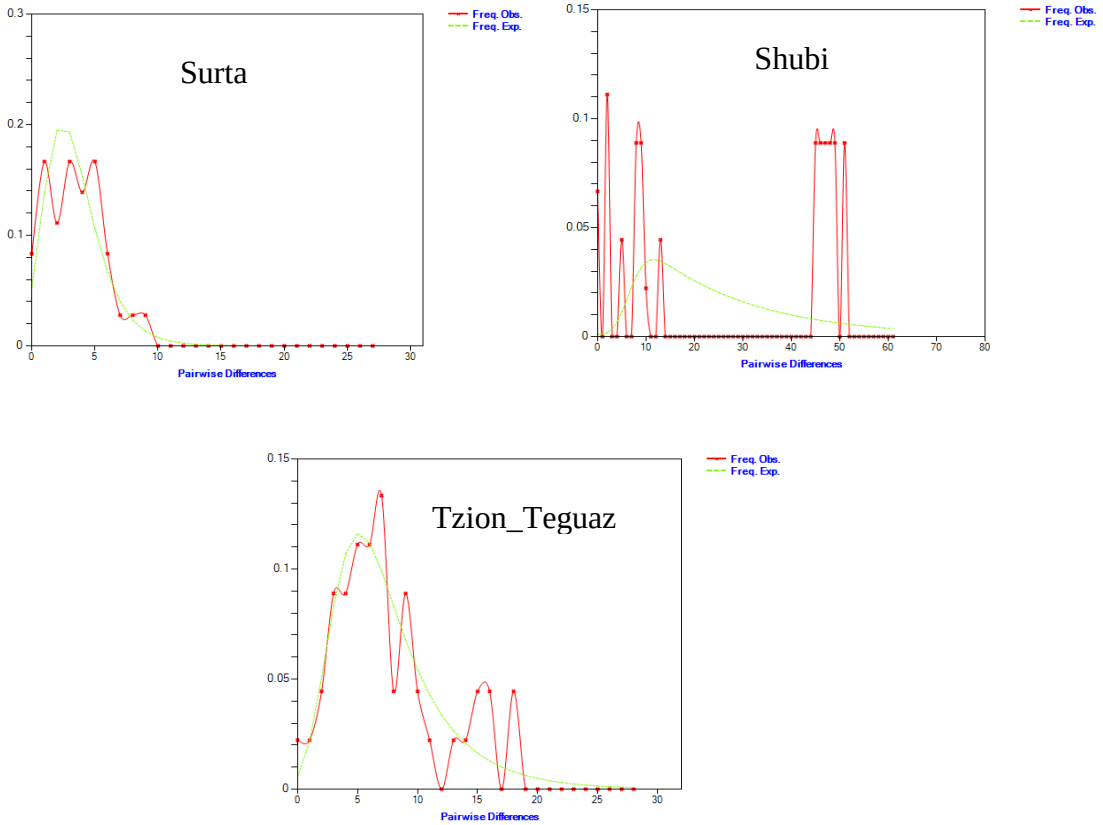
Kumato
Freq. Obs.
Freq. Exp.
Pairwise Differences


Midir
Freq. Obs.
Freq. Exp.
Pairwise Differences

Meseret
Freq. Obs.
Freq. Exp.
Pairwise Differences


Negasi
Freq. Obs.
Freq. Exp.
Pairwise Differences

Mihquan
Freq. Obs.
Freq. Exp.
Pairwise Differences



Supplementary Figure S2: Mismatch distribution patterns of populations included in this study.

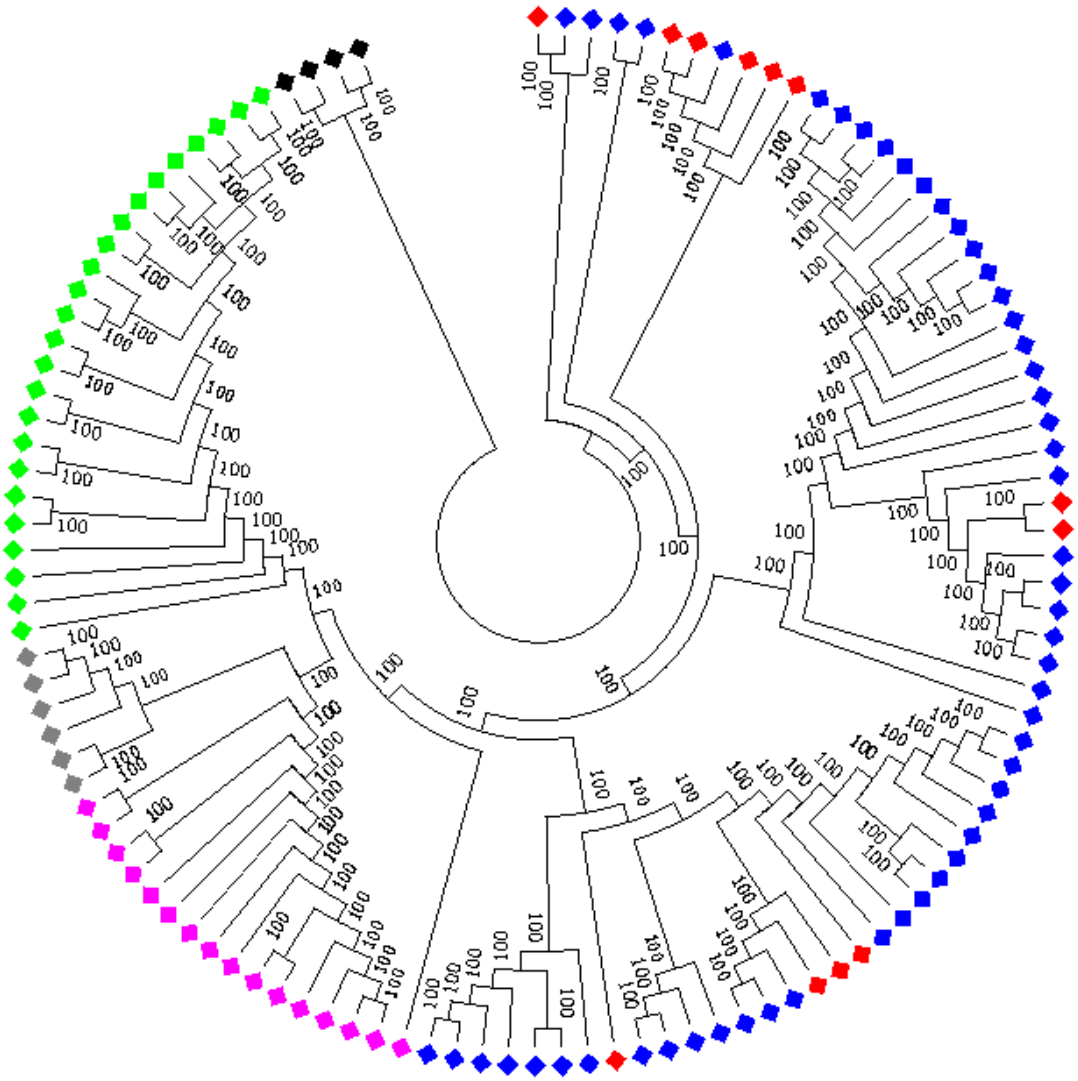


Figure S1: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 9460001-9480000 of Iraqi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

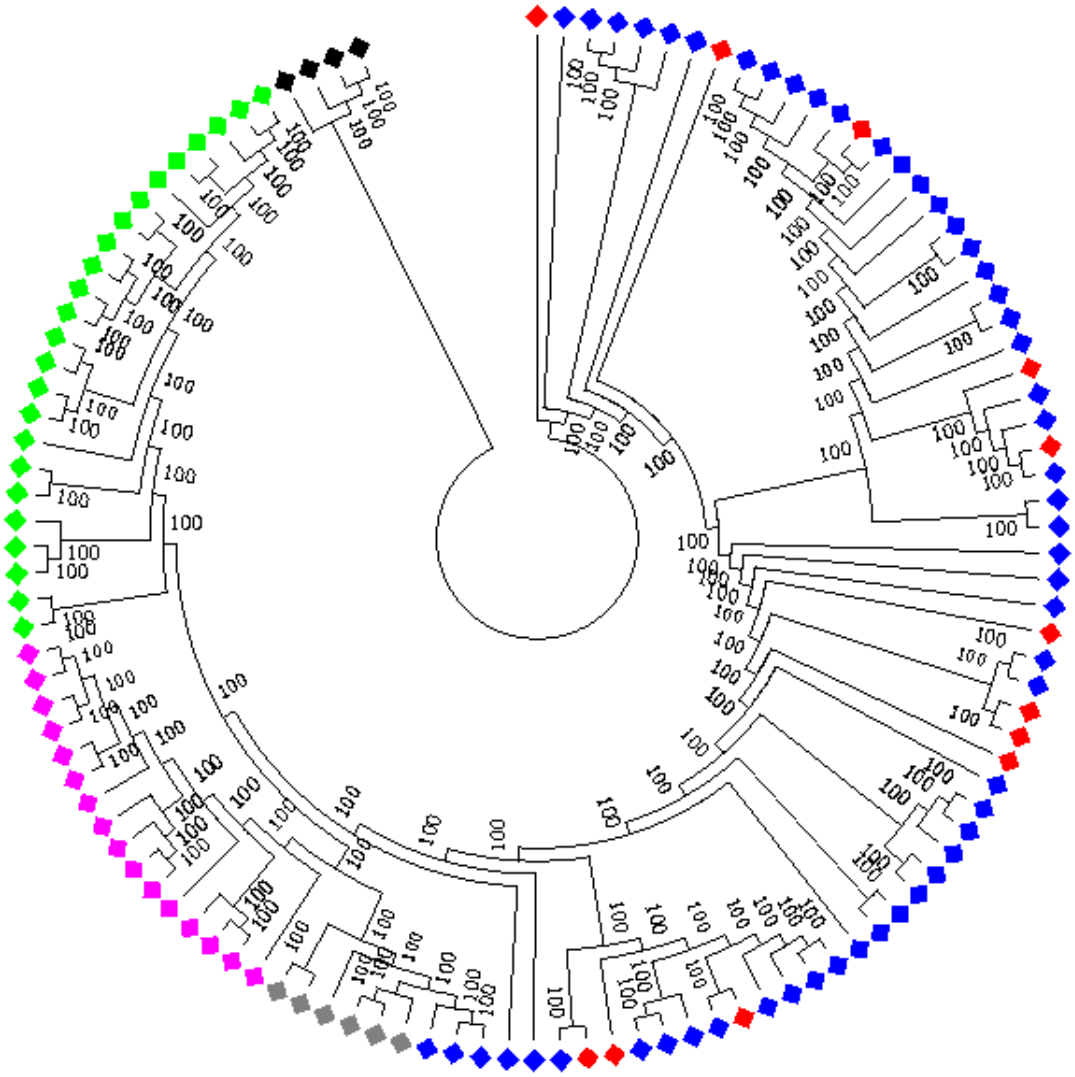


Figure S2: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 101080001-101100000 of Iraqi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

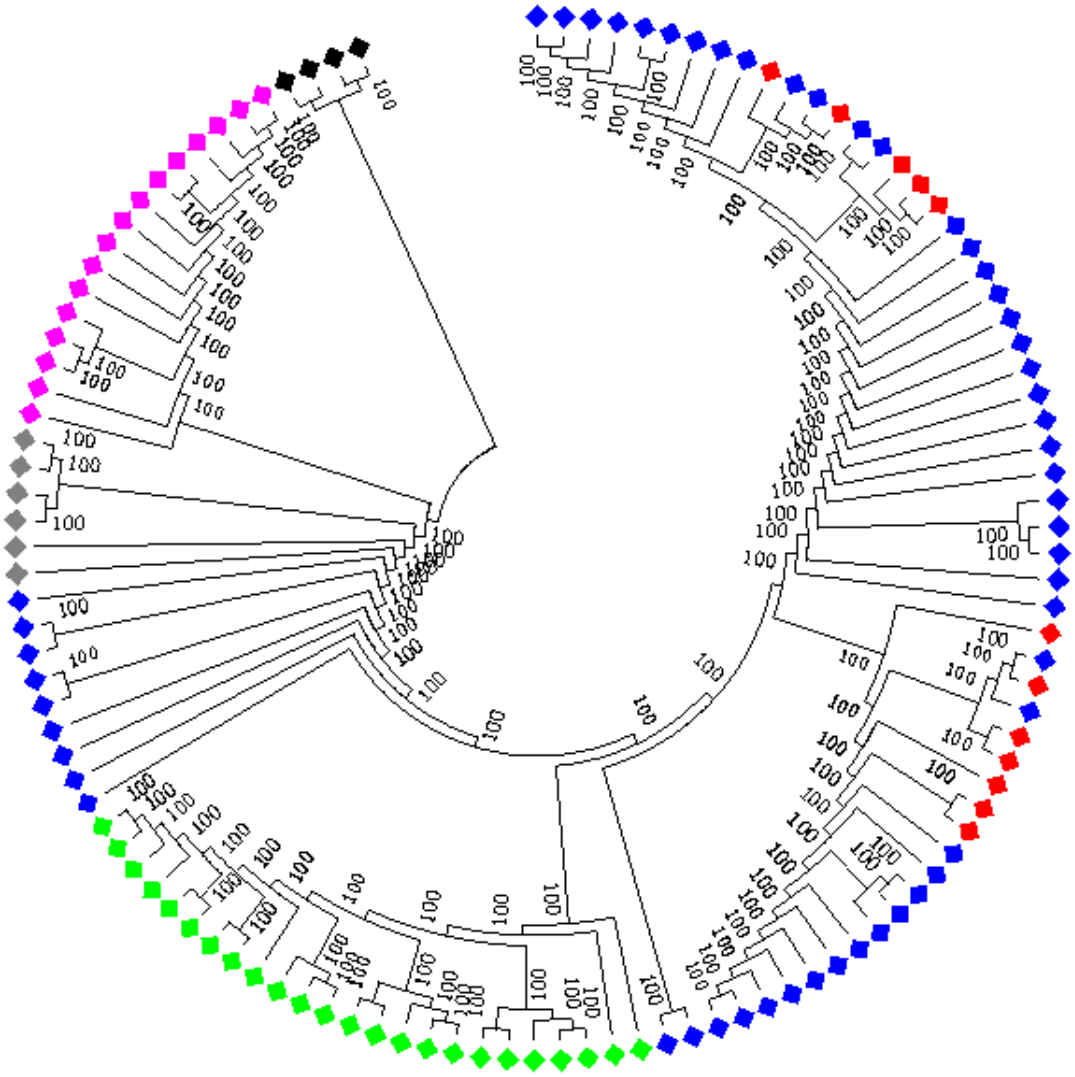


Figure S3: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 101100001-101120000 of Iraqi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

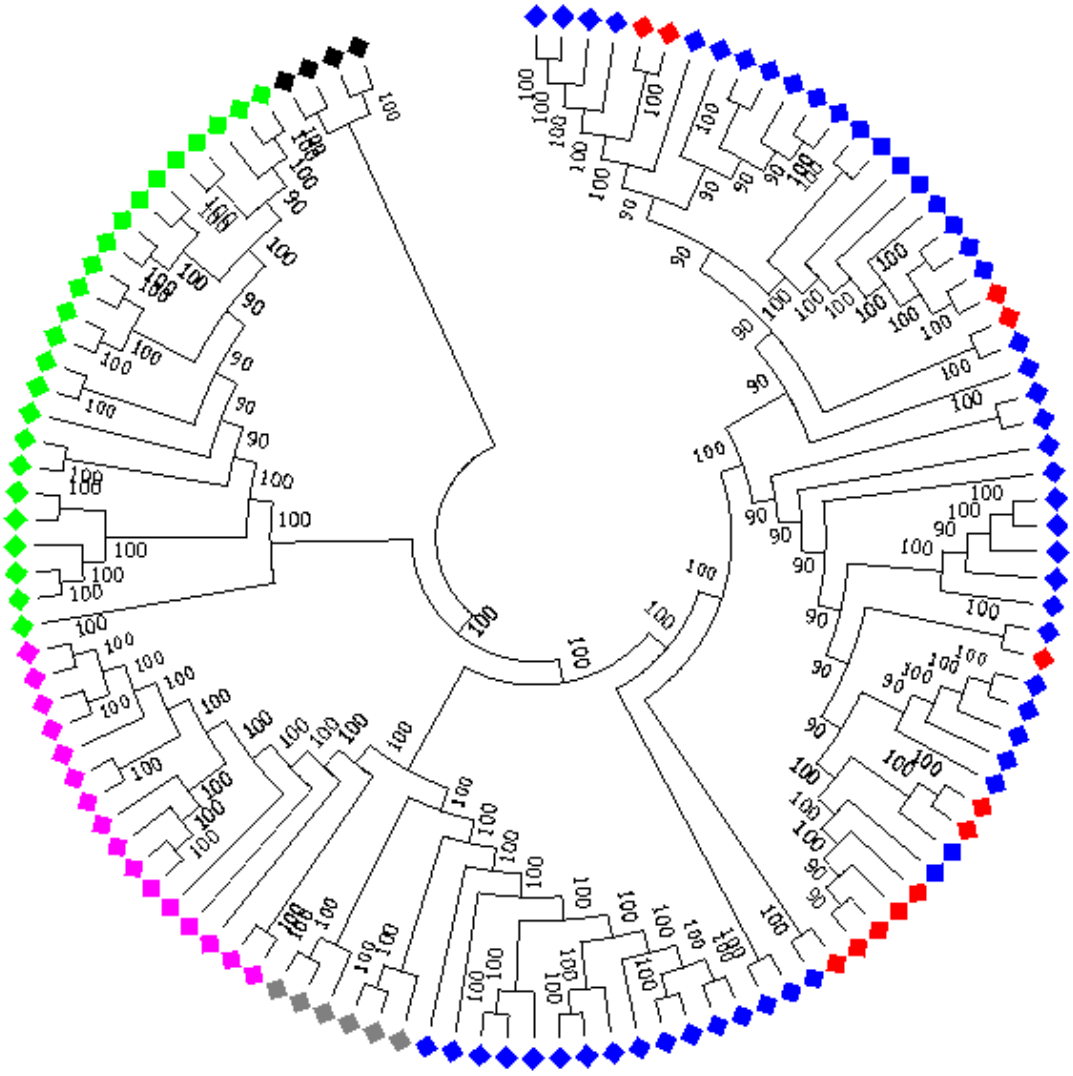


Figure S4: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 101120001-101140000 of Iraqi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

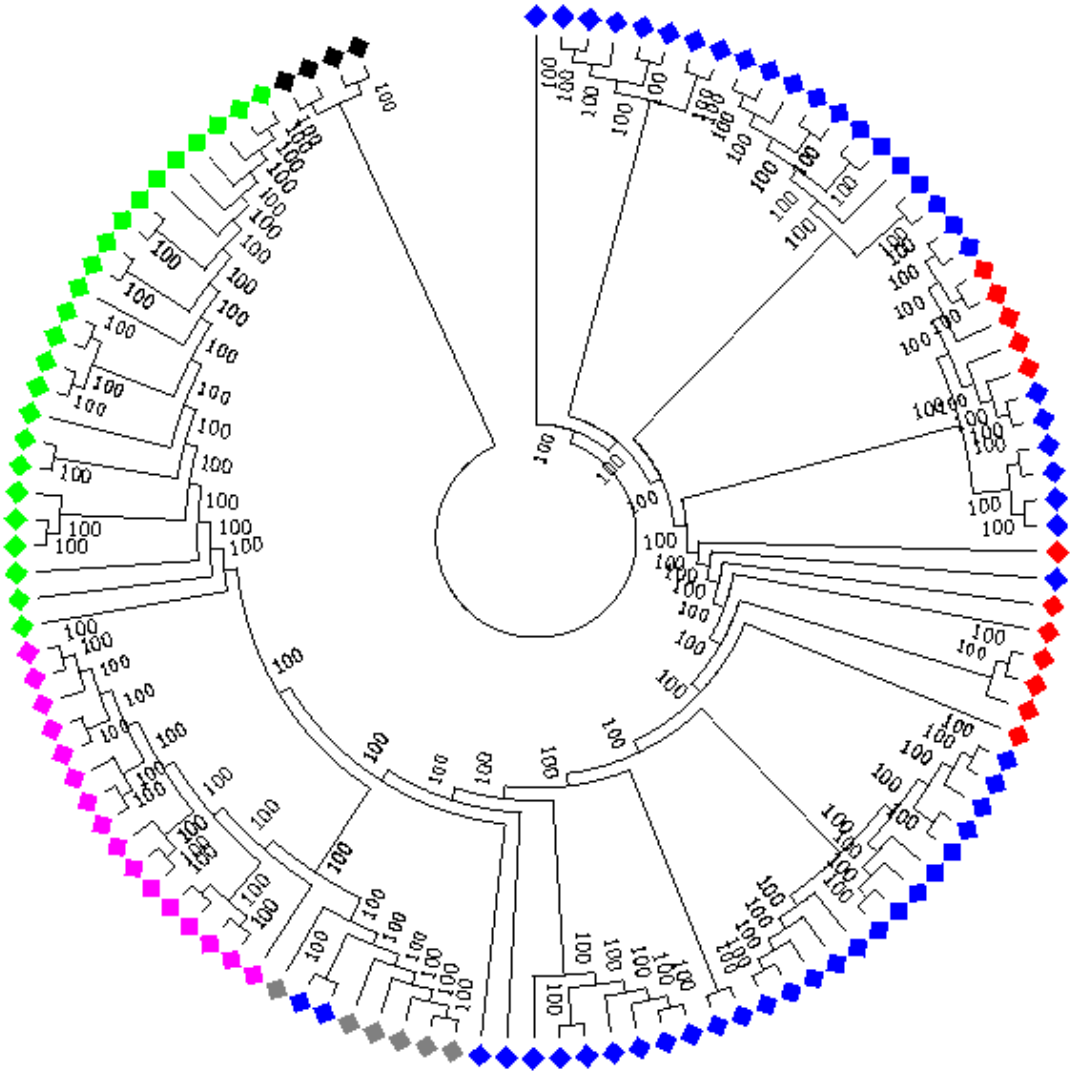


Figure S5: Haplotype maximum likelihood tree for the non-introgressed region on chr2: 57480001-57500000 of Iraqi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

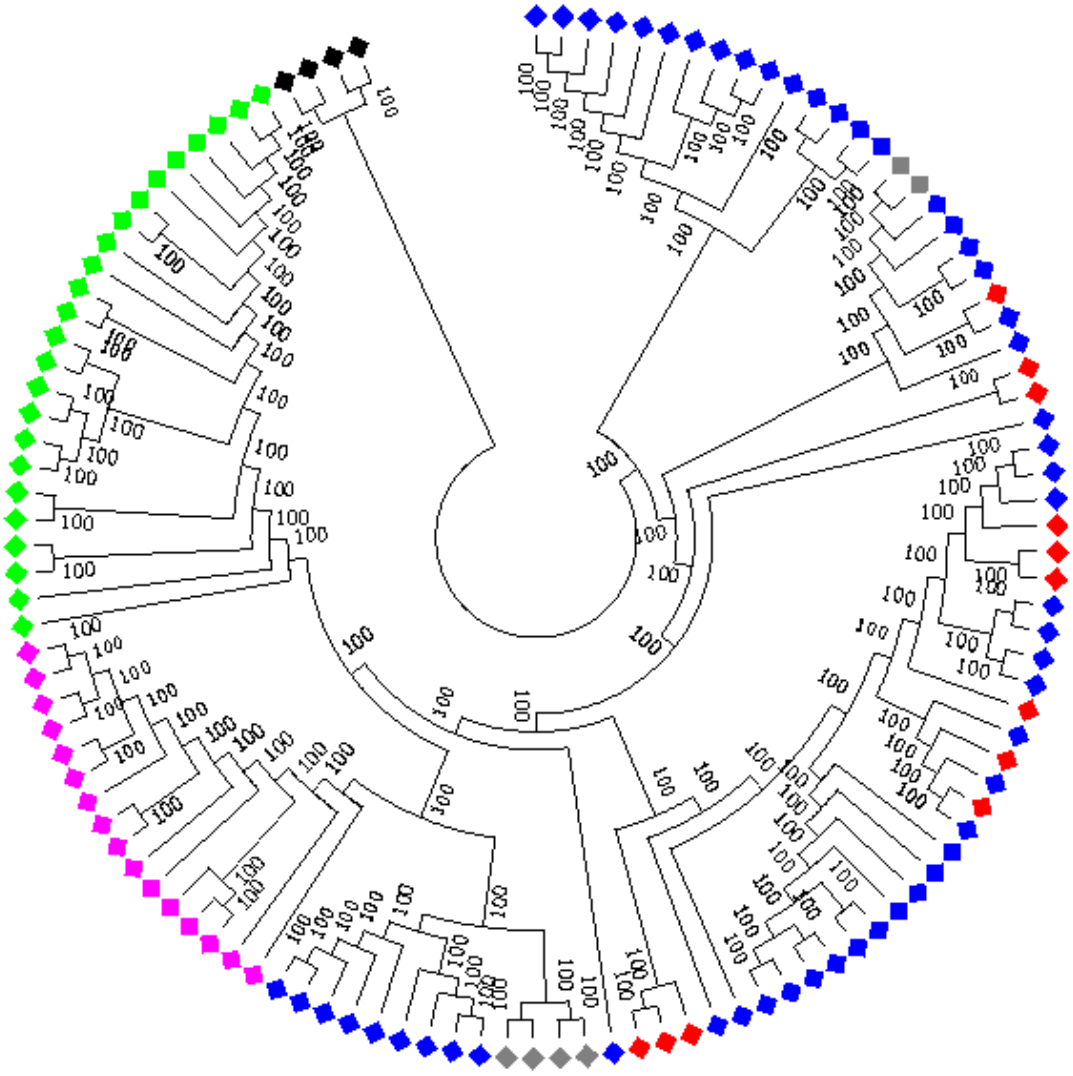


Figure S6: Haplotype maximum likelihood tree for the non-introgressed region on chr2: 57600001-57620000 of Iraqi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

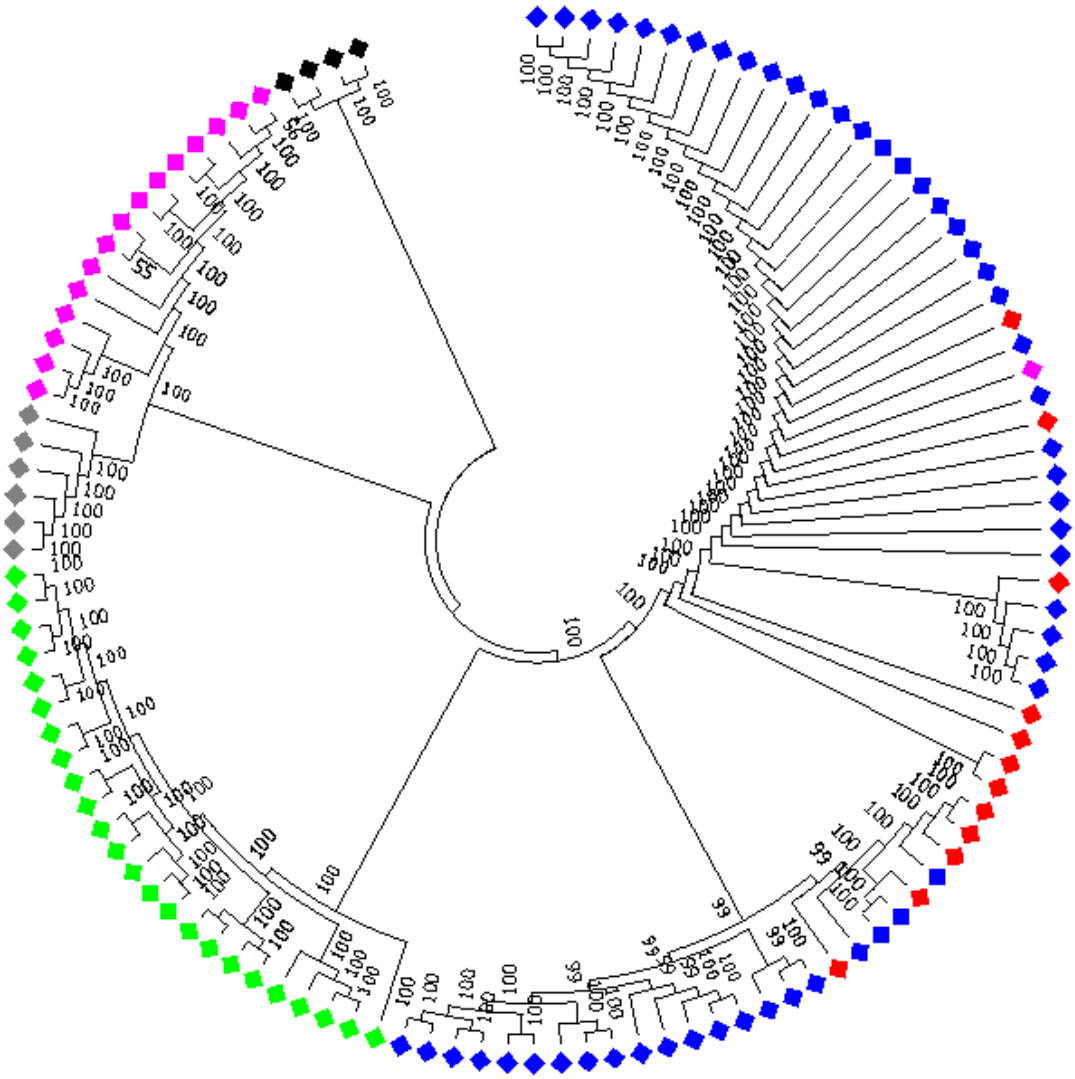


Figure S7: Haplotype maximum likelihood tree for the non-introgressed region on chr2: 79720001-79740000 of Iraqi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

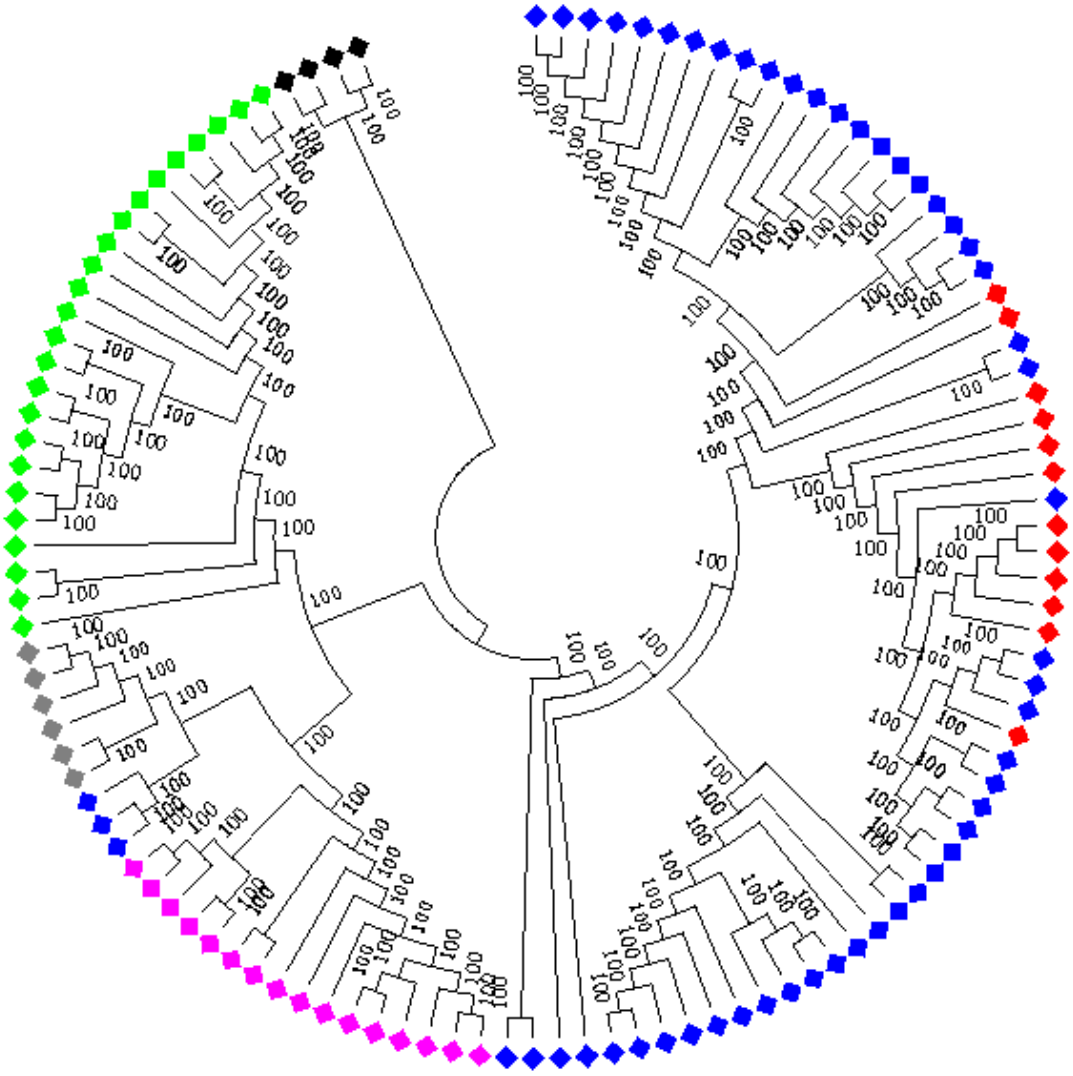


Figure S8: Haplotype maximum likelihood tree for the non-introgressed region on chr3: 70780001-70800000 of Iraqi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

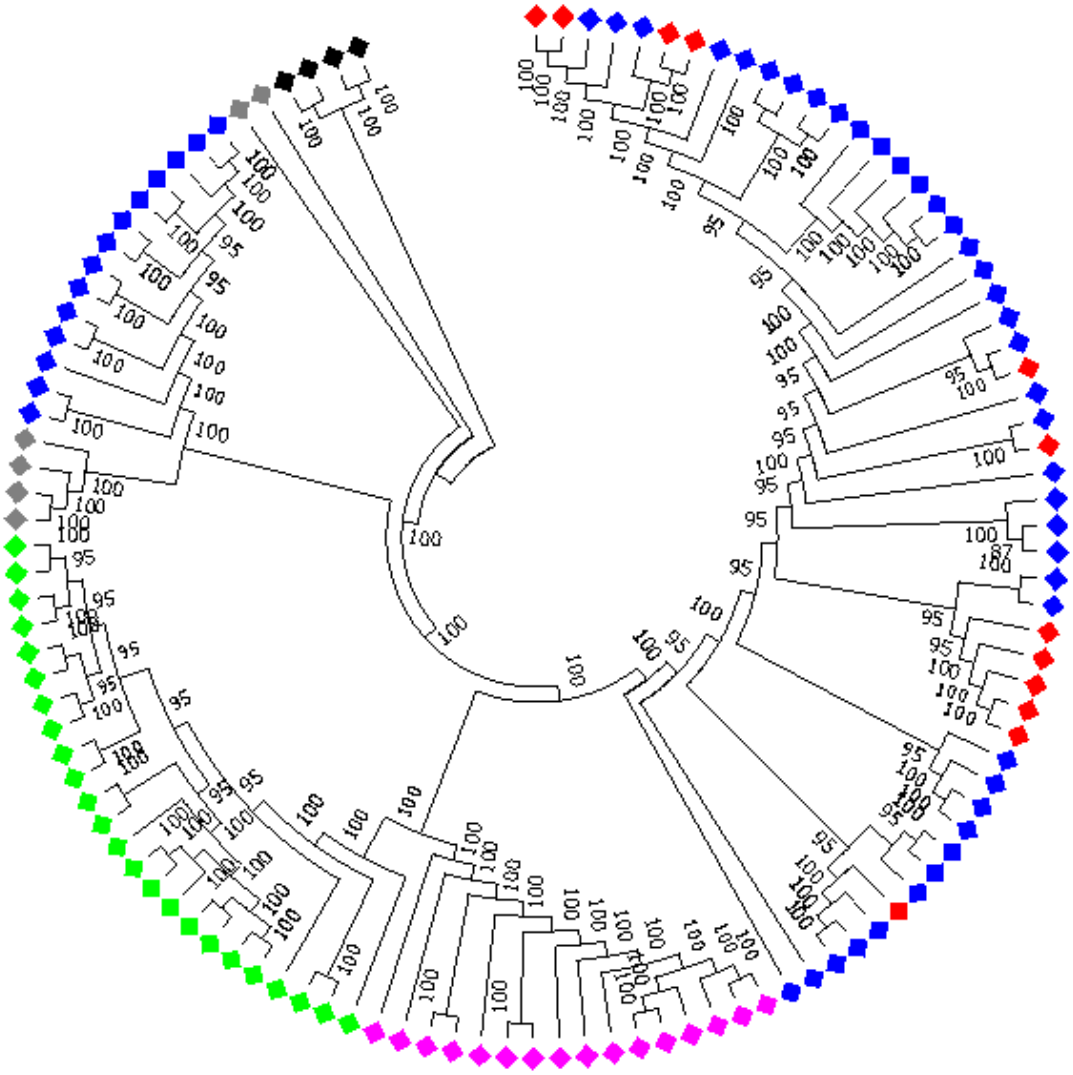


Figure S9: Haplotype maximum likelihood tree for the non-introgressed region on chr4: 22840001-22860000 of Iraqi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

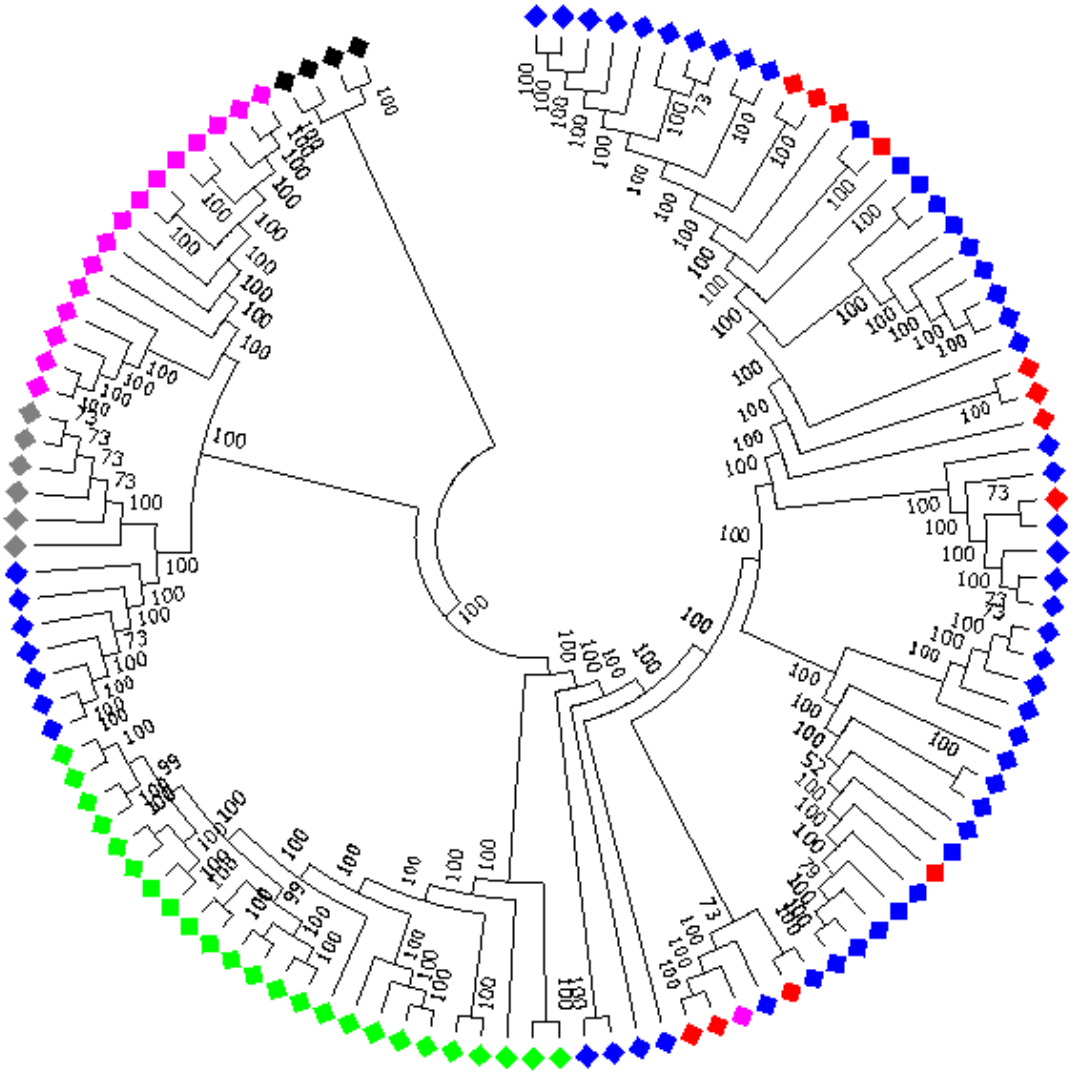


Figure S10: Haplotype maximum likelihood tree for the non-introgressed region on chr4: 26720001-26740000 of Iraqi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

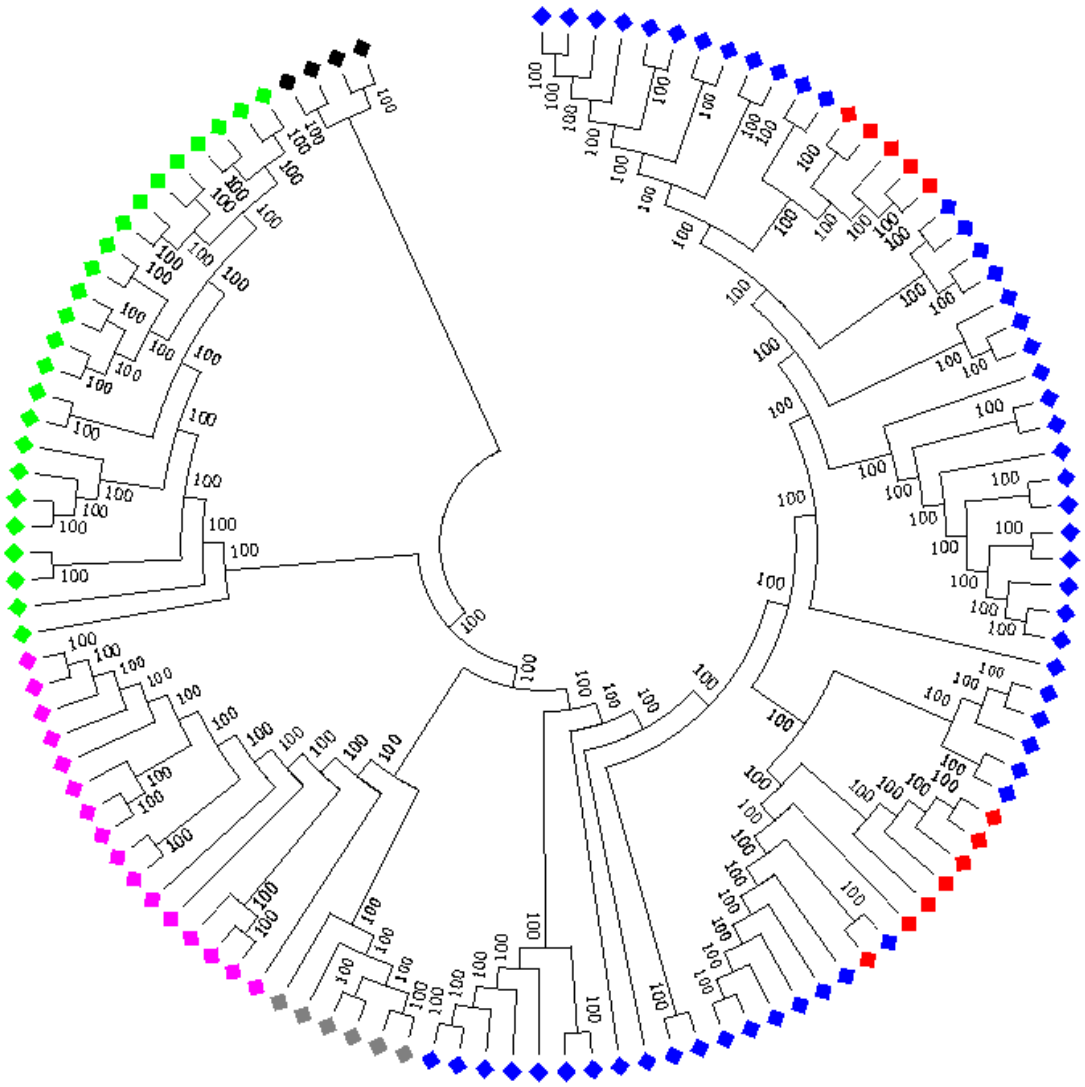


Figure S11: Haplotype maximum likelihood tree for the non-introgressed region on chr4: 26840001-26859999 of Iraqi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

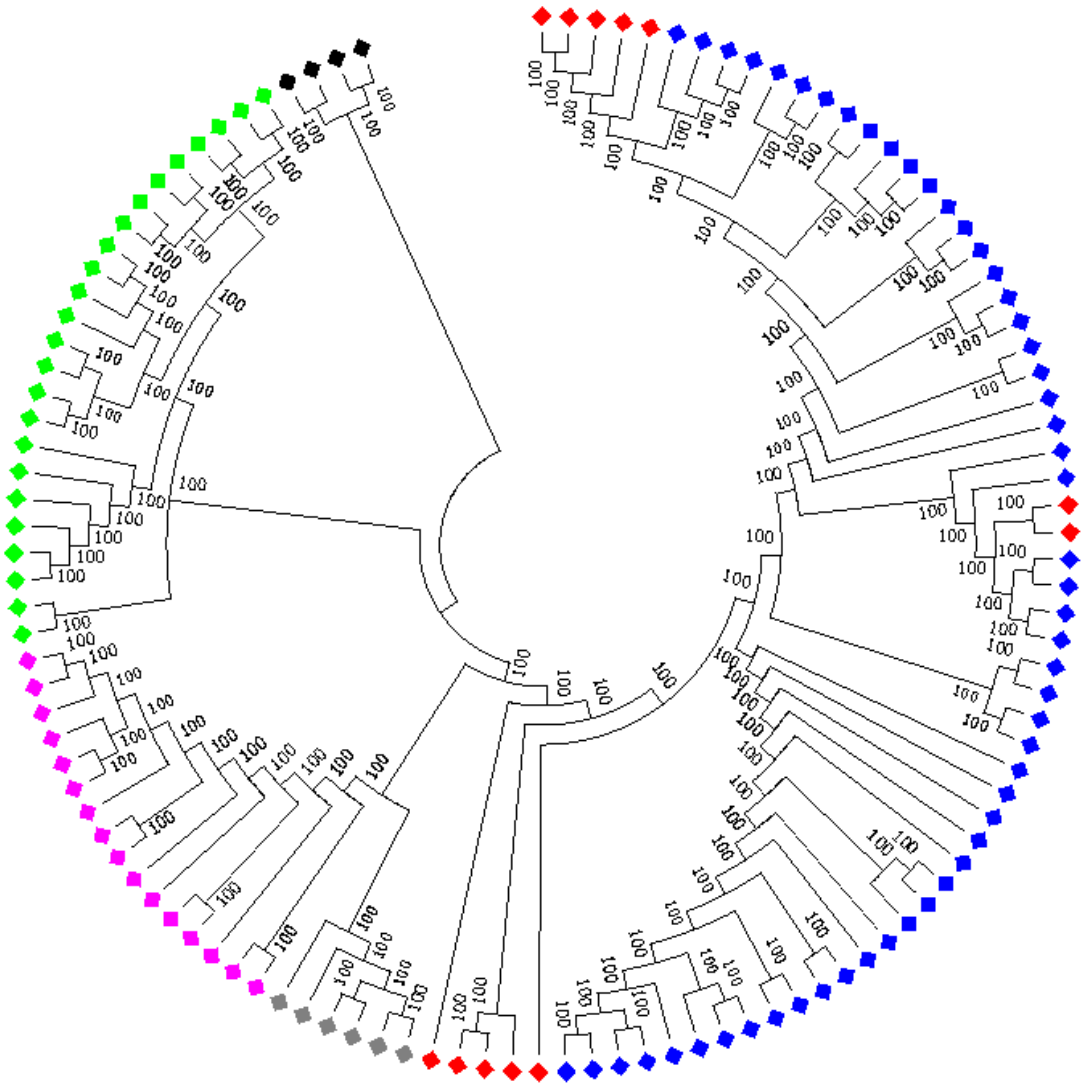


Figure S12: Haplotype maximum likelihood tree for the non-introgressed region on chr4: 73160001-73180000 of Iraqi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

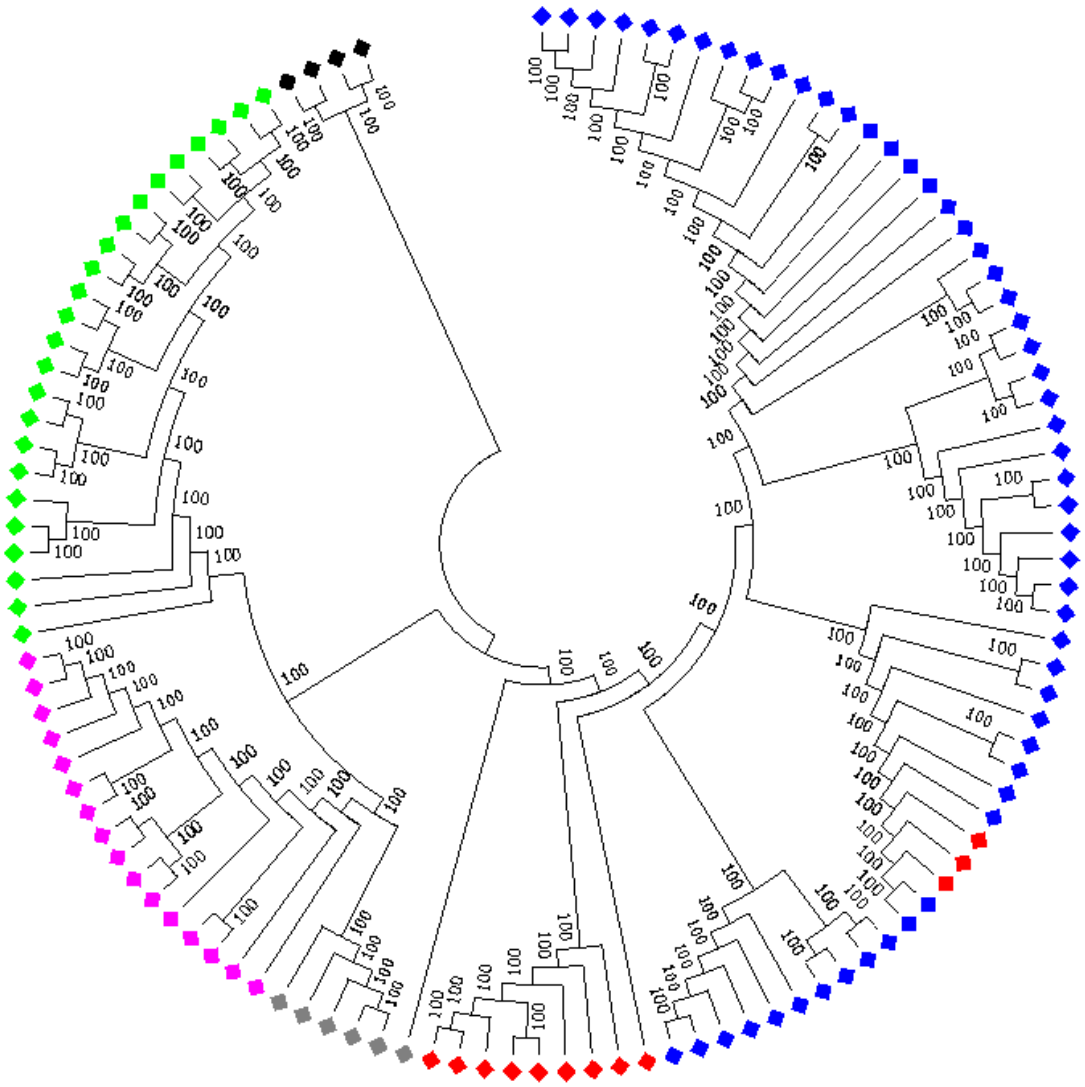


Figure S13: Haplotype maximum likelihood tree for the non-introgressed region on chr11: 3980001-4000000 of Iraqi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

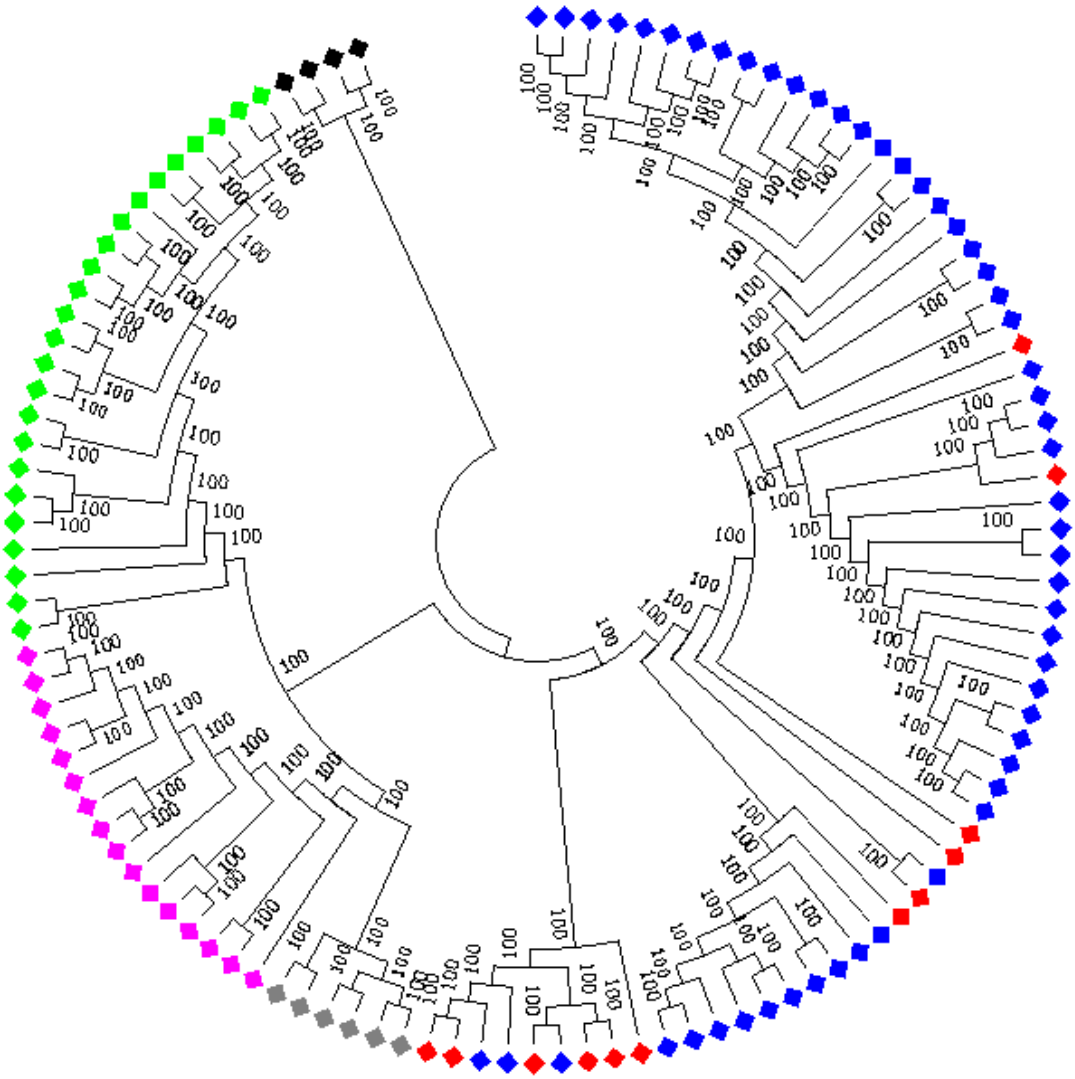


Figure S14: Haplotype maximum likelihood tree for the non-introgressed region on chr14: 14280001-14300000 of Iraqi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

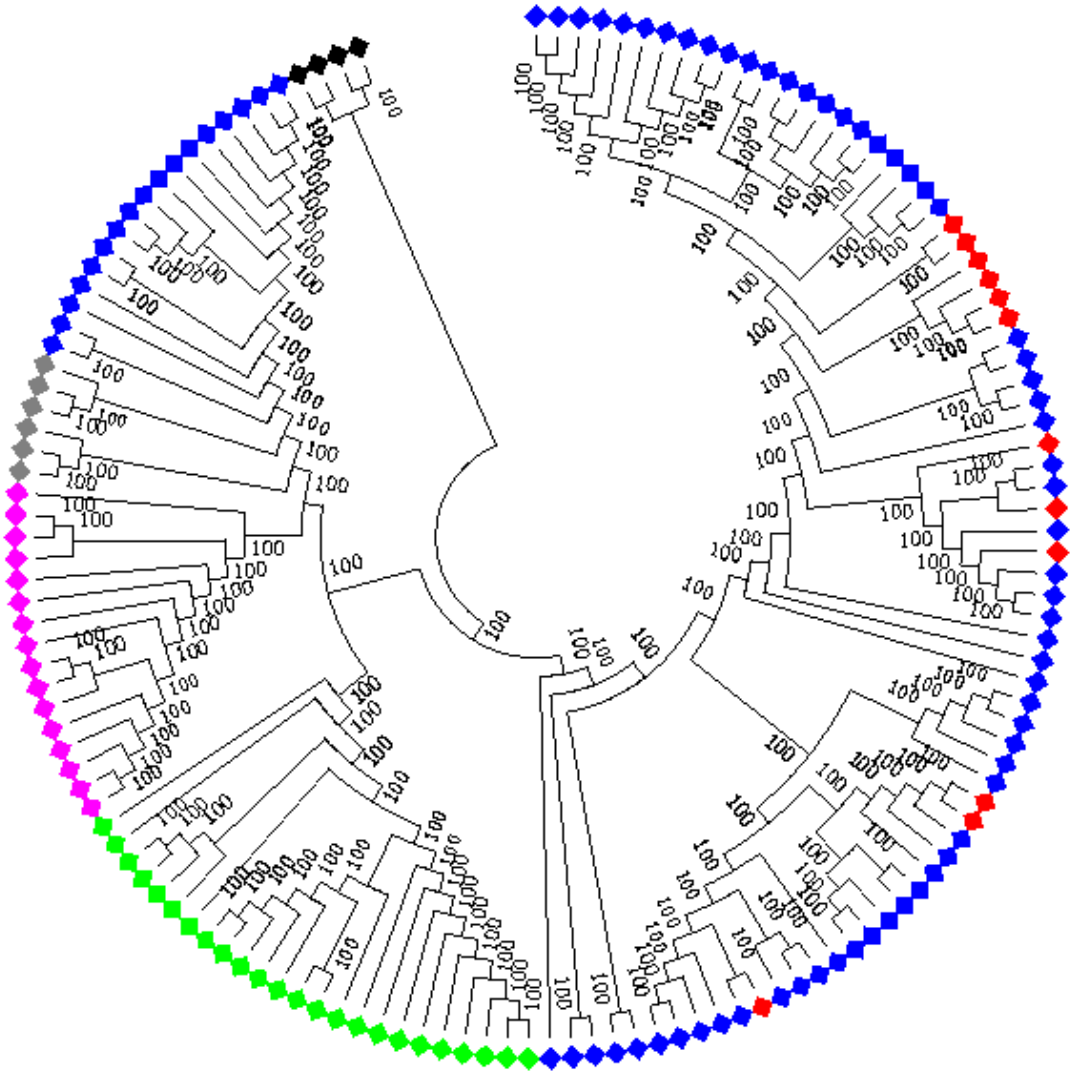


Figure S15: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 9020001-9040000 of Ethiopian domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

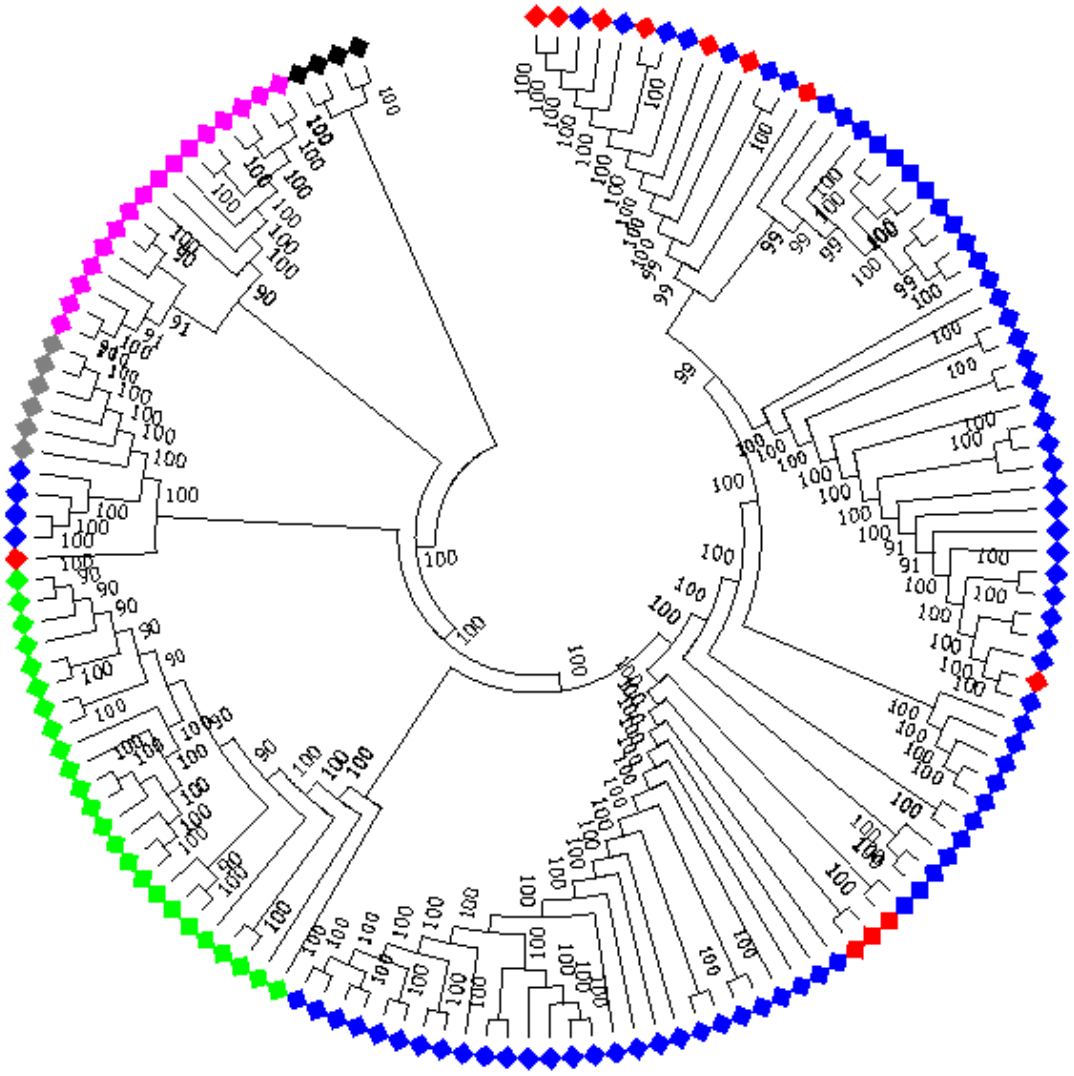


Figure S16: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 101300001-101320000 of Ethiopian domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

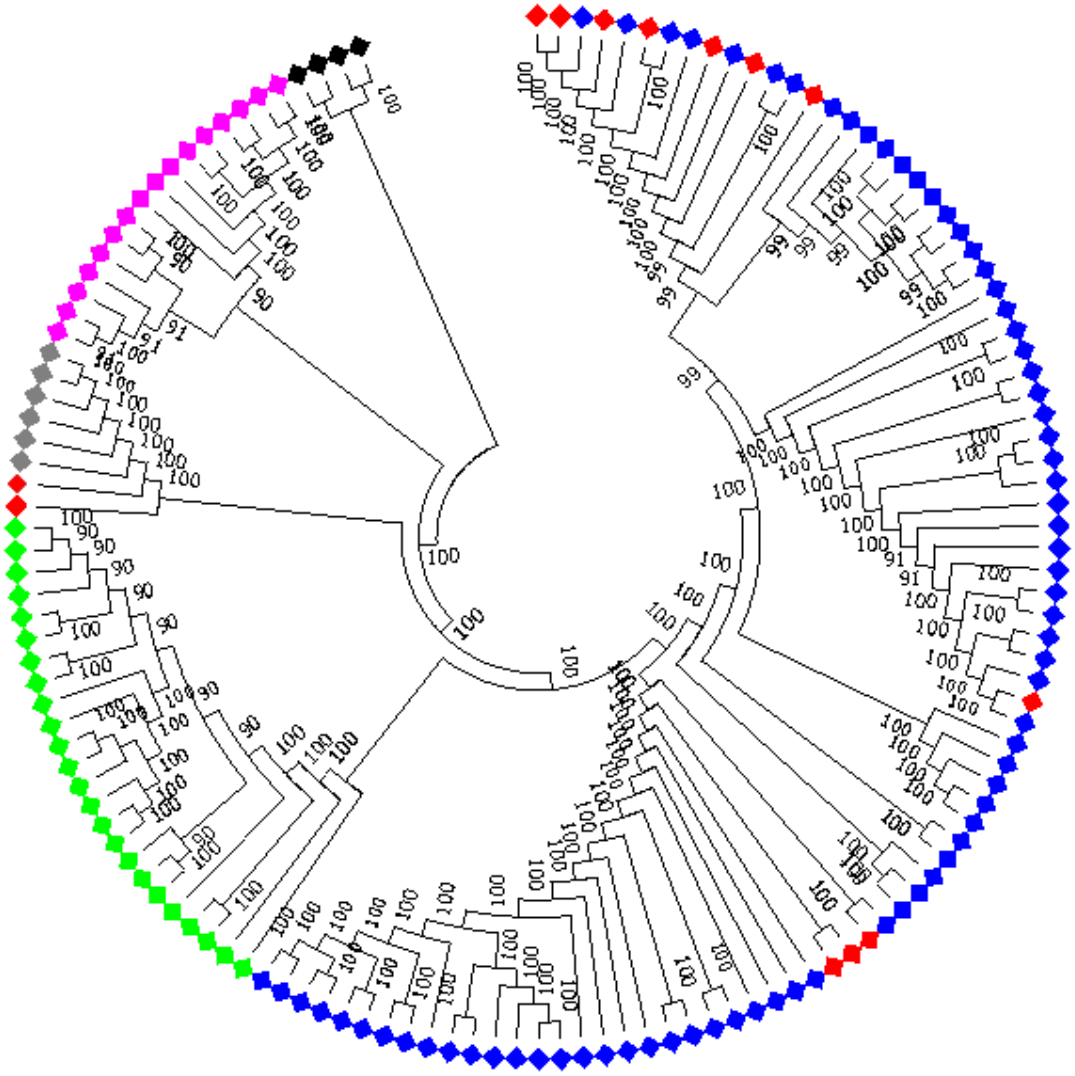


Figure S17: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 101420001-101440000 of Ethiopian domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

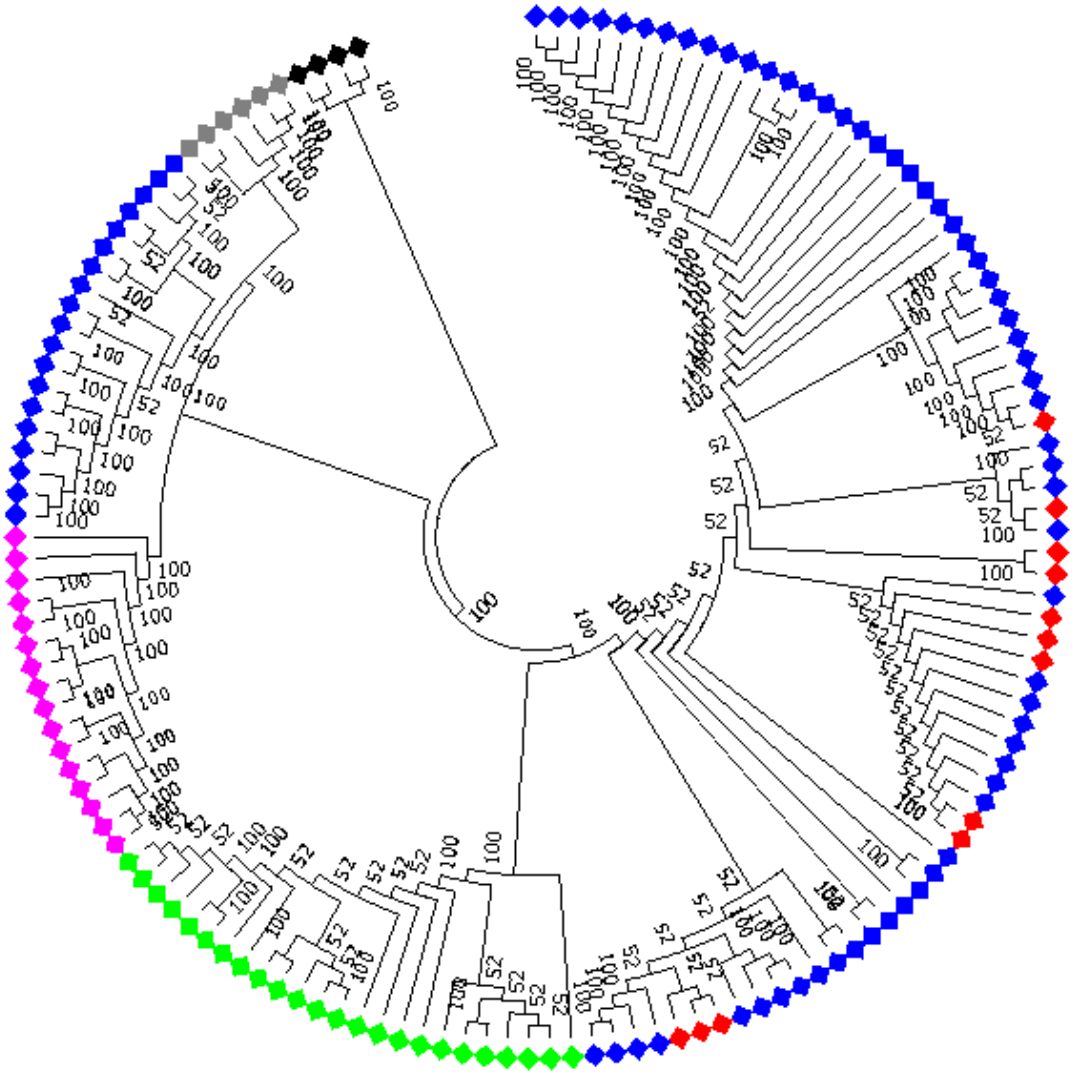


Figure S18: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 101440001-101460000 of Ethiopian domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

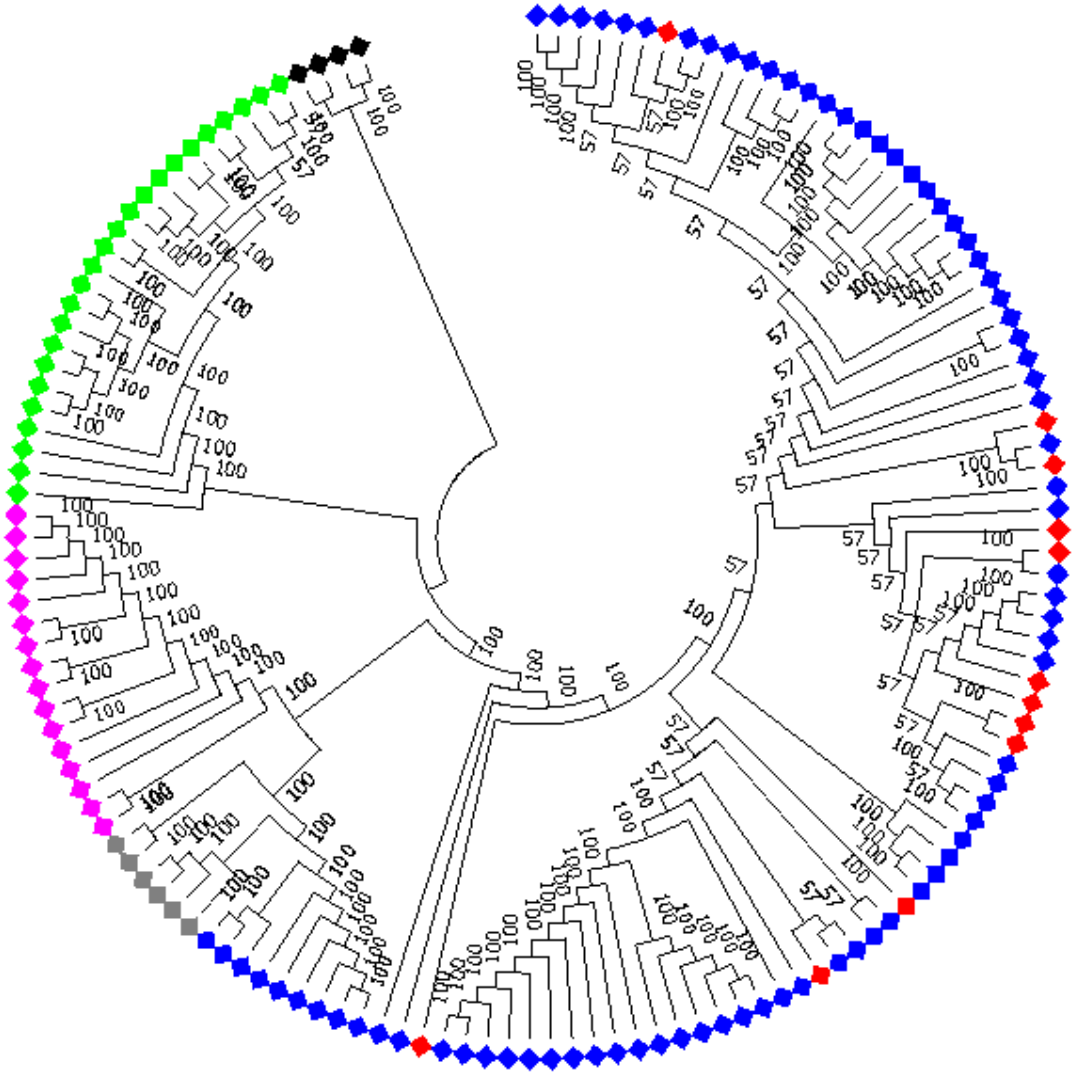


Figure S19: Haplotype maximum likelihood tree for the non-introgressed region on chr2: 72940001-72960000 of Ethiopian domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

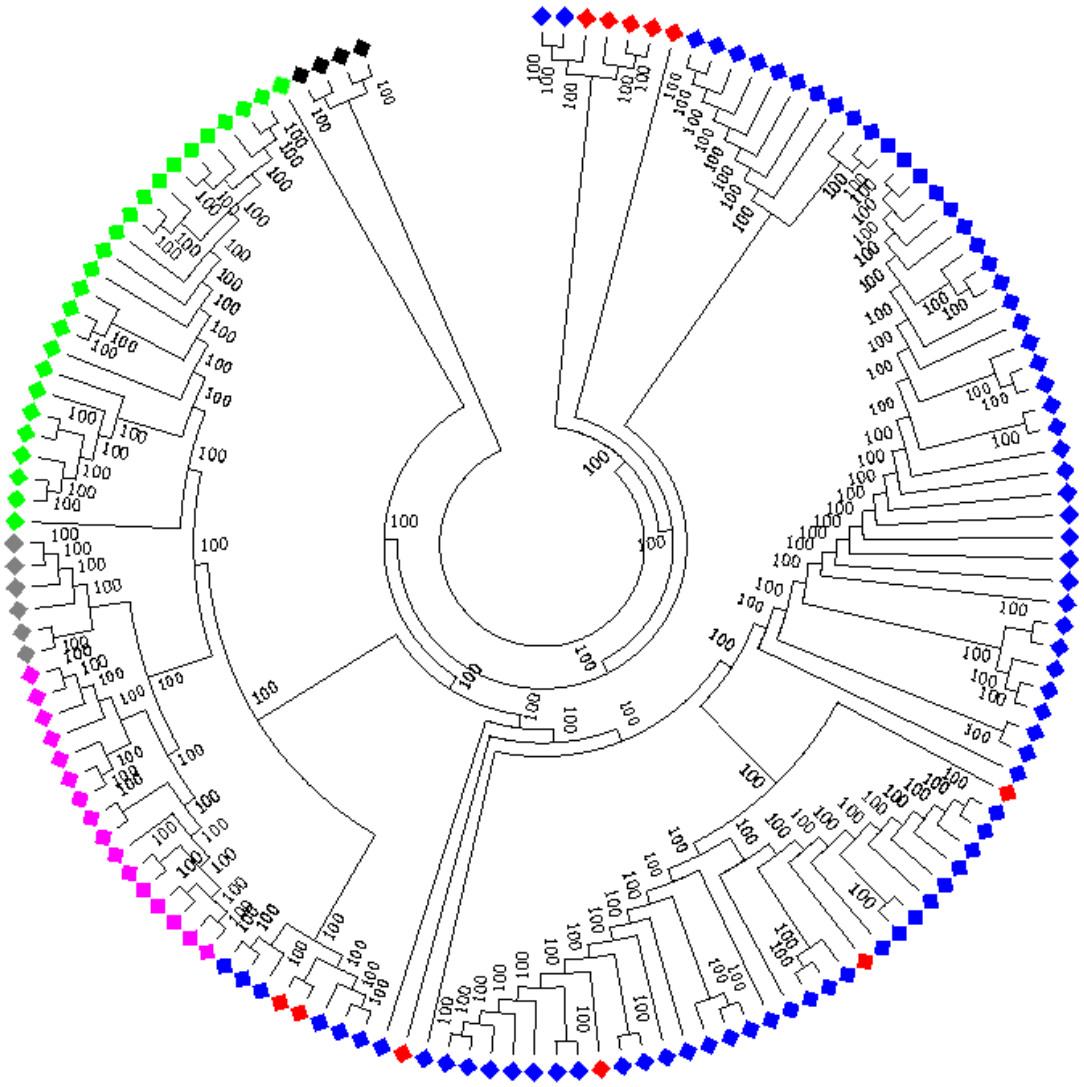


Figure S20: Haplotype maximum likelihood tree for the non-introgressed region on chr2: 72980001-73000000 of Ethiopian domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

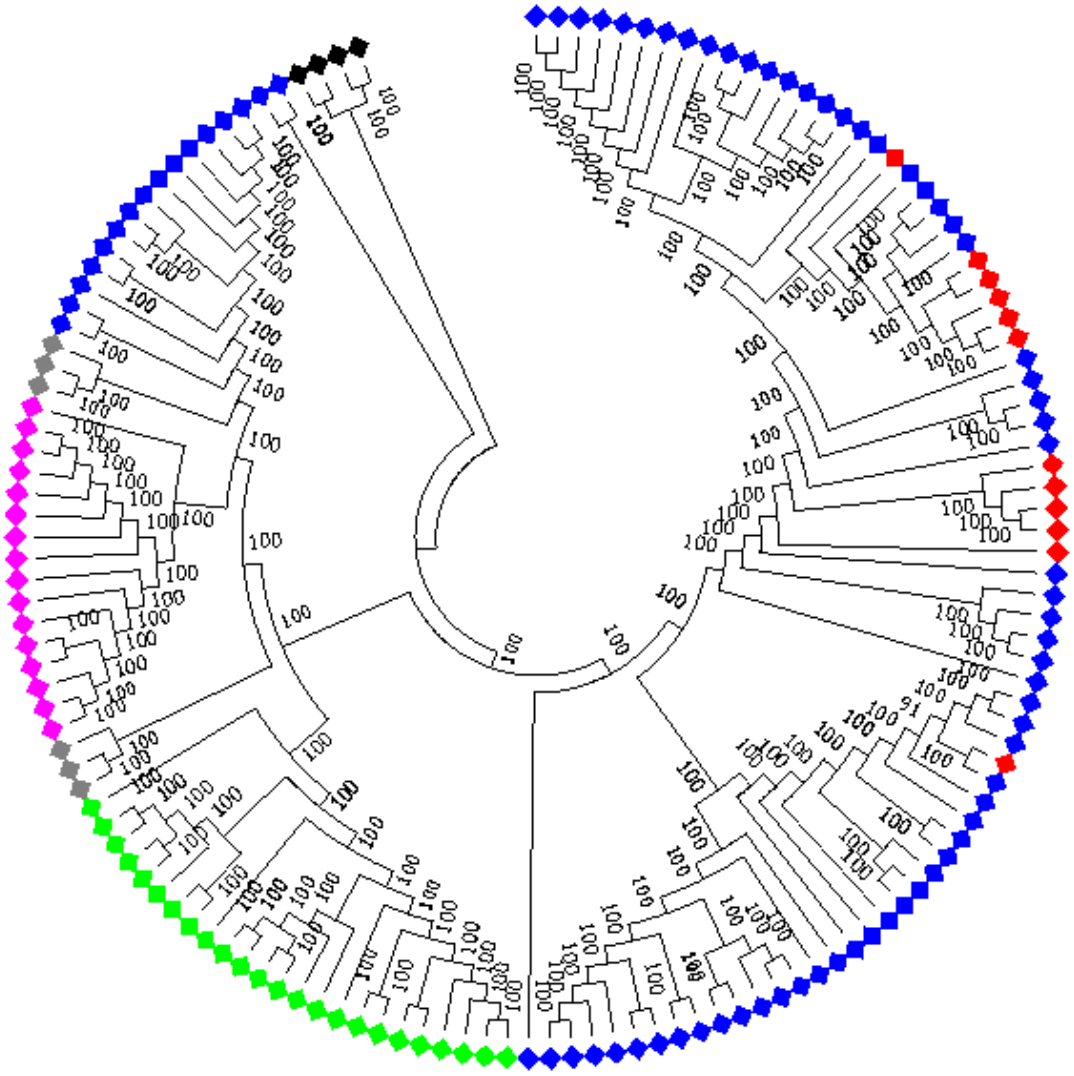


Figure S21: Haplotype maximum likelihood tree for the non-introgressed region on chr2: 149100001-149120000 of Ethiopian domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

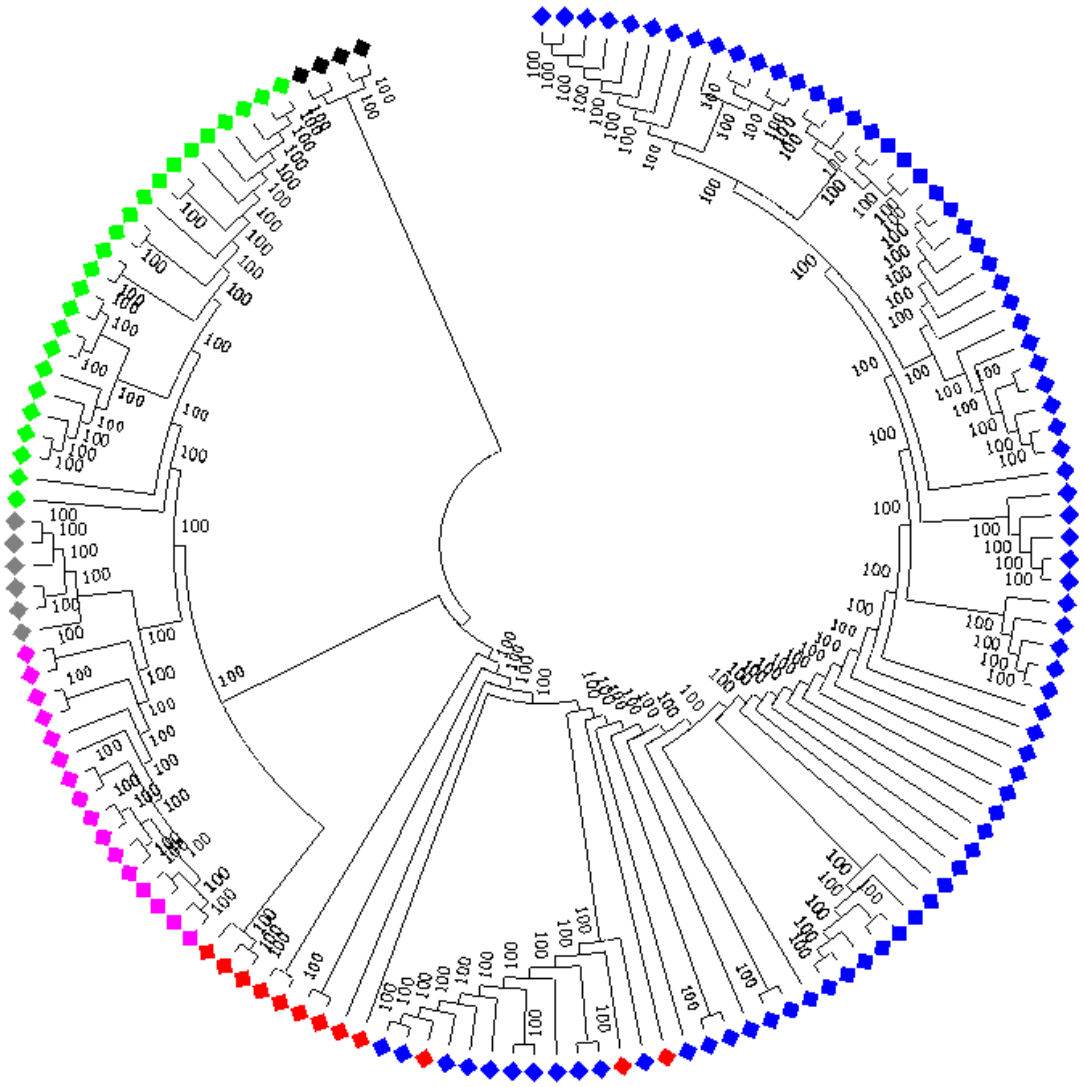


Figure S22: Haplotype maximum likelihood tree for the non-introgressed region on chr2: 149220001-149240000 of Ethiopian domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

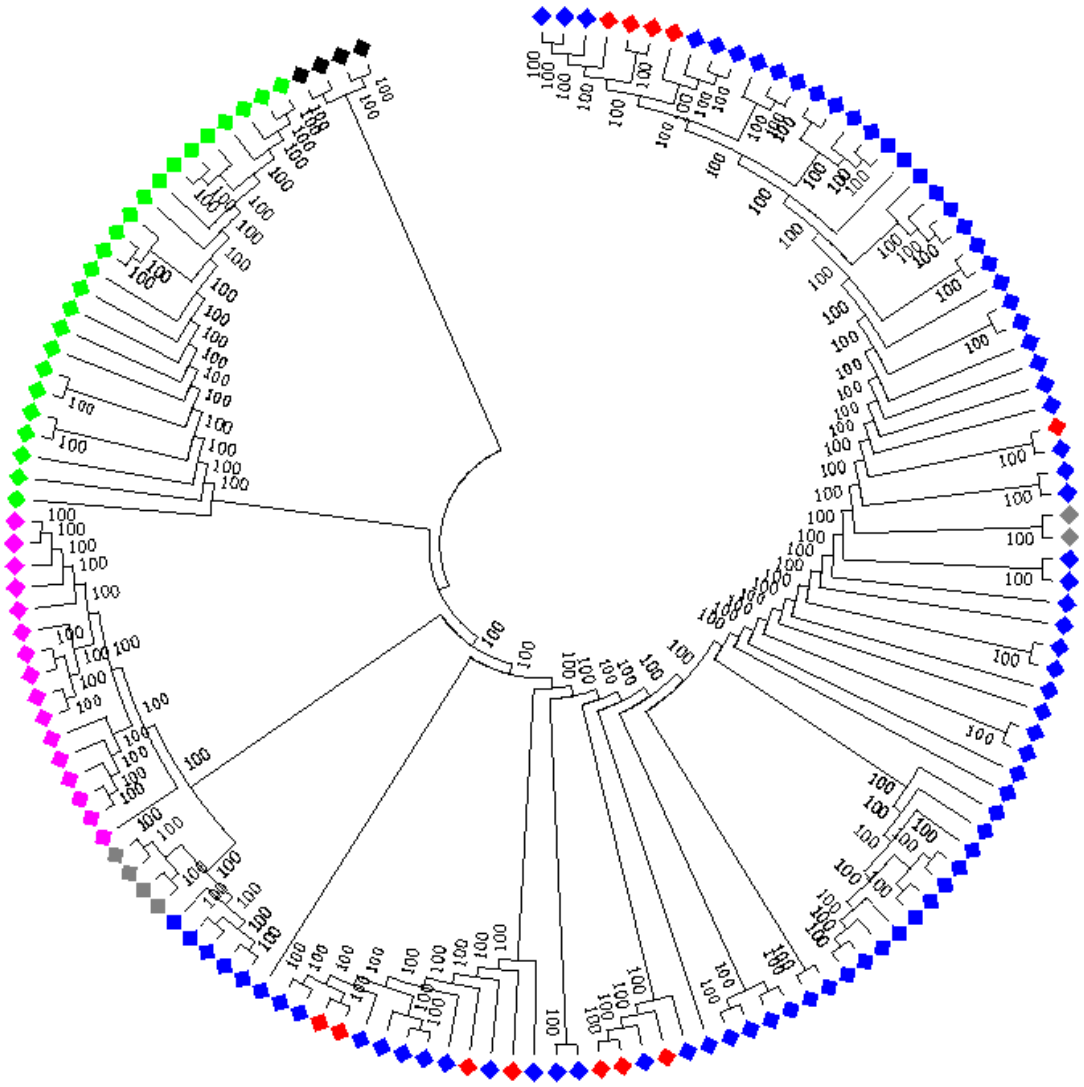


Figure S23: Haplotype maximum likelihood tree for the non-introgressed region on chr4: 47370001-47390000 of Ethiopian domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

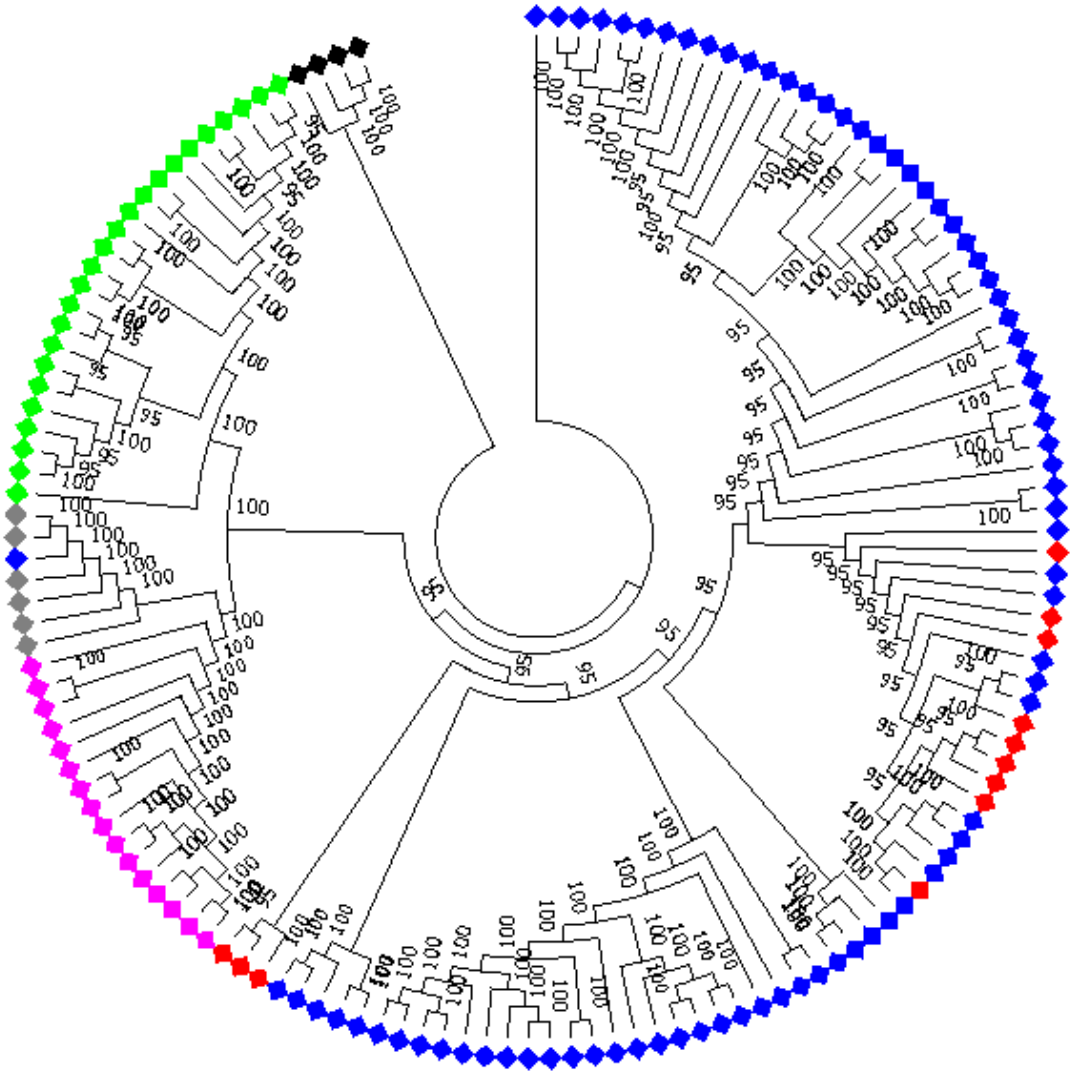


Figure S24: Haplotype maximum likelihood tree for the non-introgressed region on chr4: 47375001-47395000 of Ethiopian domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

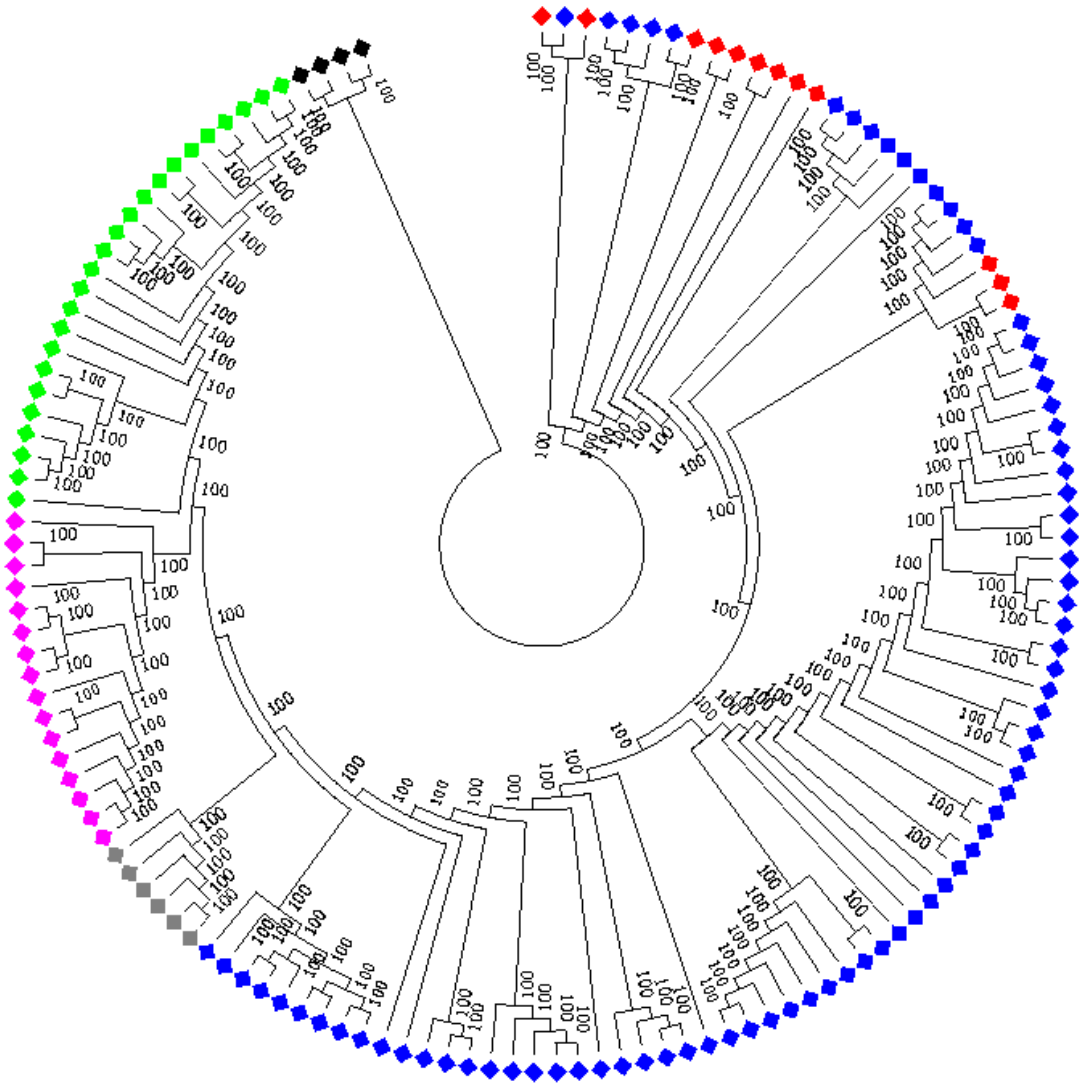


Figure S25: Haplotype maximum likelihood tree for the non-introgressed region on chr4: 47380001-47400000 of Ethiopian domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

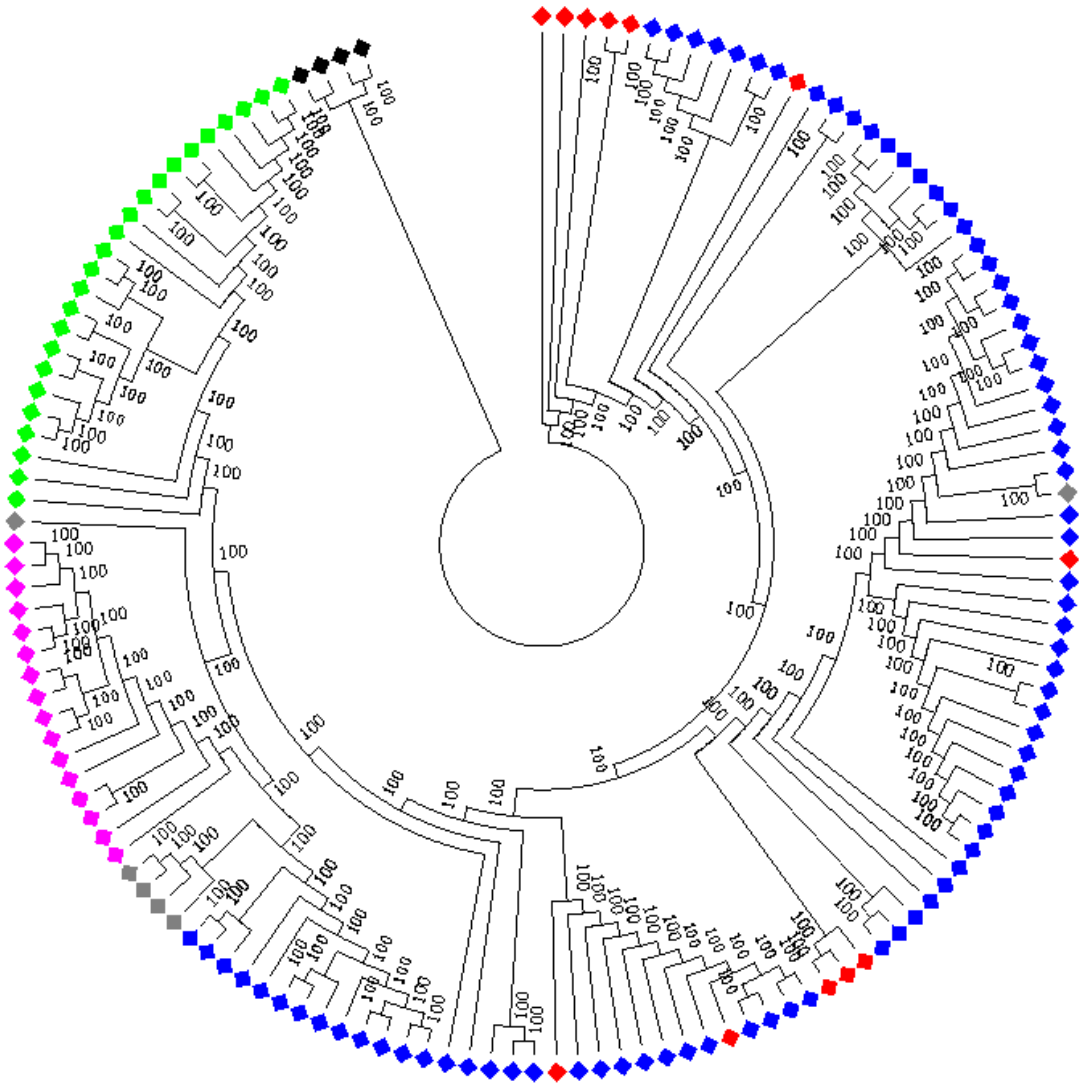


Figure S26: Haplotype maximum likelihood tree for the non-introgressed region on chr6: 21560001-21580000 of Ethiopian domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

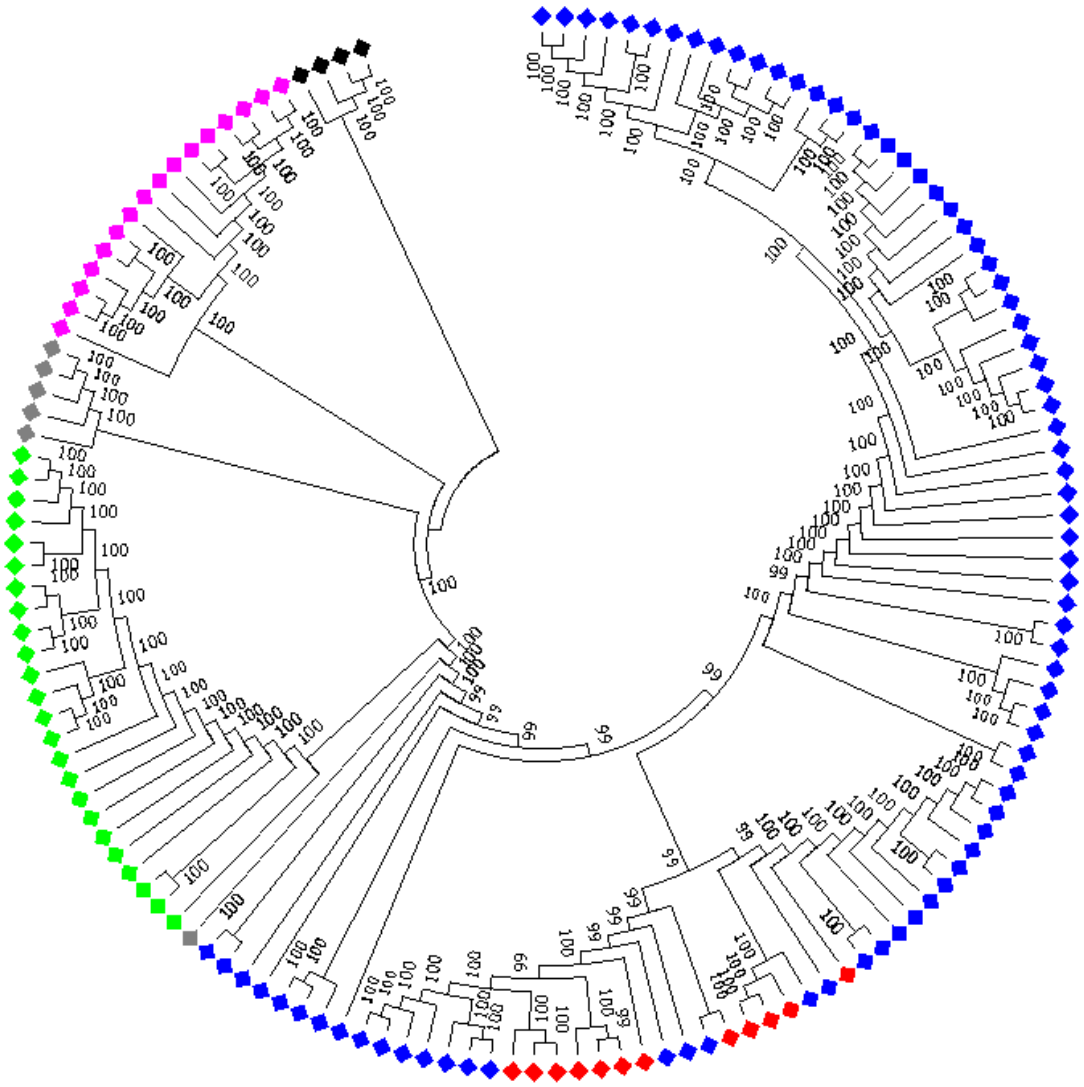


Figure S27: Haplotype maximum likelihood tree for the non-introgressed region on chr7: 21260001-21280000 of Ethiopian domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

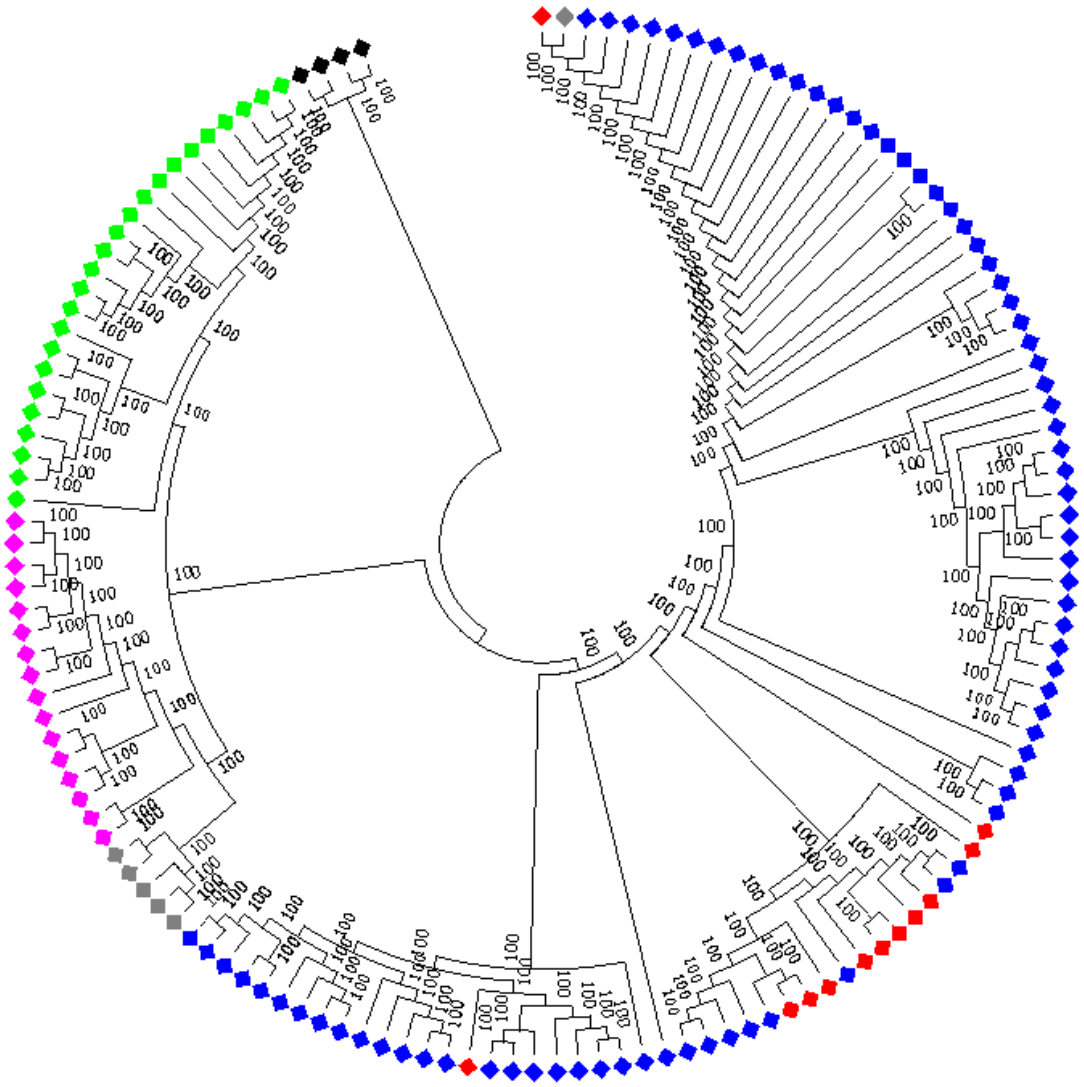


Figure S28: Haplotype maximum likelihood tree for the non-introgressed region on chr8: 18200001-18220000 of Ethiopian domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

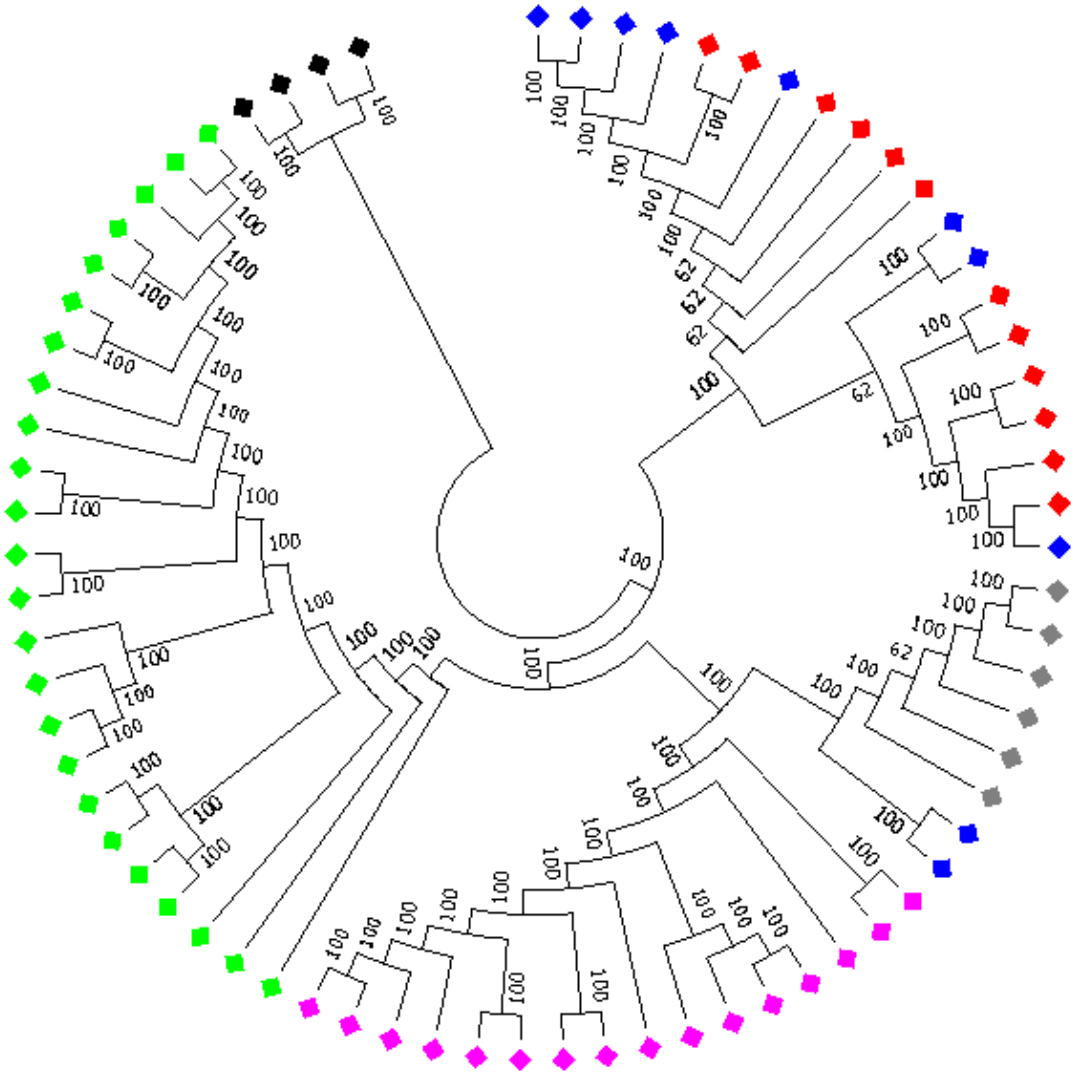


Figure S29: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 101080001-101100000 of Saudi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl.

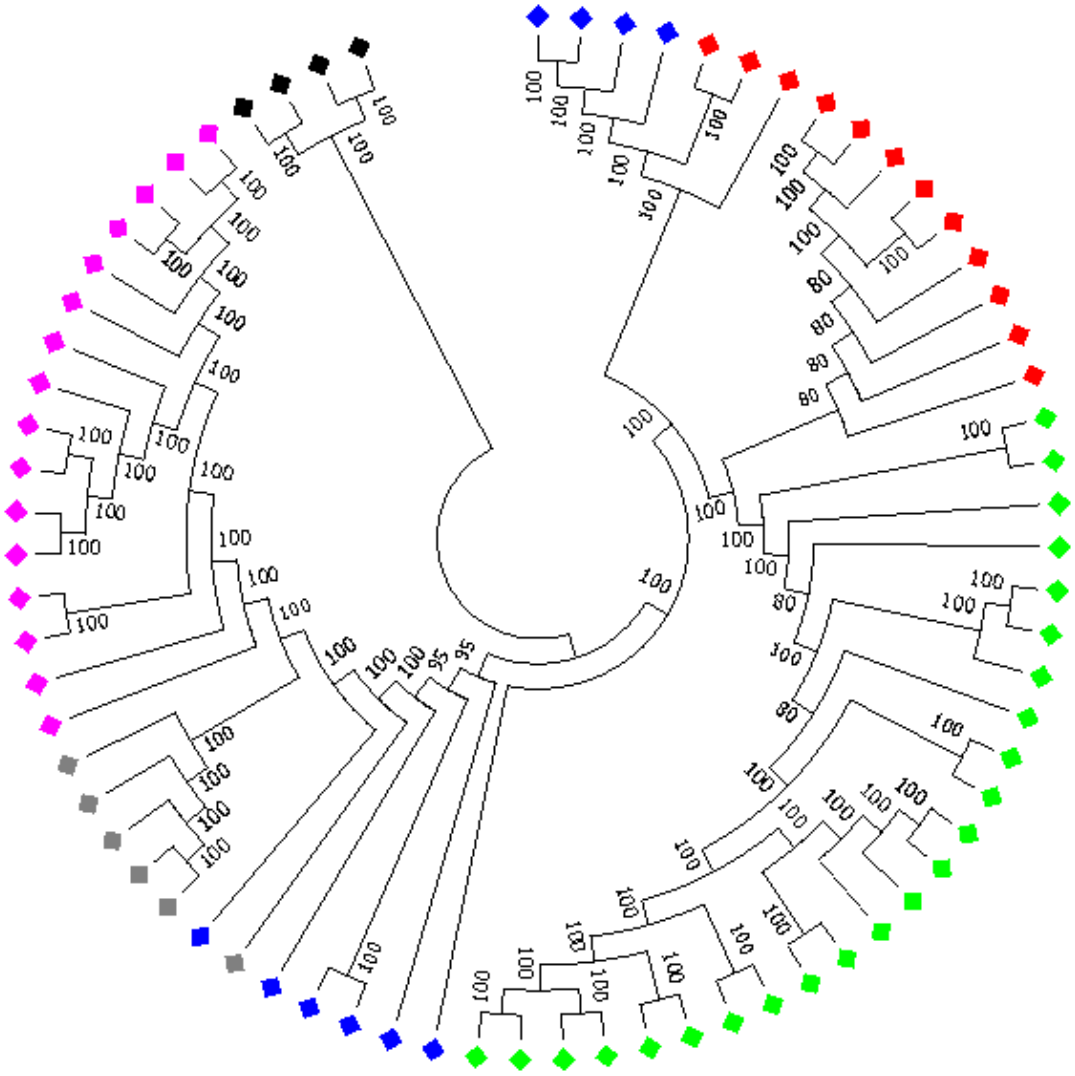


Figure S30: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 101220001-101240000 of Saudi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl.

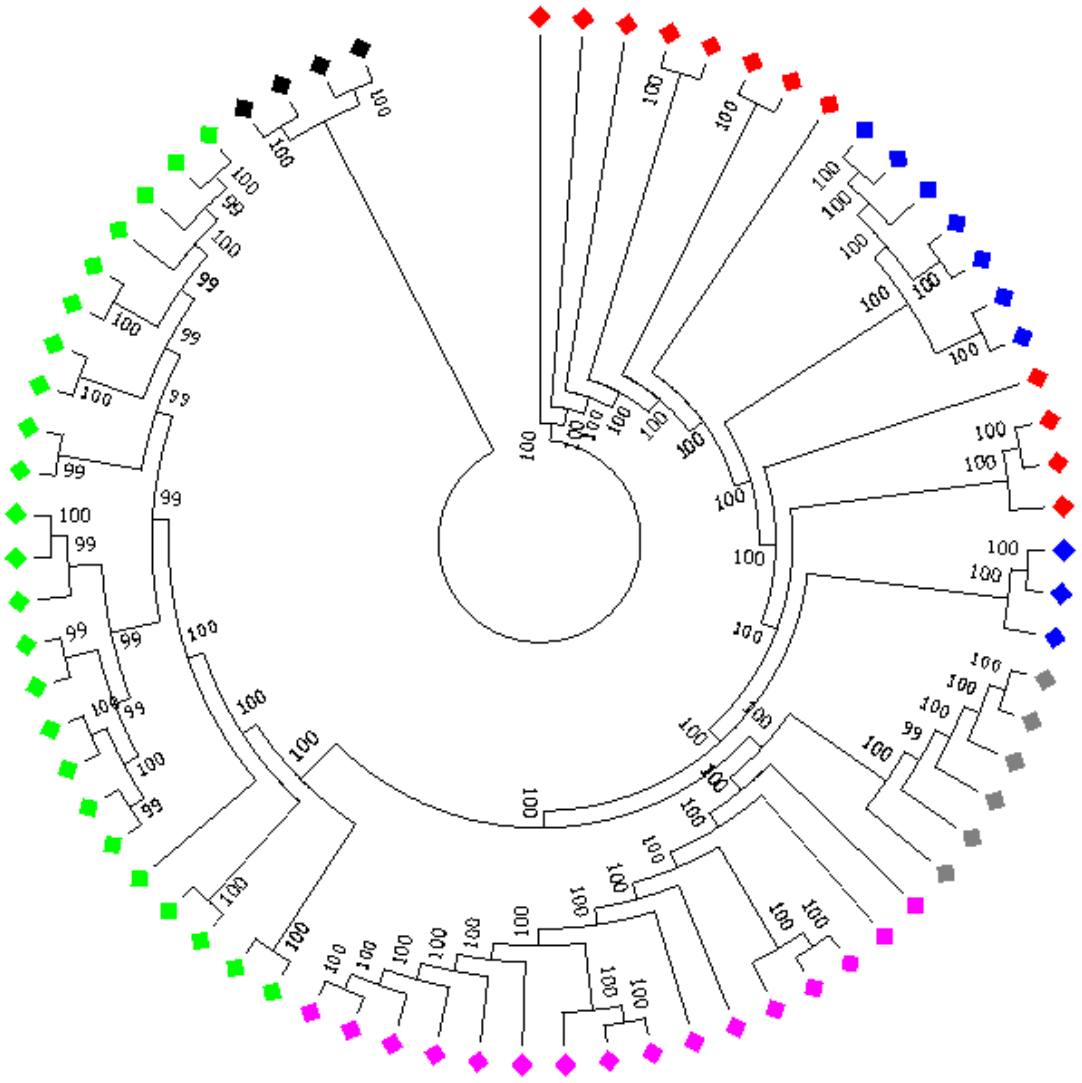


Figure S31: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 158020001-158040000 of Saudi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl.

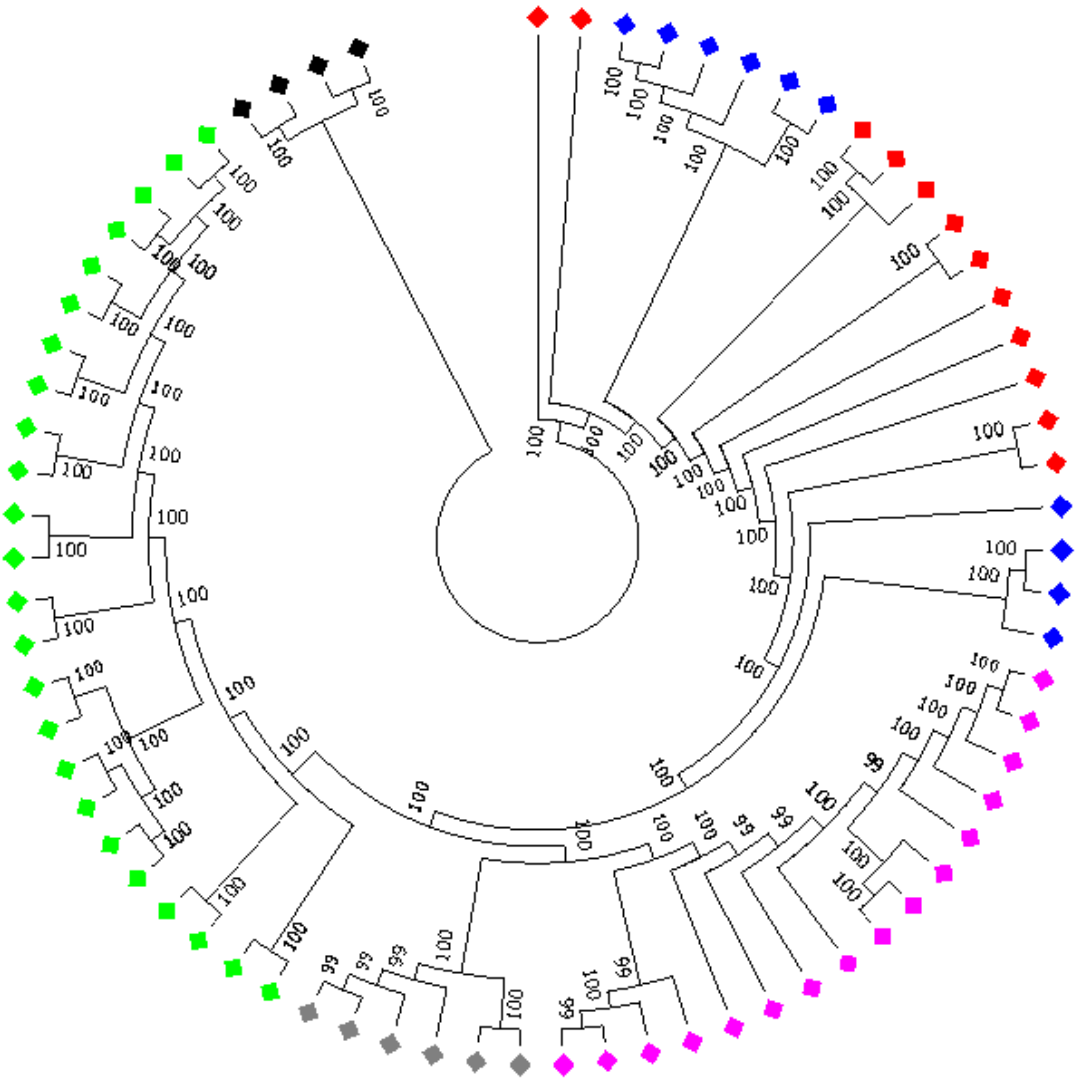


Figure S32: Haplotype maximum likelihood tree for the non-introgressed region on chr2: 120140001-120160000 of Saudi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl.

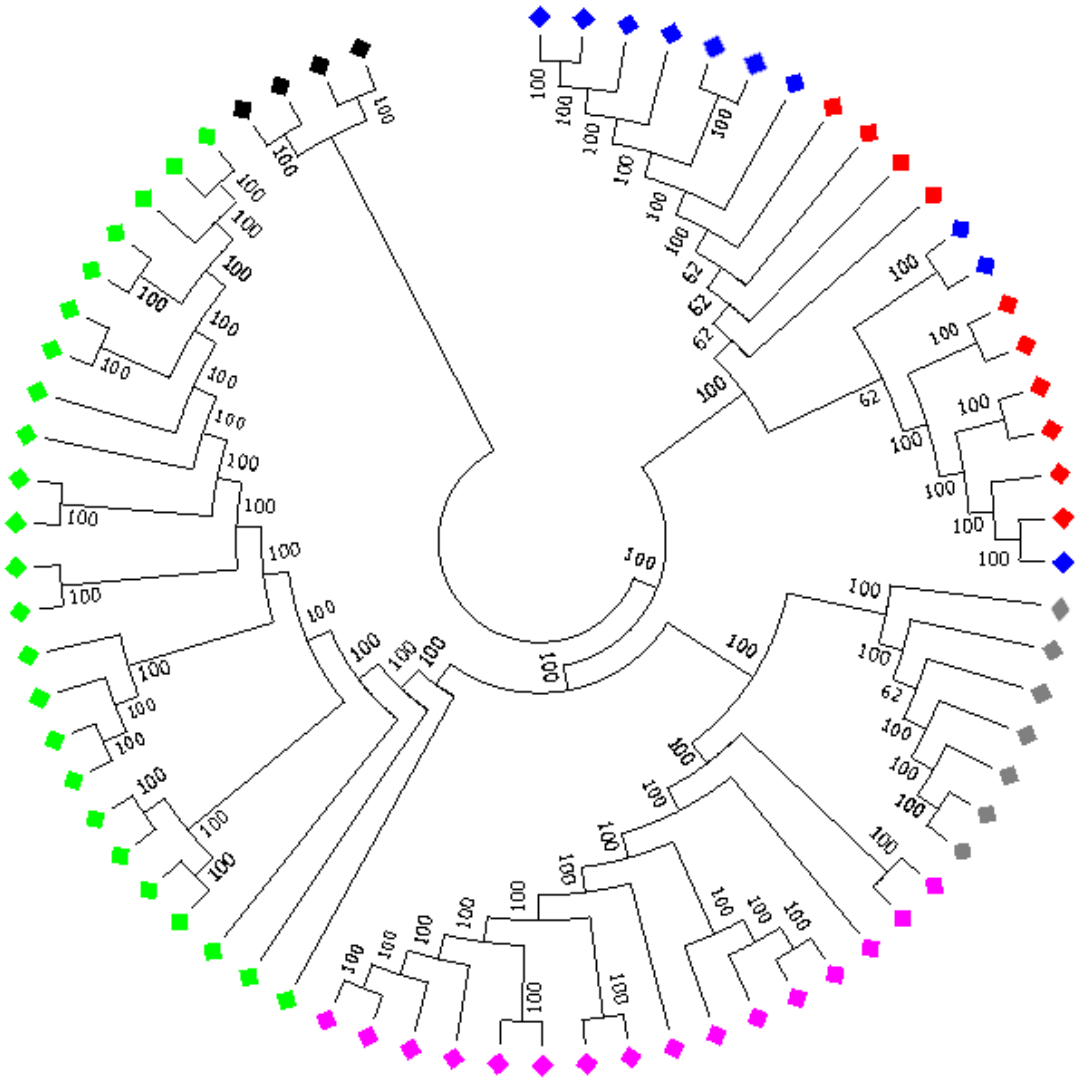


Figure S33: Haplotype maximum likelihood tree for the non-introgressed region on chr5: 30200001-30220000 of Saudi domestic chicken and Ceylon junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl.

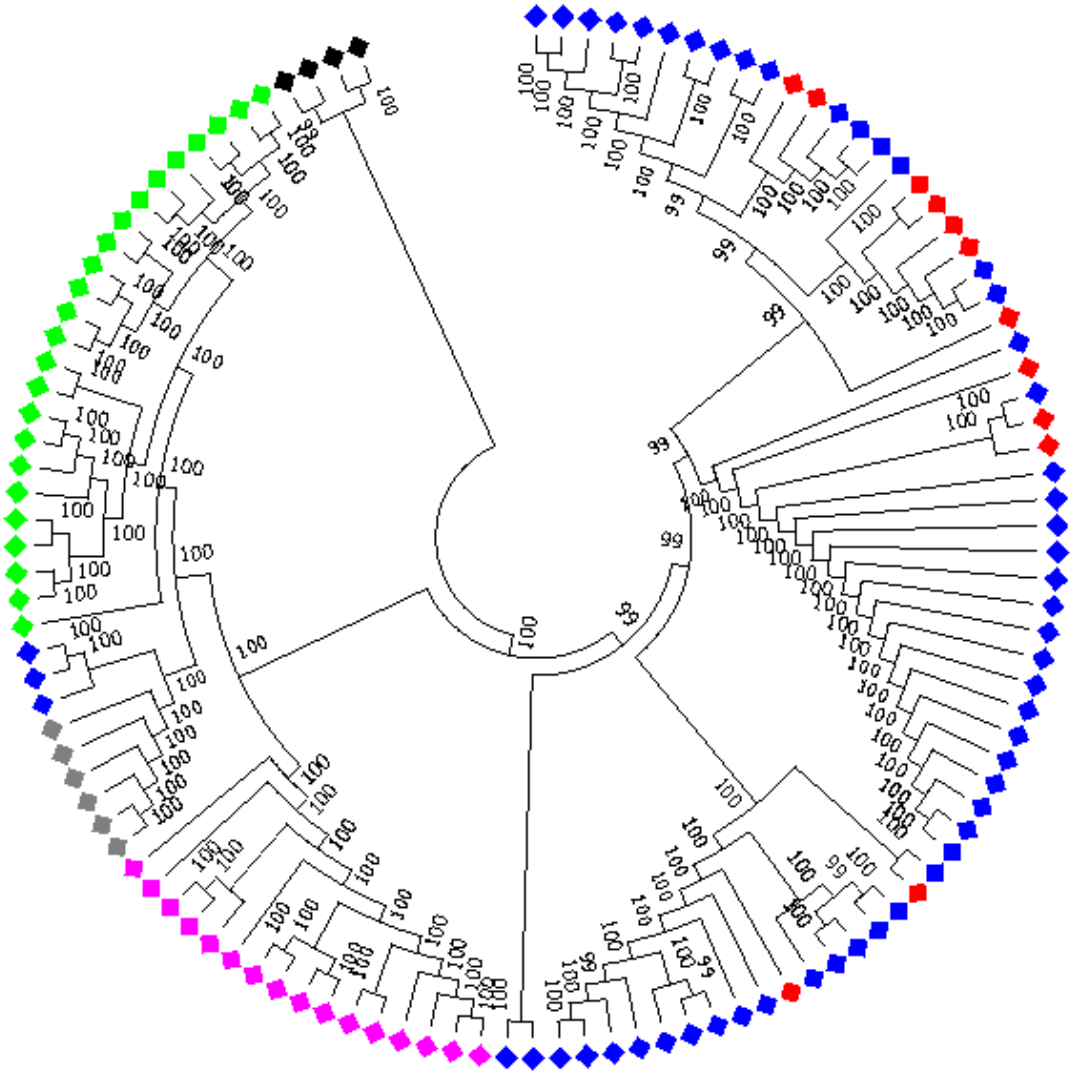


Figure S34: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 9440001-9460000 of Iraqi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

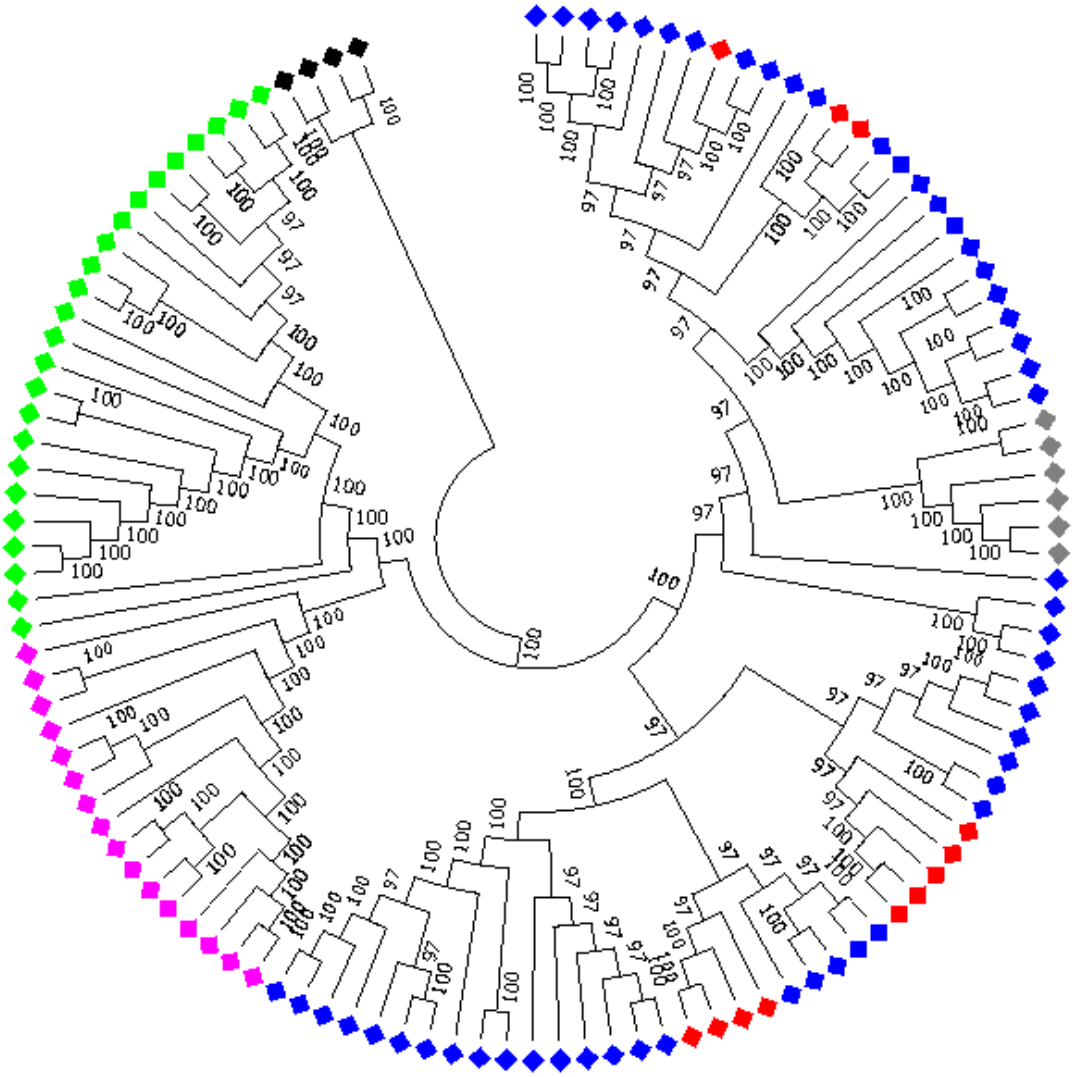


Figure S35: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 9500001-9520000 of Iraqi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

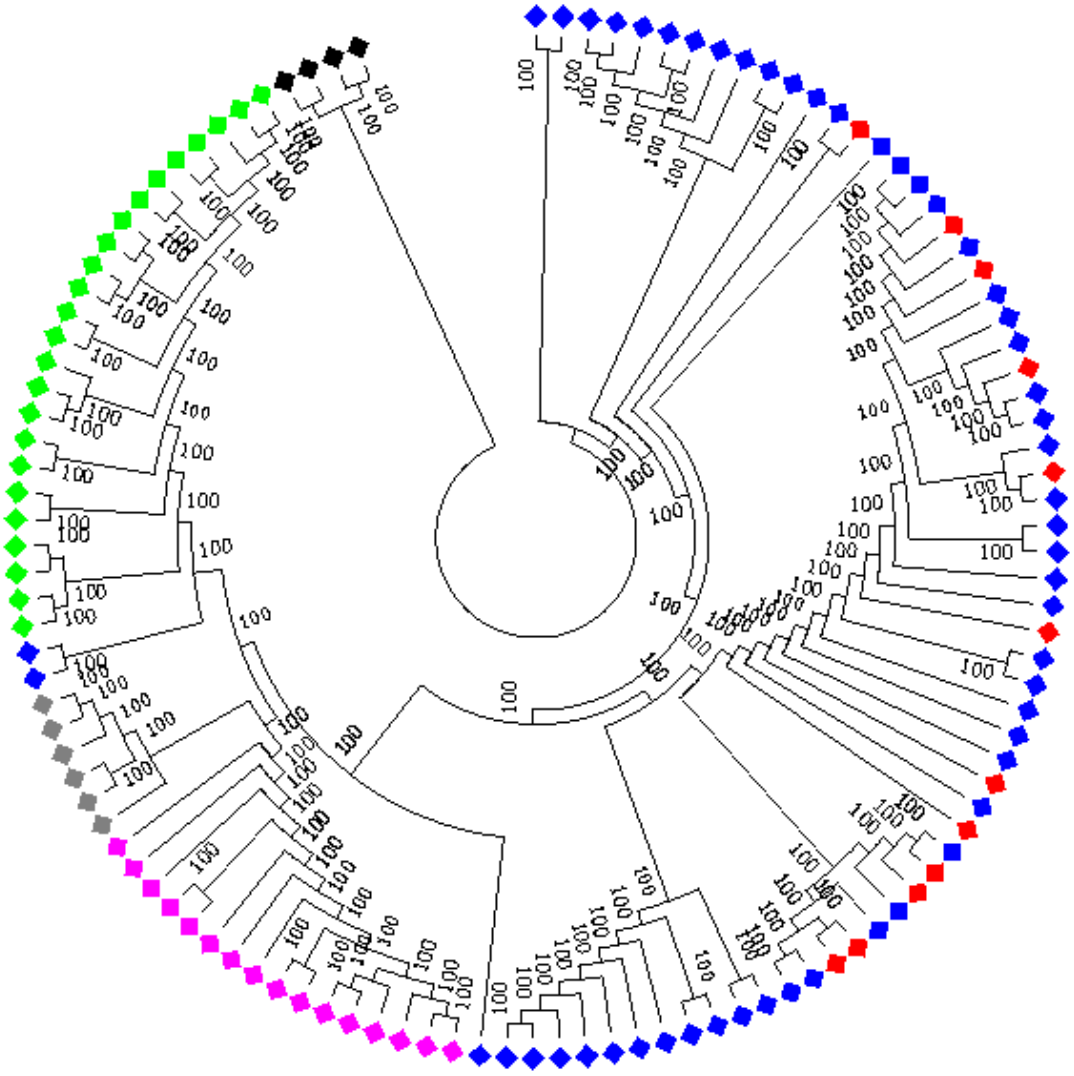


Figure S36: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 18400001-18420000 of Iraqi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

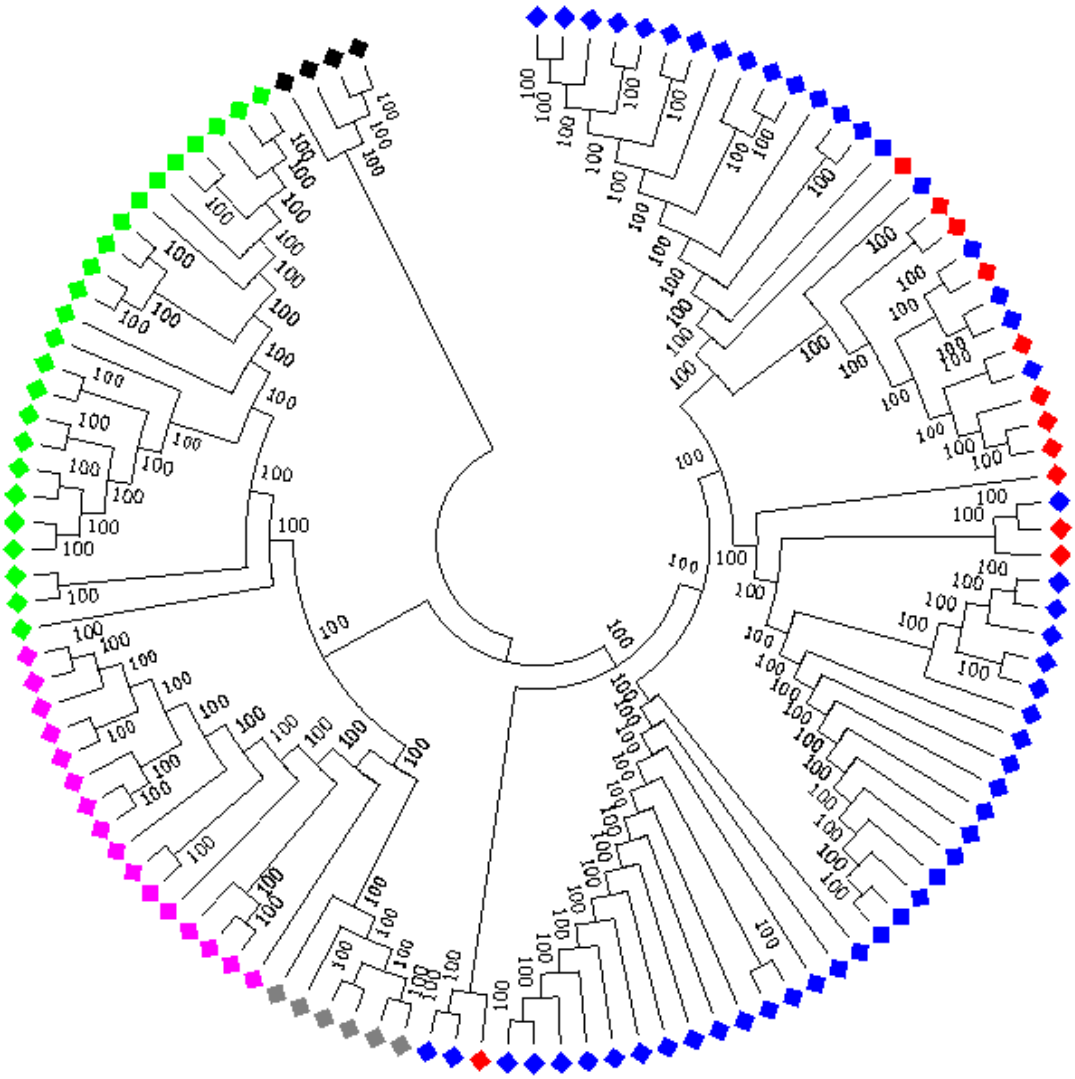


Figure S37: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 103580001-103600000 of Iraqi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

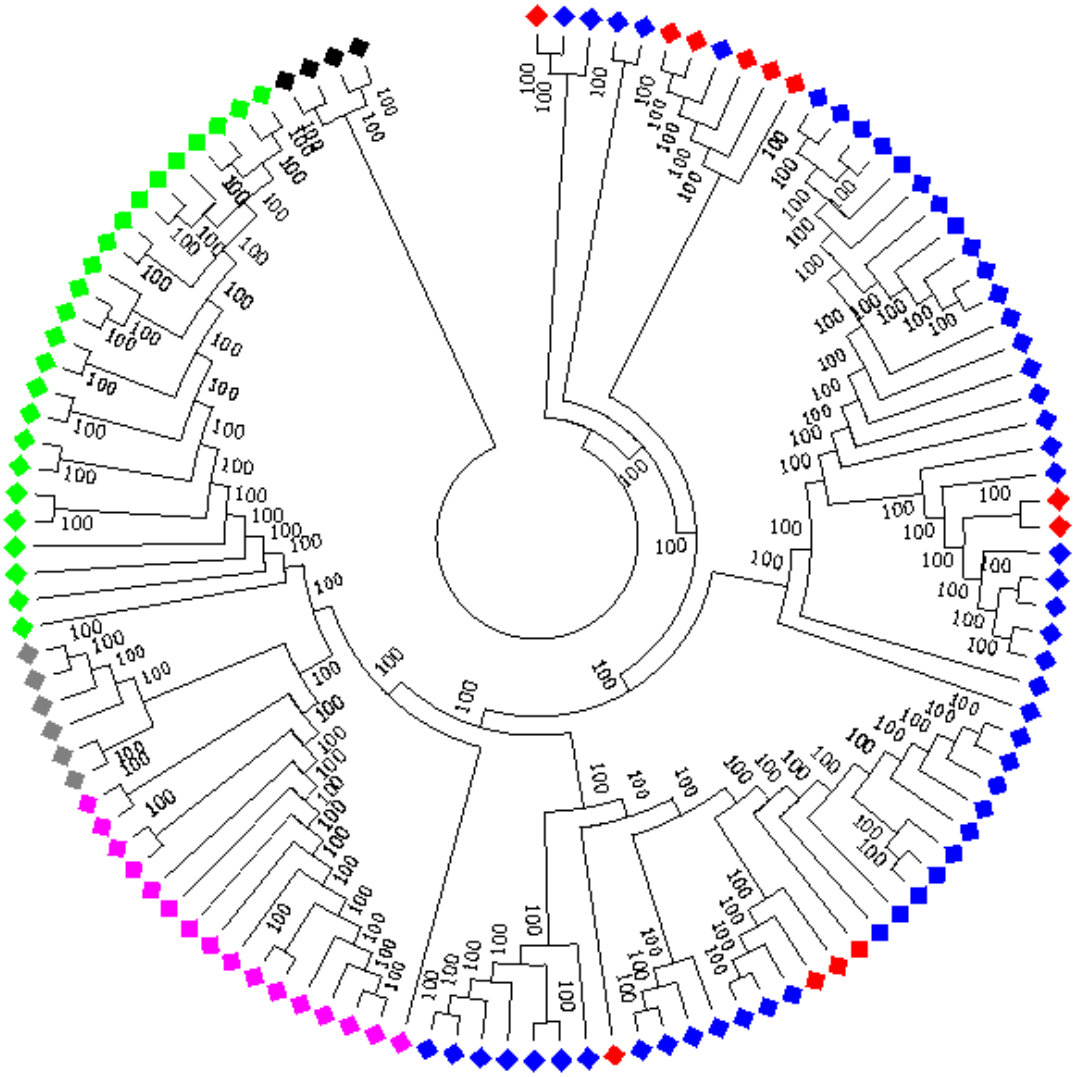


Figure S38: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 103620001-103640000 of Iraqi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

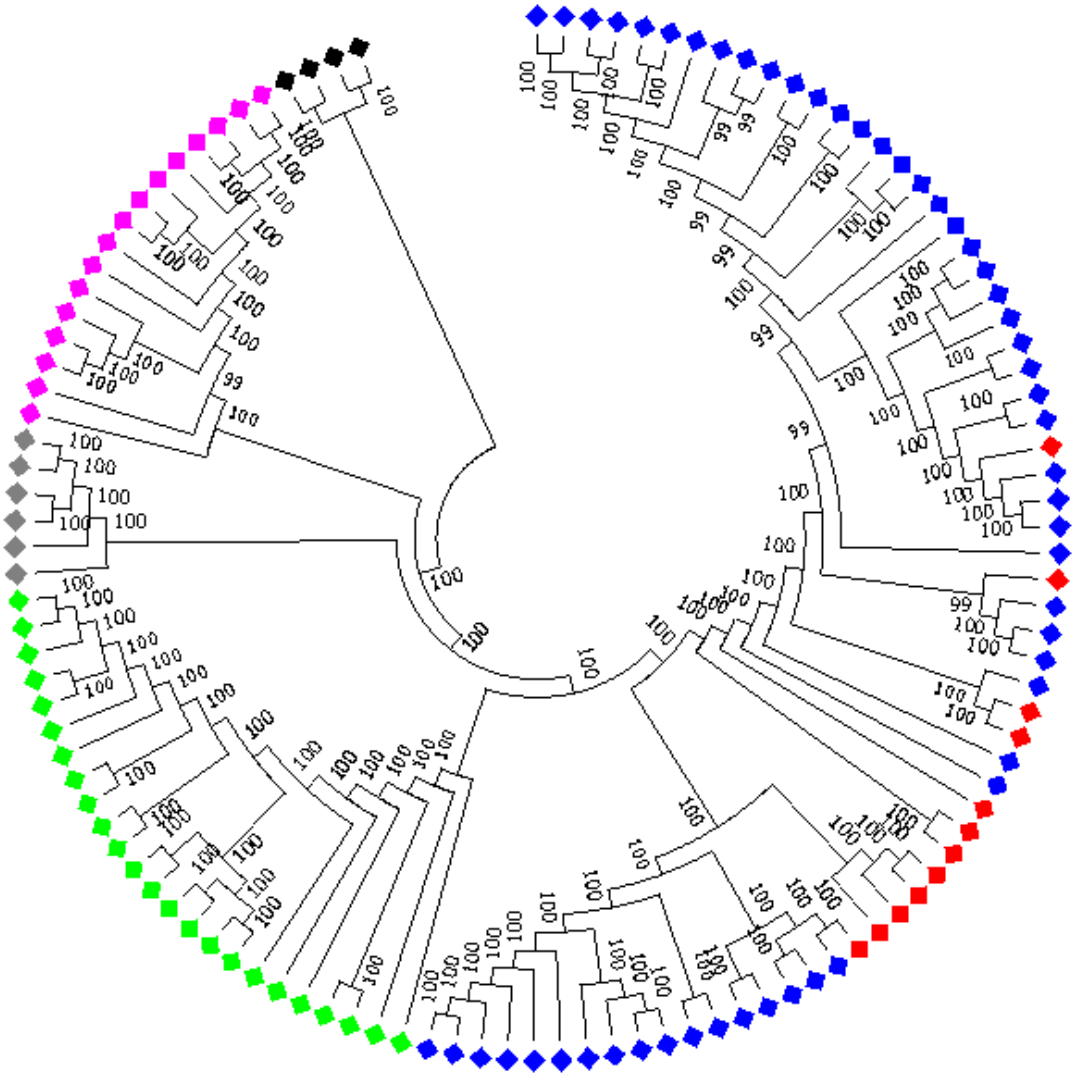


Figure S39: Haplotype maximum likelihood tree for the non-introgressed region on chr2: 54820001-54840000 of Iraqi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

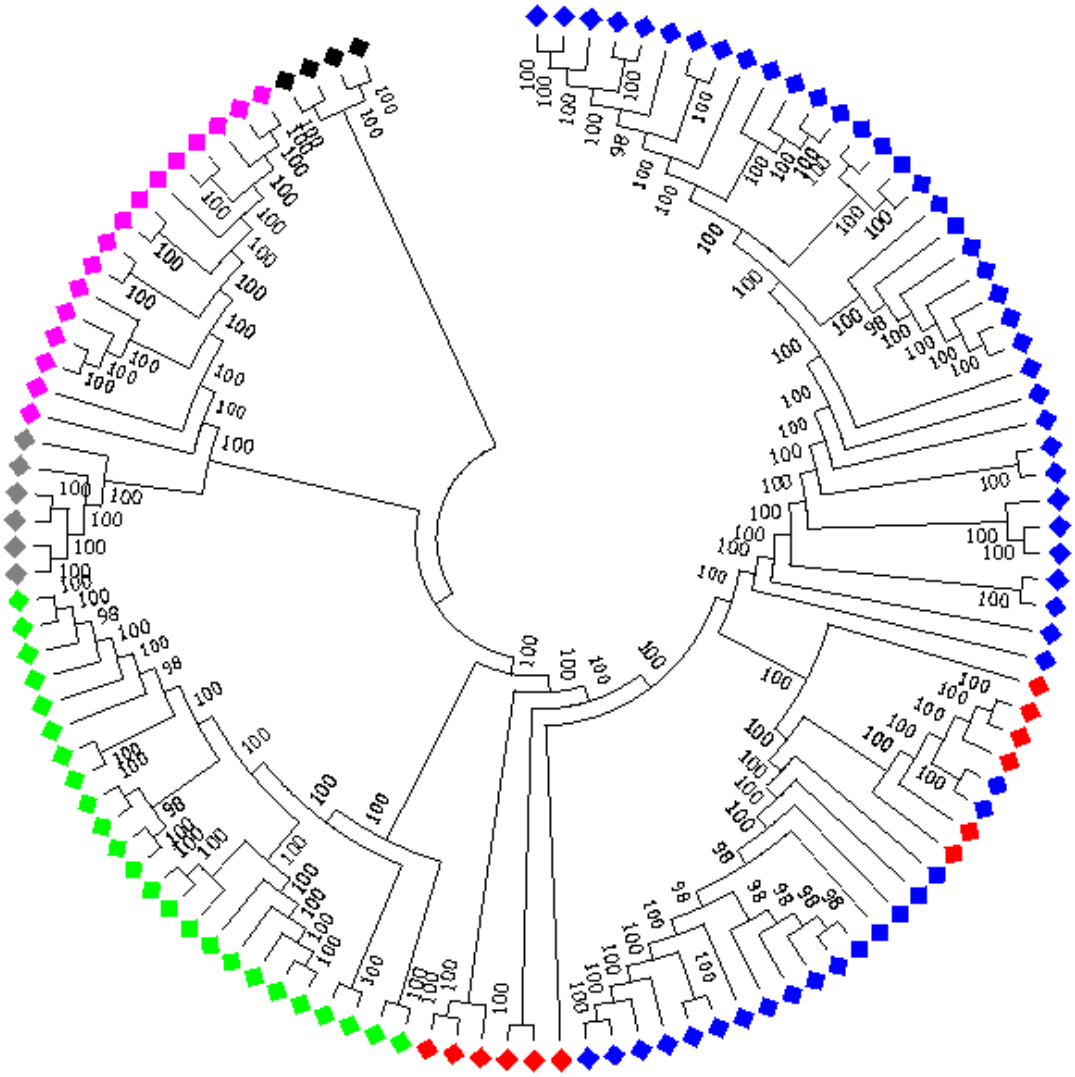


Figure S40: Haplotype maximum likelihood tree for the non-introgressed region on chr3: 70780001-70800000 of Iraqi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

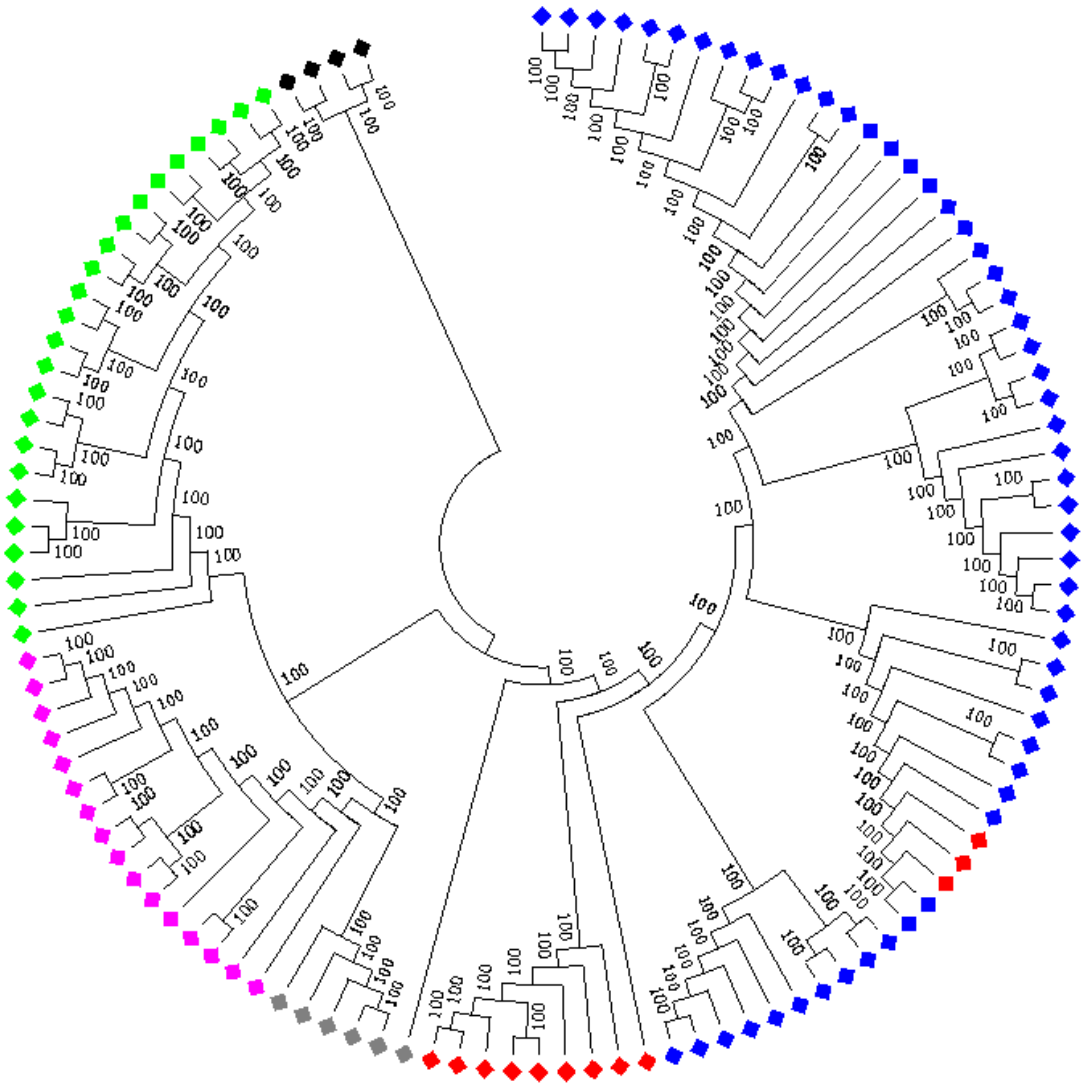


Figure S41: Haplotype maximum likelihood tree for the non-introgressed region on chr3: 100400001-100420000 of Iraqi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

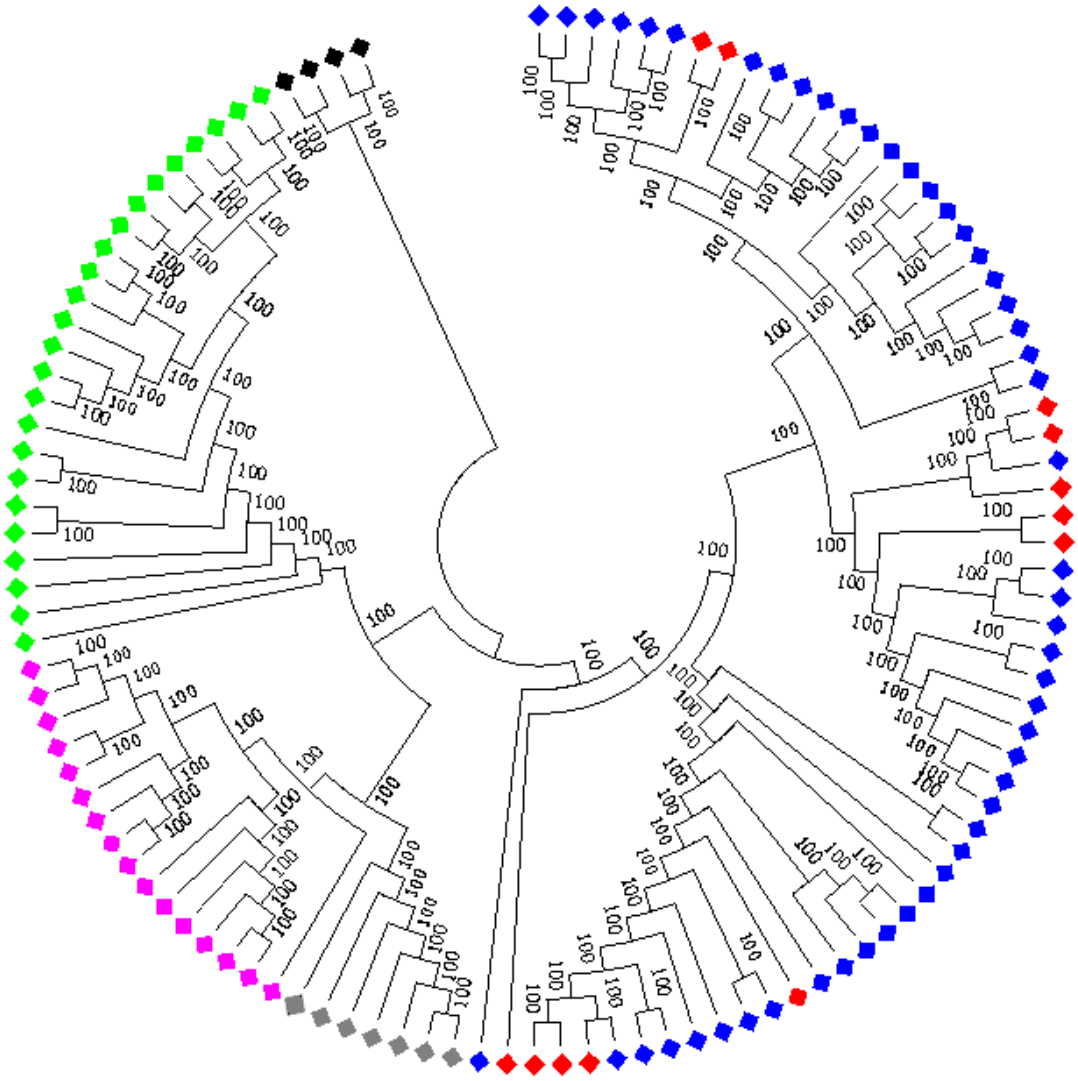


Figure S42: Haplotype maximum likelihood tree for the non-introgressed region on chr4: 26720001-26740000 of Iraqi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

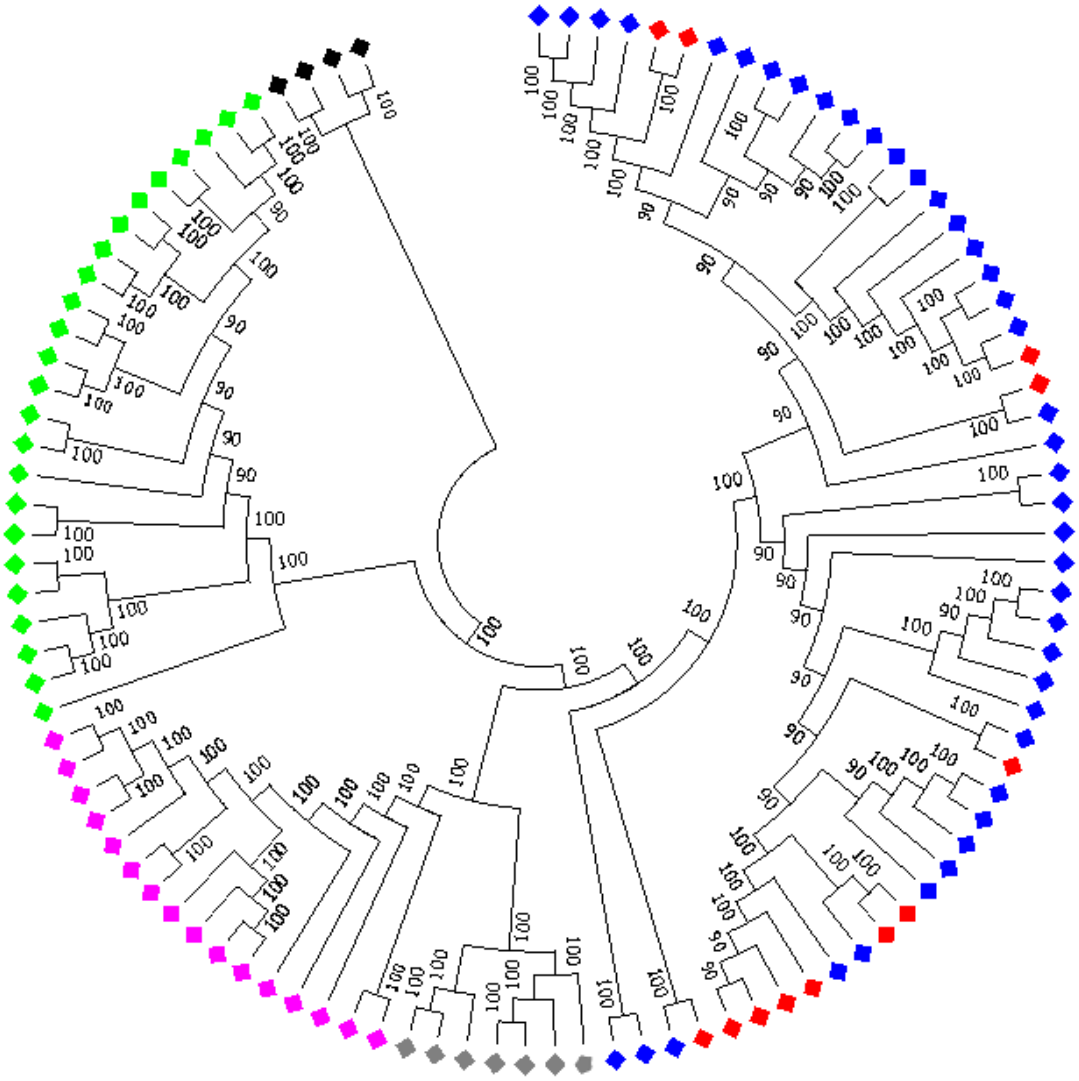


Figure S43: Haplotype maximum likelihood tree for the non-introgressed region on chr4: 72720001-72740000 of Iraqi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

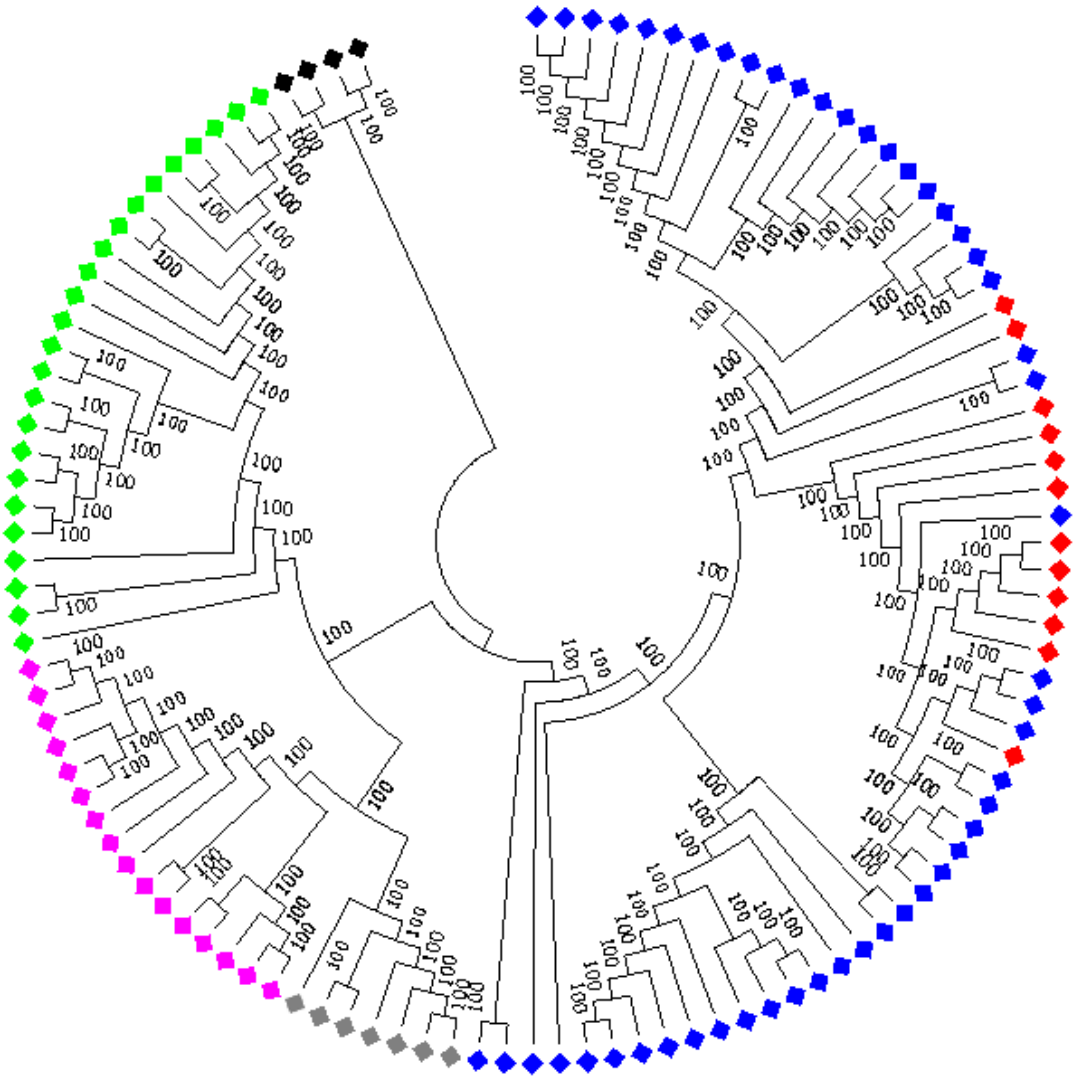


Figure S44: Haplotype maximum likelihood tree for the non-introgressed region on chr4: 73160001-73180000 of Iraqi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

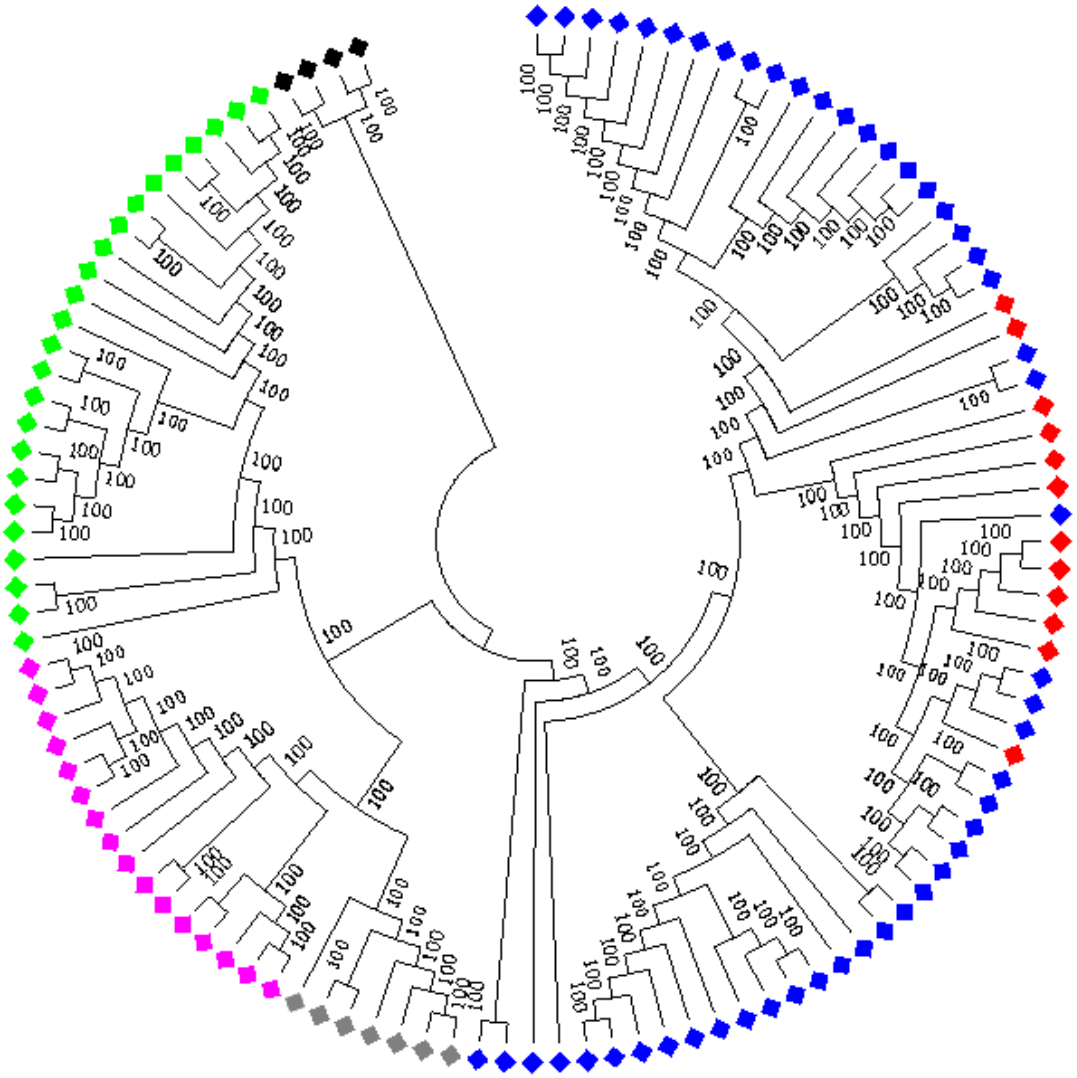


Figure S45: Haplotype maximum likelihood tree for the non-introgressed region on chr5: 33200001-33220000 of Iraqi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

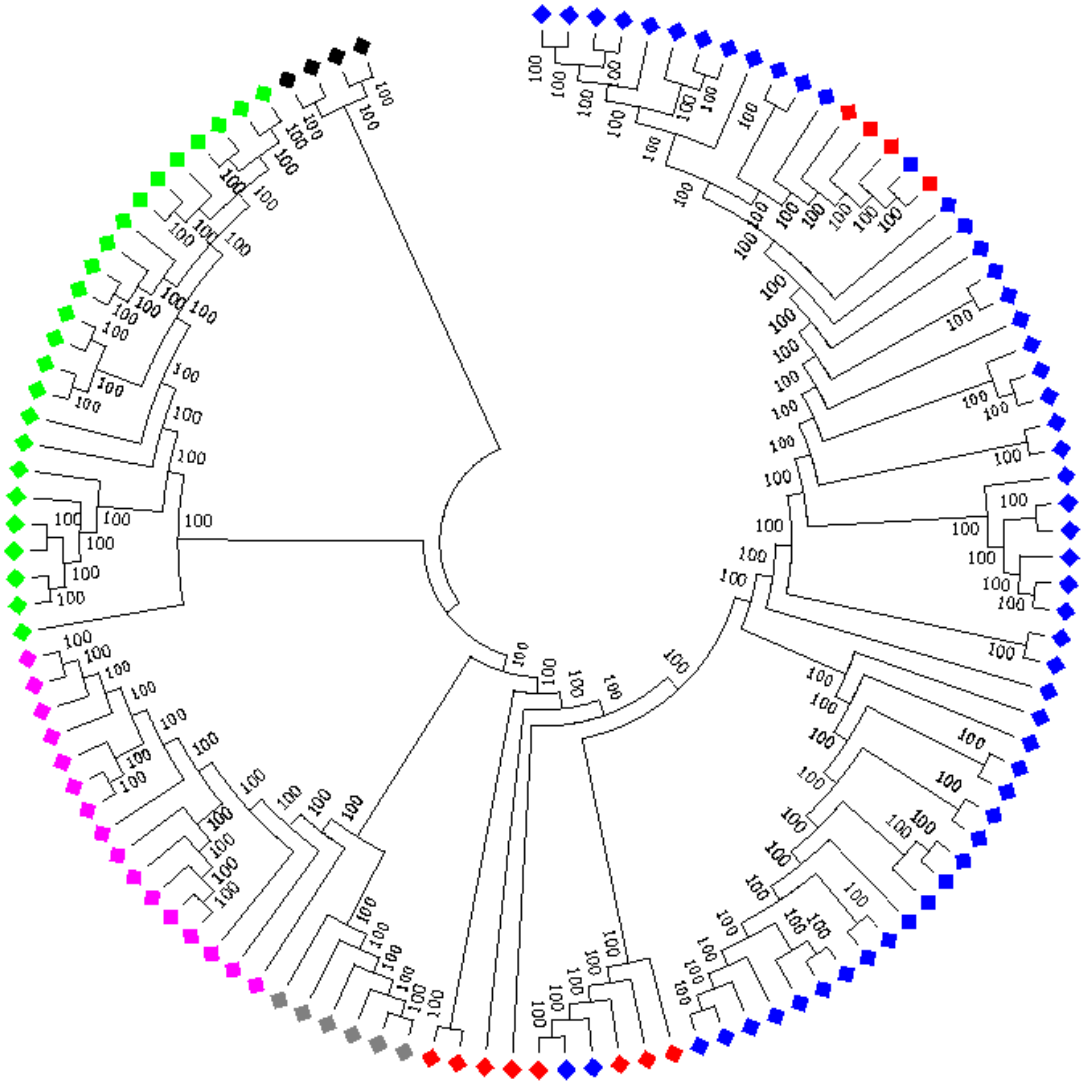


Figure S46: Haplotype maximum likelihood tree for the non-introgressed region on chr11: 3980001-4000000 of Iraqi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

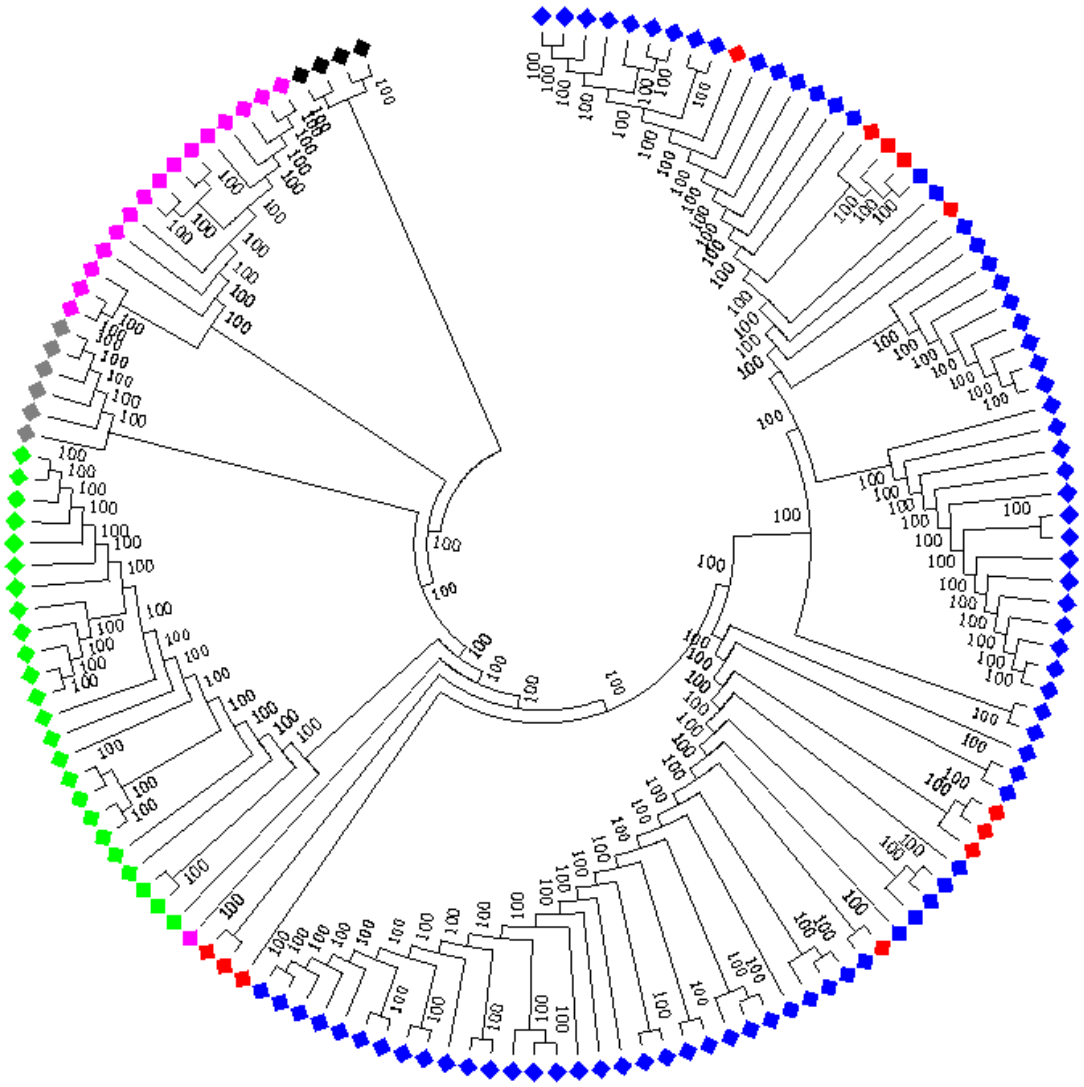


Figure S47: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 56600001-56620000 of Ethiopian domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

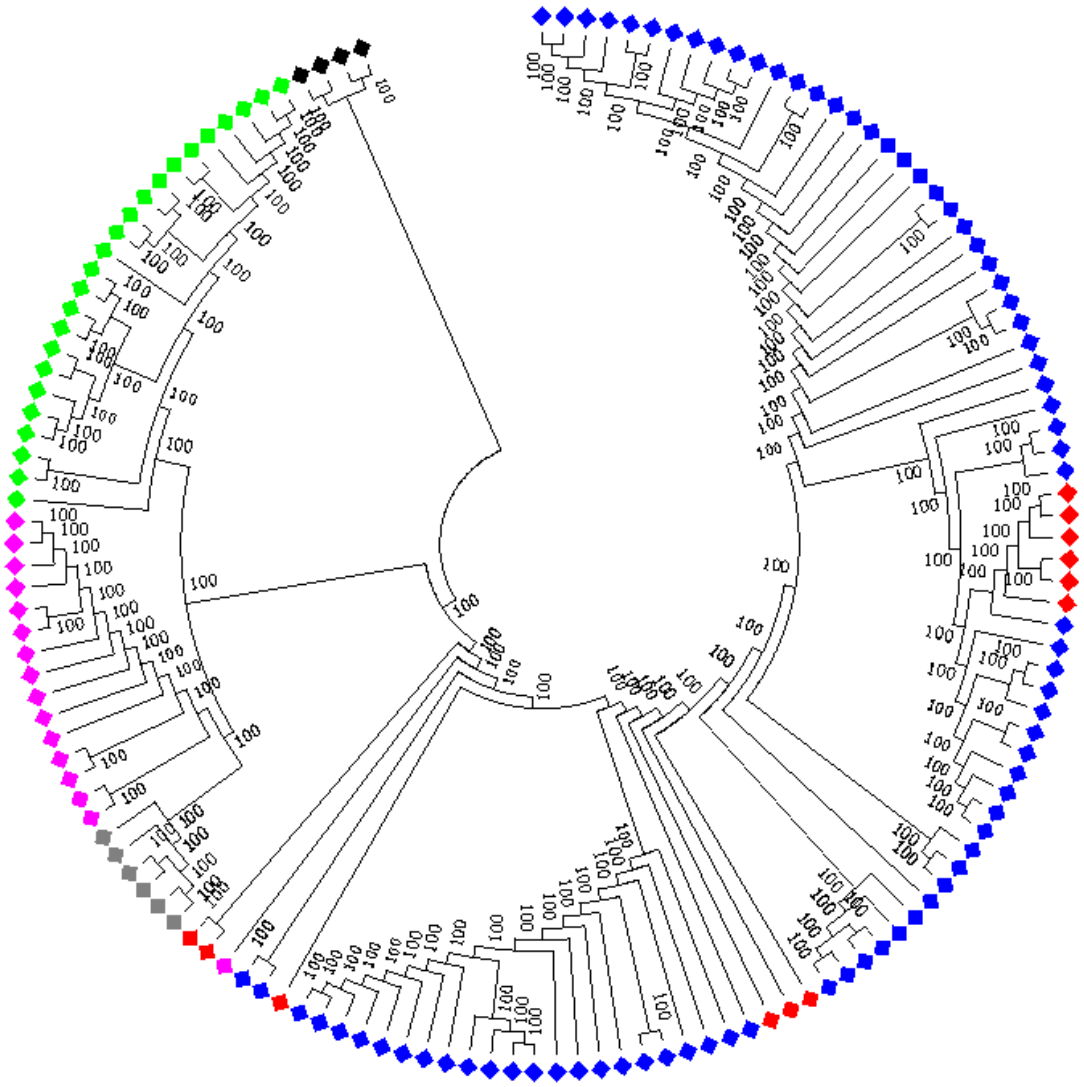


Figure S48: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 56680001-56700000 of Ethiopian domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

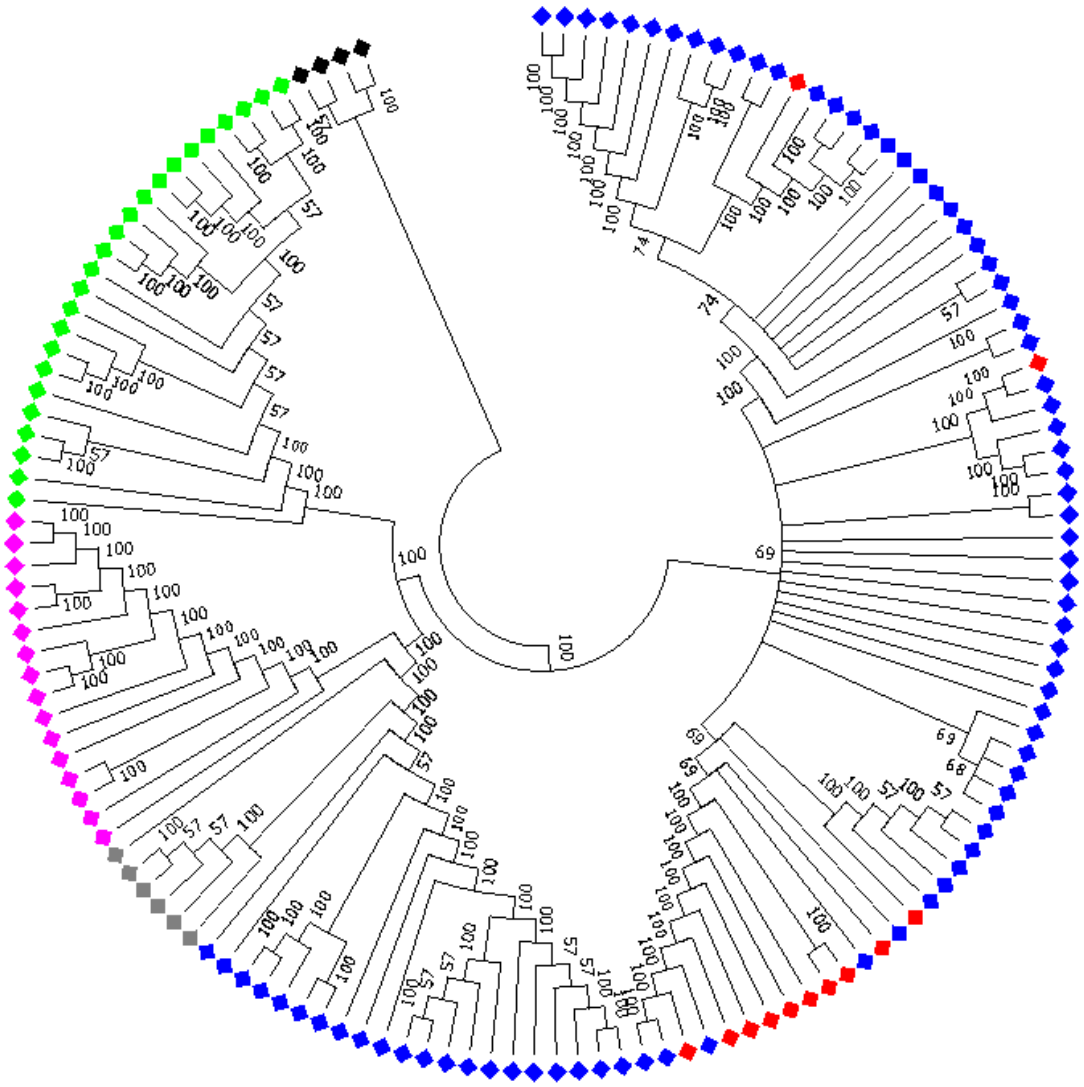


Figure S49: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 56720001-56740000 of Ethiopian domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

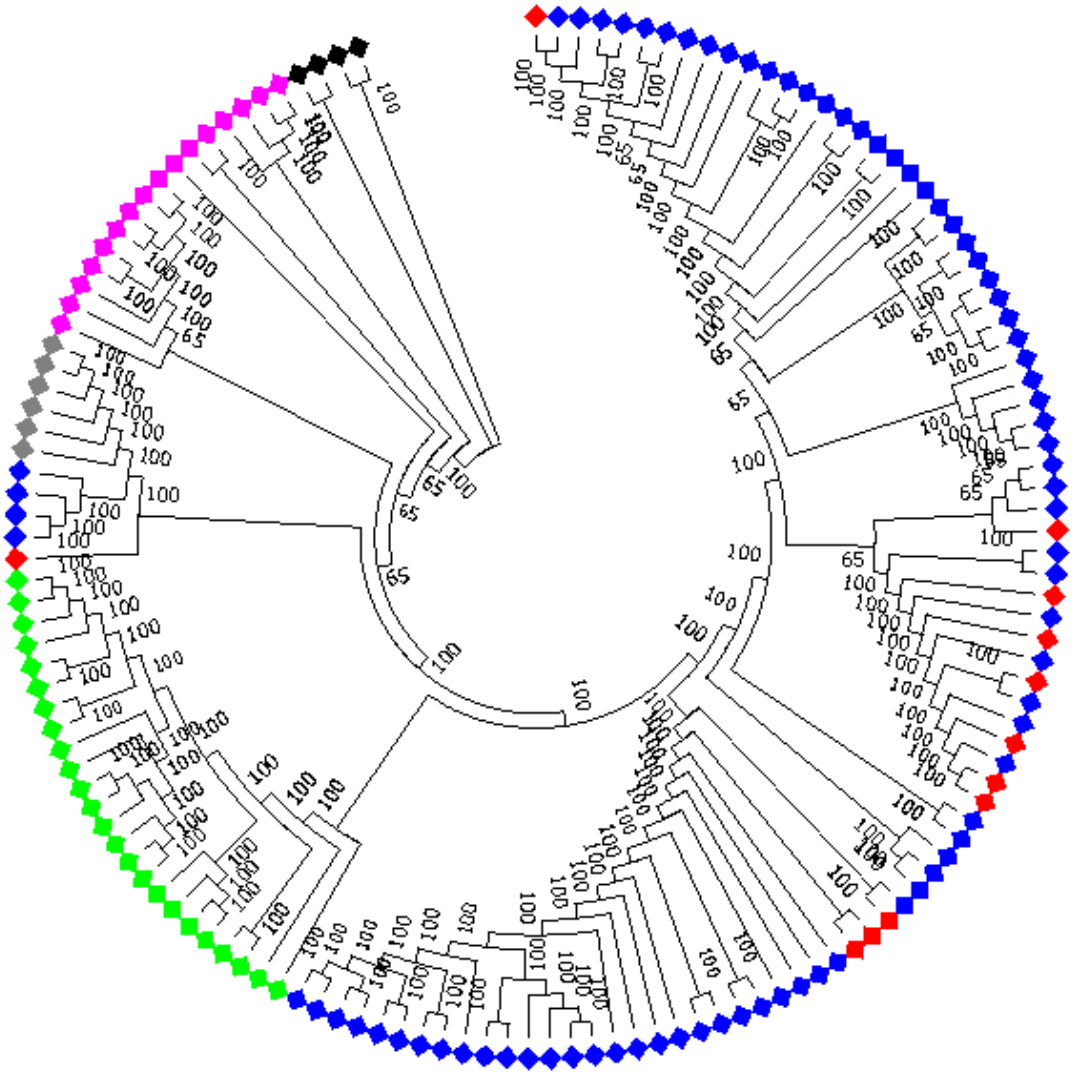


Figure S50: Haplotype maximum likelihood tree for the non-introgressed region on chr2: 88280001-88300000 of Ethiopian domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

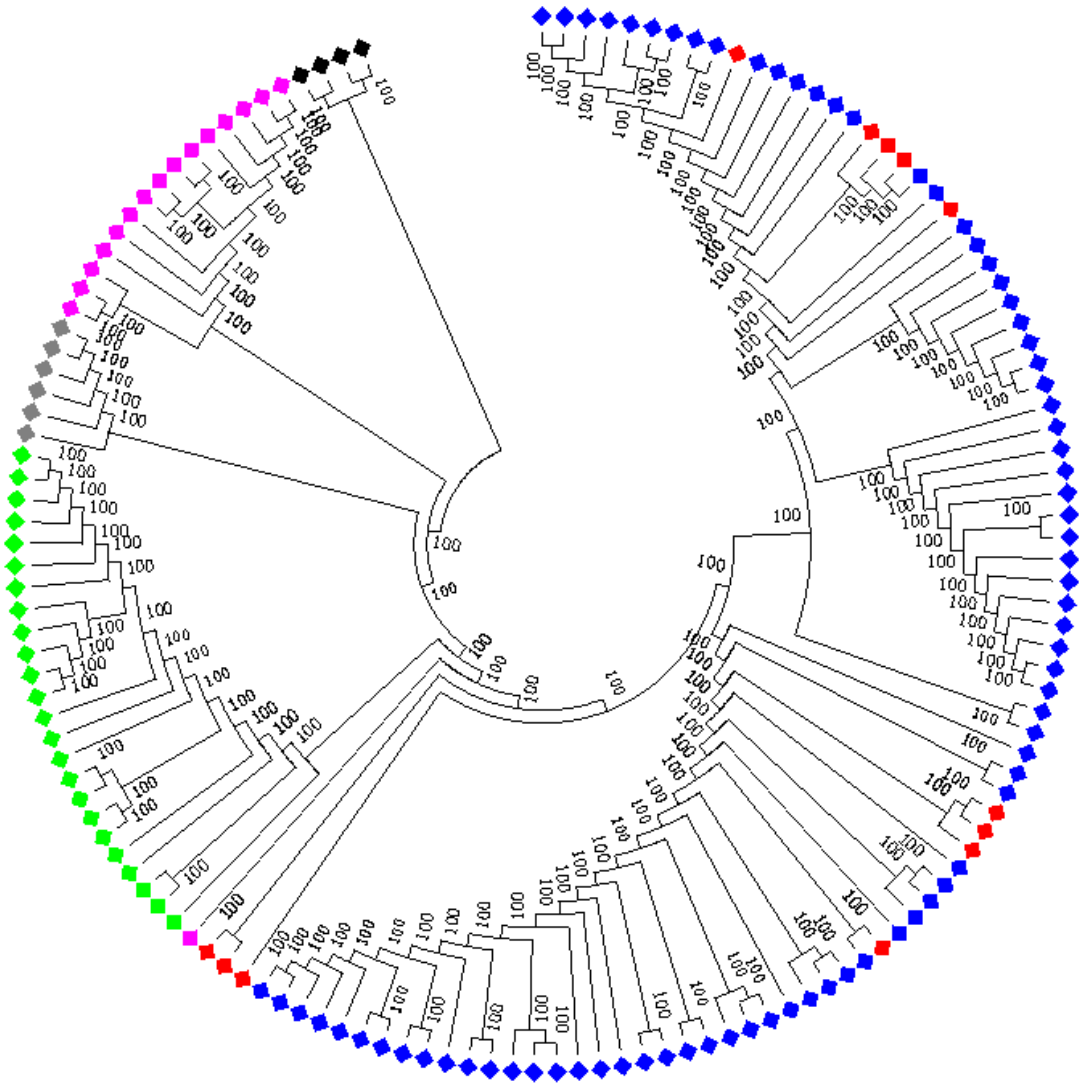


Figure S51: Haplotype maximum likelihood tree for the non-introgressed region on chr11: 33400001-33420000 of Ethiopian domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

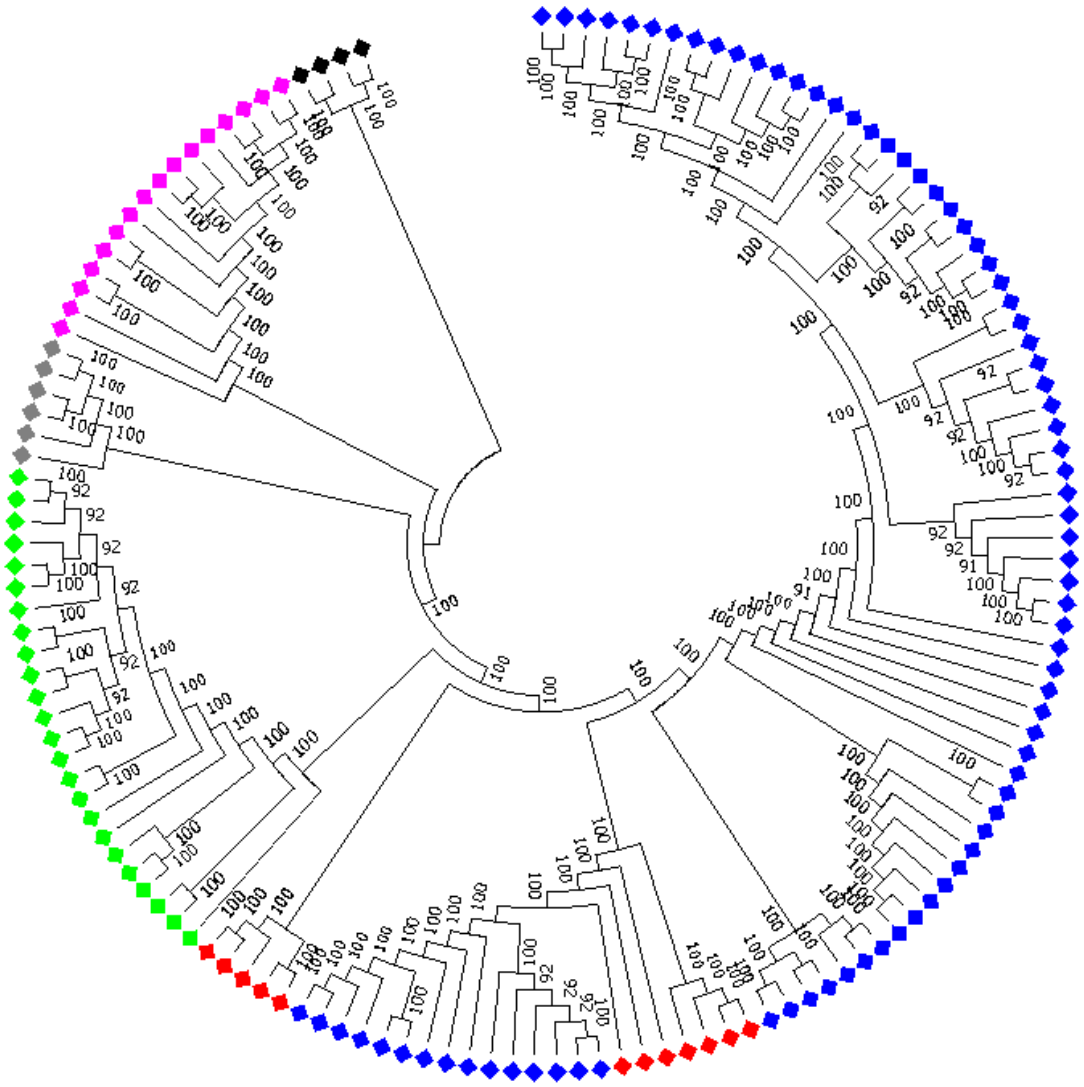


Figure S52: Haplotype maximum likelihood tree for the non-introgressed region on chr11: 3920001-3940000 of Ethiopian domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

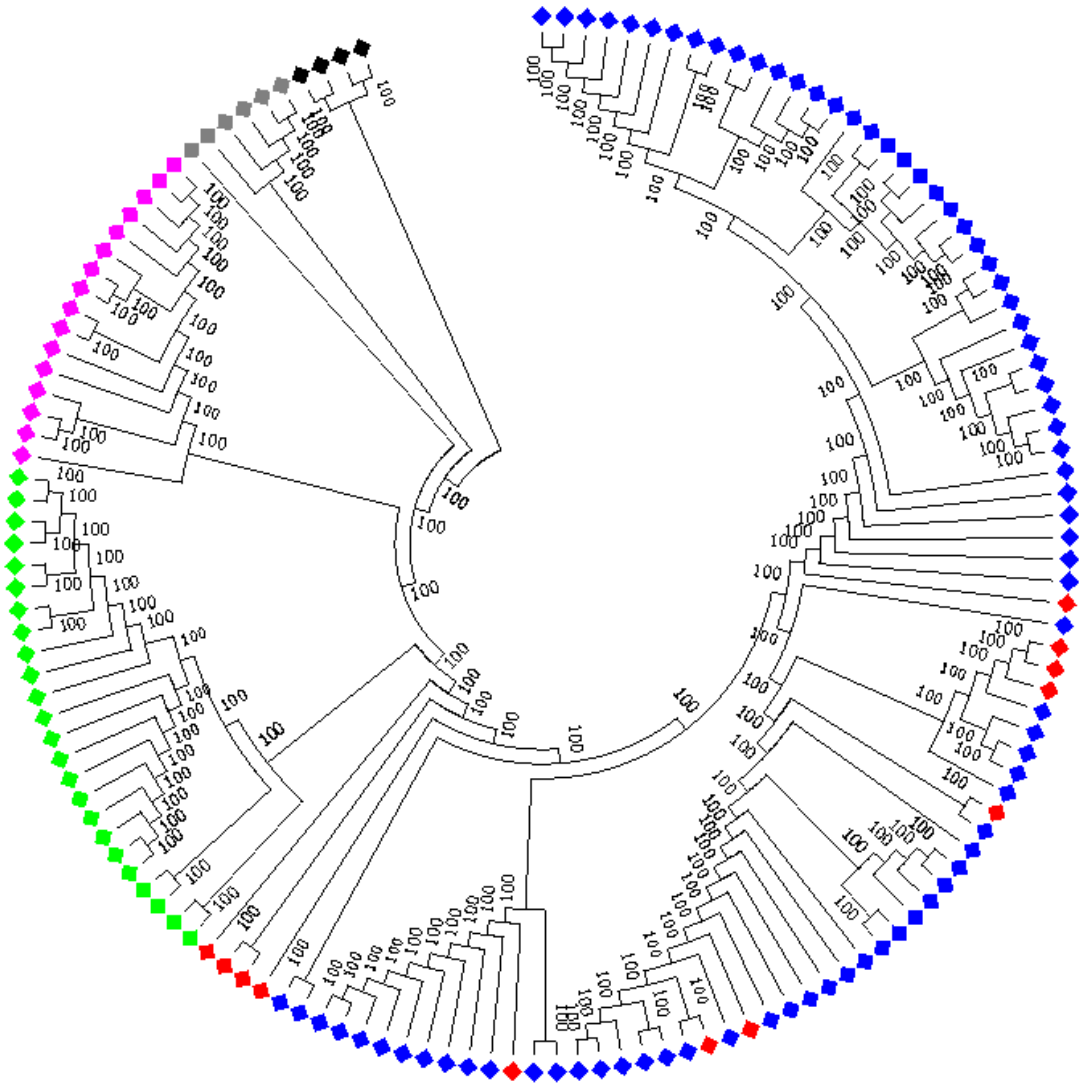


Figure S53: Haplotype maximum likelihood tree for the non-introgressed region on chr12: 9020001-9040000 of Ethiopian domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

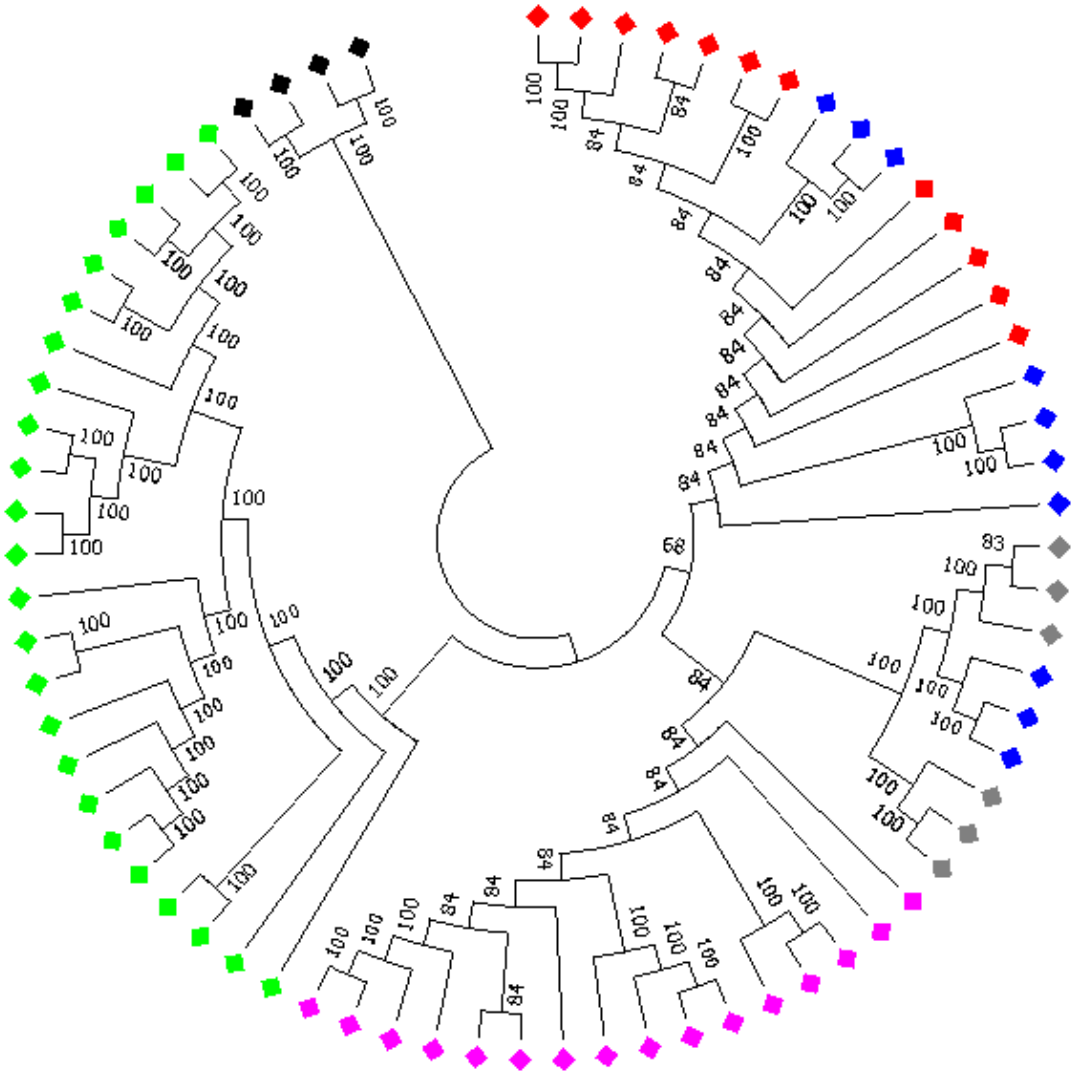


Figure S54: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 148180001-148200000 of Saudi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl.

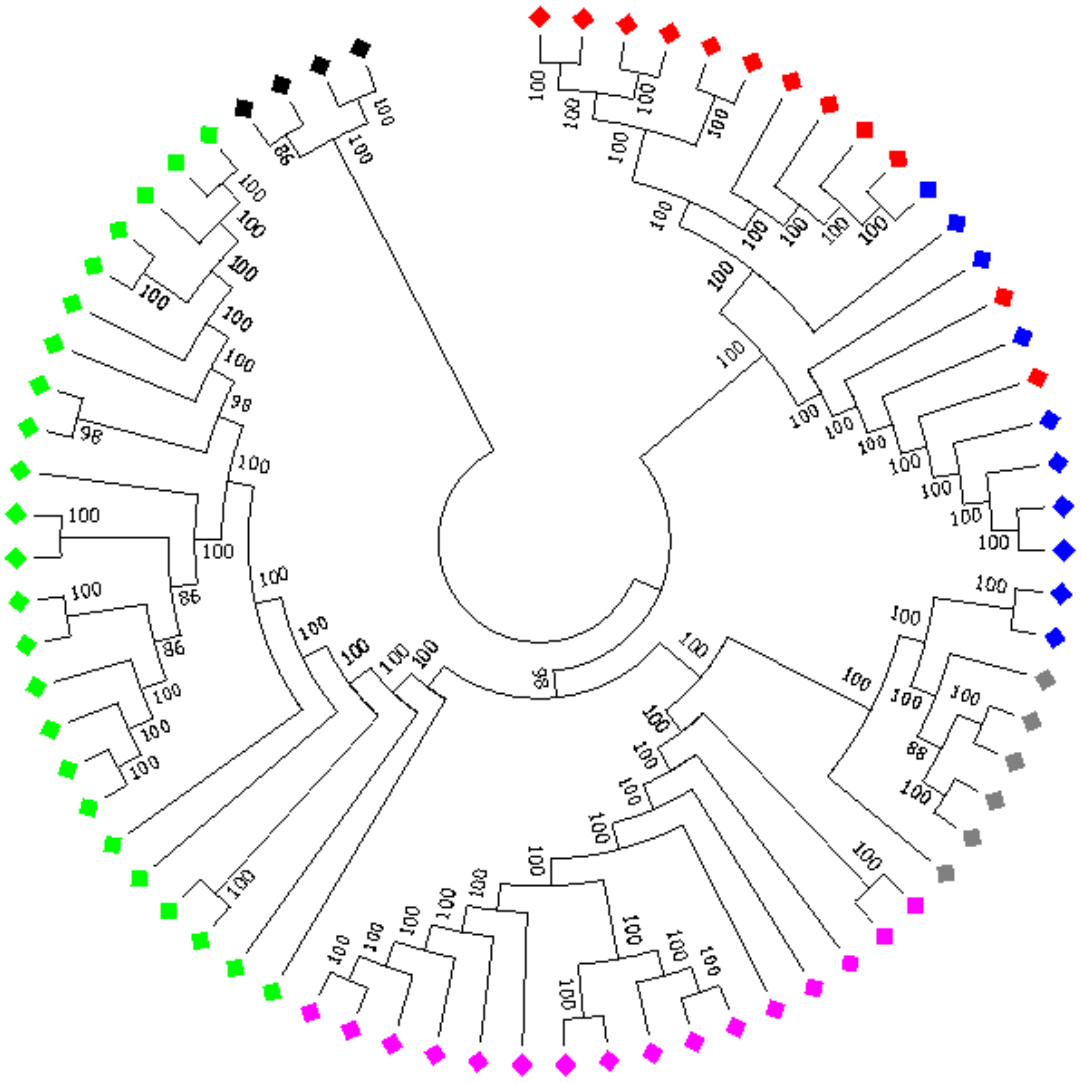


Figure S55: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 158060001-158080000 of Saudi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl.

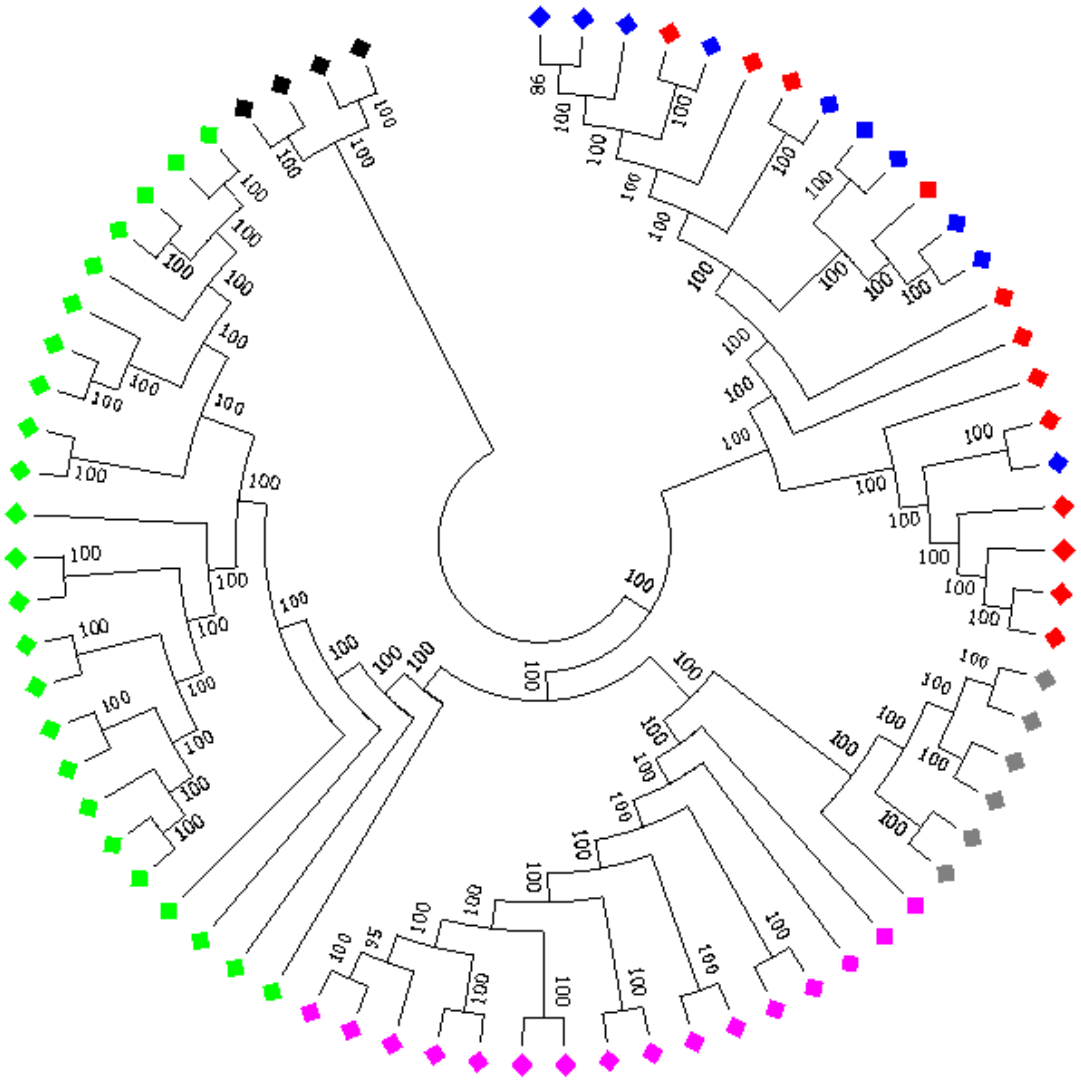


Figure S56: Haplotype maximum likelihood tree for the non-introgressed region on chr1: 191280001-191300000 of Saudi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl.

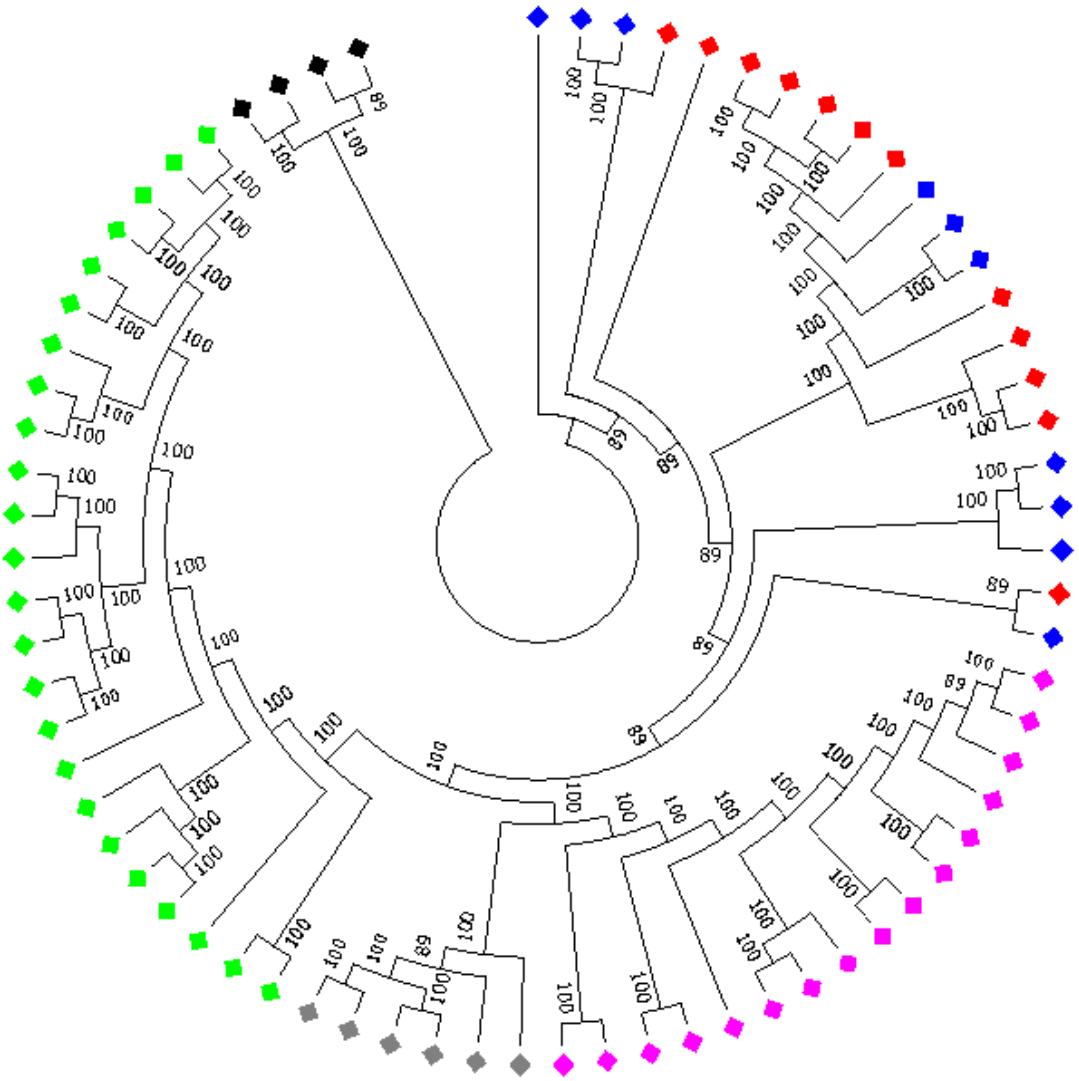


Figure S57: Haplotype maximum likelihood tree for the non-introgressed region on chr3: 70800001-70820000 of Saudi domestic chicken and Green junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl.

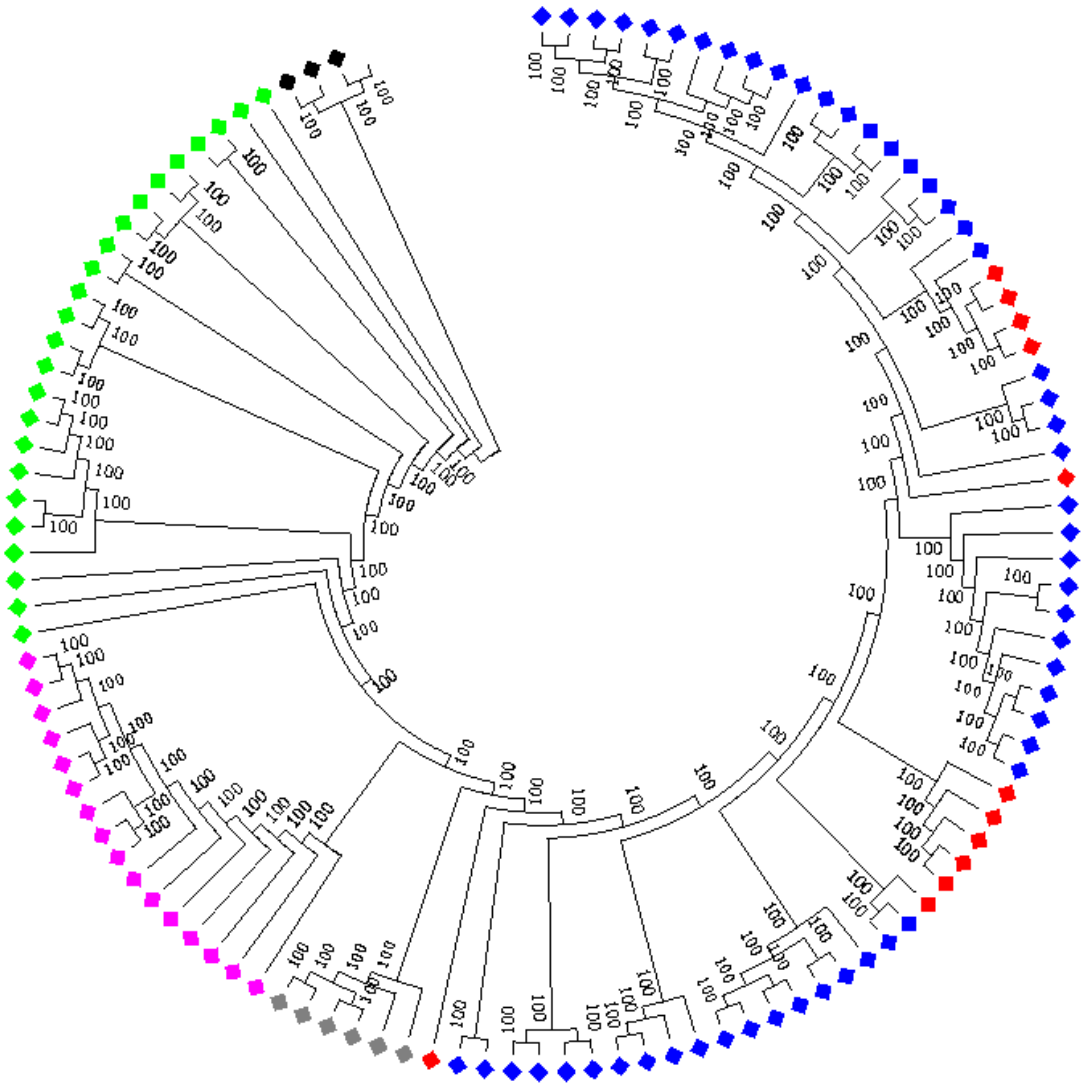


Figure S58: Haplotype maximum likelihood tree for one of six common regions on chr2: 120120000-120400000 of Iraqi domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

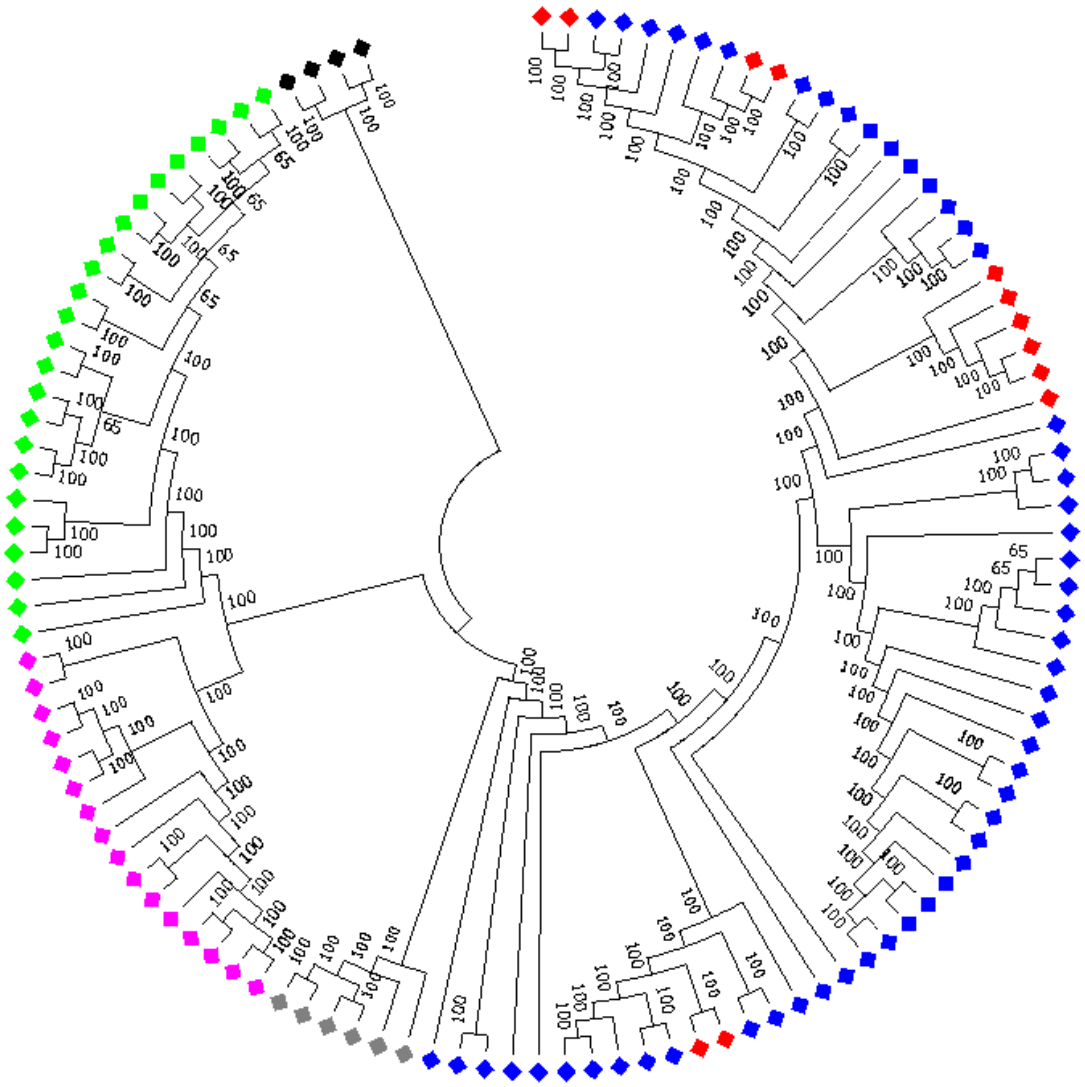


Figure S59: Haplotype maximum likelihood tree for one of six common regions on chr3: 50820000-50920000 of Iraqi domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

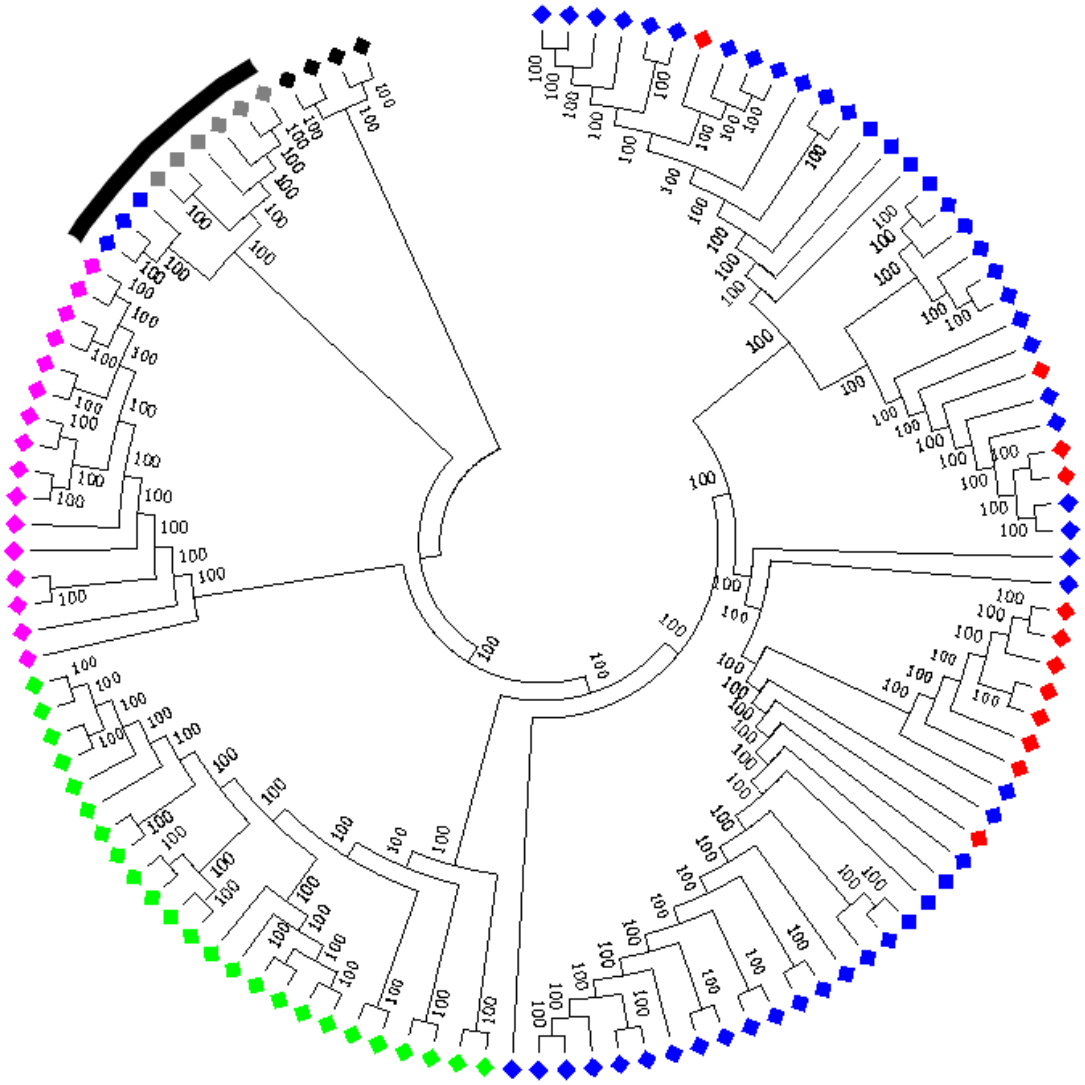


Figure S60: Haplotype maximum likelihood tree for one of six common regions on chr4: 62500000-62700000 of Iraqi domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Iraqi chicken marked by black.

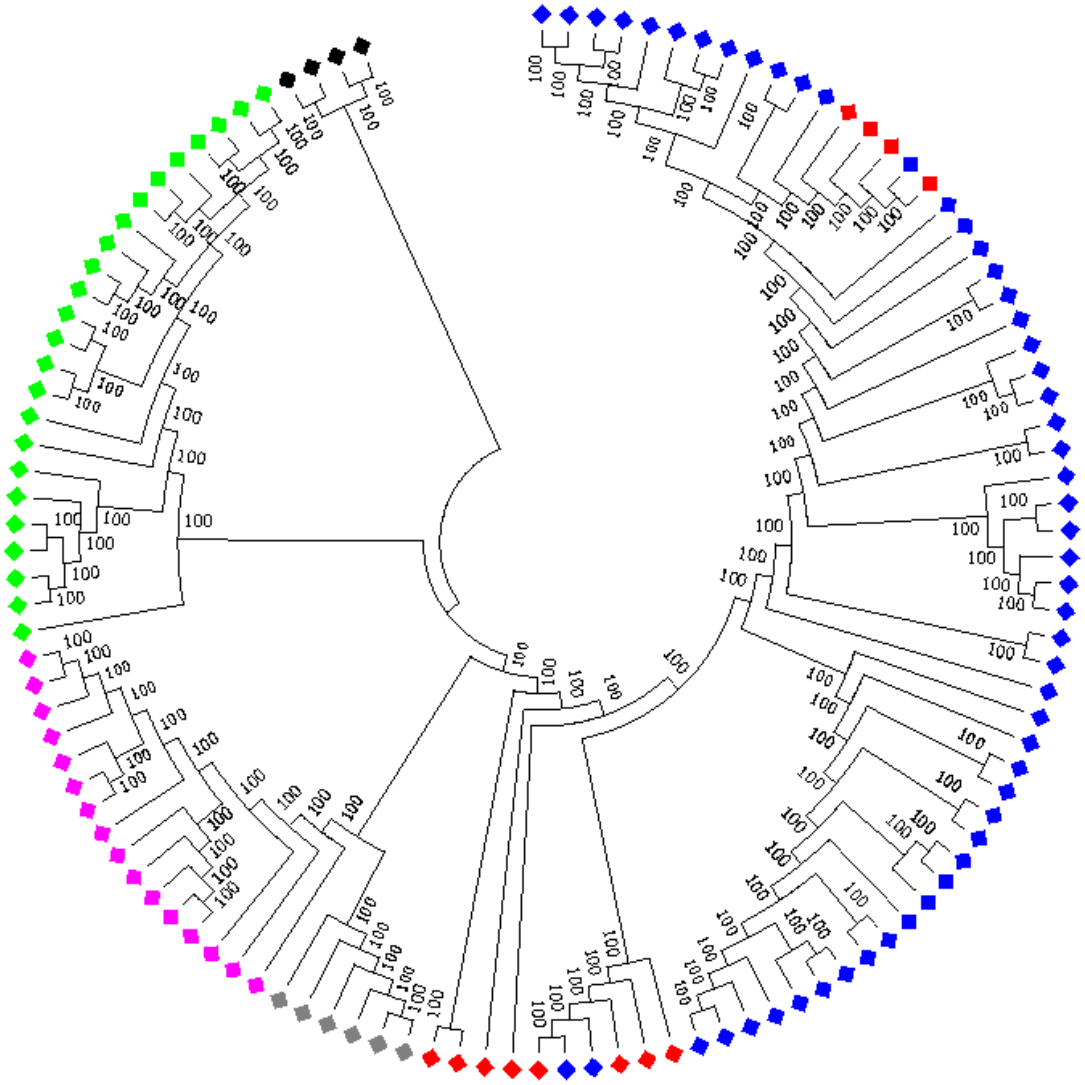


Figure S61: Haplotype maximum likelihood tree for one of six common regions on chr6: 20760000-20880000 of Iraqi domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

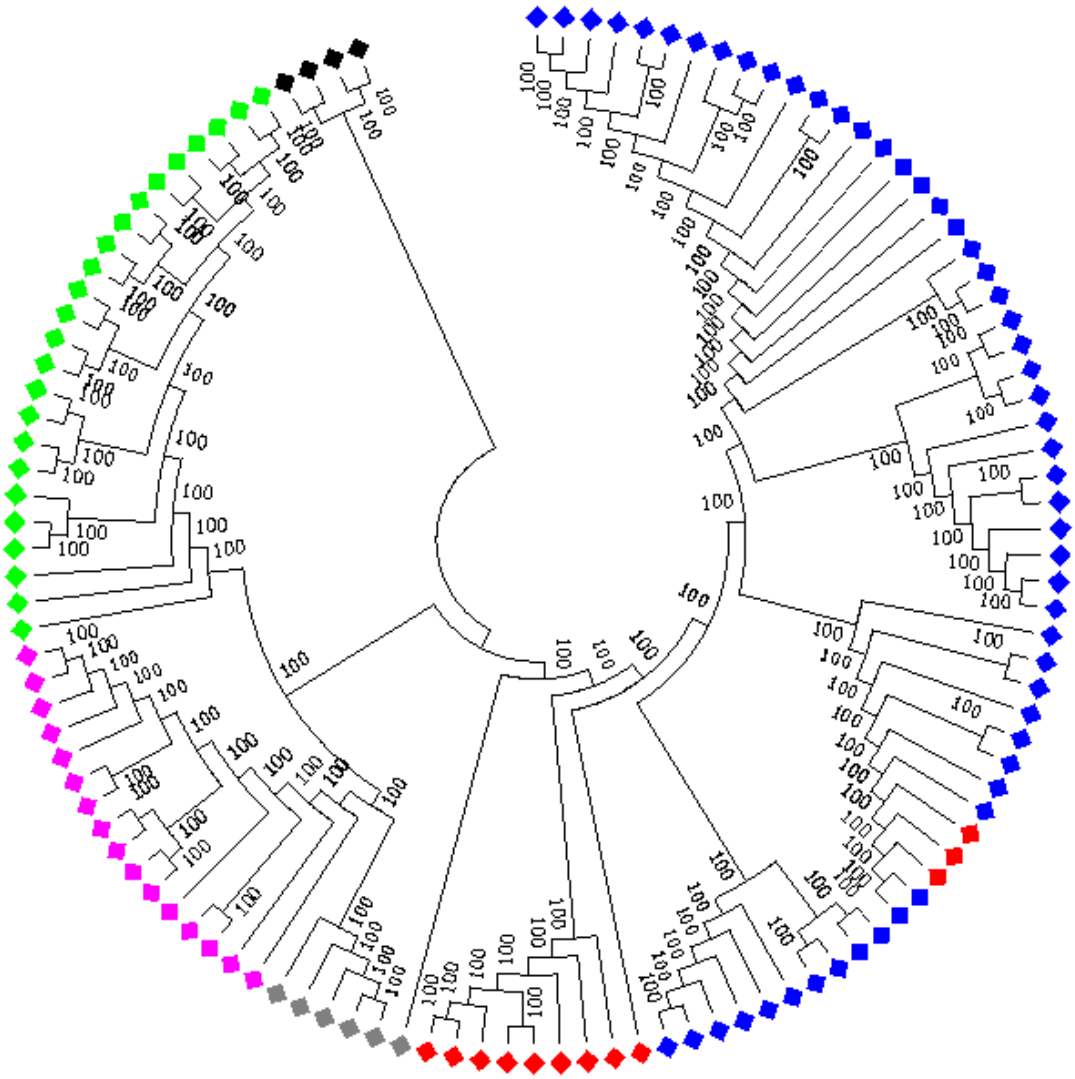


Figure S62: Haplotype maximum likelihood tree for one of six common regions on chr12: 12540000-12640000 of Iraqi domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Iraqi domestic chicken and Red = Red junglefowl.

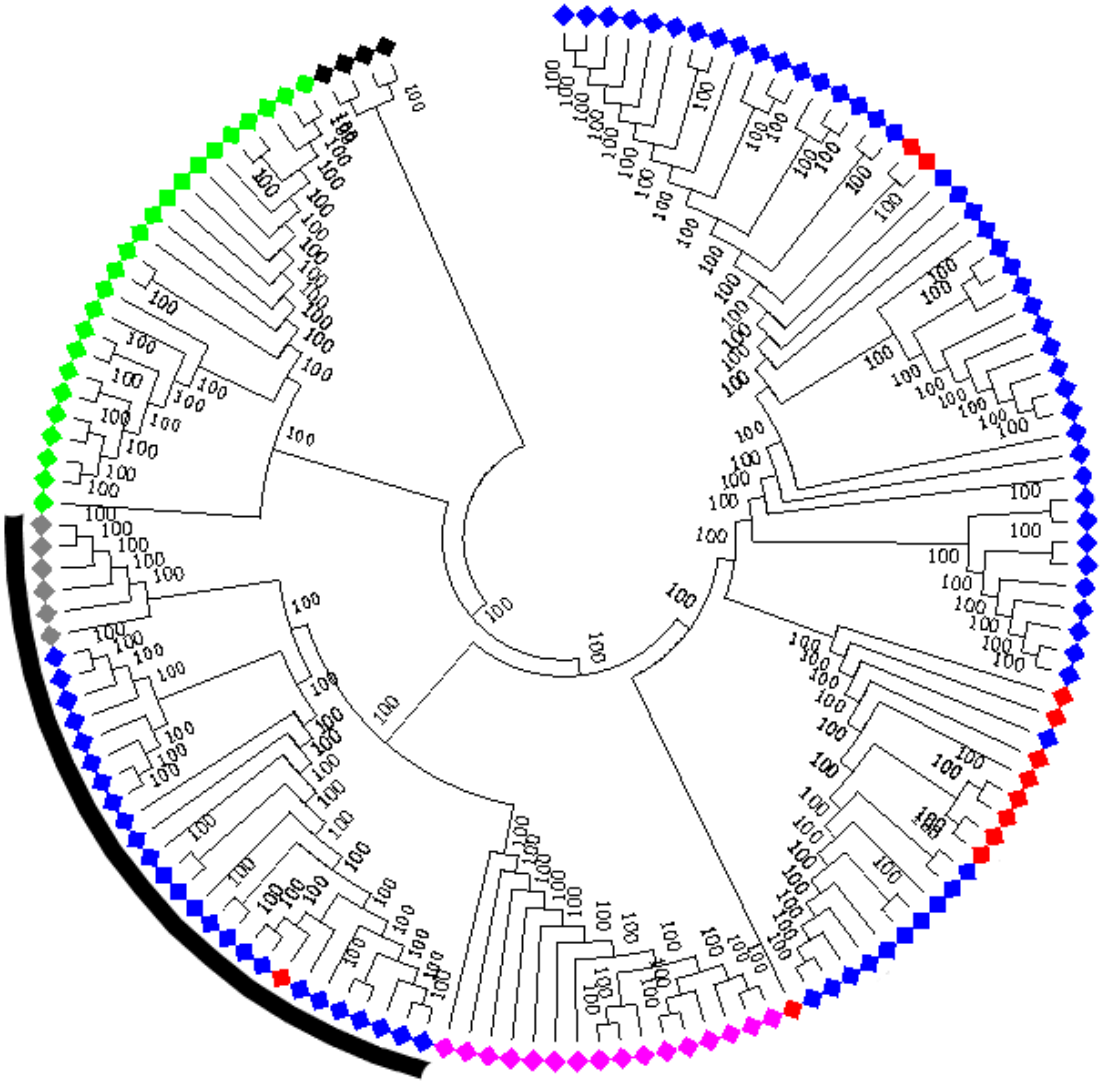


Figure S63: Haplotype maximum likelihood tree for one of six common regions on chr2: 120120000-120400000 of Ethiopian North region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

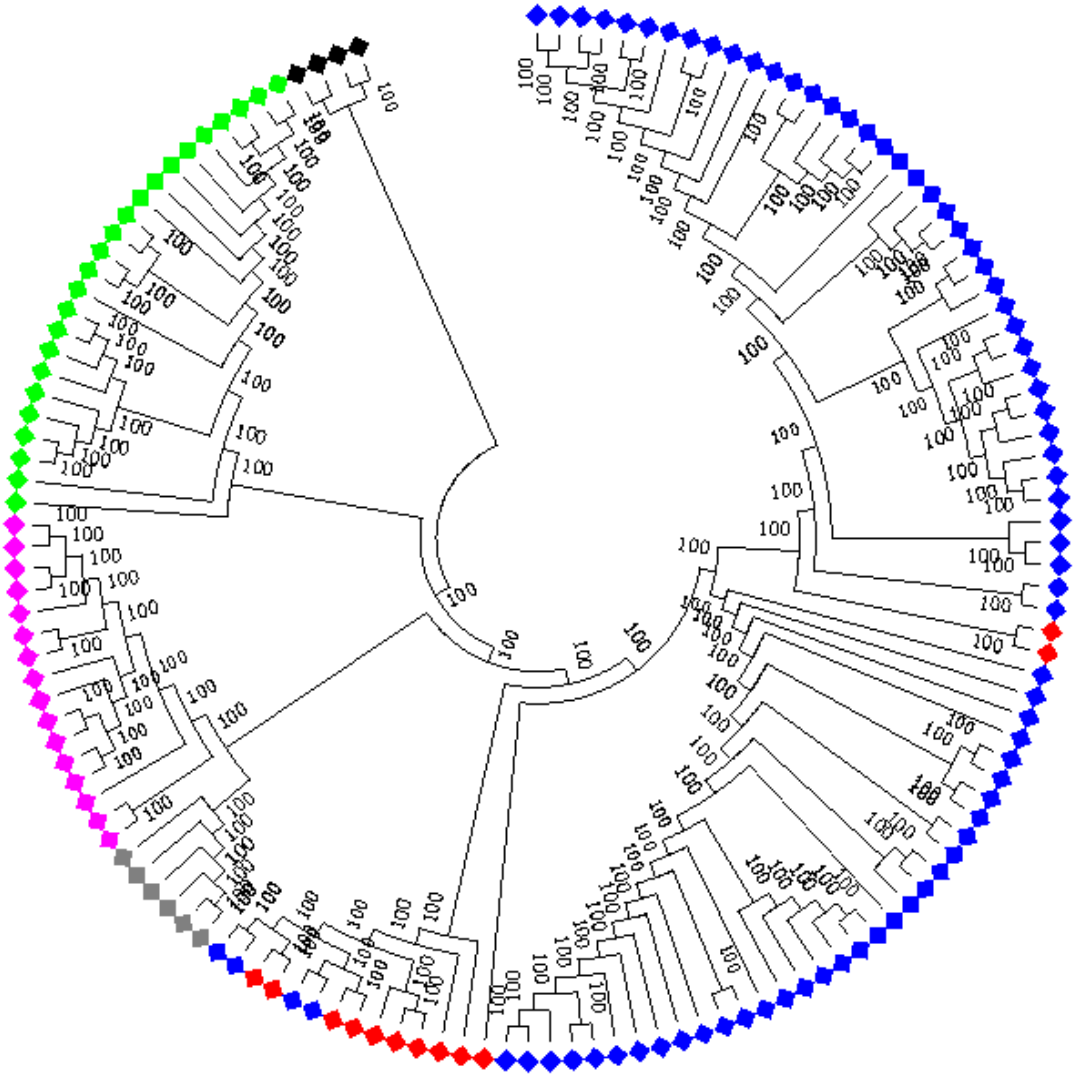


Figure S64: Haplotype maximum likelihood tree for one of six common regions on chr3: 50820000-50920000 of Ethiopian North region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

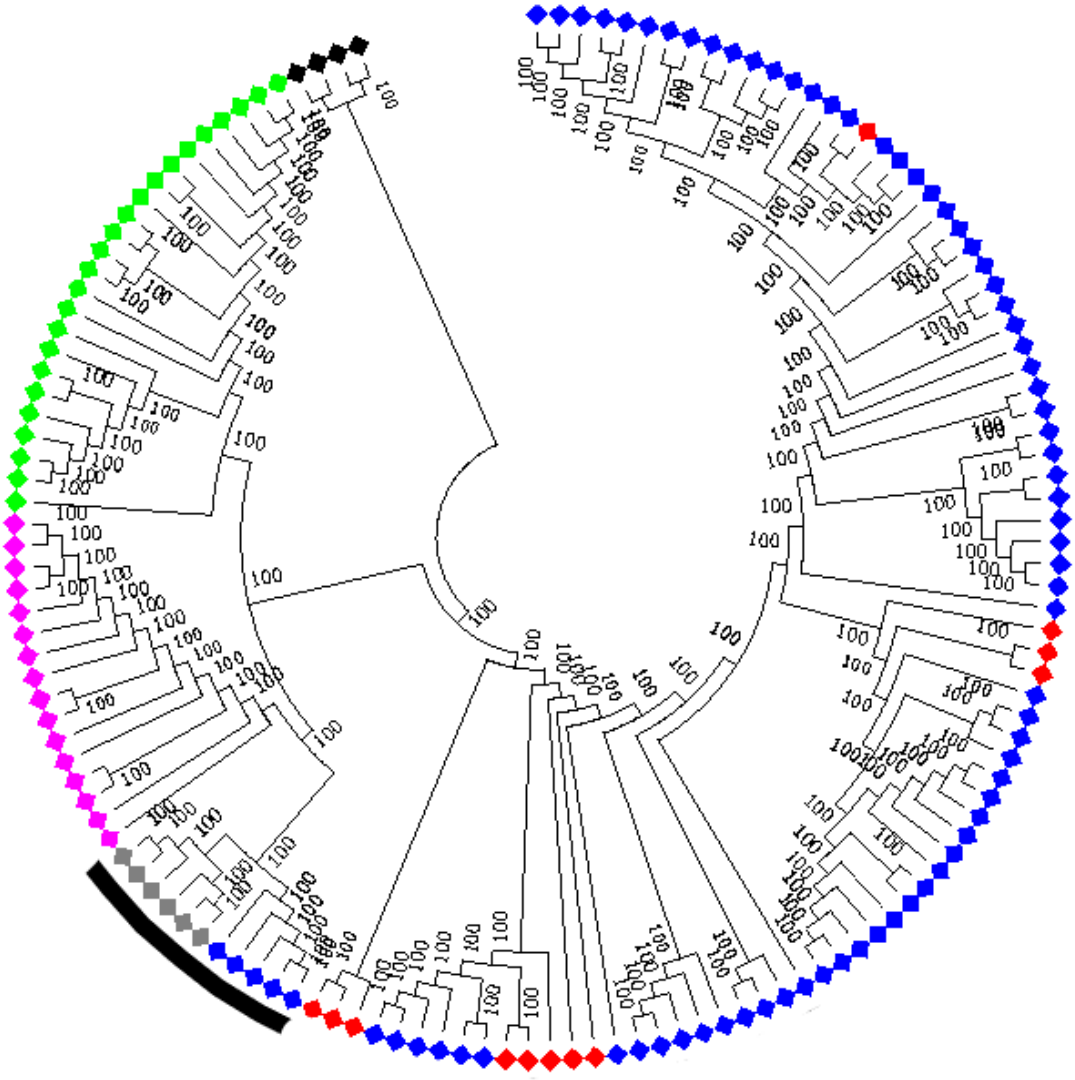


Figure S65: Haplotype maximum likelihood tree for one of six common regions on chr4: 62500000-62700000 of Ethiopian North region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

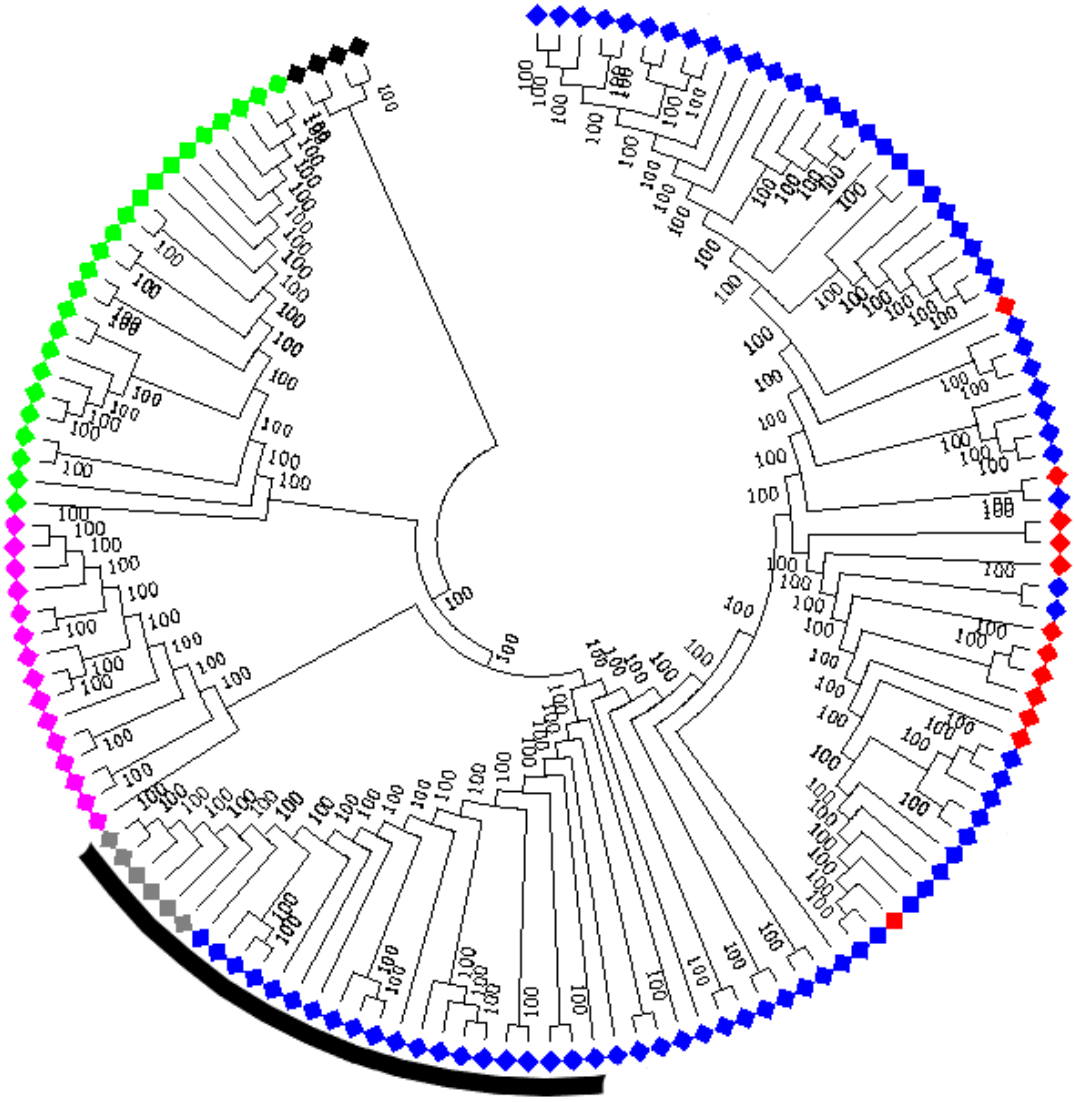


Figure S66: Haplotype maximum likelihood tree for one of six common regions on chr5: 45560000-45840000 of Ethiopian North region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

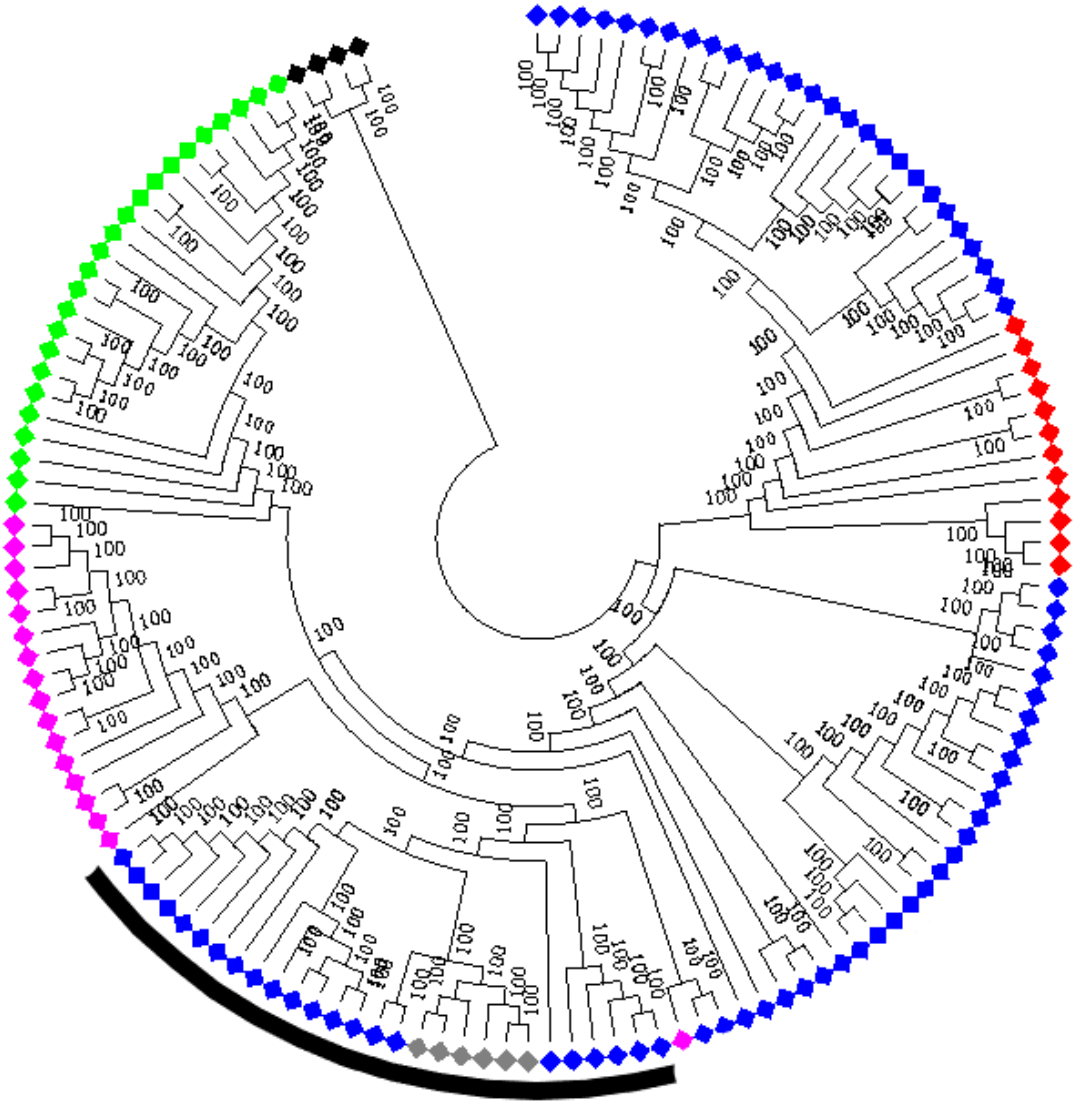


Figure S67: Haplotype maximum likelihood tree for one of six common regions on chr6: 20760000-20880000 of Ethiopian North region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

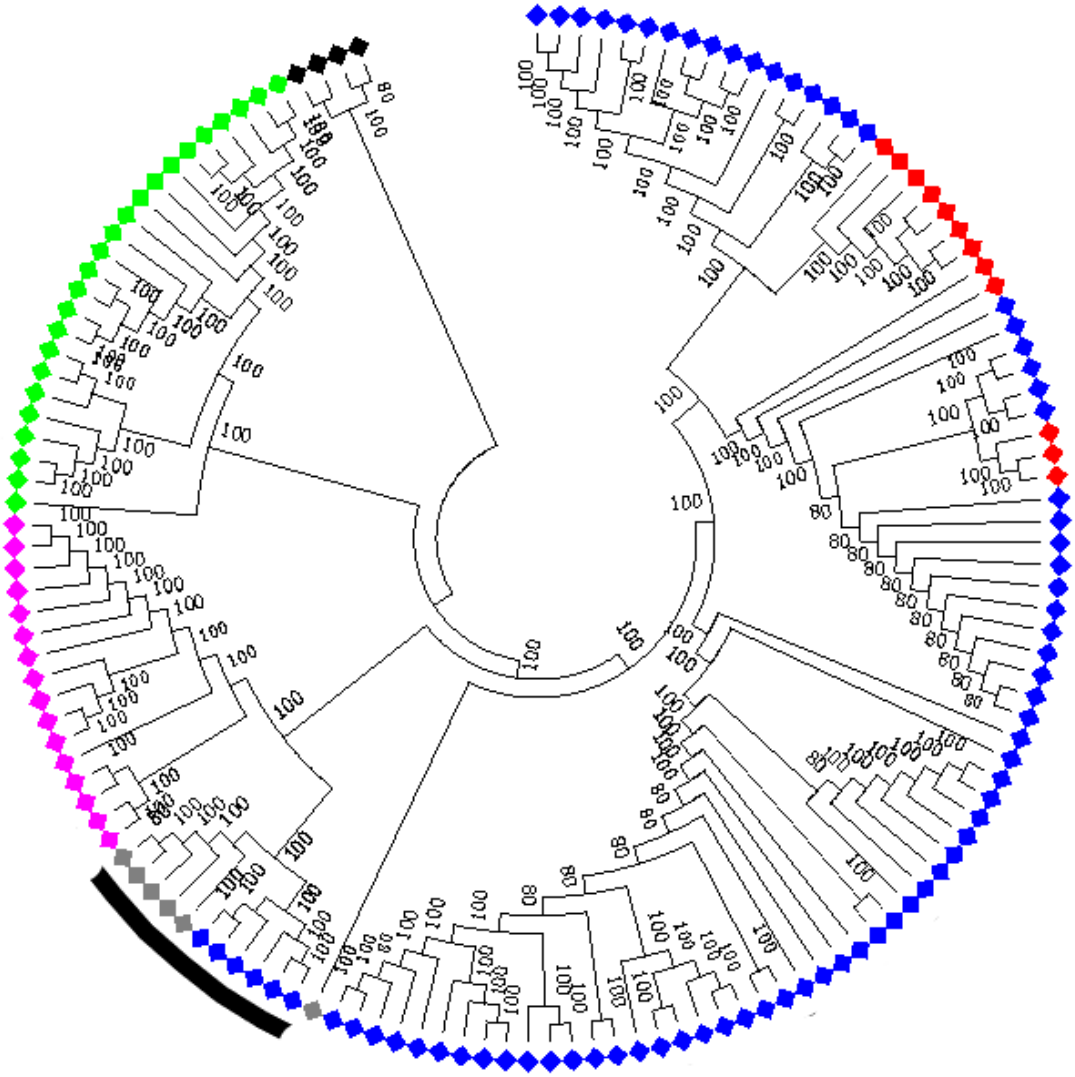


Figure S68: Haplotype maximum likelihood tree for one of six common regions on chr12: 12540000-12640000 of Ethiopian North region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

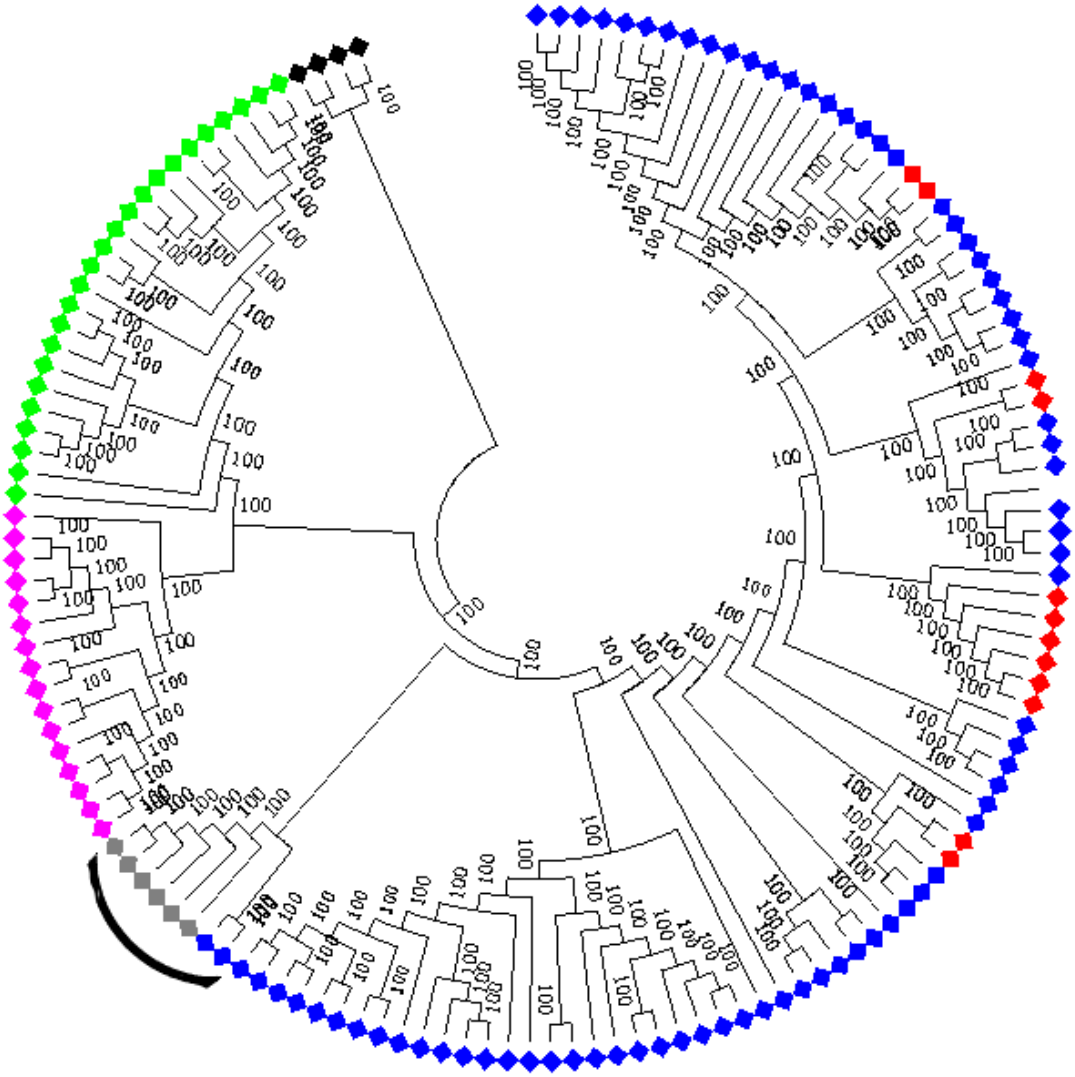


Figure S69: Haplotype maximum likelihood tree for one of six common regions on chr3: 50820000-50920000 of Ethiopian Central-East region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

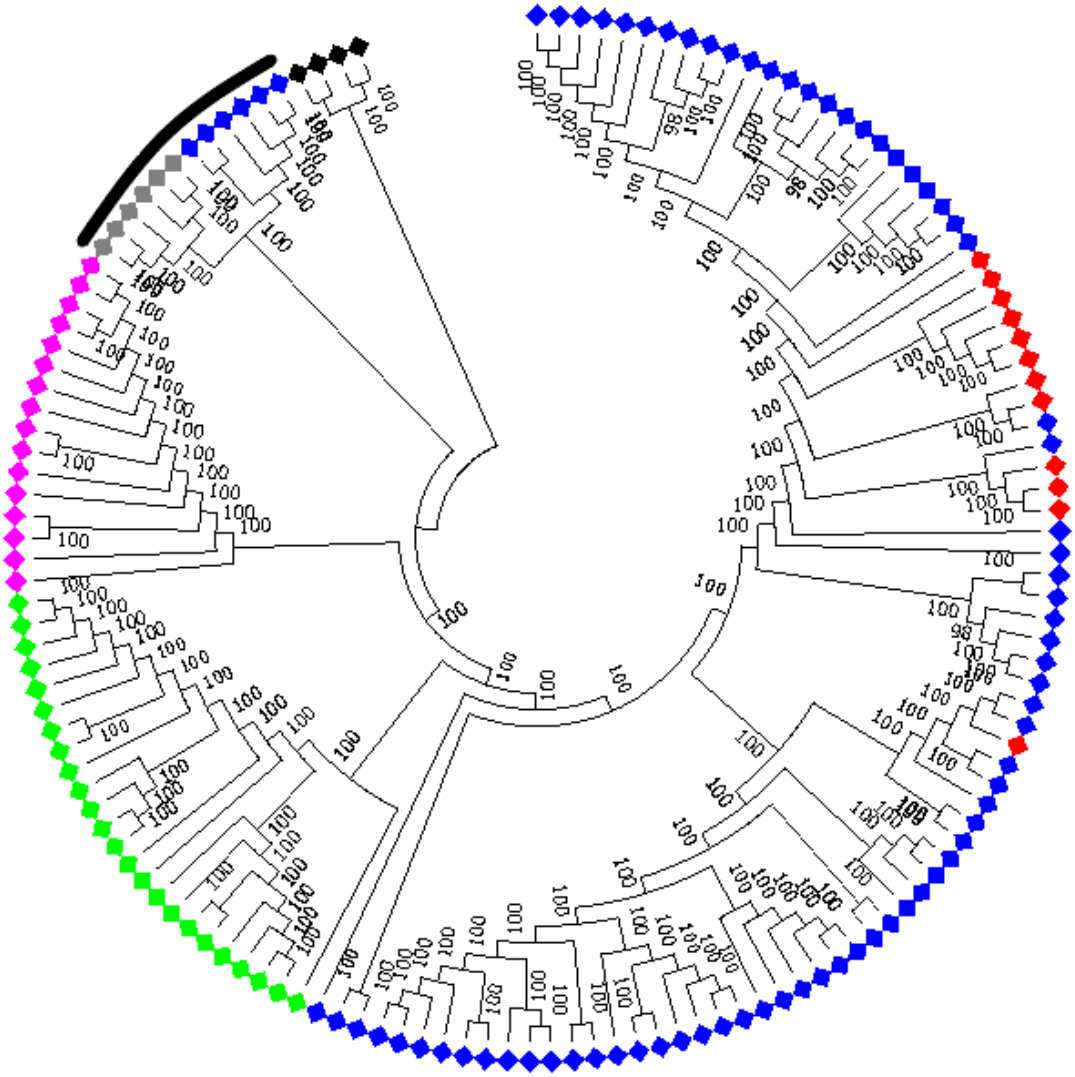


Figure S70: Haplotype maximum likelihood tree for one of six common regions on chr4: 62500000-62700000 of Ethiopian Central-East region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

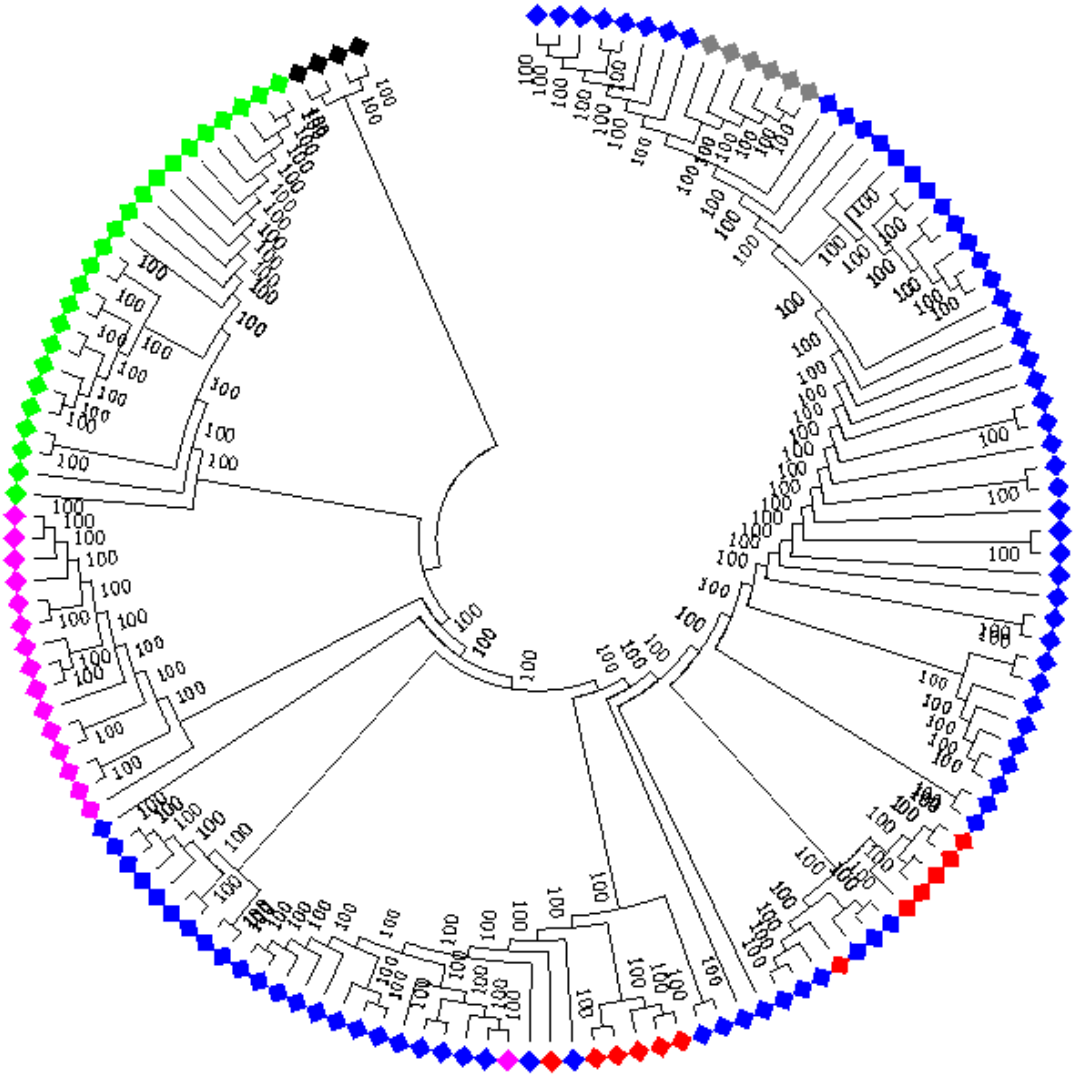


Figure S71: Haplotype maximum likelihood tree for one of six common regions on chr5: 45560000-45840000 of Ethiopian Central-East region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

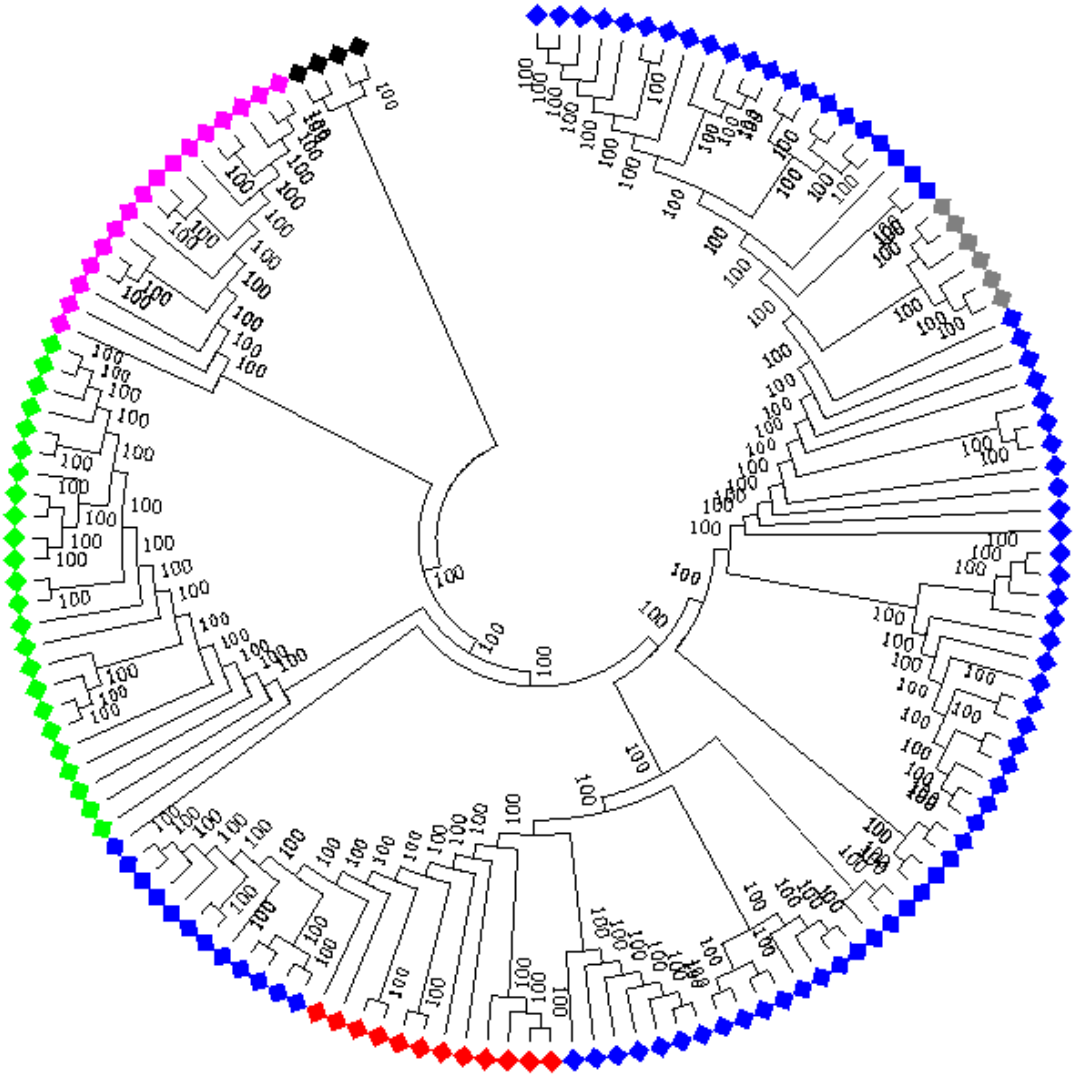


Figure S72: Haplotype maximum likelihood tree for one of six common regions on chr6: 20760000-20880000 of Ethiopian Central-East region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

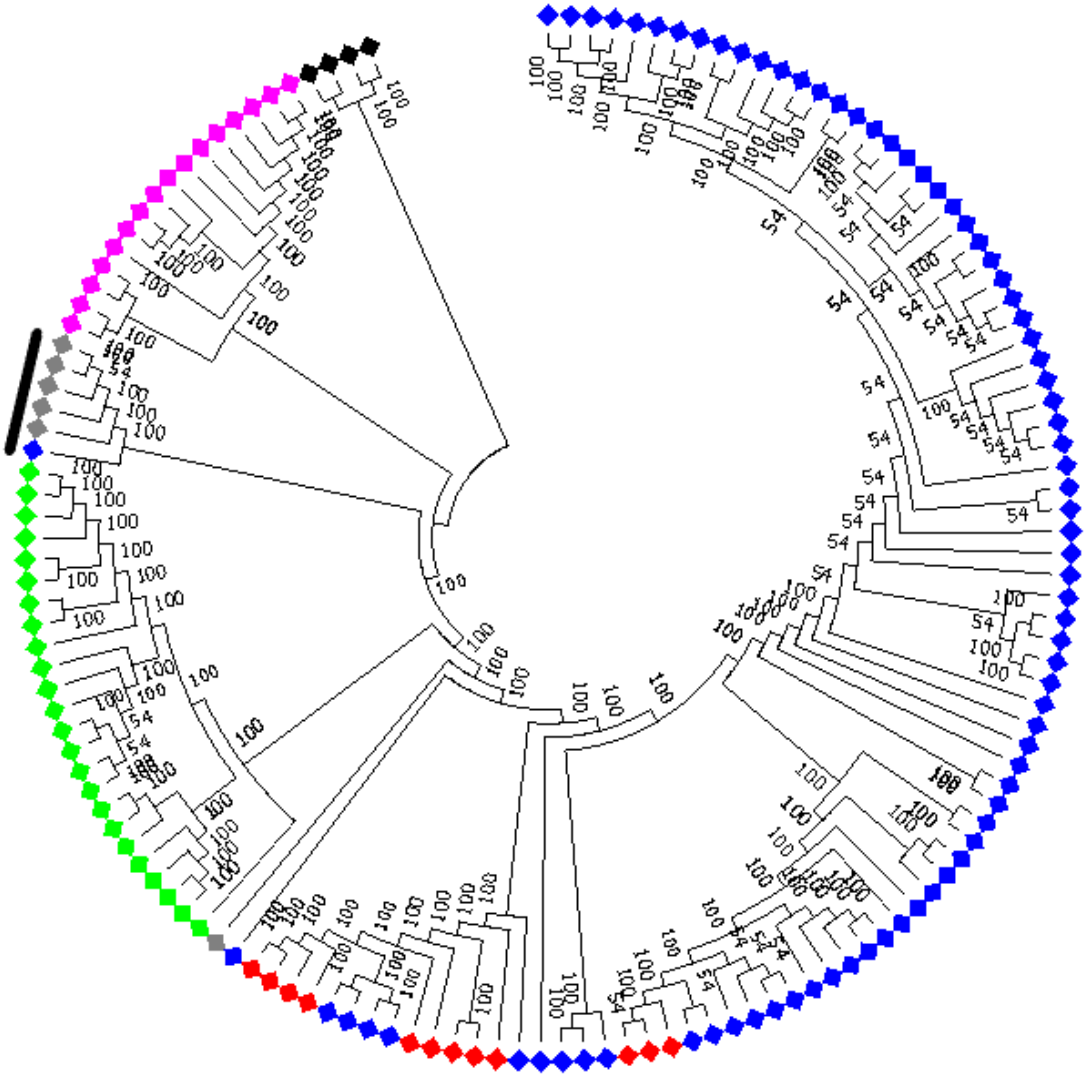


Figure S73: Haplotype maximum likelihood tree for one of six common regions on chr12: 12540000-12640000 of Ethiopian Central-East region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

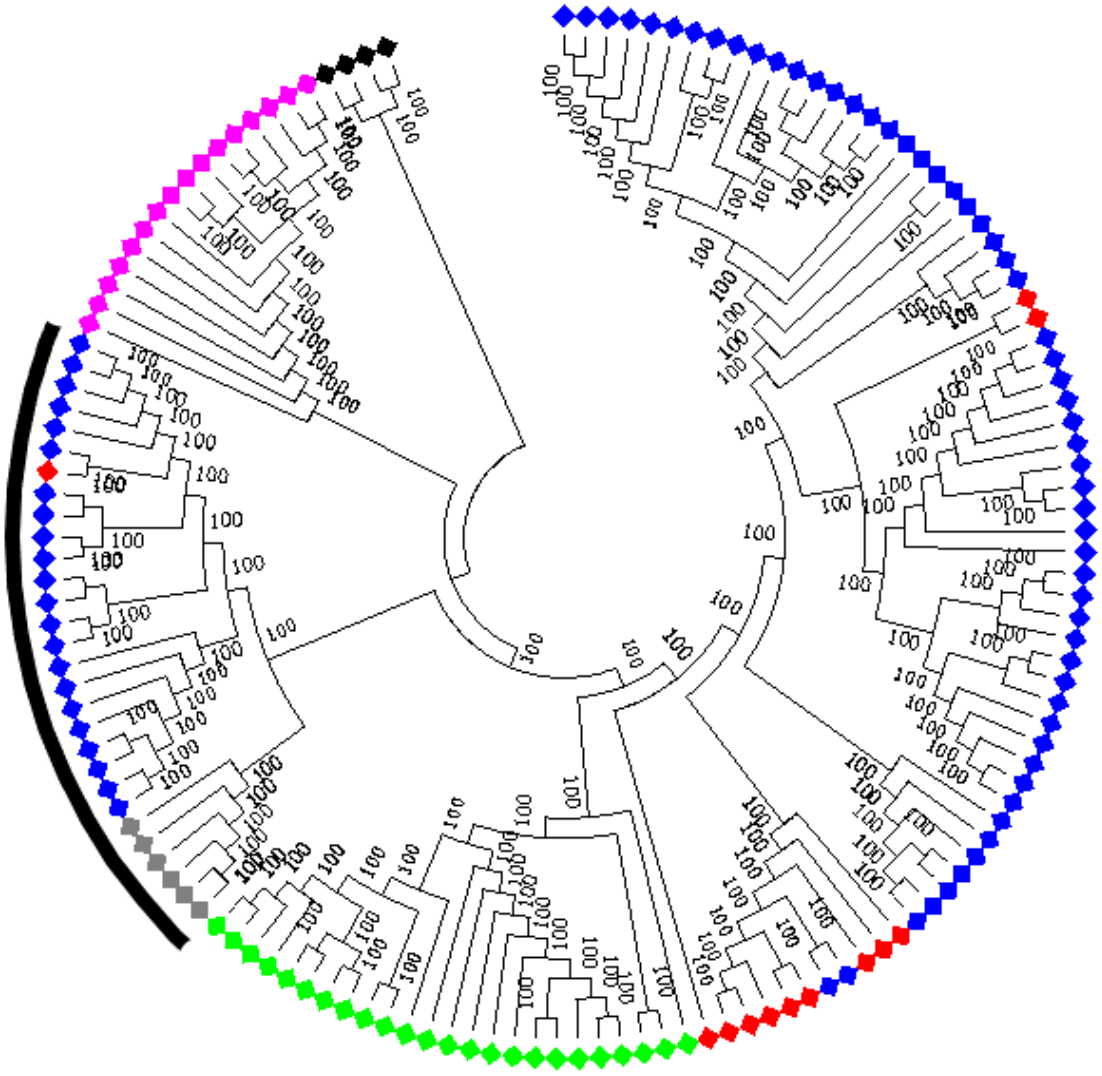


Figure S74: Haplotype maximum likelihood tree for one of six common regions on chr2: 120120000-120400000 of Ethiopian West region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

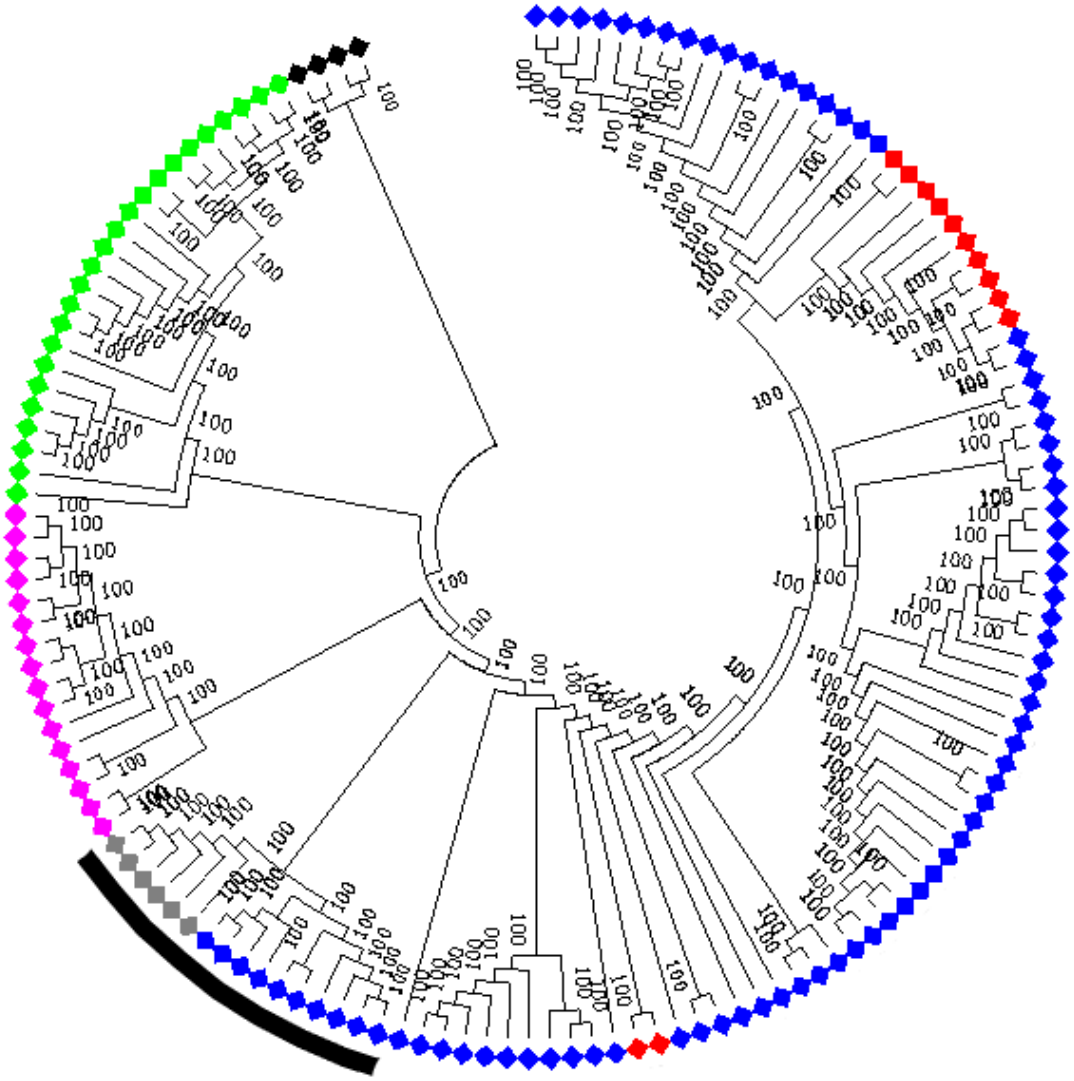


Figure S75: Haplotype maximum likelihood tree for one of six common regions on chr3: 50820000-50920000 of Ethiopian West region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

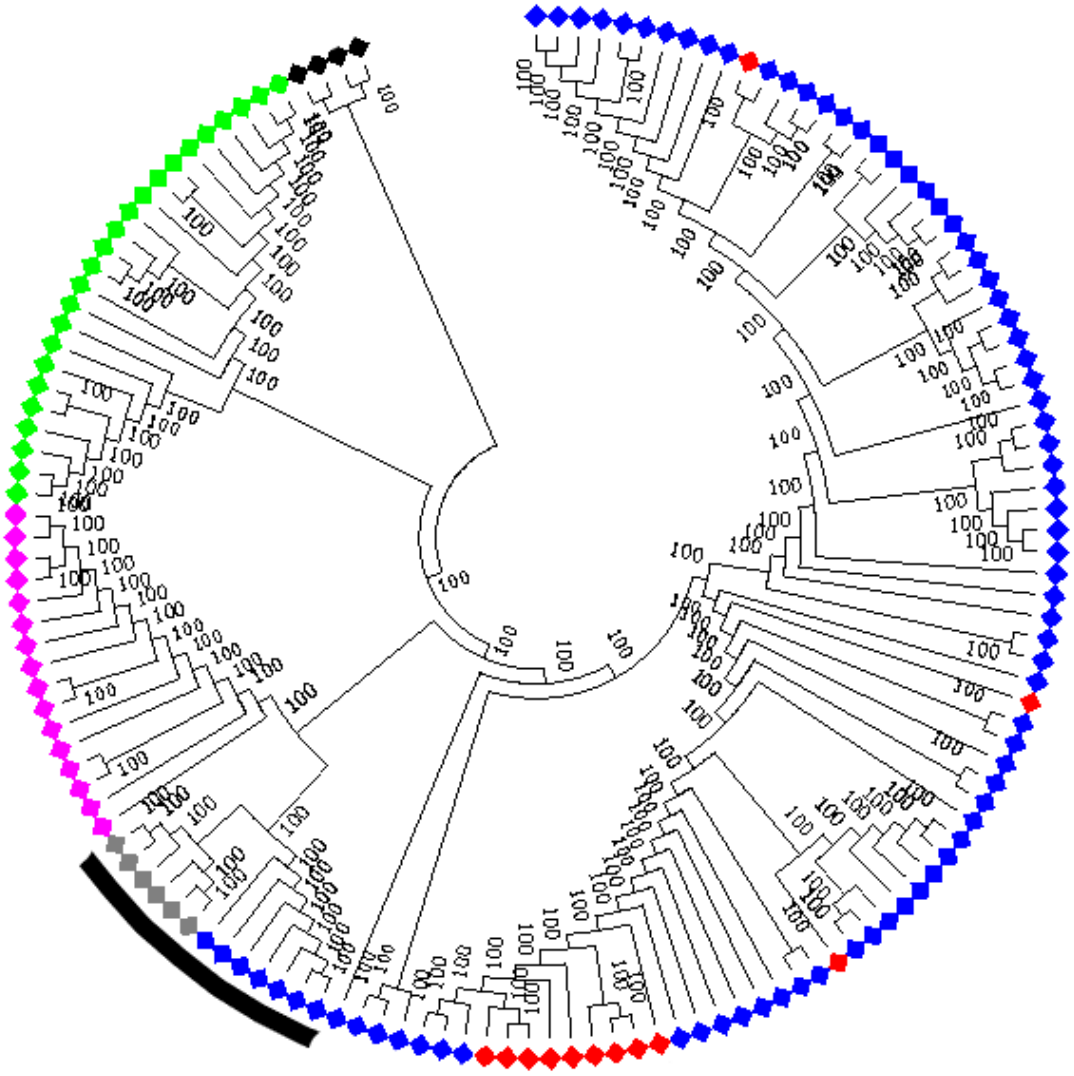


Figure S76: Haplotype maximum likelihood tree for one of six common regions on chr4: 62500000-62700000 of Ethiopian West region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

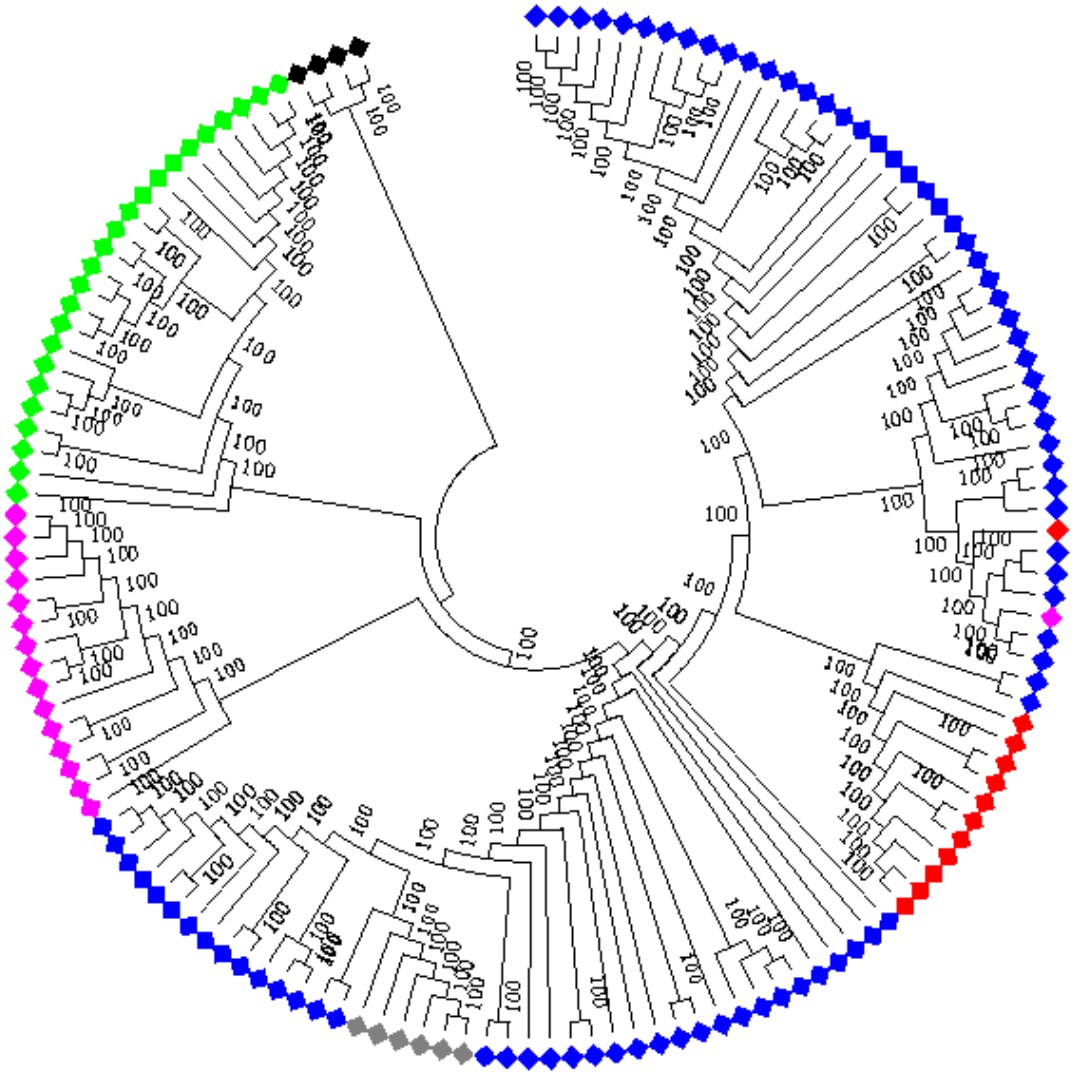


Figure S77: Haplotype maximum likelihood tree for one of six common regions on chr5: 45560000-45840000 of Ethiopian West region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

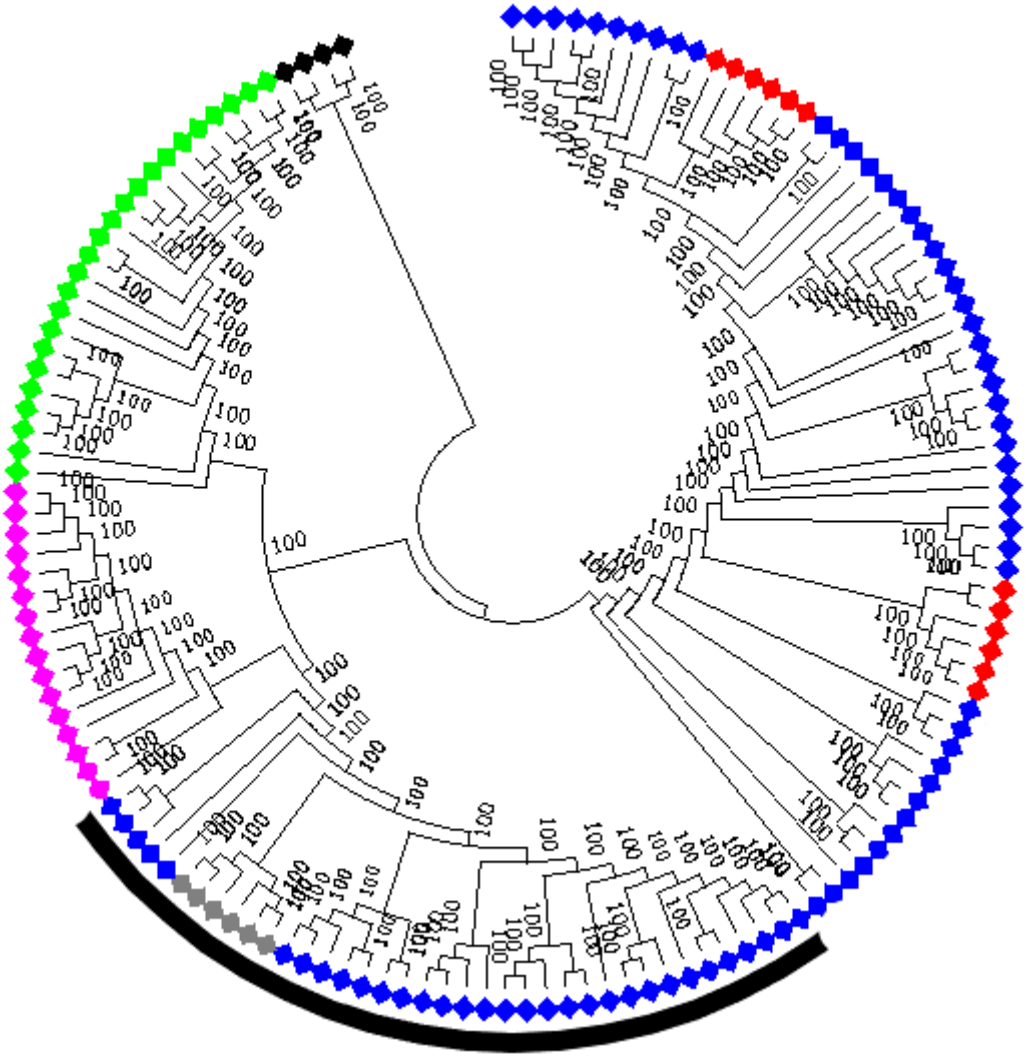


Figure S78: Haplotype maximum likelihood tree for one of six common regions on chr6: 20760000-20880000 of Ethiopian West region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

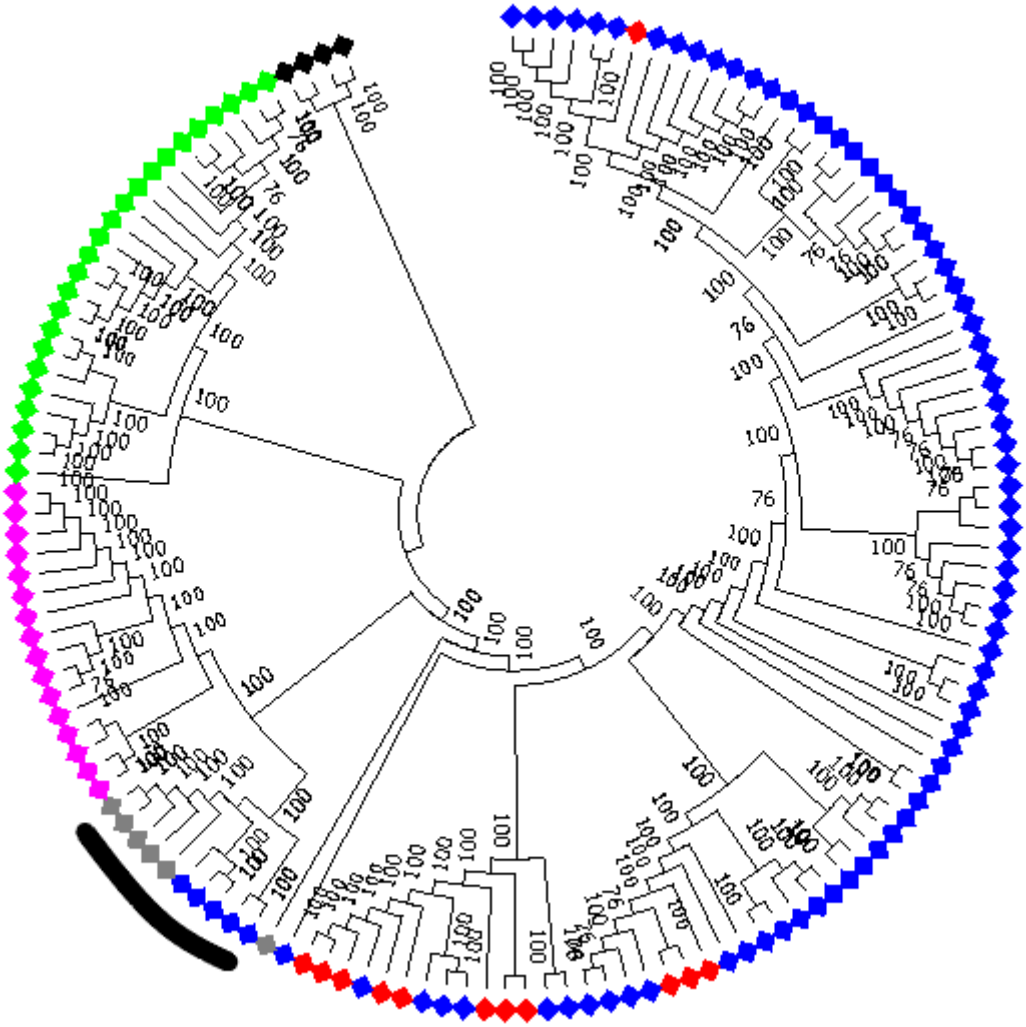


Figure S79: Haplotype maximum likelihood tree for one of six common regions on chr12: 12540000-12640000 of Ethiopian West region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

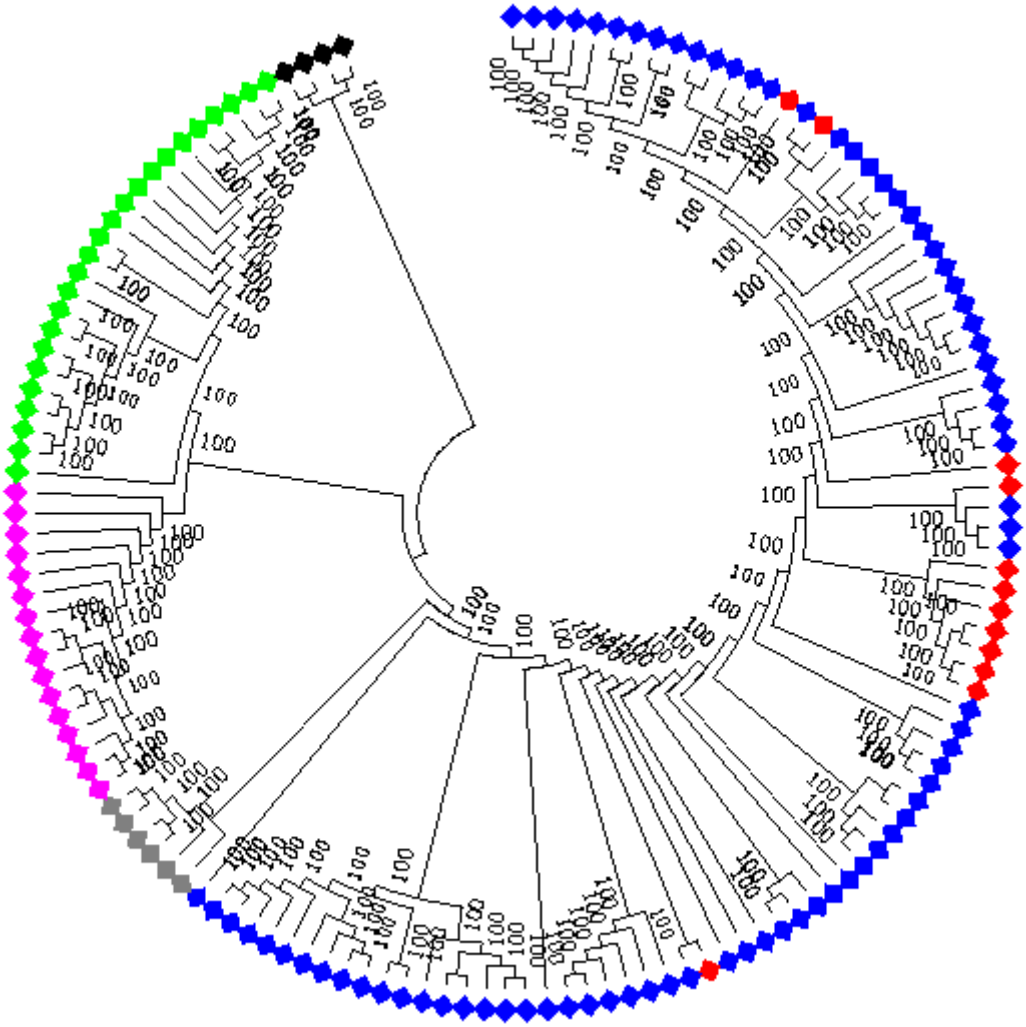


Figure S80: Haplotype maximum likelihood tree for one of six common regions on chr2: 120120000-120400000 of Ethiopian South region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

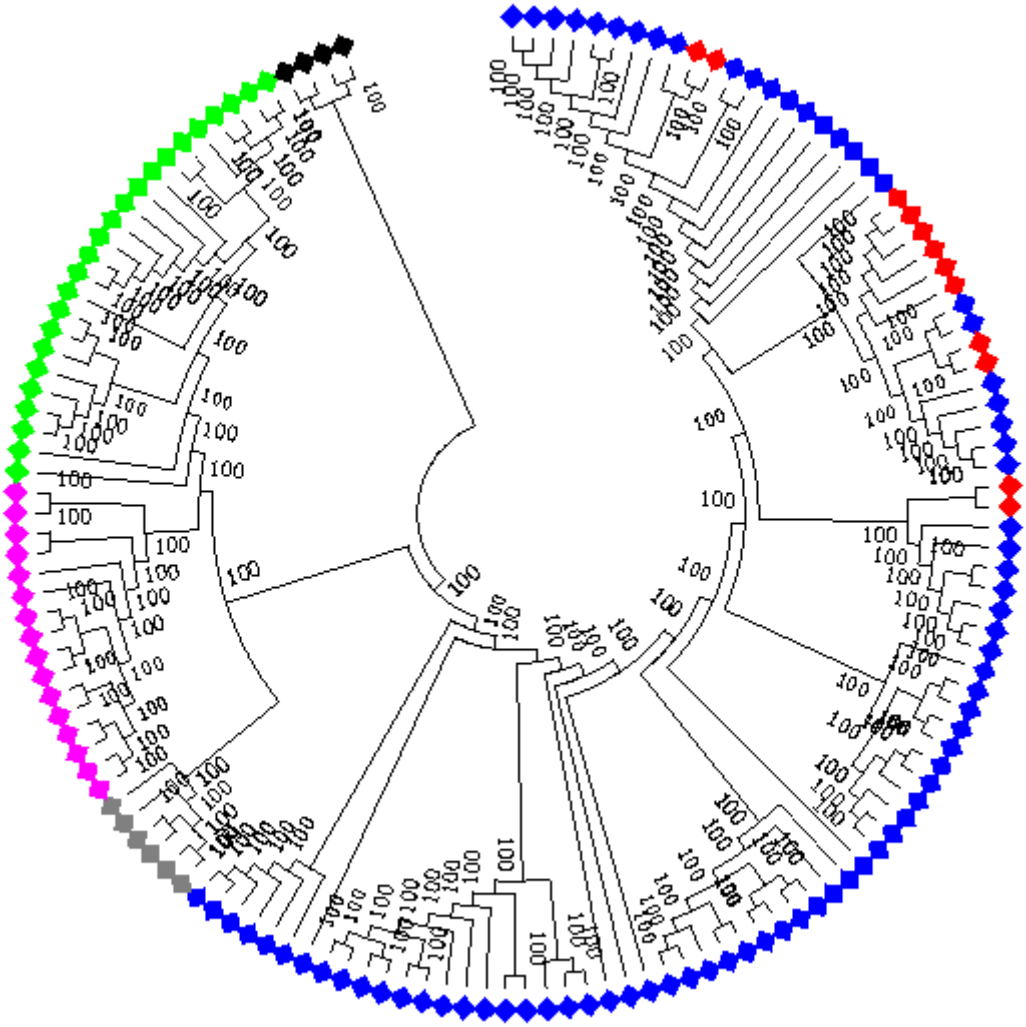


Figure S81: Haplotype maximum likelihood tree for one of six common regions on chr3: 50820000-50920000 of Ethiopian South region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

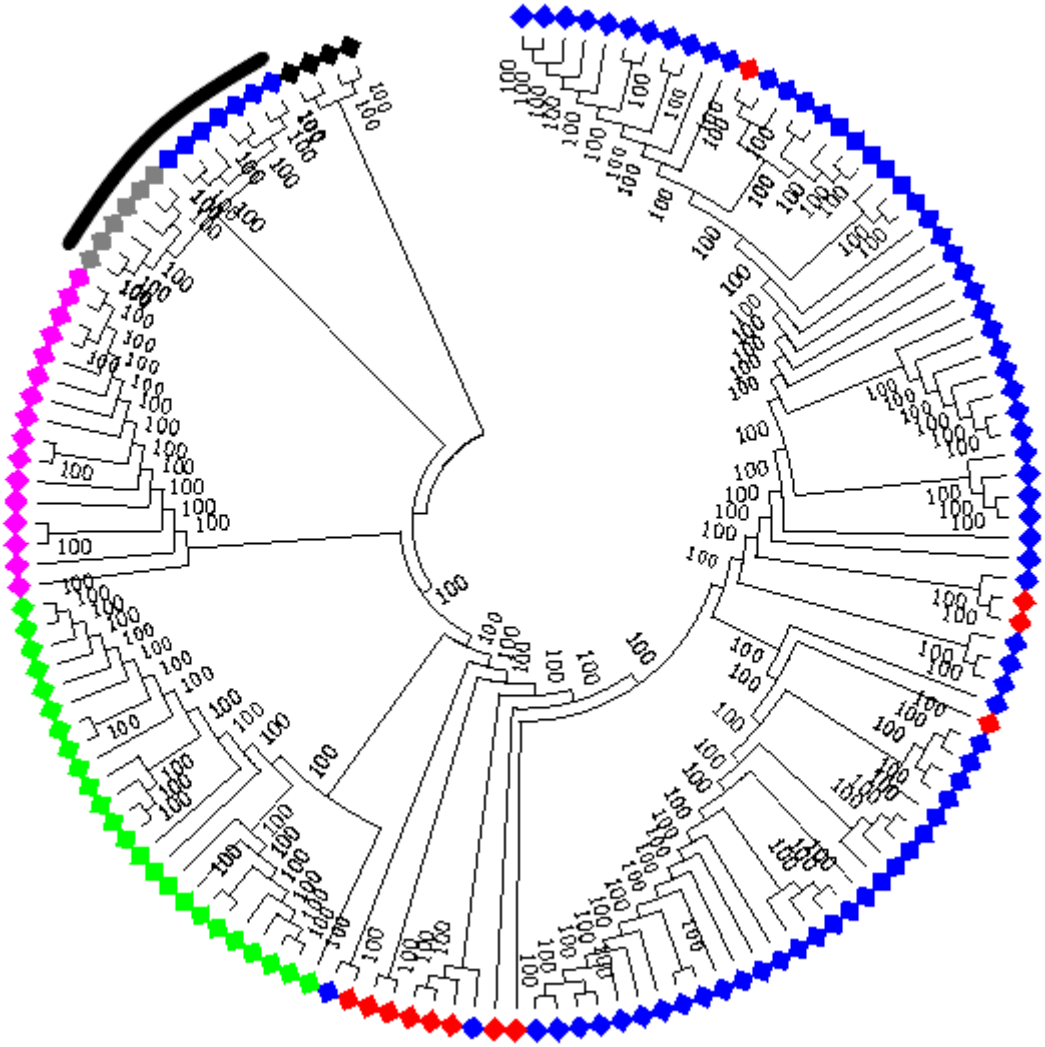


Figure S82: Haplotype maximum likelihood tree for one of six common regions on chr4: 62500000-62700000 of Ethiopian South region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

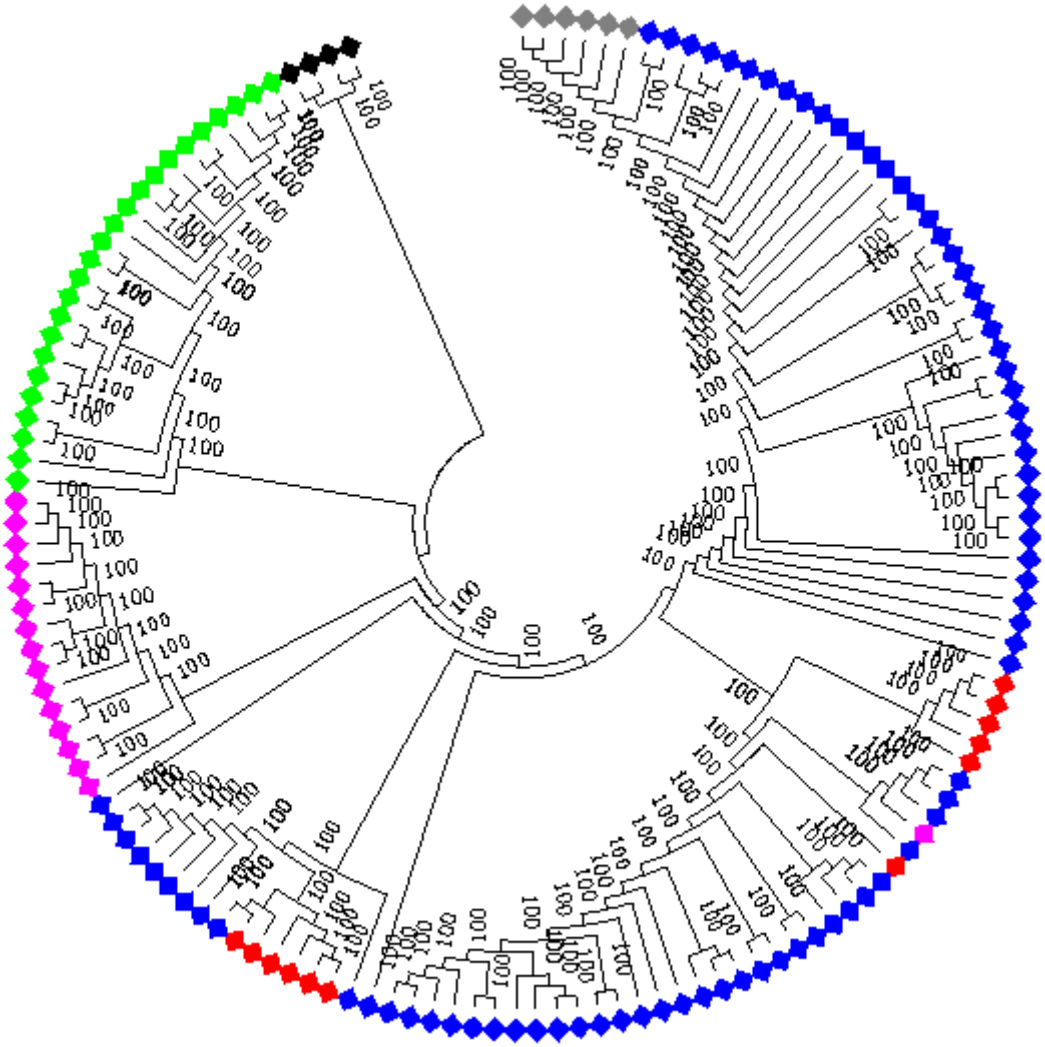


Figure S83: Haplotype maximum likelihood tree for one of six common regions on chr5: 45560000-45840000 of Ethiopian South region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

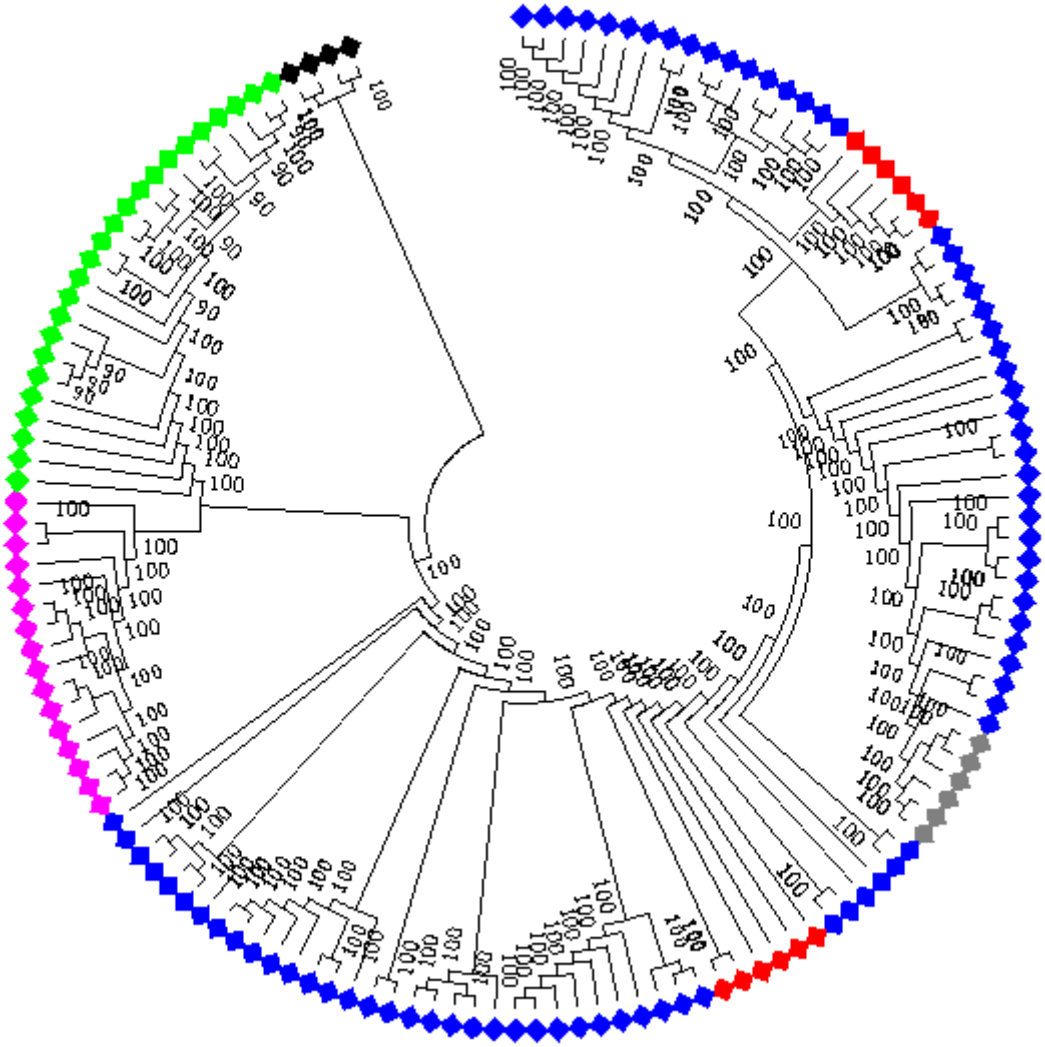


Figure S84: Haplotype maximum likelihood tree for one of six common regions on chr6: 20760000-20880000 of Ethiopian South region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl.

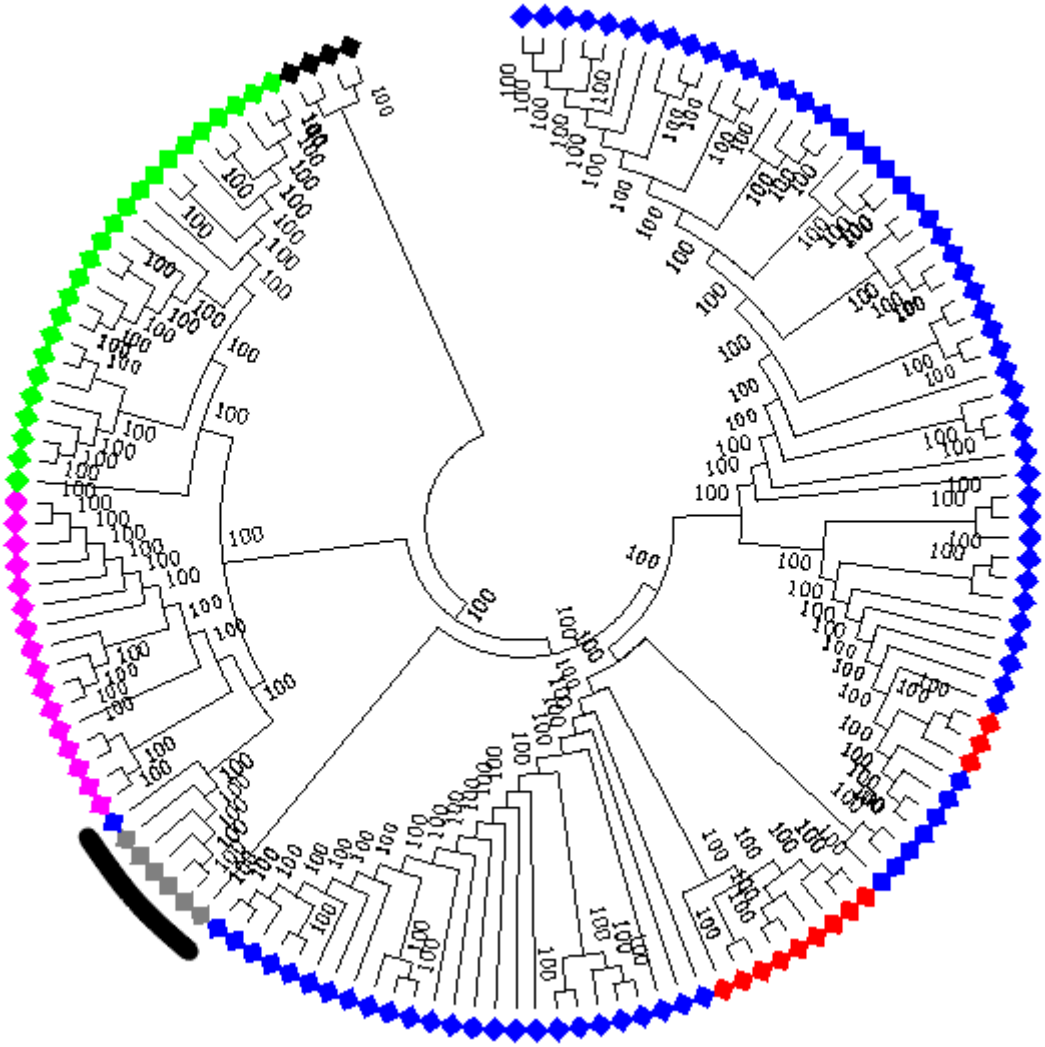


Figure S85: Haplotype maximum likelihood tree for one of six common regions on chr12: 12540000-12640000 of Ethiopian South region domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Ethiopian domestic chicken and Red = Red junglefowl. Introgressed haplotypes from Grey junglefowl to Ethiopian chicken marked by black.

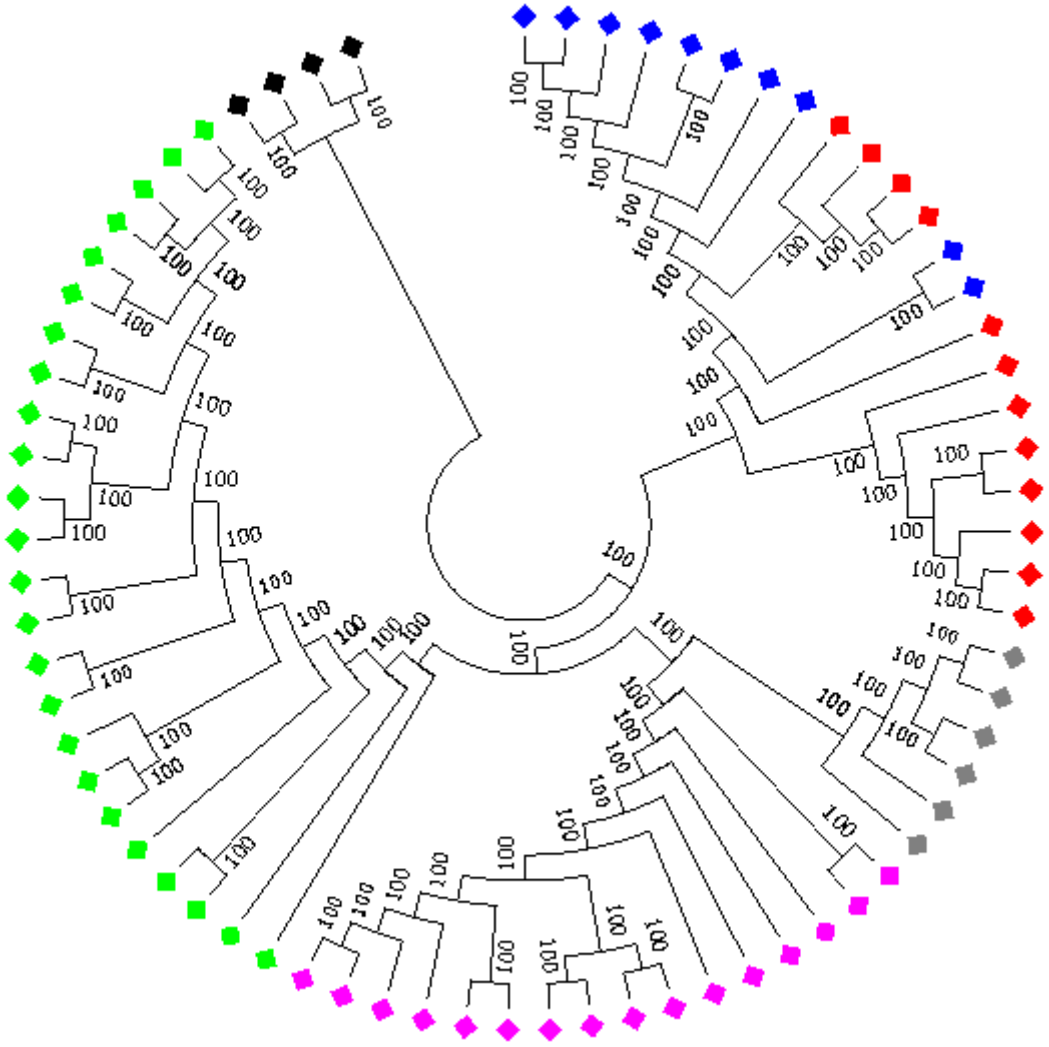


Figure S86: Haplotype maximum likelihood tree for one of six common regions on chr3: 50820000-50920000 of Saudi domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl.

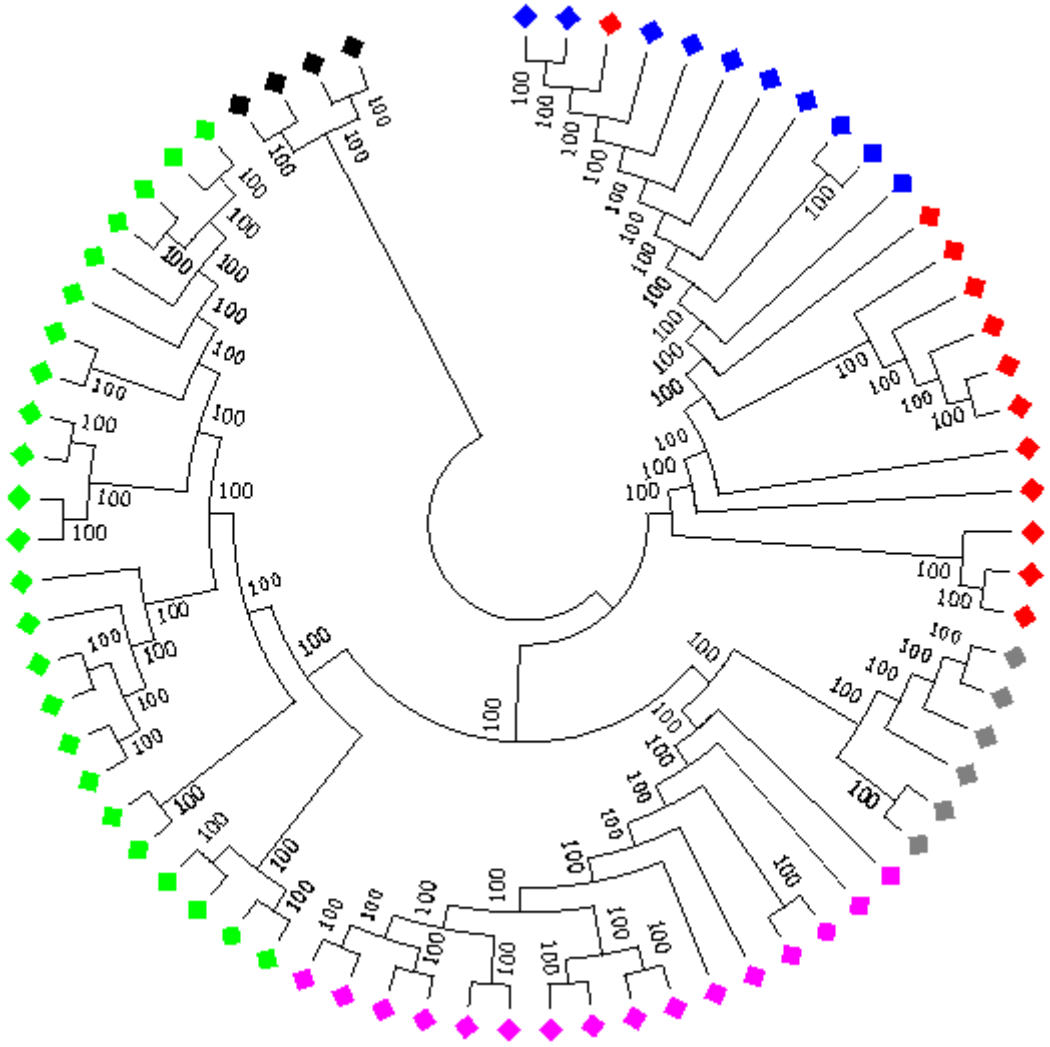


Figure S87: Haplotype maximum likelihood tree for one of six common regions on chr4: 62500000-62700000 of Saudi domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl.

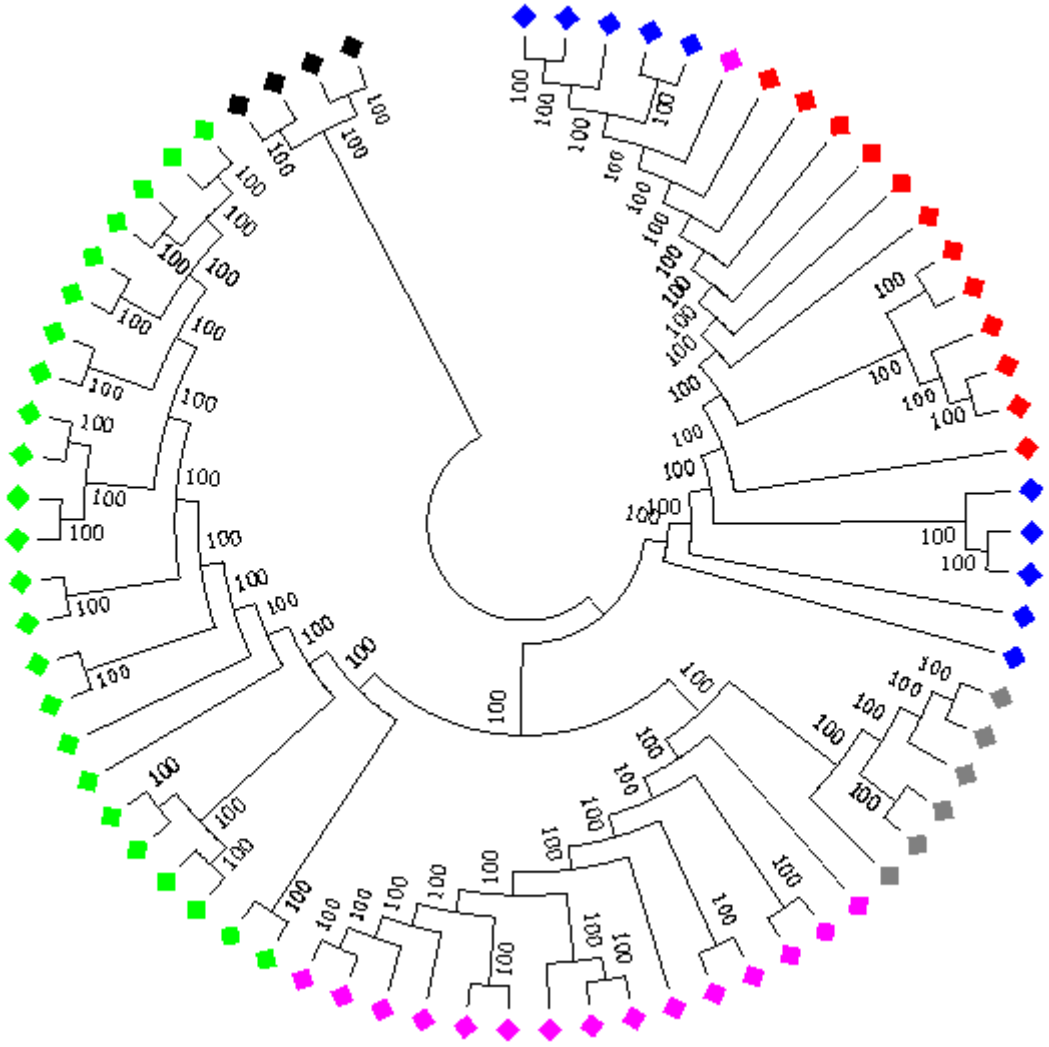


Figure S88: Haplotype maximum likelihood tree for one of six common regions on chr5: 45560000-45840000 of Saudi domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl.

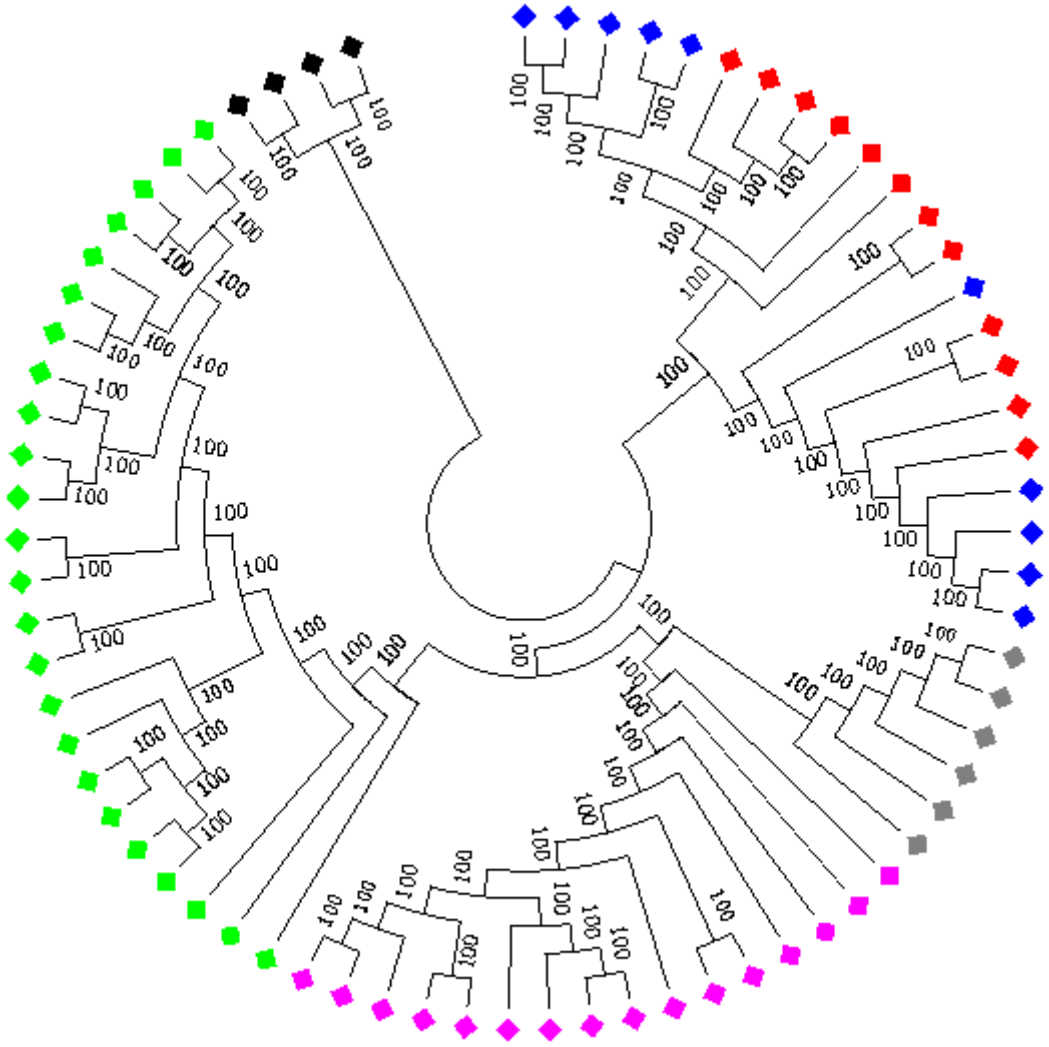


Figure S89: Haplotype maximum likelihood tree for one of six common regions on chr6: 20760000-20880000 of Saudi domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl.

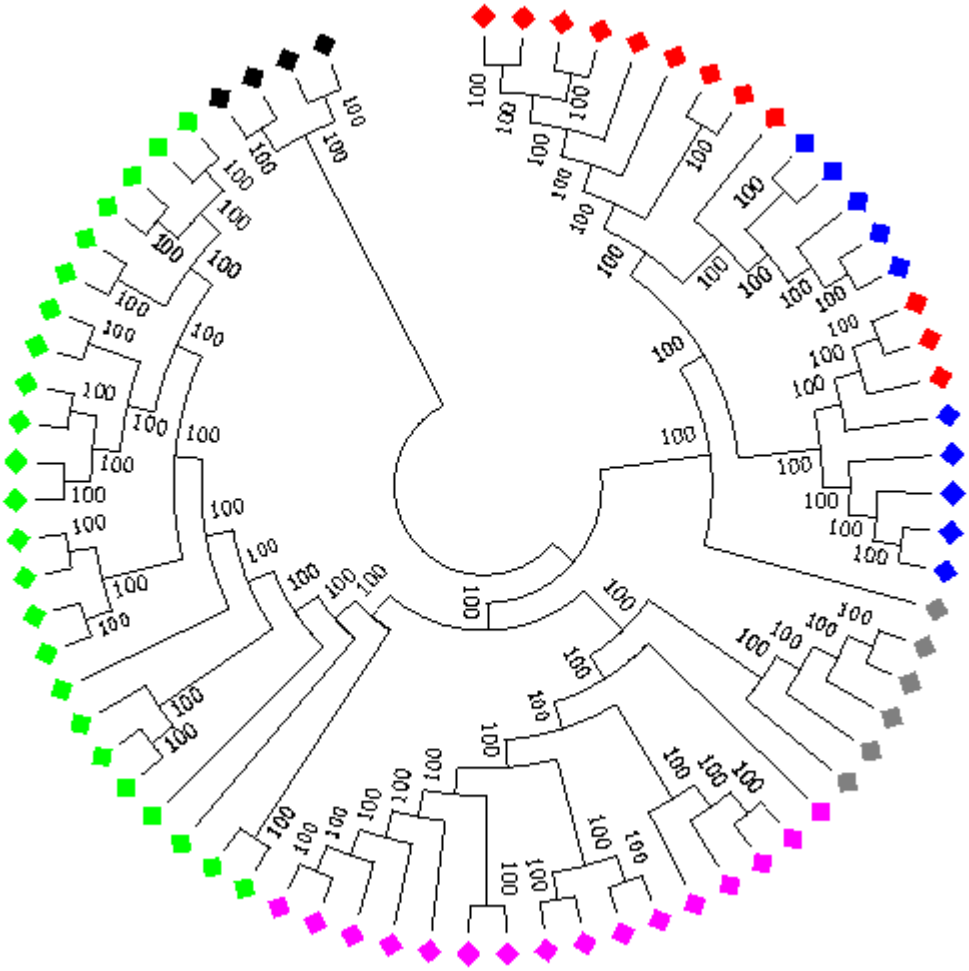


Figure S90: Haplotype maximum likelihood tree for one of six common regions on chr12: 12540000-12640000 of Saudi domestic chicken and Grey junglefowl introgression analysis. Black = Pheasant, Green = Green junglefowl, Pink = Ceylon junglefowl, Grey = Grey junglefowl, Blue = Saudi domestic chicken and Red = Red junglefowl.

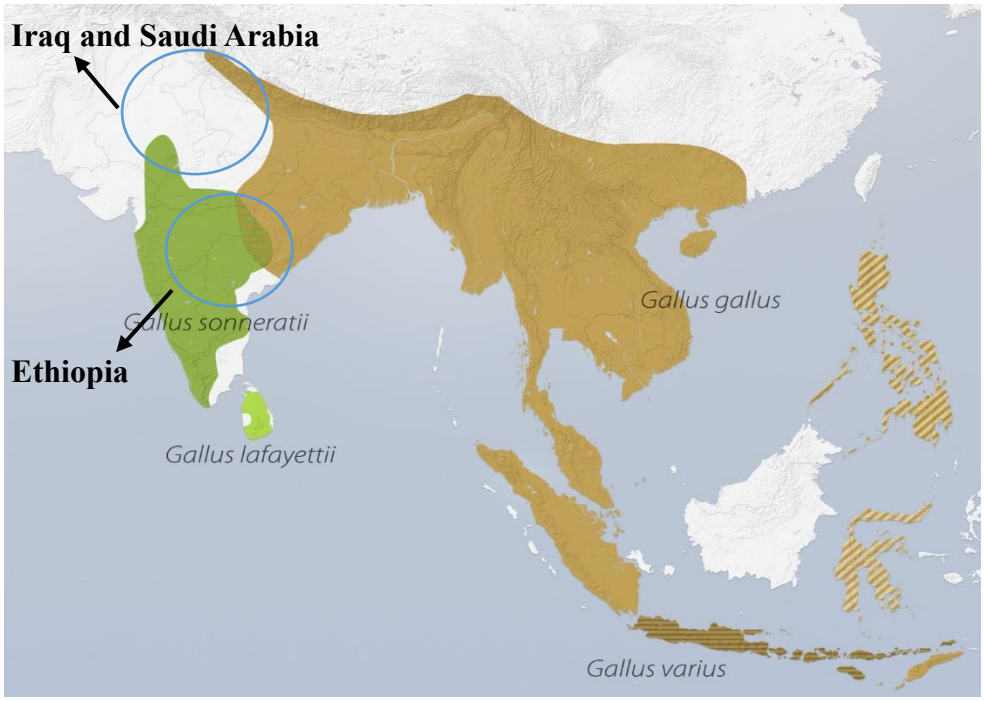


Figure S91: Geographic distribution of *Gallus* species with dispersion routes of chicken to Iraq and Ethiopia. The map was obtained from 'Handbook of the birds of the world alive' (https://www.hbw.com/).