REPUBLIQUE TUNISIENNE

MINISTERE DE L'AGRICULTURE, DES RESSOURCES HYDRAULIQUES ET DE LA PECHE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

INSTITUT NATIONAL AGRONOMIQUE DE TUNISIE

Ecole Doctorale Sciences et Techniques de L'Agronomie et de l'Environnement

THESE DE DOCTORAT EN SCIENCES AGRONOMIQUES

Spécialité : Sciences de la Production Animale

Genetic characterization of dromedary camel (*Camelus dromedarius*) population in Tunisia

Soutenue publiquement par

Ghada NOUAIRIA JEBAH

02 Août 2018 à INAT

Devant le jury composé de

M. Nizar MOUJAHED, Professeur, INAT	Président
M. M Habib YAHYAOUI, Professeur, IRA	Directeur de thèse
M. Abderrahman BEN GARA, Professeur, ESAM	Rapporteur
Mme Sonia BEDHIAF, Professeur, INRAT	Rapporteur
M. M El-houcine OTHMAN, Professeur, CRRA	Examinateur

Genetic characterization of dromedary camel (*Camelus dromedarius*) in Tunisia

ABSTRACT

The dromedary camel (*Camelus dromedarius*) has large socio-cultural and economic importance in the arid lands of Tunisia. This zoogenetic resource has been managed in a traditional mode, with a feeble selective pressure. In the insufficiency of genetic studies describing this population in Tunisia, the present research work affords foundations for the genetic improvement of this livestock species. It offers a comprehensive description of the genetic variability, the structure and the recent history of this population.

The genetic variability was investigated on the autosomal DNA, by amplifying 20 microsatellite markers, and on the mitochondrial DNA, by sequencing two mtDNA regions (the control region and *cytochrome b* gene). The nucleic markers analysis demonstrated a high level of variability and a weak substructure (F_{ST} = 0.02) of the Tunisian dromedary population. The known ecotypes, defined on the basis of phenotypic and morphologic characters, were faintly observed at the genetic level. The population showed a large admixture, close genetic distances and an important gene flow (Nm = 7) between the ecotypes and between the geographic areas. The population also exhibited a moderate inbreeding level (F_{IS} = 0.2).

The mitochondrial DNA analysis allowed the evaluation of the phylogeographic origin of the studied population. It proved that two maternal lineages are present at differential proportions. The most abundant haplotype (> 80 %) was identical to that largely present in most dromedary populations in Asian and African continents. The second lineage (5 %) was more genetically close to haplotypes recorded in the relatively isolated Eastern African dromedary population.

In a final part of this study, we designed an efficient and affordable parentage assay to check camel pedigree recordings. This assay will afford the possibility of tracing back the genealogy of animals of high performance.

Implications of the present genetic evaluation in the camel development plans were addressed. The study emphasized that the cameline genetic resources in Tunisia provide promising potential that ought to be exploited for sustainable food security in arid lands, through strategic and judicious management and development plans.

Key words: *Camelus dromedarius*, genetic variability, genetic resources, microsatellites, mitochondrial DNA, parentage testing.

Caractérisation génétique du dromadaire (*Camelus dromedarius*) en Tunisie

RESUME

Malgré l'importance socio-culturelle et économique des dromadaires (*Camelus dromedarius)*, ces ressources génétiques sont généralement conduites selon un mode d'élevage traditionnel, caractérisé par une faible pression sélective. De plus, les connaissances sur la structure de cette population restent limitées et l'origine et l'histoire de cette population n'ont pas été abordées auparavant. Dans ce cadre, le travail présent consiste à évaluer la diversité génétique, les origines et l'histoire récente des camelins en Tunisie, sur la base d'une analyse moléculaire. Pour ce faire, nous avons utilisé des marqueurs nucléaires (20 microsatellites) et des marqueurs mitochondriaux (la séquence de la région de contrôle et celle du gène *cytochrome b*).

Les analyses basées sur les marqueurs microsatellites ont mis en évidence une variabilité génétique importante au sein de la population étudiée. L'étude a aussi révélé que la diversité génétique observée est expliquée par la différence entre individus et que les écotypes connus chez les éleveurs sont peu différenciés au niveau génétique. Une large proportion du matériel génétique est partagée par les différents écotypes, reflétant fidèlement les pratiques de reproduction non controlée et les zones d'élevage à la fois limitées et rapprochées. Un risque modéré de consanguinité est aussi décelable (F_{IS} = 0.2), ce qui nécessite un travail de sensibilisation chez les éleveurs. Les implications de cette évaluation génétique dans la mise en place de programmes de gestion ainsi que dans la mise en œuvre d'un programme d'amélioration génétique des dromadaires en Tunisie ont été abordées.

L'analyse du polymorphisme de l'ADN mitochondrial a permis de confirmer les résultats sur la diversité génétique et a révélé l'existence de deux lignées maternelles. La lignée abondante provint de l'origine de domestication, le Moyen Orient, la deuxième est génétiquement proche de la population de l'Est de l'Afrique.

Enfin, vu l'absence des pratiques de contrôle et de suivi dans la plupart des troupeaux de dromadaires en Tunisie, on a développé un test de parenté efficace, basé sur les marqueurs microsatellites. Ce test permettra le traçage des relations généalogiques, notamment des individus à potentiel productif important.

Mots clés : *Camelus dromedarius*, variabilité génétique, ressources génétiques, microsatellites, ADN mitochondrial, test de paternité.

التّوصيف الوراثي للجمل العربي بتونس

ملخَّص

يتميّز الجمل العربي وحيد السّنام (*Camelus dromedarius*) بتكيّفه مع الطّبيعة الجافّة والقاحلة وقدرته غير المسبوقة على تحمّل العطش والتّنقّل على الرّمال. و لم يحض هذا الحيوان، على اهميّته، بالقدر الكافي من دراسة مخزونه الوراثي حيث تعتبر التطوّرات في مجال البحوث الجينيّة في الإبل محدودة مقارنة مع بقيّة الحيوانات المربّاة.

نتطرّق في هذه الدّراسة الى الخصائص الوراثيّة للجمل العربي بتونس من خلال نوعين من الواسمات الوراثيّة: الواسمات النواتيّة من نوع الواسمات الوراثيّة الصغريّة (marqueurs microsatellites) و قراءة تسلسل الحمض النووي للميتوكندريا (mitochondrial DNA) في منطقة توجيه التّناسخ الجيني (control region) و جين السيتوكروم ب (*cytochrome b*).

خلصت هذه الدراسة الى تأكيد التّنوّع الوراثي الكبير للجمل بتونس من خلال الإحصائيّات الجينية. أمّا دراسة هيكلة الجمال بتونس، فقد أظهرت أنّ هذا التّنوّع لا يعكس بدقّة الأصناف المعروفة لدى المربّين (المرزوقي، العرضاوي، الترقي، الغيلوفي، والقعودي) والّتي تمّ توصيفها حسب صفاتها الظّاهرة.

و من جهة أخرى، فإنّ الجمال الموجودة اليوم في تونس تحمل الوصمة الجينيّة لاثنين من أصول نسب الأمومة (maternal lineage).

النّسب الأول، و هو الأكثر انتشارا في تونس، يتطابق مع الوصمة المرجع لجينوم الميتوكندريا للجمل العربي (Reference mitochondrion genome) والمنتشرة في مختلف قطعان الجمل العربي في العالم. نسب الأمومة هذا يرجع غالبا الى مصدر تدجين الجمل وهو منطقة الخليج العربي وهذا يؤكّد ما وُصف في دراسات سابقة بأن الجمل العربي المتواجد اليوم يرجع إلى بؤرة تدجين واحدة. أما النّسب الثّاني فهو متواجد بصفة أكبر في منطقة إفريقيا الشّرقيّة حيث تطوّر الجمل بصفة منعزلة نسبيّا عن بقيّة المناطق.

أخيرا، يقدّم هذا العمل وسيلة تطبيقيّة لتجاوز معضلة غياب متابعة عمليّات التّزاوج في معظم قطعان الجمال في تونس وذلك بتوفير اختبار الحمض الجيني للنّسب ممّا يمكّن من تتبّع أصول الجمال ذات القيمة الجينيّة المرتفعة.

الكلمات المفاتيح: الجمل وحيد السنام، الواسمات الجينيّة، نسب الأمومة، الواسمات الوراثيّة الصغريّة، اختبار الحمض النّووي للنّسب.

إهداء

إلى من أدين لها بحياتي وأكثر — زليخا

إلى من آمن بي وساندني منذ ولادتي — عـلّالة

إلى من أناصفه هذا العمل كما أناصفه حياتي — عبد الواحد

إلى من ضحّى معي منذ رأى النّور — أحمد يوسف

إلى أخواتي بجيناتٍ مختلفة — سميرة، مفيدة، نزيهة، إيمان وعائشة

ACKNOWLEDGEMENTS

In the first place, I would like to thank my supervisor **Dr. Mohammed Habib YAHYAOUI** for planning the present research work and for his immense guidance, support and input over the past five long years. I acknowledge the financial support of the Laboratory of Livestock and Wild Life (LEFS) in the Arid Lands Institute in Medenine (IRA).

The experimental chapters would not have been possible without the input of several persons from the different institutes I worked in. Data and part of the samples were available from the LEFS. I thank everyone who contributed directly or indirectly to the realization of this research. By this occasion, I address my gratefulness to **Pr. Houcine KHATTELI**, director of IRA Medenine, **Pr. Mohammed HAMMADI** and **Pr. Touhami KHORCHANI,** current and former head of the LEFS. In this laboratory, I found guidance, help, and human qualities from each and every professor, staff member and fellow. I recognize their assistance and their moral support.

I would like to address my sincere gratefulness to the members of the jury; **Pr. Nizar MOUJAHED**, president of the jury, **Pr. Abd Rahman BEN GARA** and **Pr. Sonia BEDHIAF** and **Pr. Mohamed El-houcine OTHMAN** for agreeing to give a critical reviewing to this humble work and for contributing to its finalization.

A part of this research work has been performed and financed by the Department of Agronomy, Food, Natural resources, Animals and Environment (DAFNAE) in the University of Padova. I would like to thank **Pr. Martino CASSANDRO** for his availability and guidance in my six month stay under his supervision.

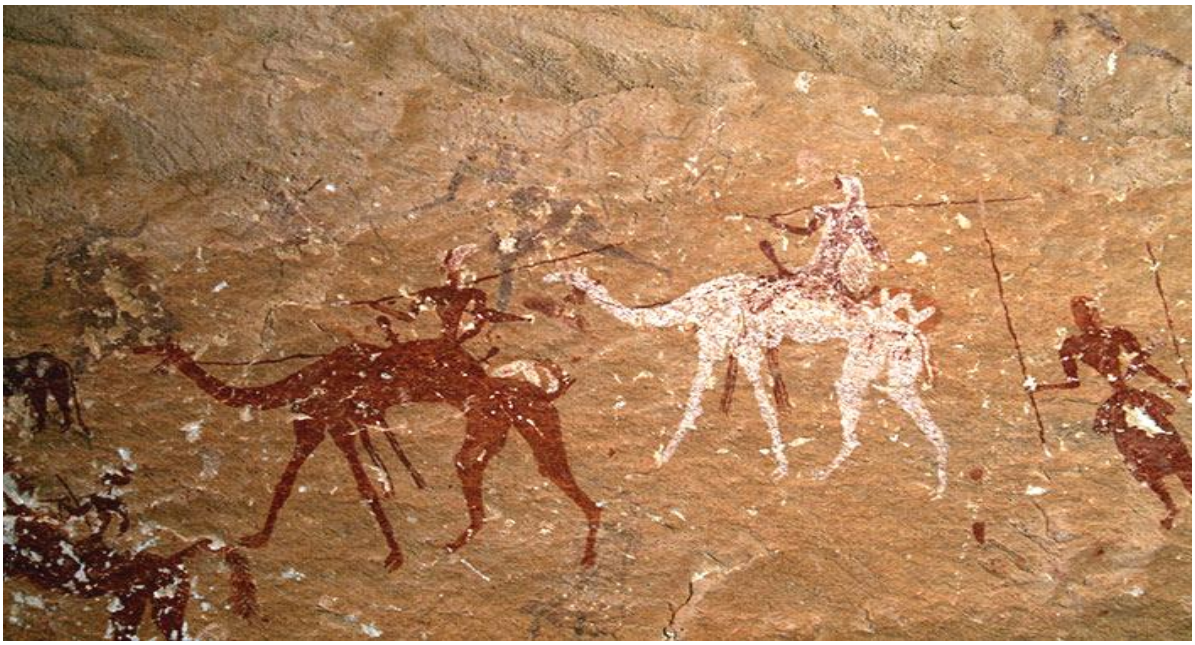
As well, I would like to thank **Pr. Pamela BURGER** for the valuable training I had in the University of Veterinary Medicine in Vienna and **Pr. Steffen WEIGEND** and **Pr. Ulrich BAULAIN** for all what I learned from them during my training at the Frederish-Loeffler Institute, Mariensee.

Last, but not least, I thank every single professor to whom I am indebted for any piece of knowledge and critical thinking I today possess. I would especially site **Pr. Brahim HADDAD**, for my esteem for him is that of a father, and **Pr. Moncef BEN M'RAD**, **Pr. Gley KHALDI** and **Pr. Boulbeba REKIK,** for their loss deeply scarred me.

TABLE OF CONTENT

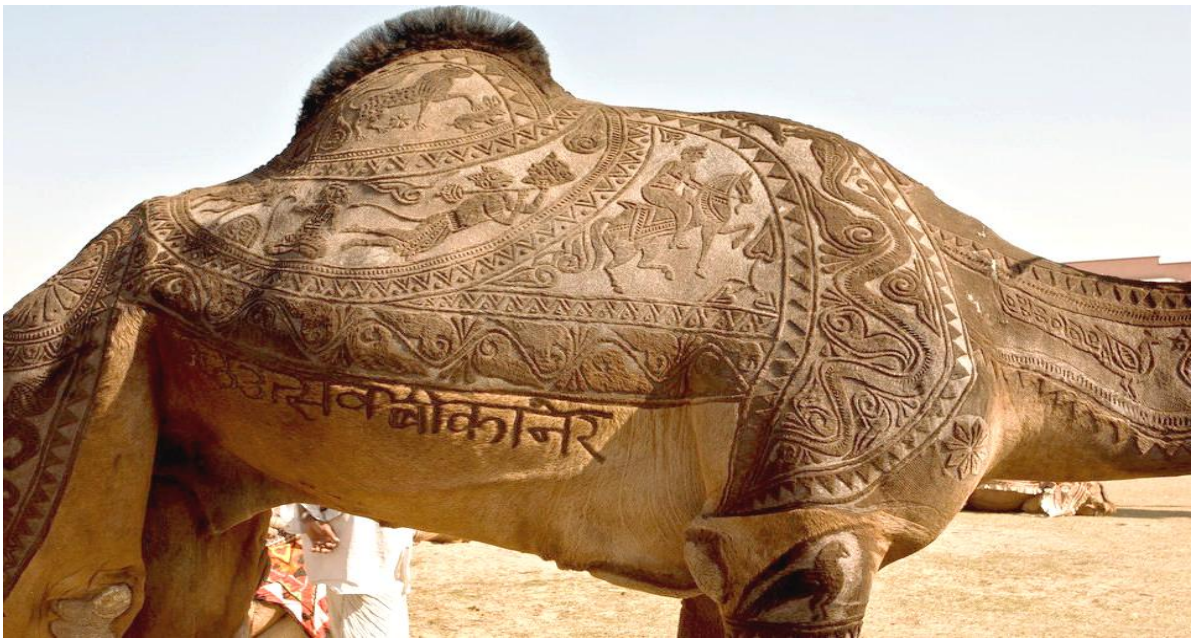
TABLE LIST

FIGURE LIST



Painted scene of Amazigh warriors riding camels

(African Rock Art image database, David Coulson)



Camel hair art at Bikaner camel festival in India

(Photographer: Osakabe Yasuo)

INTRODUCTION

More voices are constantly arising to stress the vital role of preserving global diversity, in the reality of the climate change and the global warming. In addition to *Camelus dromedarius* being zoogenetic resources in need of such preservation, this animal offers promising research axes; a unique adaptability to arid lands and rough environment, an exceptional immunity system (heavy-chain antibodies), unmatched response to dehydration, and much more. Thus, their genetic potential is of extreme importance to ensure agriculture durability when meeting future challenges.

Although this animal was one of the last domesticated animals (Heide, 2010), Camels had a significant historical role in human movements and trade. Today, their economic and socio-cultural role is well established in different cultures and countries, as they ensure food, transportation and other features, in environments where no other animal species can.

Despite of all these characteristics and while genetic research entered the genomic era, knowledge on camel populations and camel genome is still in its early stages compared to other livestock species. Lately, attention towards this animal increased as its whole genome was sequenced (Wu *et al.,* 2014; Fitak *et al.,* 2016). In the last decade, several populations of dromedaries have been characterized and their history have been partially revealed (Almathen *et al.,* 2016; Burger, 2016). Even though, a lot is still to be done to unravel origins, history and genomic specificities of this animal.

On the country's level, camels were present in Tunisia at least since 500 Before Christ (Richard, 2003). Their presence was solidified after the Islamic conquest of North Africa, which allowed new and constant trade exchange with the Middle East, using camels. The extant population is most probably the result of several introduction waves, plausible natural and human selection, and other evolutionary factors. Dromedaries are of extreme importance in rural and arid areas in Tunisia and hold great potential for economic development. This livestock animal addresses multiple human needs in these areas; food supply (milk and meat), transportation, and traditional industry (leather and hair). Beyond that, they adapted to touristic activities and have particular socio-cultural status. Improving the share of dromedaries in livestock production will ensure a sustainable development in arid lands.

On the other hand, Tunisia importation of red meat climbed to 7 million US$ in 2016 (FAO, 2018; OEC, 2018). As a part of the government plan to recover food security in red meat, the "Office de l'Elevage et de Pâturage" defined promoting camel meat production as one of the main actions in its program for agricultural development in 2017 (OEP, 2016). However, to ensure optimal exploitation of camel genetic potential, any strategy should be based on perceiving knowledge of the history and the current state of these genetic resources.

In this context, this research work tried to explore the diversity of dromedaries in Tunisia, to investigate its genetic structure, and to understand the phylogenetic relations with other populations, using nuclear and mitochondrial genetic markers. This thesis was structured into three main parts; (i) a literature review of the studied species in general and its particularities in Tunisia, and of the used molecular and statistical means, (ii) a comprehensive view, based on results from microsatellite markers and mtDNA amplification of the genetic aspect of dromedary population in Tunisia, and (iii) the development of a parentage assay that aim to promote the genetic progress in this species.

LITERATURE REVIEW

CHAPTER I: *Camelus dromedarius*

I. *Camelus dromedarius*

I.1. *Camelus dromedarius,* history and origins

I.1.1. Ancient species

The most ancient known species in the Camelidae family was baptized *Camelops hesternus* (Tuomey and Joseph, 1854). Weighing around 800 Kg and most likely one humped, Camelops arose in the late Early Miocene in North America around 17.5 to 16 Ma, then diverged to Camelus lineages prior to the late Late Miocene (11.6 to 5.3 Ma) (Heintzman *et al.*, 2015).

The *Paracamelus aguirrei* is the known common ancestor of the Old World camels (Pickford *et al.*, 1995). By 7.5 to 6.5 Ma, it crossed the "Bering Land Bridge", a drop in sea level in Beringia (figure 1) during the Ice Age that connected America to Eurasia. After 1.4 Ma, sea level completely submerged this bridge and animals from the two continents evolved separately.

First aPearing in the Asian continent, the direct progenitor of *C. dromedarius* and *C. bactrianus* may have gone extinct long before humans began to pay attention to these useful animals (Ji *et al.*, 2009). The *C. dromedarius* species evolved separately from its common ancestor with other Camelus species (*C. bactrianus* and *C. ferus*) by 5 to 8 Ma (Mohandesan *et al.*, 2017).

I.1.2. Domestication and recent history

An archeological wild dromedary from southeast Arabian Peninsula was found to be among the founders of domesticated dromedary gene pool (Almathen *et al.*, 2016). Diffeent studies (Mohandesan *et al.*, 2017, Almathen *et al.*, 2016) agreed that domestication most probably occurred in the Arabian Peninsula, between 2000 and 1000 BCE, and was followed by introgressions from individual wild dromedaries. Nonetheless, the coexistence of early domesticated and wild individuals was brief (Burger, 2016). The introduction of dromedaries to North Africa probably began before these lands became a part of the Islamic empire, in 632 AD (Richard, 2003). But the trade and exchange through the Silk Road fortified its presence in these areas.

The map in figure 1 summarizes the main steps of camel aParition and dispersion in their respective geographical areas.

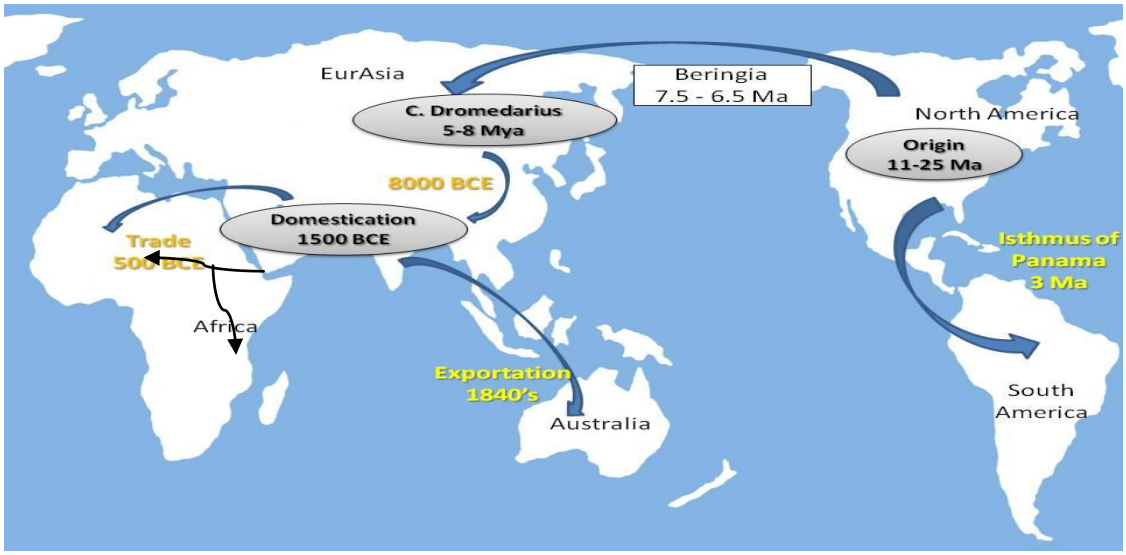


Figure 1. Main historical events in ancient and extant *Camelus* species

I.2. *Camelus dromedarius* populations in the world

The world counted over 28.4 million camels in 2016, evolving from hardly 16.6 million in 1970 (FAO, 2018). *Camelus dromedarius* is more numerous than his Bactrian cousin and represents almost 90 % of the *Camelus* genus (Al-swailem *et al.*, 2010). Over 85 % of this population is located in Near Asia and Africa (FAO, 2018).

By their unique adaptability to drought, high temperatures, desert lands, and scarce, poorly nutritious food, dromedary camels are widely popular in deserts and arid lands of Northern Africa, Eastern Africa, Arabian Peninsula, and Near Asia. They were also introduced in Australia in the middle of the 19th century (Burger, 2016).

I.3. *Camelus dromedarius* in Tunisia

I.3.1. Evolution of the population

According to the last national report from the « Office de l'Elevage et du Pâturage (OEP) », the dromedary camel population in Tunisia counted roughly 80,000 productive female heads in 2016 (OEP, 2016). They are located mainly in the southern of the country in addition to some herds in the center. For the same year, the FAO estimates reached 236,640 heads. In fact, individual identification being poorly practiced, any estimation of the camel population in Tunisia can only be approximate, which explains the inconsistency of the different estimations. In 2011, thanks to the OEP's efforts to establish individual identification practice, 41,500 heads were identified (OEP, 2016). Even so, a large proportion of the population is thought to be omitted.

Nonetheless, the size of the dromedary's population in Tunisia is dramatically decreasing. Reports from the 50s indicated about 225,000 heads that droPed down to 80,000 in the early 90s (Khorchani *et al.*, 1996). This dramatic decrease may be explained by (i) the remarkable socio-economic transformation in the Tunisian society after the independence (mechanization of agriculture and transport and the orientation towards an urban life style) and (ii) the significant decline of pasture areas. Lately, the shortage of camel shepherds, the high cost of caretaking (80 to 120 DNT/ head/ year) and the reduced profitability especially in the extensive mode are the main causes of this noticeable decrease.

Thanks to national regeneration efforts, massive importation of dromedaries from Libya (Nasr *et al*., 2000) helped restoring the population. Despite of the instability and fluctuations of the dromedary population (figure 2), the latest OEP statistics indicated 80,000 female heads in 2016.

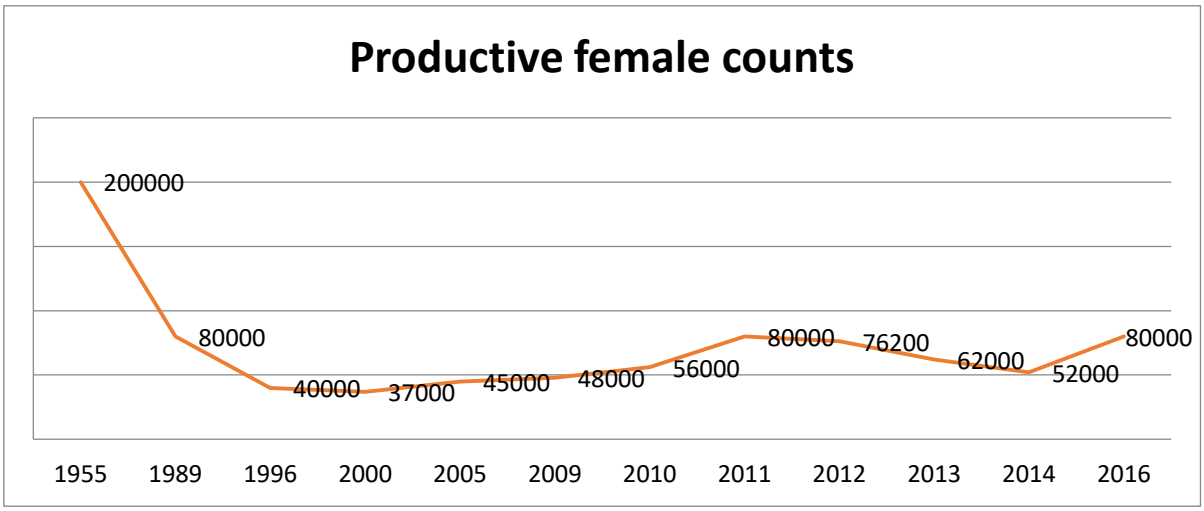


Figure 2. Fluctuations of *Camelus dromedarius* population in Tunisia in the last two decades. Adapted from OEP, 2016; OEP, 2017; Khorchani *et al.*, 1996.

I.3.2. Distribution through the country

The dromedary population is –unsurprisingly- more condensed in the southern of Tunisia where the environment affords the halophyte plants and the arid lands.

Three governorates; Medenine, Kebili and Tataouine are home for about 85 % of the whole population (figure 3, table 1). About 2,300 dromedary herders are located in the south and center of the country (OEP, 2016). While some thousands of heads inhabit each southern governorate, this number decreases to some hundreds in the middle of the country, located mainly on the coasts. In the North, their presence limited to very small herds belonging to nomads or used in tourism activities.

Figure 3. Distribution of the Camel herd through the Tunisian territory. Adapted from OEP, 2016.

Table 1. Distribution of the whole population and of the productive females in the different districts in Tunisia

District	Dromedary heads	Productive females heads
Medenine	19,000	12,500
Kebili	13,000	9,800
Tataouine	9,745	8,500
Tozeur	4,893	3,500
Gafsa	2,500	2,000
Gabes	1,660	1,300
Mahdia	1,500	900
Sfax	1,000	400
Sidi Bouzid	700	500
Kasserine	100	50
Monastir	100	50
Total	**55,000**	**40,000**

Source : OEP, 2016

During the last fifteen years, the distribution of the camel population changed greatly among the southern regions. In 2000, CARDN and OEP reports (OEP, 2016) indicated that Tataouine held the largest camel counting, reaching 45 % (22,091 heads) of productive females. Now, roughly 20 % of all productive females inhabit this governorate, handing over the first place to Medenine where the female count evolved from 19 % in 2000, to 40 % in 2010, and then 31 % in 2015 (figure 4).

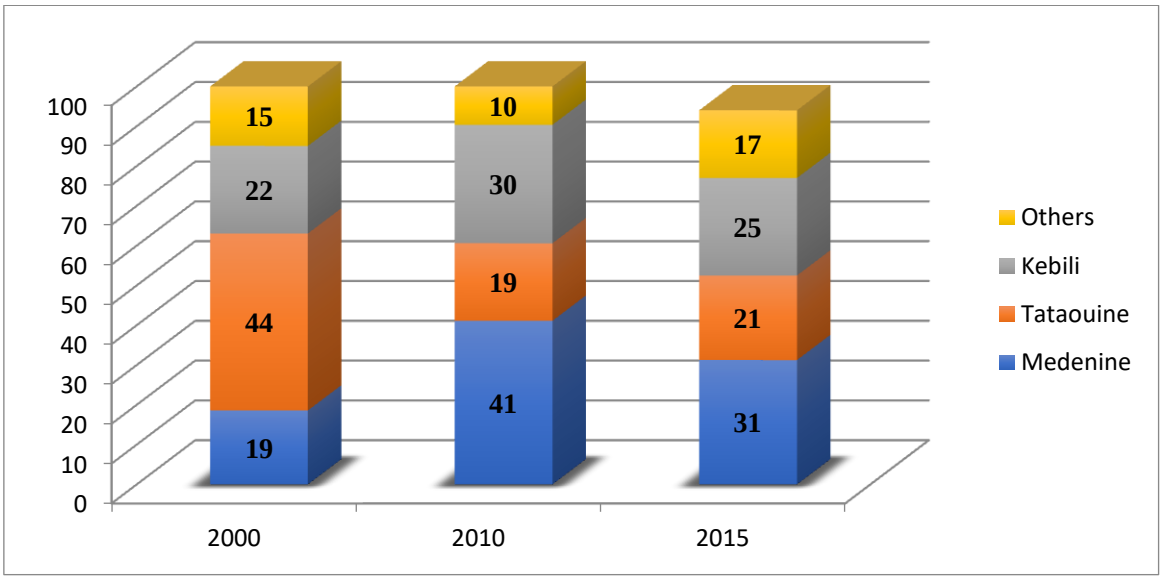


Figure 4. Evolution of the distribution of the productive camel female (Naga) count in Tunisia. Adapted from OEP, 2016; OEP, 2017

I.3.3. *Camelus dromedarius*, the livestock species

As the most adapted species to hot weather, arid conditions and Sahara, dromedary camels are of more than a usage in these difficult environments.

I.3.3.1. Breeding environment

In Tunisia, two types of lands are valorized by this animal;

- Sebkhas, which are salty, shallow lakes and mashes located in the center and south of the country whereby the vegetation is mainly of halophyte plants (primarily *Nitraria retusa* and

Salsola sieberi). Camels frequent these areas typically from the beginning of the autumn and along the winter (Khorchani *et al.*, 1996),
- Arid, desert lands in the far south of the country where the spiny plants are not valued by small ruminants (Belgacem *et al*., 2011).

Dromedary camels are essentially bred in an extensive mode; the animals travel several areas in a phenomenon called transhumance to graze and valorize different types of vegetations. Herds from Medenine, Tataouine and Kebili, gather in the "Ouaera" and "Dhahar" desert lands during spring, in search of exploitable vegetation. Other points, such as "Behayer" in Gabes and "Chott Djerid" in Tozeur are areas for transhumance of nearby herds. About 1.235 million hectares of Tunisian desert fields are exploited by camel (Ould Ahmed and Djemmali, 2011). In dromedary alimentation, the complementation with bran is as low as 10 %, if any (OEP, 2016).

I.3.3.2. Breeding characteristics

The herd's size is extremely variable from few heads for most herders to some hundreds for few of them. In average, herds are of about 80 heads in which only one mature male is kept to ensure reproduction. Dromedary camels in Tunisia are characterized by a rather belated puberty of the females, at the age of 3 to 4 years. The average prolificacy is one offspring every two years. Thus, annual reproduction average is limited to 50 %. Unfortunately, youth mortality in Tunisia is fairly high, reaching 11 % (Sghaier, 2003). The average weight of an adult dromedary male is of 600 kg and that of a female is of 400 kg (Djemali, 2005; Faye *et al.*, 2004).

Traditionally, the animals used to be identified by a tattoo indicating the tribe to which they belong. This identification system is still widely practiced, even though a national identification system is now available. Systematic identification at birth and records of the animal's performance, genealogy and health history are uncommon if not absolutely absent in the largest part of herds.

I.3.4. Social and economic importance

In very harsh environments, where resources are extremely limited, dromedary camels produce milk, meat, leather and hair, and offer other usage. Thus, this animal was of great interest to the Bedouin and traditional communities inhabiting southern Tunisia. Today, this animal is still treasured to the utmost in the communal heritage and the large panel of its useful products gave it an important economic status.

I.3.4.1. Milk production

Dromedary milk is of high nutritional value, is low in lactose, and has a characteristic acid taste. The female or "*Naga*" requires the presence of her offspring to secrete prolactin and to eject milk (can be substituted by a shot of occytocine). *Nagas* produce about 1.5 to 6 liters/day. In total, 300 to 1200 liters/year/female can be attained (Sghaier, 2004). Dromedary milk is of high economic value -as its price exceeds 5 times that of cow milk- and is believed to have healing and preventive ability against a number of diseases. However, the total production in Tunisia is estimated to only 11,100 tons and is –in its majority- reserved to familial consumption or local selling (OEP, 2016). The share of camel milk production is limited to 0.8 % of the whole milk production in Tunisia (figure 5). Lately, studies about the adaptability of dromedaries to milking machines tried to identify the optimal milking parameters, in order to automate this task as a part of intensifying milk production in dromedaries (Atigui *et al.*, 2016).

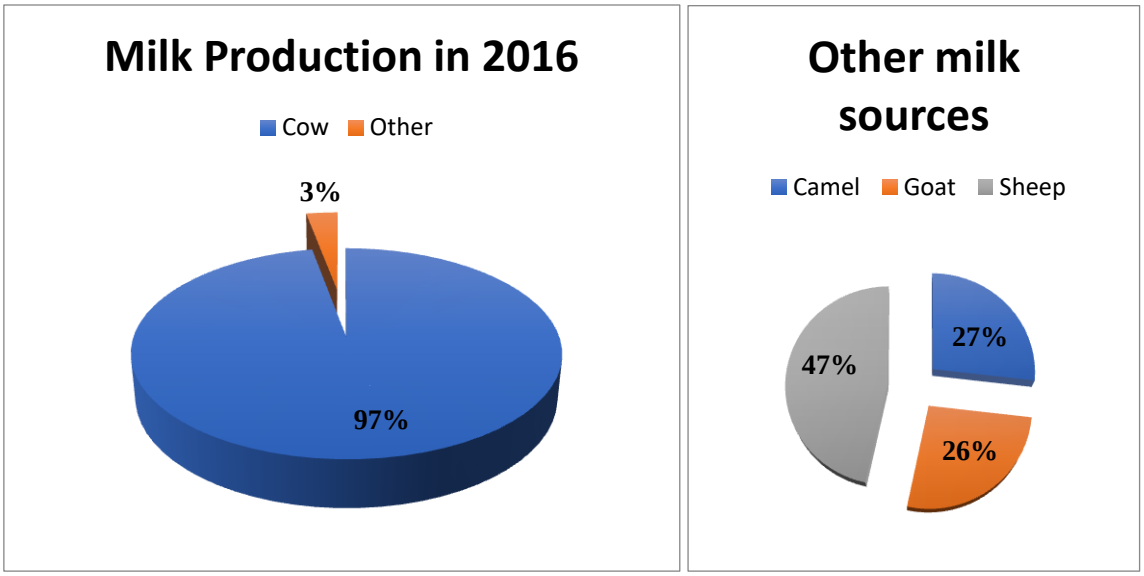


Figure 5. Low contribution of camel milk in the whole milk production in Tunisia. Adapted from FAO-STAT, 2018 and OEP, 2017 data.

I.3.4.2. Meat production

Dromedary meat is known for its low fat content, especially that the slaughtered animals are in general young males. At slaughtering, the animals are of about 18 months of age and an average of 200.3 ± 38.3 kg of weight (Ould Ahmed and Djemmali, 2011), even though the legislation requires a minimum weight of 250 kg. This product is mainly popular in the South and in the Sahel regions of the country. A total of 4,000 tons is produced annually according to the last report of OEP (OEP, 2017), representing about 3 % of the whole red meat production in Tunisia (figure 6).

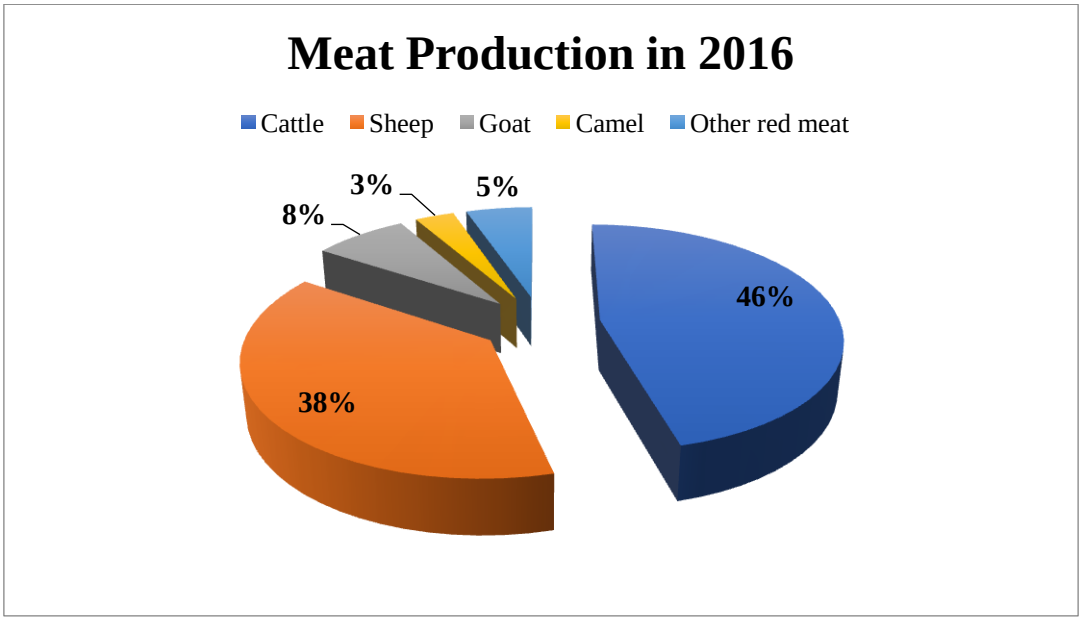


Figure 6. Red meat production by species in 2016. Adapted from FAO-STAT, 2018 and OEP, 2016 data.

I.3.4.3. Other products

The dromedary hair is exploited in textile, mostly in artisanal products like “Ouazra” and “Barnous” used as a coat in cold weather. The average production is approximately of 750

g/animal/year (Faye *et al*, 2004; Sghaier, 2004). As well, camel leather is exploited in traditional products like "Belgha" and other forms of high quality shoes. Although, lately these products became rare because of the laborious and delicate effort they require. Table 2 illustrates how these secondary products decreased in quantity throughout the last two decades.

Table 2. Evolution of different dromedary products (in tons) between 2000 and 2014

Product	1996	2000	2005	2009	2013	2014
Hair	35	40	48	50	23	28
Leather	577	651	677	701	500	575

Source OEP, 2016

Before 2010, and thanks to continuous efforts from OEP to develop this sector, slow but steady increase of camel products was observed, as plainly illustrated in figure 7. Following the revolution in Tunisia in 2011, political instability led to discontinuity in these efforts. Consequently, since 2010, different dromedary products have been in decrease, reflecting the decline in the whole dromedary population count. In fact, during the first years after the revolution, there was a lack of control on the borders. It is thought that illegal trade of camels occurred in that period, but official data between 2009 and 2013 are missing due to the same political conditions. After 2013, the stability in the political situation permitted a relative recovery of the camel population.

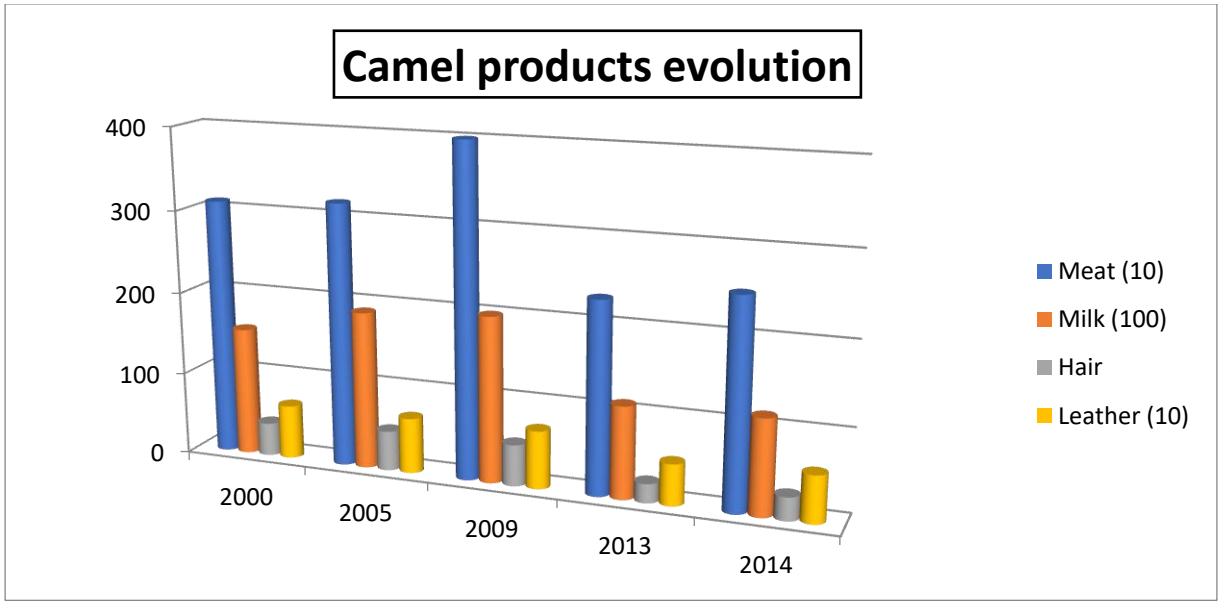


Figure 7. Evolution of different dromedary productions (in tons) between 2000 and 2014

I.3.4.4. Other usage

Although camel caravans transporting merchandise and connecting the East to the West through the so-named Silk Road is now history, dromedaries still ensure conveyance in the deep Sahara and arid lands, where no other modern transportation mean can carry humans and goods in the same efficiency that camels afford.

In addition, the economic importance of dromedary camels comes from its use in touristic activities, local private ceremonies, animal racing, folkloric shows and other cultural activities

of the sort. Furthermore, owning dromedaries was and in some regions is still considered a symbol of high status in the society.

I.3.5. Challenges for development

Camel livestock is facing a number of constraints on more than a level. On the social level, the Tunisian society changed towards the urban lifestyle, while maintenance of camel herds is related to living in arid lands and transhumance. Lately, camel shepherds became more and more scarce. On the logistics level, the infrastructure in desert areas (watering points, shepherds shelters and other equipments) is unsatisfying and needs improvement. As well, more efforts are needed to improve veterinary assistance and to generalize vaccination and animal's individual identification.

Other challenges are related to the livestock itself; (i) the low productivity of the camel; carcass cutting yield, daily weight gain (Hammadi *et al.*, 2001; Kamoun M. *et al.*, 2009) and (ii) the difficult conservation and transformation of camel milk (Kamoun, 1995) are barriers to promoting investment in this sector. Looking for alternatives for product evacuation and better food resources that are specifically developed for camel livestock are of interest.

In addition, genetic improvement would permit a steep upgrade in camel livestock as it can target the animal's performance as well as the products quality.

I.3.6. Official strategy for camel development

Recently, increased importance has been given to camel husbandry. In fact, consumption of red meat in Tunisia is up to 117.2 mille tones yearly, from which 96 % are locally produced, while 4 % are imported (OEP, 2017). Thus, one of the objectives intended in the last OEP plan for agriculture development was to promote camel livestock in the southern regions in order to improve its contribution to red meat production. The OEP report defined a number of acts including maintaining and setting up basic infrastructure like watering points, enhancing investment in this sector, and improving the animal's performance.

I.4. *Camelus dromedarius*, the genetic resources

During the last twenty years, camels in general and dromedaries in particular became an interest for genetic studies. Shortly after the development of genetic markers for camels by cross-species amplification (Sasse *et al.* 2000) and by direct isolation in Camels (Mariasegaram *et al.*, 2002), several genetic characterization research projects were conducted on different populations in order to estimate the genetic variability in this species and to identify distinct genetic entities.

I.4.1. Genetically characterized sub-populations in the world

The main large study on extant dromedary populations was published in 2016 by Almathen and his collaborators. It concluded that little phylogeogaphic signal is detectable in the modern populations of *C. dromedarius*. Indicative of an extensive gene flow affecting all regions except East Africa, where this animal remained relatively isolated. Little population structure was thus found in modern dromedaries. This was considered as the consequence of cross continental back and forth movements.

Consistently, even though local breeders identify -based on morphological traits- different "breeds" of dromedary camels, the genetic data from different populations did not support these classifications in most cases. In India, for instance, Vijh *et al.* (2007) used a large number of microsatellite markers (23) but detected no more than 12 % of variability between the 4 breeds recognized in the country, while 88 % of the observed variability was among the individuals. They considered that two breeds; Jaisalmeri and Mewari showed genetic differentiation and were worthy of selection for specific traits. In Kenya, the genetic distance between the 4 known breeds was at the limit of 0.1 (Mburu, 2010). Two separate genetic

entities were identified; one grouping Gabbra, Rendille and Turkana breeds and one composed only of Somali camels. Nevertheless, breed assignment did not reach 50 % and the authors concluded that the genetic data did not support the extant recognized ecotypes.

In Saudi Arabia, a possible origin of domestication, camel population is traditionally classified into four ecotypes; Magaheem, Wadha, Hawara and Omani. They exhibited low genetic differentiation ($F_{ST} \leq 0.01$) in a study by Mahmoud *et al.* (2012), using microsatellite markers. Other studies concerning populations in Pakistan (Shah *et al.,* 2014), Canary Islands (Schulz *et al.,* 2010), South Africa (Nolte *et al.,* 2005) and Egypt (Mahrous *et al.,* 2011) showed moderate to low genetic differentiation in the concerned camel populations.

Finally, concerning the countries where camel populations are known under the aPellation "Maghrebi camel", very low differentiation was observed. Among the Moroccan ecotypes (5), Nei's genetic distances ranged from 0.02 to 0.4 (Piro *et al.,* 2011). Likewise, "no clear genetic structure was observed" in the Algerian population, stated Cherifi and his collaborators (2017).

I.4.2. Classification of camels in Tunisia

I.4.2.1. Phenotypic characterization and ecotypes

The population of *Camelus dromedarius* in North Africa, regardless of its heterogeneity, is grouped under the term; « Maghrebi camel ». Many classifications of the population in Tunisia have been reported, which makes it confusing to define a clear one. Four groups have been reported by FAO and UNEP (2000); Camel of the South, Camel of the Cap-Bon, Camel of Sahel and Camel of Kairouan. Given that camels in the north hardly constitute 10 % of the whole population today, camels in the south are far more morphologically diverse. More recent classifications (Ould Ahmed *et al.,* 2007; Chniter *et al.,* 2013; Bedhiaf *et al.,* 2015) - based on the traditional, common classification among camel breeders in Tunisia- were conflicting about the number of existing ecotypes.

Bedhiaf and collaborators (2015) defined 6 ecotypes; Merzougui, Gueoudi and Chaambi in Kebili, Ardhaoui Medenine and Ardhaoui Beni-Zid in Gabes, and Khawar in Djerba. In another study, five main ecotypes were identified by Ould Ahmed *et al.* (2007). Three of them are encountered in Kebili; (i) Merzougui, (ii) Gueoudi and (iii) Mehari (or Targui) that were introduced by Algerian shepherds in Douz. In the other southern regions, two additional ecotypes were defined; (iv) Ardhaoui (or Arbi), localized in Medenine, Tataouine and Gabes, and (v) Khaouar, sparsely bred in different herds in the South of the country.

Chniter *et al.* (2013) work relied on the tribal affiliation as a basis for ecotype identification. While they agreed to the two first groups (i and ii) defined by Ould Ahmed *et al.* (2010), they defined another third group called Ghiloufi in Kebili, and also distinguished two ecotypes in Ardhaoui (Table 3). Khaouar was not mentioned as an ecotype in their publication but as a *Naga* phenotype characterized by sleek hairs. Based on phenotypic and morphological characteristics; 9 body measurements and two quantitative traits (categorical classification), the authors reported a clustering of the studied animals consonantly with their tribal affiliations and they provided clear description of these ecotypes. Ecotypes situated in Kebili region (Gueoudi, Ghiloufi and Mersougui) were very close ($R^2 > 0.7$). Considered as a unique group; Kebili camels, Ardhaoui Tataouine and Ardhaoui Medenine exhibited less than 50 % of similarity between each other.

Table 3. Nomenclature of camel ecotypes and their localization

Nomenclature	Tribue/ Geographic zone	Governorate
Ardhaoui Medenine	Touazine/ Ben Guerdane	Medenine
Ardhaoui Tataouine	Wederna	Tataouine
Ghiloufi	Ben ghilouf	Kebili
Gueoudi	Ouled gharib	Kebili
Merzougui	Merazigues/ Douz	Kebili
Targui or Mehari	From Algeria	Kebili

Adapted from Chniter *et al.,* 2013, Ould Ahmed *et al.,* 2010

In general, ecotype recognition is rather intricate and subjective in most cases. Most herds include more than one ecotype. Furthermore, during transhumance, animals from different regions come together and breeding males are most likely exchanged between herds (Ould Ahmed and Djemali, 2011).

I.4.2.2. Genetic characterization

From a genetic prospect, only two works studying the genetic resources of dromedary camel in Tunisia were published. Ould Ahmed and his collaborators (2010) genotyped 90 dromedaries using 6 microsatellite markers. The population exhibited a deficit in heterozygotes (H_0 = 0.46) while the expected heterozygosity was 0.6. Only samples from Tataouine did not display this deficit (H_0 = 0.53). Nei's unbiased genetic distance estimates showed an average differentiation of Kebili's camels (0.2) and a close similarity between Tataouine and Medenine's animals (0.1). Using neighbor joining clustering, Ould Ahmed *et al*. defined two main groups divided each into subgroups. However, this grouping was independent both from the geographic origin and from the ecotypes, indicating a common genetic basis for the whole population. All F_{IS}, F_{ST} and F_{IT} statistics were significant at a 0.05 threshold. Based on F_{ST}, only 9% of the observed variability was attributed to difference between populations. The authors considered this value as indicative of a moderate differentiation over all the population. Pairwise F_{ST} estimates demonstrated that the most remarkable differentiation was in Kebili population. Gene flow between Tataouine and Medenine was 3 times more than that between them and Kebili, supporting previous observations. The two genetically distinct groups -defined by the study- were identified as Nefzawa (referring to all Kebili's ecotypes) and Aaradh (grouping Medenine and Tataouine). Nevertheless, the author noticed the limits of their work as they used a restraint number of markers. They recommended contemplating a larger study using more genetic markers in order to be more conclusive regarding the definition of the selection objectives and the genetic improvement planning.

The second work (Bedhiaf *et al.,* 2015) was based on Amplified Fragment Length Polymorphism (AFLP) markers and concerned only Ardhaoui Medenine, Ardhaoui Gabes and Merzougui. The authors estimated a similarity of 45 % over all the samples. Two groups were obtained by clustering; the first one enclosed only Ardhaoui samples, and the second included Ardhaoui and Merzougui samples. Thus, they concluded that a common genetic basis was shared among these ecotypes.

In general, these studies were not irrefutable, as the number of samples and that of the genetic markers were not sufficient to conclude about the whole camel population in Tunisia. Furthermore, a new publication of a large study including samples from Algeria and Egypt, suggested that the entire North African dromedary population is feebly variable and detected weak structure in it (Cherifi *et al.,* 2017). This study concerned 331 camel samples from genotyped using 20 microsatellites. Weak population differentiation was observed in Bayesian clustering of these samples. The neighbor joining clustering showed that distance

between the two populations; Algerian and Egyptian was not greater than distances between known subpopulations in the two countries. Although pairwise F_{ST} values were significant at a 0.01 threshold, they were very low.

As for the origins of the Tunisian camel population, no previous scientific publication specifically addressed this research question, to our knowledge. Ould Ahmed *et al.* (2010) estimated -based on historical information and phenotypic characters- that this population was the outcome of inter reproduction of a number of genetic types; mainly of "Maghrebi" and "Chargui" from Lybia, "Mehari" or "Targui" from Algeria, "Sudaneese" from Sudan, and probably also other types.

I.4.2.3. Concerns on inbreeding risk

As the total count of the dromedary population in Tunisia is fairly limited and 85 % of it is located in the southern regions that extends to roughly 65,500 km² (desert areas), concerns about the risk of elevated inbreeding overall the cameline livestock are reasonable.

In fact, some practices of camel husbandry -that are common among breeders in Tunisia- may favorite inbreeding. Breeding mates belong, in 85 % of cases, to the same herd and are generally used for an average of seven consecutive years (Ouled Ahmed and Djemmali, 2011) which favorites reproduction between related animals. Studies on the cameline genetic resources in Tunisia reported an inbreeding rate of 8 to 15 % inside the regions; Kebili, Medenine and Tataouine (Ould Ahemd *et al.*, 2010). Being under 25 %, the average similarity between half siblings, these rates are moderate and not alarming.

In case of high inbreeding rates in animal populations, number of visible issues can arise; decrease in animal performance, elevated risk of genetic disease and malformation, and decrease in adaptability and natural resistance to diseases. From a genetic standpoint, inbreeding leads to inbreeding depression of dominant traits and loss of genetic variability. Thus, such issue must be controlled in order to maintain a long-term, sustainable livestock production (Kristensen and Sorensen, 2005). When needed, a number of recommendations may be issued for breeders in order to control the situation and to preserve the genetic variability among the livestock population.

LITTERATURE REVIEW

CHAPTER II: Molecular analysis of genetic diversity

II. Molecular analysis of genetic diversity

II.1. Tools for molecular analysis

In the last few decades, technologies and means in the genetic field sprout up and allowed genetic research to take new dimensions in genome-scale analysis, epigenetics, and others. To make the utmost benefit of this development, genetic characterization of livestock populations is the groundwork for any insightful genetic improvement whether it is as simple as selection, genetically assisted selection, or even genome editing.

In the last decades, genetic characterization evolved from the usage of few to hundreds of genetic markers, then to thousands of SNPs (Single Nucleotide Polymorphism), then to the complete genome sequencing. This last technique is not only the ultimate resolution for detecting genetic variability, but it also became rapid and low cost. Lately, a group of researchers sequenced a whole plant's genome within hours (Parker *et al.,* 2017). Yet, genetic markers are still broadly used as they are more suitable for population genetic characterization, they are less expensive, or for other convenience reasons (availability of equipments, etc.).

II.1.1. Usage of molecular data for population variability studies

Being the ultimate level of diversity that can be studied, DNA variability became the main target for genetic studies in livestock populations. It allowed a wide range of research such as population structure, breed identification, and more.

II.1.1.1. Usage of microsatellite markers

Microsatellites and SNPs are the main genetic markers that are used for population structure, nowadays. Other molecular markers are less popular, such as AFLP (Amplified Fragments Length Polymorphism), RFLP (Restriction Fragments Length Polymorphism), and RAPD (Randomly Amplified Polymorphic DNA). Hereafter, we describe the interest and weakness of using microsatellites for population characterization.

II.1.1.1.1. Why microsatellites?

Being present in most eukaryotic organisms and all livestock species, microsatellites are neutral and versatile, offering applications in population genetics, conservation biology, and evolutionary biology. They consist of a tandem repeat of 2 to 6 bp motif. The variation in the motif's number of repetitions gives multiple alleles. Their interesting usability comes from their high polymorphism, codominant interaction and Mendelian inheritance. On a practical point of view, microsatellites are fairly rapid, convenient and efficient with a reasonable cost. Hence, they are largely used in discerning the population structure, establishing paternity tests, and inquiring inbreeding.

The recent development of SNPs markers was revolutionary for different types of genetic analysis. Coates *et al.* (2009) noted that the differentiation coefficient (F_{ST}) was generally higher in SNP analyses. An overestimate of migration rates between subpopulations, due to homoplasy in microsatellite alleles, may be at the origin of this ascertainment. Homoplasy occurs when a microsatellite allele mutates to another existing allele. Therefore, they become identical but not because of a common origin. It is typically not possible, from the data in hands, to infer homoplasy.

However, it had been demonstrated that an adequate number of microsatellites gives analogous results to SNPs for detecting population structure, and that microsatellites are able to delimit fine-scale structuring (Herràez *et al.,* 2005; Coates *et al.,* 2009; DeFavri *et al.,* 2013).

II.1.1.1.2. What limitations?

Microsatellites can be subject to different phenomena interfering with their potential of reflecting the accurate population structure. First, the "Slippage" is a mutation process resulting in an addition or subtraction of the repeated motif. Analogous to slippage occurring in nature, polymerase in PCR can produce sequences of different size from the copied DNA by up to 5 repeat units (Ciofi *et al.*, 1998). Slippage effect is more accentuated when using mono and dinucleotide repeats (Wan *et al.*, 2004). The second source of incongruence is the affinity of Taq polymerase to insert an extra Adenine nucleotide to the 3' end of the DNA copy. Error in allele size determination occurs when this nucleotide is only occasionally added, not when it is always (or never) added.

Another issue, which may lead to biased estimation of statistical parameters, is the presence of size homoplasy. Finally, point mutations are relatively frequent in flanking sequences. When they occur in the primer binding region, they may inhibit annealing and result in a reduced or even absent amplification. This phenomenon is known as "null alleles". The direct consequences are false homozygotes that can affect heterozygosity estimation in variability studies and mismatches between parent and offspring in parentage tests. Both size homoplasy and null alleles lead to an underestimation of genetic diversity but will not cause significant biases in genetic analysis when a large number of loci is used.

Wrapping up, Microsatellites are still broadly used in a large panel of genetic analyses ranging from population genetics to genetic mapping. Their automation and availability for most -if not all- livestock and wild species justify their convenience.

II.1.1.2. By sequencing

Developed by Frederick Sanger (1977), Sanger technique is based on selective incorporation of detectable (fluorescent) dideoxynucleotides (ddNTPs) by DNA polymerase throughout *in vitro* DNA replication. Nowadays, its automation offers rapid and accurate genetic data. Sanger sequencing produces read-lengths of up to 1,000 bp and its accuracy is close to 100 %.

In the beginning of the new millenary, Next Generation Sequencing (NGS) technologies were specifically developed for parallel (high throughput) sequencing. Assembled under the denomination of NGS, they are actually based on different physicochemical technologies: Sequencing by Synthesis (SBS), Pyrosequencing, Semiconductor, Ligation, etc.

New technologies, baptized Third Generation Sequencing, emerged lately. PacBio RS (based on real-time sequencing), Oxford Nanopore, and Illumina TruSeq perform the sequencing by reading the nucleotide sequence at the single molecule level which does not require DNA fragmentation. Thus, they produce longer sequences but also offer support for epigenetic analyses. Recently, hybrid methods were developed where NGS and third generation sequencing are combined to take benefit of both technologies while overcoming their disadvantages. A coarse comparison of the three types of technologies is presented in table 4.

Table 4. Characteristics of different sequencing technologies (Miyamoto *et al.*, 2014; Lee *et al.*, 2016 ; Allali *et al.*, 2017; NIH, 2017)

	Sanger sequencing	**Next Generation**	**Third Generation**
Output per run	1 to 84 Kb	10 Mb to 1.5 Gb	3 Gb
Read length	400 to 900 bp	50 to 700 bp	3 to 15 Kb
Sequencing time	20 mn to few hours	2 h to 27 h	20 mn
Preparation time	Few hours	Few hours	No amplification required
Cost per Mp In US dollars	~ 500 $	0.05 $	0.5 – 2 $
Error rate (%)	10^{-3}	10^{-2}	10 to 30
Advantage	High quality Long reads	Fast, high throughput Several applications	Rapid, Long reads Portability
Weakness	High cost Low throughput	High error rate Short reads	Increased error rate

For population characterization studies, Sanger sequencing is suitable for SNP detection, phylogeny reconstruction and for estimating the different variability parameters.

II.1.1.3. Specificities of using mitochondrial DNA

Mitochondrial DNA (mtDNA) is one of the most sequenced DNA fragments. It has been utterly the most used molecular estimator of biodiversity and one of the mitochondrial fragments (COX1) is now the standardized tool for molecular taxonomy and identification (Galtier *et al.,* 2009). MtDNA structure consists of highly conserved gene content lacking introns and flanked by short, variable intergenic regions (Gissi *et al.*, 2008).

Mitochondrial DNA (mtDNA) has three main features, making it very informative about the degree of differentiation between the genetic groups of a population. First, being inherited exclusively from the dam, the whole mtDNA is transmitted with no recombination. Thus, all sites virtually share identical genealogical history. Second, the evolution of mtDNA is mainly neutral. As the genes are essential for the functions of the mitochondria (respiration), they are considered less likely to be prone to adaptive processes (Galtier *et al.,* 2009). The third and last main feature is that the evolutionary rate of mtDNA is assumed to follow a clock-like substitution rate. In addition to all these features, mtDNA is experimentally easy to amplify due to its presence in multiple copies in each cell.

However, reports of paternal inheritance of mtDNA in several livestock animals (Chicken; Alexander *et al.*, 2015, Sheep; Zhao *et al.*, 2001), and those of inserted mtDNA material in nuclear genome (Numts) (Bensasson *et al.*, 2001) challenged the classic view of mtDNA as a neutral and clonally transmitted marker. Nonetheless, it is still a powerful molecular asset that allows tracing back the origins and the phylogeographic relationships inside an animal population.

II.1.2. Molecular biotechnology and *Camelus dromedarius*

Camelus dromedarius has 74 chromosomes (Samman, 1993). Three draft genome assemblies were released between 2010 and 2016. Dromedary's genome counts over 2000 Mb, and the mitochondrion complete genome is of 16,643 bp. In 2010, Al-swailem and his collaborators realized the annotation and assembly of camel's Expressed Sequence Tags (ESTs), revealing over 4,500 potentially specific or fast evolving camel gene sequences with no homology to other known genomes. Wu *et al.* (2014), in a comparative analysis between

camelids genomes, exposed a large number of adaptation related gene families that are specific to *C. dromedarius*.

Comparatively to other livestock animals or even some wild species, the genetic aspects of dromedary camels are not satisfactorily explored. The physical map is unknown, the number of genetic markers (microsatellites and SNPs) is insufficient, and little has been done in the genetic improvement of this animal.

During the last two decades, an increased interest in this animal emanated worldwide thanks to its developing economic importance, its high quality products (milk and meat), its unique properties (presence of heavy chain antibodies, water loss tolerance, etc.), and its unique evolutional and adaptive particularities that may hold promises for food security facing the future challenges of global warming and climate change.

This late interest in Camel genetics opened several perspectives in scientific research. One of the developing axes is the design of a species specific SNP chip to allow a wide genome covering while minimizing both the time of the experiment and its cost. Other axes are the finalization of the draft genome of *Camelus dromedarius*, the assembly of complex genomic regions, and the construction of the genome map. These progresses will significantly boost the genetic improvement of dromedary camels.

II.2. Foundations of statistical analyses for genetic studies

II.2.1. Microsatellite genotyping data analysis

In order to adequately analyze microsatellite markers data in population genetics, different statistical parameters have been defined, using theoretical models that follow the particularities of the evolution of microsatellite alleles.

II.2.1.1.Mutation models for microsatellites

Several mutation models were proposed: the Infinite Alleles Model (IAM) (Kimura and Crow, 1964), the K-Allele Model (KAM) (Crow and Kimura, 1970), and the Stepwise Mutation Model (SMM) (Ohta and Kimura, 1973; Valdes *et al.,* 1993).

While IAM supposes that each mutation results in a new allele in the population, KAM model supposes that only a finite number of alleles (K) are possible in a population, with an equal probability of mutating from an allele to another. SMM –on the other hand- is a powerful model and is the most commonly used mutation model for microsatellites. It assumes a symmetric forward and backward random steps leading to distinct new allele differing by a single repeat unit. The previous state is conserved after one step. Under this model an evolution of the microsatellite length to an infinite or a null length is possible (Dirienzo *et al.,* 1994). Moreover, it had been demonstrated that mutations of multiple repeat units are common in microsatellites (Huang *et al.,* 2002).

These models are not complete regarding the evolutionary mechanisms by which microsatellites may evolve. Nevertheless, statistical methods based on these models, namely Wright's (1931) F statistics and Nei's genetic distance (Nei and Roychoudhury, 1974), showed a strong potential of accurately estimating population genetics parameters. Refinements to these models have been later developed.

To explain the finite number of step mutations under SMM, another model; the Two-Phase Model (TPM), was described by Kruglyak and his collaborators in 1998. They stated that two types of mutations can occur; (i) strand sliPage mutations that add or delete a single repeat unit while depending on allele length, and (ii) point mutations interrupting microsatellite repeats and generating two short alleles from an initial, long one. Microsatellite alleles in this model can evolve by step mutations but when attaining a large length, they would be interrupted by point mutations. Another model to explain how alleles cannot take infinite length under SMM was proposed by Garza *et al.* (1995), where a mutational bias pushes microsatellite towards a focal length. Nevertheless, biological mechanisms allowing such phenomena were not determined. Other microsatellite properties, such as their high mutation rate, homoplasy (Estoup *et al.,* 2002) and the presence of multiple alleles, are still to be integrated to the evolutionary mechanism of microsatellites (Csilléry, 2009).

II.2.1.2. Classic statistical parameters

The main evolutionary forces that affect the structure of a population are; mutations, genetic drift, migration, and selection (natural or artificial). The presence of one or more of these forces can be inferred by comparing the heterozygote's frequency to that when Hardy Weinberg equilibrium (HWE) is supposed. Other statistical parameters describe the genetic variability both inside genetic groups and between them.

II.2.1.2.1. Intra population diversity

Several parameters can estimate the genetic variability; (i) number of alleles per locus, (ii) effective number of alleles (the number of equally frequent alleles that would take to achieve the same expected heterozygosity), (iii) the observed rate of heterozygosity, (iv) the expected heterozygosity (estimated when HWE is assumed), and (v) the Polymorphic Information Content (quantifying the discriminative power of a locus based on the number of alleles and the distribution of their frequencies). In addition, the mean unbiased heterozygosity (Hnb)

was proposed by Nei (1978), for low sample size. It presents the probability of randomly pulling two different alleles in a locus.

As well, it is also important to estimate the inbreeding level in the population. Wright's F_{IS} quantifies the gap between observed and expected heterozygosity due to non-random mating. In addition, pairwise relatedness (Wang, 2007) can estimate genetic relationships between two individuals (dyadic) in a population, or also with the use of a third individual as a control (tryadic).

II.2.1.2.2. Inter population diversity

In one species, different populations may be under the action of different evolutionary forces. The genetic structure is thus differentially evolving into new allelic frequencies. The main parameters that quantify the diversity between populations are:

(i) The genetic distance: the proportion of different genetic elements between two samples. Demonstrated by Nei (1978), it is based on allele frequencies.

(ii) Wright's F statistics: a measure of departures from HW equilibrium according to three levels of variability; (i) between individuals and subgroup of the population (F_{IS}), (ii) between an individual and the whole population (F_{IT}), and (iii) between a subgroup and the whole population (F_{ST}), which estimates the differentiation between subpopulations.

(iii) The gene flow between populations counterbalances their differentiation. The number of migrants (Nm) is inversely proportionate to F_{ST}. Nm can be estimated from F_{ST} under the assumption of infinite number of subpopulations (Wright, 1951) or with an infinite number of subpopulations (Li, 1976).

II.2.1.3. STRUCTURE software

Based on the SMM model, Pritchard and Feldman (1996) developed the coalescent model based on the pairwise differences in the number of repeats between randomly chosen microsatellite alleles. They developed STRUCTURE software, based on the Bayesian approach and using Markov Chain Monte Carlo (MCMC) algorithm that performs individual analyses for a number of assumed population number (K). From a simple and basic perspective, Markov Monte Carlo Chain is a simulation of data (a finite number of samples) that can reconstruct more accurate distribution of the outcome values.

In their software, for a test of K populations, four main models for ancestry are available: (i) no admixture, where individuals' genomes are supposed to be unconnected, (ii) admixture model, where each individual's genome is a mixture of all K populations, (iii) linkage model, where admixture is assumed but linked loci are considered to come from the same population, and (iv) prior model, where information about the location is considered in a Bayesian model. The prior is used to pre-define some populations or to assist clustering with weak data or also to detect migrants.

The linkage model is not fitted for studies where the genetic markers are not mapped, which is the case for dromedary camel microsatellite markers. The no admixture model is more fitted for distinct populations, as it generates the posterior probability -for each individual- to be originated from the K^{th} population (the prior is in equal probability of originating from any of the K populations; 1/K).

The admixture model is useful as an exploratory first step because of its flexibility for dealing with the complexity of real data. To end with, integration of relevant information to the clustering in the prior model may be useful especially when the structure is relatively weak. It supposes that individuals coming from the same sampling location are from the same population. The model LOCPRIOR tests this hypothesis on the data before allowing the main analysis to use this information as prior. Thus, when the genetic information is discordant with the sampling information, this last one is ignored.

II.2.1.4. Multivariate analysis

In general, multivariate analysis captures the multidimensional nature, or simply the complexity, of the genetic data and allows the visualization of its main variability.

Following this concept, the Principal Component Analysis (PCA) defines the multidimensional space in which the data presentation shows the widest variance. Recently, this method has been considered as an alternative to Bayesian clustering algorithms, such as STRUCTURE, as it does not require any pre-assumptions about the population genetic model (Lee *et al.,* 2009). Nevertheless, PCA lacks the ability to offer group assessment. Furthermore, it considers the whole data variability, i.e. both the structured genetic variability (distance between the groups) and the random one (variation inside a group).

Another method of multivariate analysis is the discriminate analysis (DCPA) that uses multiple functions, i.e. linear combinations of alleles, to maximize the distance between the groups while minimizing the variation within groups. DCPA focuses on between-group variability and tries to choose the parameters that maximize this distance. The advantage over STRUCTRE is that this method is suitable in all population models. Furthermore, it does not hold any assumptions on the presence or absence of HW equilibrium or linkage disequilibrium. Another powerful feature of DCPA is that the user selects the number of principal components to be used. Nevertheless, Jombart and his collaborators (2010) recognized that "this procedure (selection of number of components) is largely *ad hoc*". To assist the user in this procedure, DCPA algorithm provides a scatter plot presenting the correlation between the number of PCAs and the data variance. The authors recommended retaining the number of components that holds at least 80 % of the variance.

Another powerful multivariate analysis method is the Factorial Analysis (FCA). Like other data reduction methods, FCA describes the observed variability of the measurements (alleles) and their correlations, using a lower number of unobserved variables called factors. This method is more suitable for complex parameters, i.e. those that result from a large number of factors. Unlike PCA, the FCA explicitly accounts for measurement errors and –like DCPA- does not hold any assumptions of the distribution of the data nor the genetic model.

All these multivariate analysis methods share the risk of inferring artifact clustering or groups that have no biological significance. Though, DCPA has the advantage of providing a graphical distribution of the elements (individuals) to the defined groups.

II.2.1.5. Bottleneck tests

When the effective size of a population goes through a recent reduction, the allelic diversity and the heterozygosity in this population are observably reduced. The number of alleles decreases faster than the number of heterozygotes, since low frequency alleles are more vulnerable to bottleneck events. Under particular allele mutation models, an excess of heterozygotes can be observed and compared to the reduced number of alleles (Comuet and Luikart, 1996). Numerous software were developed to detect the pattern of allelic and heterozygosity change resulting from demographic growth or reduction. They perform simulations of the data to determine the likelihood of past demographic change and to infer probabilities of the initial population size and the date of such events. One of the most used software is MSVAR, a full-likelihood Bayesian method for microsatellite data using SMM model.

II.2.2. Sequence analysis

II.2.2.1. Population variability parameters

Alignment of the sequences allows calculating different statistical parameters that are informative about the population variability:

(i) Number of polymorphic sites,

(ii) Parsimony informative sites: polymorphic sites occurring more than once in the samples,

(iii) Nucleotide diversity (Pi): the average number of nucleotide differences per site between two sequences (Nei, 1987),
(iv) Average number of nucleotide differences (k) between two DNA sequences, in all possible pairs (Tajima, 1983),
(v) Watterson estimator (Theta): estimates the genetic diversity based on the effective population size and its mutation rate (Watterson, 1975).

In addition, comparative parameters can inform about inter population variability such as (i) gene flow (Nm) (Nei, 1982), (ii) F_{ST} (Hudson et al., 1992) that are both comparative to those calculated for microsatellite data, (iii) the average number of nucleotide substitutions per site between populations (Dxy) (Nei, 1987), and (iv) the number of net nucleotide substitutions per site between populations (Da) (Nei, 1987).

II.2.2.2. Demographic change

Much like Hardy Weinberg equilibrium, the neutral theory of molecular evolution (Kimura, 1983) defines the state where mutations and genetic drift are in equilibrium. Tajima's D test (1989) measures the difference between the mean number of pairwise difference and the number of segregating sites. When the population is evolving neutrally, these two parameters are correlated. Tajima's D test identifies sequences that do not fit to the neutral theory model evolving under non-random process. Another parameter is Fu's (1997) Fs, which estimates the probability of observing a certain number of haplotypes from a certain value of theta. As the fluctuations in population effective size impacts the distribution of allelic frequency, these measures can inform about past demographic changes.

In addition, the distribution of the pairwise differences between the sequences (mismatch distribution) can visualize changes in the population's effective size. If the population is expanding, this distribution is unimodal, and the graph shows the frequency of pairwise difference decreasing when the number of differences rises.

II.2.2.3. Phylogeny

One of the first developed methods for phylogeny reconstruction was the Neighbor Joining (NJ) clustering method (Saitou and Nei, 1987). This method is simple and relies on the distance matrix between the sequences to infer their phylogeny. Its main virtue is its rapidity and it usually gives satisfying results when used in a low level of taxonomic difference, such as in intra-species or intra-population level.

The most statistically powerful method for phylogeny estimation is the Maximum Likelihood method (ML). It estimates the probability of observing the data, given a certain model. Thus, from a set of possible phylogenetic trees, the highest ML probability defines the most probable one. It is the statistical model that is the least affected by sampling error, it considers all possible tree topologies (unlike NJ method) and, it does not hold any assumption in the evolutionary model (unlike parsimony).

Another phylogeny inferring method is the Median-joining Networks that is more suitable for intra-specific data and proved to be mainly useful for recombination-free (ex. Mitochondrial DNA) phylogeny analysis (Bandelt *et al.*, 1999). It relies on the maximum parsimony heuristic; i.e. the number of changes (mutations) should be as low as possible. The particularity of the median joining network method is that it searches the minimum spanning trees by favoring short connections.

II.3. Parentage test

II.3.1. Parentage testing for Livestock species

The history of paternity testing goes back to 1920s. Its first form was based on blood typing with a power of exclusion was reduced to only 30% (Douris, 1922). Other forms, namely serological testing and Human Leukocyte Antigen (HLA) typing were developed during the 18th century with increased exclusion powers (Lee, 1975). However, starting to use the Restriction Fragment Length Polymorphism (RFLP) markers in the 1980s, DNA testing became available and took paternity testing to a whole new level by a power of exclusion exceeding 99.99%. Ever since, molecular biotechnology was developing rapidly with the development of Polymemrase Chain Reaction (PCR), Short Tandem Repeats (STR), and Single Nucleotide Polymorphism (SNP) arrays, and so did the paternity testing techniques. Nowadays, the very definition of paternity testing is: the biotechnology technique determining the paternity relations between individuals of a same species. Today, paternity testing can be practiced for fetus as early as 8 weeks' gestation in human, using Next Generation Sequencing (NGS) (Yang *et al.,* 2017).

For livestock animals, the development of parentage tests traced the same path. By the 1960s, the blood groups were already being used for parentage exclusions in more than a livestock species (Rendel, 1957; Buschmann, 1965; Stormont, 1967), then serology based tests were used (Rendel and Gahne, 1961), and then the RFLP technique (Theilmann, 1989). Now, STR and SNP based parentage tests are available for most -if not all- livestock species, and for a large number of natural animal species. Their cost is affordable by most livestock owners and the used techniques are customary in most laboratories.

For parentage analysis, microsatellites are more suitable and discerning than older genetic markers. They are also not less conclusive than SNPs. In a comparative study, Herraez and his collaborators (2005) found that the usage of 17 microsatellites gave paternity exclusion estimations higher than 99 %, while using 43 SNPs gave roughly 98 %. The choice depends mainly on the available equipments. A number of paternity tests based on SNPs (sheep and goat: Tourtereau *et al.*, 2014; Cattle: Werner *et al.*, 2003, Hara *et al.*, 2010, Panetto *et al.*, 2017) and others based on microsatellites (Cattle: Heyen *et al.*, 1997; Goat: Luikart *et al.*, 1999; da Silva *et al.*, 2014; Sheep: Al-Atiyat, 2015) -to name few- are available.

For Camelids, Agapito and his collaborators (2008) used 10 microsatellite markers to develop a parentage test for *Vicugna pacos*. For *Camelus dromedarius* and since they are used for racing in Middle-East, there is an extended need for parentage detecting in that population, Al Swailem and his collaborators (2010) first developed a RAPD-based test, then a microsatellite based test was described by Spencer *et al.,* 2010.

II.3.2. Statistical basis for parentage testing

Several approaches have been developed for parentage analysis and they are used in different sampling situations. The simplest method is exclusion, relying on the one to one comparison of the alleles in the candidate parent-offspring pairs. This method can only exclude decidedly non plausible pairs. In addition, mismatches due to mutations and typing errors are not –automatically- accounted for. The parental reconstruction approach can be used when known groups of full or half siblings are available. It reconstructs the parental genotypes from the offspring ones. The "Sibship" reconstruction is less decisive but handles the situation where no parents and no known groups of full and half siblings are available. Other approaches, namely fractional allocation, full probability parentage analysis and categorical assignment are statistically more powerful as they use estimations of likelihood and posterior probability for each parent-offspring pair.

CERVUS (Marshall *et al.,* 1998) is the most popular software using the categorical allocation method (Jones *et al.,* 2010). This software estimates a number of statistical

parameters: Logarithm of Odds scores (LOD), Delta scores (D) and other parameters detailed hereafter.

II.3.2.1. Features of CERVUS

Based on maximum likelihood, CERVUS calculates a LOD score for each parent-offspring pair. It is the natural logarithm (log to base e) of the overall likelihood ratio of the hypothesis: parent-offspring relationship is true. A positive LOD reflects a high possibility of the hypothesis to be verified. LOD is null if the candidate is at equal probability to be or not to be the offspring's parent. Negative scores reflect the mismatches between the two animal's genotypes. The best LOD from a list of candidate parents is considered to be the assigned parent. The LOD expression accounts for allele frequencies in the population in a genotype replacement model (Kalinowski *et al.,* 2007). This replacement model corrects for mismatches from mutations or experimental errors.

The second parameter is delta score (D). It reflects the difference between LOD scores of the best two candidates (parents for an offspring). When multiple candidates have positive LOD, they are classified following increasing D.

CERVUS runs a simulation to identify critical LOD and D values for which the assignment is the most accurate (maximal tolerated error rate = 0.05). Error rates are estimated by a unique method (Marshall *et al.,* 1998). Populations of parent-offspring pairs are simulated given a rate of genotyping error and a proportion of sampled candidates. These two parameters can be defined by the user. Alternatively standard values can be automatically defined.

In fact, after LOD-based parentage assignment, the statistic test D is calculated for each assignment. The distribution of D values for correct assignments in the simulation is compared to that for false assignments. A critical D value is then defined and used to determine a confidence in assignment of the parent. Consequently, the error rate is an experiment dependant parameter. Each parent-offspring pair is then associated to a confidence rate. A 95 % confidence is a minimum value for the population confidence level set in CERVUS.

Thus, the singularity of CERVUS resides in the experiment-specific estimated error rate, based on the replacement model described above. Other software use Bayesian posterior probabilities (Nielsen *et al.*, 2001; Koch *et al.,* 2008) or simulation-based expected error rate (Jones *et al.*, 2010) but CERVUS offers a simple and easily understandable approach and is unique in providing a confidence rate.

Another parameter CERVUS software can estimate is the inbreeding rate based on F_{IS} (Wright, 1987). Thus, if parents are inbred, information about their relatedness may be incorporated.

II.3.2.2. Limits of CERVUS

While null allele presence and frequencies are estimated in CERVUS based on departures from Hardy-Weinberg equilibrium, there is no correction for them in the parentage assignment. The objective is then to evaluate the usage of the affected loci in parentage analysis.

CERVUS is most powerful if all candidate parents are sampled (Jones *et al.,* 2010) but it accommodates when a proportion of the potential parents are missed.

But the major limit in the usage of CERVUS is that the calculations of the critical parameters; namely the confidence and error rates, are dependent of the correct choice of the model's parameters. These parameters: the genotyping error rate and the proportion of the sampled candidates, can be critical to evaluate. The simplest way to evaluate the genotyping error rate is to use known pairs as reference. For a one parent known with certainty, the mismatches between offspring and the second parents can serve to estimate the genotyping error rate (Jones *et al.,* 2010). The error rate should be reduced as much as possible by

optimizing the experimental techniques, as this parameter, even when accounted for, can substantially reduce the power of parentage analysis (Marshall *et al.,* 1998). The second parameter can be overcome when the parentage test is adequately powerful markers (highly variable and with minimal null alleles) and the error rate is sufficiently reduced.

II.3.3. Need for a parentage testing for Tunisian camels

Camel breeding in Tunisia is practiced mainly in a traditional way. Individual animal identification is not systematic and pedigrees are rarely recorded.

In fact, camels in Tunisia are bred following an extensive mode where the breeders travel with their herds, along the year, to make benefit of the different pasturing areas. This phenomenon is called transhumance. Usually, during the spring, a number of camel herds are transported to the Ouara and Dhaher regions, in the West-South of the country. Shepherds from three governorates: Medenine, Tataouine and Kebili come with their animals seeking vegetation. Likewise, El Behayer and El Chareb in Gabes, herders from Gabes and Kebili mainly but also from other regions come seeking exploitable grasslands. Exchange of sires is common between these herds, during this period. In general, this action is rather random. A parentage assay may offer a way to choose the best sire, i.e. displaying the least inbreeding coefficient with the concerned herd.

In addition, camel mating is not controlled nor recorded, thus no selection could be done on the animals. While camel breeding in Tunisia is evolving into more productive forms, high milk potential –for instance- is one of the selection objectives. For tourism activities, hefty animals and docile character are also sought for. But, no serious animal selection can be efficient when animals of high performance cannot be traced to their parents with certainty.

In addition, parentage test can be used to prevent inbreeding risk. Since some herds suffer from increased inbreeding rate, the use of a less inbred male is a necessity. In critical cases, there is a necessity to determine which male is less relative to the herd, as in some regions, inbreeding was noticed in camel population.

THESIS MOTIVATION AND OBJECTIVES OVERVIEW

The present work was interested in the evaluation of the diversity of dromedaries in Tunisia. Advance in molecular biotechnology for livestock and -lately- for this species is a significant asset that was largely exploited by using different types of genetic markers. This will allow a comprehensive view of the genetic aspect of dromedary camel population in Tunisia.

The research work in this thesis consists in three parts;

(i) Genetic characterization based on nuclear markers; aiming to assess variability estimators and explore the population structure either following the geography or the ecotypes. We aim to answer the specific question: Are the Tunisian camel ecotypes described in literature supported on the genetic level?

(ii) Genetic characterization based on maternal lineage markers; aiming to give -for the first time- an insight on the genetic variability at the mitochondrial level and to elucidate the genetic origins and the recent history of the local population.

(iii) Designing a parentage assay aiming to offer an efficient and affordable way to trace the genealogy of dromedaries, in the absence of recordings.

MATERIAL AND METHODS

III. Material and Methods

The genetic variability was studied on two levels; on autosomal chromosomes and maternal lineage using different molecular techniques; microsatellite genotyping and DNA sequencing.

III. 1. Preliminary study

III. 1.1. Available data

In a previous research project, camel DNA samples from Medenine, Tataouine and Kebili were amplified using seven microsatellite markers (VOLP03, CVRL10, CVRL02, VOLP08, VOLP10, VOLP32 and CVRL05). A total of 62 dromedary camels (20 Merzougui, 21 Ardhaoui Tataouine and 21 Ardhaoui Medenine) were successfully genotyped using 4 microsatellite markers (VOLP08, VOLP10, VOLP32 and CVRL05), and the genotyping database was thus available. This database was analyzed in order to construct a preliminary view of (i) the genetic variability of the camel population in Tunisia and (ii) of the genetic differentiation of some of the known and most common ecotypes (Ardhaoui Medenine, Ardhaoui Tataouine and Merzougui).

III.1.2. Statistical Analysis

The software GenA1Ex V6.501 (Peakall and Smouse, 2006) was used to estimate different statistic parameters including observed and expected heterozygosity and F statistics (Weir et al., 1984). We also used GENEPOP V4.1.2 (Raymond and Rousset, 1995) to perform exact tests of Hardy Weinberg Equilibrium, using Markov Chain Monte Carlo (MCMC) simulations (100 batches of 1000 iterations each, with a dememorization number of 10000). Based on Nei's genetic distances (Nei et al., 1983), a neighbor joining tree was inferred using POPULATION V1.2.32 (Langella, 1999).

III.2. Protocols for the large study of population genetics analysis

III.2.1. Animals' sampling

III.2.1.1. Sampling strategy

As previously cited in the bibliography review, 85 % of the total dromedary camel population in Tunisia is hosted by only three administrative divisions; Tataouine, Medenine and Kebili, situated in the south of the country. Being the case, the sampling process covered these regions in addition to Gabes, another southern administrative division, as shown on the following map.

Figure 8. Geographic localization of the sampled regions

Finally, 248 animals of different sex and age were sampled, including 36 related animals forming 17 trios (17 distinct mothers, 17 offspring and 2 breeding males) from the experimental herd of IRA Medenine.

III.2.1.2. Sampling protocol

The camel ought to be caught and completely immobilized in order to safely draw blood from the jugular vein in 5 ml vacutainer tubes containing EDTA (Ethylene-diamine-Tetra-acetic Acid), to prevent coagulation, then conserved at -20°C for later use.

III.2.2. Extraction of the whole DNA

After thawing, 250 µl of blood sample were used to extract DNA following the salting-out protocol (Miller *et al.,* 1988). Quantity and quality control were performed by optical density (OD) measurements at two wave lengths: 260 nm and 280 nm, using a spectrophotometer. The OD260/OD280 ratio estimated the purity of the product. If needed, additional purification by phenol-chloroform was performed.

III.2.3. Genotyping using microsatellites

III.2.3.1. Choice of microsatellites

Among 25 microsatellites recommended by the joint Food and Animal Organization of the United Nations (FAO) and International Society for Animal Genetics (ISAG) for dromedary camel population studies (FAO-ISAG, 2001), 18 have been chosen to be used in this study, in addition to two other microsatellite markers. The markers are autosomal, yet their positions on the 74 dromedary camel chromosomes are undetermined. Their size ranges and repeat motifs are summarized in the following table, while more details on the primer sequences and annealing temperatures figure in annex 1.

Table 5. Microsatellite markers description

Microsatellite	Repeat motif	Gene Bank Accession number
CMS9[a]	$(GT)_{24}$	AF305231
CMS121	$(TG)_{24}$	AF305234
CMS13[a]	$(AC)_{27}$	-
CMS15[a]	$(TG)_{23}$	-
CMS17[a]	$(AT)_{38}$	AF217608
CMS18[a]	$(GT)_{14}$	AF091125
CMS25[a]	$(CT)_{33}$	AF217606
CMS32[a]	$(CA)_{30}$	AF217602
CMS50[a]	$(GT)_{27}$	AF217604
CVRL1D[b]	$(GT)_{27}(GC)_6(GT)_9$	AF217601
CVRL4D[b]	$(GT)_{19}$	AF329149
CVRL5D[b]	$(GT)_{25}$	AF329146
CVRL6D[b]	$(TA)_5(CA)_{35}(TA)_4$	AF380345
CVRL07[b]	$(GT)_{14}(AT)_{14}$	AF329148
CVRL08[b]	$(CA)_{10}(GA)_5$	AF329147
LCA66[c]	$(CA)_{13}$	AF329151
VOLP10[d]	$(TG)_2TA(TG)_7TA(TG)_7$	AF329158
VOLP32[d]	$(TG)_{20}$	AF329159
YWLL44[e]	AC=TG repeats	AF329160
YWLL59[e]	$(CA)_{17}$	AF091125

[a] Evdotchenko *et al.,* 2003. [b] Mariasegaram *et al.,* 2000. [c] Penedo *et al.,* 1999. [d] Obreque *et al.,* 1998. [e] Lang *et al.,* 1996.

III.2.3.2. Samples

The 126 blood samples used for microsatellite genotyping touched 87 herds in four regions, with a maximum of two samples from each herd. The following table details the distribution of the samples by region. A complete survey was conducted, and different information about the sampled animals were recorded: sex, herd of origin, geographic location, ecotype, robe color and photo.

Table 6. Distribution of the samples in the regions

Region	Dromedaries headcount[a]	Number of Samples	Sampled herds
Medenine	19,000	41	25
Kebili	13,000	41	35
Tataouine	9,745	38	24
Gabes	1,660	6	3
Total	**43,405**	**126**	**87**

[a]OEP, 2016

III.2.3.3. DNA Amplification

The 20 microsatellites were divided into three multiplexes; two containing 7 microsatellites each and one containing only 6. In order to normalize the DNA for multiplex PCR reactions, all samples were diluted to a final concentration of 30 ng/µl.

The multiplex polymerase chain reactions (PCR) were conducted in 10 µl volumes. All DNA samples being of the same concentration, a standardized protocol was applied to all samples, using the Qiagen multiplex PCR kit. This kit provides a "Master mix" containing Qiagen PCR Buffer, $MgCl_2$, dNTP mix and HotStarTaq DNA polymerase. It had been chosen to minimize optimization phase and pipetting steps, and it was used to a final concentration of 1X. HotStarTaq DNA polymerase is modified polymerase held inactive at ambient temperatures which improve the PCR specificity by preventing misprimed products and primer-dimers. The primer mixes of the three multiplexes are detailed in annex 2.

The reaction program included 35 cycles of (i) denaturation (30 s) at 95 °C, (ii) hybridization (90 s) at 60 °C for multiplexes I and II and 58 °C for multiplex III, and (iii) elongation at 72 °C (1 mn), for each cycle. An initial denaturation of 2 mn at 95 °C, and a final elongation of 30 mn at 60 °C were also applied. The PCR products were then stored at -20 °C.

III.2.3.4. Capillary electrophoresis

A control of PCR product quality was performed using NanoDrop ND-2000 spectrophotometer (Thermo Scientific).

For capillary electrophoresis, 1 µl from each sample (a minimum of 10 ng/marker is required) were transferred in a 96 wells plate and dried at 65 °C for 4 mn as recommended by the sequencing laboratory. If necessary, the fragment detection was repeated using 2 µl of the samples. DNA fragment detection was outsourced to an external laboratory, BMR Genomics (http://www.bmr-genomics.it), affiliated to Università degli Studi di Padova and accredited by it. The capillary electrophoresis was realized in an APlied Biosystems AB3500 Genetic Analyzer.

Table 7 details the fluorescence dyes for each microsatellite marker. The used fragment marker was Liz 500, offering 16 peaks ranging from 35 bp to 500 bp.

Table 7. List of used dyes relative to the microsatellites' allele size range

Microsatellite	Dye	Allele size range (bp)
CMS9	6-FAM	225-256
CMS121	6-FAM	147-173
CMS13	6-FAM	236-265
CMS15	6-FAM	116-144
CMS17	HEX	135-167
CMS18	6-FAM	157-170
CMS25	HEX	90-110
CMS32	HEX	198-209
CMS50	6-FAM	129-190
CVRL1D	HEX	188-253
CVRL4D	HEX	156-180
CVRL5D	6-FAM	148-179
CVRL6D	HEX	155-175
CVRL07	6-FAM	265-315
CVRL08	6-FAM	191-209
LCA66	HEX	212-262
VOLP10	HEX	236-268
VOLP32	HEX	256-290
YWLL44	HEX	86-120
YWLL59	HEX	103-117

The output (.fsa files) was read and analyzed using PEAKSCANNER V1.0 (APlied Biosystems, 2006).

III.2.3.5. Data analysis

III.2.3.5.1. Descriptive analysis

Different statistics were estimated to evaluate the informative potential of the used microsatellites, such as (i) pairwise marker linkage disequilibrium (by Markov chain, based on 5000 iterations), (ii) Hardy-Weinberg equilibrium by marker (tested for statistic significance using chi-square test and likelihood ratio test), and (iii) Polymorphic Information Content. These parameters were calculated mainly using POPGENE (Yeh *et al.* 1999), or manually.

On the whole population, classic descriptive parameters in population characterization were estimated using a number of software. Table 8 summarizes the parameters, the software and the methods. The input files were either prepared manually or obtained by converting from another input form, using CONVERT 1.31 (Glaubitz, 2004).

Table 8. Statistical parameters for population description

Parameter	*Method*	Software
Allele frequency	*Simple statistics*	MSA 4.05[a]
Allele richness	*Simple statistics*	MSA 4.05
Effective number of alleles	*Kimura and Crow, 1964*	POPGENE 1.31[b]
Observed heterozygosity		POPGENE 1.31
Expected heterozygosity and unbiased heterozygosity	*Nei, 1973 and Levene, 1949*	POPGENE 1.31
Hardy-Weinberg Equilibrium	*Markov chain exact p-value*	GENEPOP 4.1.2[c]
Null allele frequency	*EM algorithm Dempster*	GENEPOP 4.1.2

	Laird and Rubin, 1977	
Inbreeding (Fis)	*Wright 1978*	POPGENE 1.31
F statistics	*Barton & Slatkin, 1986*	GENEPOP 4.1.2
Gene flow (Nm)	*Nei, 1987*	POPGENE 1.31
Pairwise relatedness	*Maximum Likelihood method Jinliang Wang, 2007*	COANCESTRY 1.0.2[d]
Genetic distance	*Nei, 1972*	POPGENE 1.31

[a] Dieringer *et al.,* 2003. [b] Yeh *et al.,* 1999. [c] Rousset, 2012. [d] Wang, 2007.

Ewens-Watterson test was performed to test the neutrality of microsatellite markers; the statistic F (sum of square of allelic frequency) and the upper and lower limits at 95% confidence region for the test were calculated using 1000 simulated samples that is implemented in Popgene software package (Yeh *et al.,* 1999).

III.2.3.5.2. Exploratory analysis

In addition to the parameters referred above, different methods are available to model the relations between individuals in the population, in order to reveal any clustering among them and to infer their phylogeny. The following table sums up the analyses conducted on the data in the current study.

Table 9. Exploratory analyses and the corresponding used software

Analysis	***Method***	**Software**
Population structure	*Bayesian method*	STRUCTURE 2.3.4[a] STRUCTURE Harvester[b]
Clustering	*Principle Component Analysis* *Factorial Analysis (ACF)* *DiscriminativeAnalysis of Principal Component*	RSTUDIO/adegenet 2.0.1[c] GENETIX 4.05.2[d] R STUDIO/pegas 1.0[e]
Phylogeny	*Neighbour joining*	DARWIN 6.0.15[f]
Bottleneck test	*Bayesian likelihood*	MSVAR 0.1.1[g]

[a] Pritchard *et al.,* 2000. [b] Earl *et al.,* 2012. [c]Jombart, 2015. [d] Belkhir *et al.,* 2004. [e] Paradis, 2010. [f] Perrier et Jacquemoud-Collet 2010. [g] Beaumont 1999.

In Bayesian clustering performed by STRUCTURE, many models were tried: admixture, no admixture, linked and independent frequencies. The use of prior was also tested with the ecotypes or the location as a prior. Each test was performed with a Burnin period length of 100000 and the Monte Carlo Markov Chain repetitions were set to 10000.

Finally, to test the hypothesis that the dromedary population in Tunisia went through a bottleneck in the past, MSVAR autoregression test was run under 4 hypotheses; two population expansion hypotheses with an initial effective number of (i) Ne = 10^3 to Ne =10^5 and (ii) Ne = 10^2 to Ne =10^4, and two population inclination starting from initial populations of (iii) Ne = 10^5 to Ne =10^3 and (iv) Ne = 10^4 to Ne =10^2. The generation time was set to 4 (age of reproduction in female camels), but another try was performed with a generation time of 5.

III.2.4. Mitochondrial DNA sequencing

III.2.4.1. Amplified sequences

Two regions of the mitochondrial DNA (mtDNA) were amplified; Displacement loop or *D-loop* also called control region (CR) which is a neutral region and the *cytochorme b* (CB), a gene implicated in the respiration process.

III.2.4.1.1. *Cytochorme b* gene

Cytochorme b is a component of the respiratory chain complex III. The protein is 380 amino acids and the gene is of 1140 bp. The gene sequence has stable parts which allows for universal primers, and variable parts that are species specific.

III.2.4.1.2. Displacement loop sequence

The *D-loop* is a DNA structure where the strands of the double-stranded DNA are separated and kept apart by a third strand of DNA. Thus, replication of mtDNA can occur starting in the *D-loop* form. Certain parts of this region are conserved but large parts are highly variable and are used in phylogeny.

III.2.4.2. PCR

The extracted DNA was used to amplify the desired regions by primers in table 10. An additional internal primer (F') for CR sequencing was designed. The reagents used in the reactions are described in annex 3. The same amplification program was applied for the PCR of both regions; an initial denaturation phase of 5 mn at 95 °C preceded 30 cycles of denaturation (30s) at 95 °C, annealing (30s) at 60 °C both for the CR and CB, and elongation at 72 °C (1 mn), followed by a final elongation of 10 mn at 60 °C. The PCR products were controlled on an agarose gel (1.5 %) before storage at -20 °C. The following photo shows the PCR product of the control region's amplification in 13 samples.

Table 10. Mitochondrial DNA primers description

mtDNA region	Primers sequences	Ta (°C)	Estimated size
Control Region	F : CCAAAGCCCCACCACCAG	60	1316 bp
	R : CCATCCAGGCATTTTCAGCG		
	F' : AACTGTGGTGTCATGCATTTGG		
Cytochorme b	F : AGCCTTCTCTTCAGTCGCACAC	60	1289 bp
	R : GCCCATGAA AGCTGTTGCT		

Ta: Annealing temperature

III.2.4.3. Sequencing

The amplified regions were sequenced by Sanger method using an APlied Biosystem AB3500 sequencer, in BMR Genomics laboratory (www.bmrgenomics.it).

The standard Sanger sequencing protocol of BMR Genomics requires 1-2 ng per 100 bp for sequencing reaction and recommends sending double the quantity. From each sample, 1 µl was transported to 96 well plates and dried at 65 °C for 4 mn. The laboratory also requires 13 pmol of each primer per sample to be sent with the samples.

In BMR Genomics, the samples are first purified then sequenced in an AB3500 automatic sequencer. We chose the ExoSAP-IT (Thermo Scientific) PCR product cleanup method. It is an enzymatic reaction aiming to decompose the trailing primers and dNTPs that can interfere with the sequencing. However, this step does not remove salts or other secondary PCR products. Samples are then transported to 96 wells plates. Table 11 summarizes the reagents and steps of this protocol.

Table 11. ExoSAP purification protocol

Protocol	PCR Product	5 µl
	Exonuclease I	5 U
	Shrimp Alkaline Phosphatase	1 U
Cycle	30 mn at 37 °C	
	30 mn at 80 °C	

III.2.4.4. Sequence analysis

Since the number of sequences and their length are restraint, the manual manipulation of the data using visualizing software can be performed. On the other hand, basic tools for sequence analysis are freely available and allow more flexible single step workflow design.

The sequencing outputs (.ab1 files) are capillary sequence trace files containing the PHRED quality scores for the base calls. To each nucleotide, the PHRED quality score Q (see the formula thereafter) encoded in ASCII code to obtain a single character.

$$Q = -10 \log_{10} P$$

Where *P* is the probability of a base-call being erroneous.

The conversion from an "ab1" file (chromatogram) to a "fastq" file -containing the DNA sequence and the PHRED scores- is possible through European Molecular Biology Open Software Suite EMBOSS 6.5.0 (Rice *et al.,* 2000) tool baptized "seqret". All the sequences were grouped in one file using a simple script written in Perl 5.12.1 (Wall, 2010).

The next step is to perform a quality check on the sequences base per base. "Fastqc" tool from EMBOSS 6.5.0 outputs a report including different summary statistics and visualizations of the sequences. Mainly, a visual presentation of the quality scores per base position is a very useful tool to decide the trimming intervals in the beginning and ending of the sequences where the quality deteriorates. The tool visualizes also a graph of the statistics on average quality scores per sequence albeit Sanger sequencing technique yields high quality sequences ($Q \geq 20$). It also traces graphs on N content per base, sequence length distribution and other statistics.

On the light of these statistics, "fastx_trimmer" from FASTX-TOOLKIT 0.0.13 (Hannon lab, 2010) was used to trim sequences' end and beginning. Then the quality was checked again by "fastqc".

The complementary and reversed sequences of each reverse sequence (R), obtained from 3'-5' DNA strand sequencing, were then inferred using "revseq" tool from EMBOSS 6.5.0. For each sample (animal), the forward sequence and reverse sequence were merged using "merger" (EMBOSS 6.5.0) to obtain a unique and complete sequence of each sequenced regions (CB and CR).

"Fastx_trimmer", "merger" and "revseq" are scripts written in Perl and Python languages. Each used script was inspected to understand the basic steps they perform and to make sure of their suitability with respect to the objectives of the performed tasks.

Finally, the multi sequence alignment was performed using Clustal O 1.2.2 (Sievers *et al.,* 2011). Clustal Omega is the last version in the Clustal family, written in C and C++ languages and based on Clustal W (Larkin *et al.,* 2007). Methodologically, this software progresses in three main steps: (i) pairwise alignment based on local alignment algorithm, (ii) construction of guide tree based on sequence embedding, and (iii) multiple sequence alignment based on Unweighted Pair Group Method with Arithmetic Mean (UPGMA) that clusters the sequences in an agglomerative way.

Through all these steps, specific Perl scripts were written in order to automate different tasks and to format input files. Using a Perl script also, the repetitive region in the CR was

identified by position and eliminated as -in general- repetitive regions have high mutation rates and can be misleading for the statistical analyses.

III.2.4.5. Descriptive analysis

All the found SNPs were inspected manually to check the quality of the base-calling. A number of parameters have been estimated using the aligned sequences to a reference of whole mitochondrion DNA (GenBank accession number: NC_009849.1) of an Arabian camel, from the National Center for Biotechnology Information (NCBI) website. Number of segregating sites, number of haplotypes, nucleotide diversity, average number of pairwise differences and Theta (Watterson, 1975) were estimated for both sequenced regions separately. They have been also estimated for a consensus sequence. The consensus sequence was defined by the longest DNA sequence -shared between samples- from CB, concatenated to that from CR. The software DNAsp v6.10.04 (Rozas et al., 2017) was used to read and visualize the alignment file produced by Clustal O 1.2.2 (.phy files) and to infer the statistical parameters.

III.2.4.6. Phylogeny

Two possible methods may be applied to infer phylogeny in the case of this study: (i) constructing a consensus sequence for all samples and infer the phylogenetic tree form it, and (ii) inferring phylogenetic trees from each of the sequenced region (CB and CR) and then estimating a consensus tree. Both methods were applied.

First, the sequences were compared to the NCBI database, using BLAST. The identical and the best hits of the new haplotypes were downloaded in order to perform the sequence alignment (table 12). Mitochondrion genomes of dromedaries from different geographical regions were used. As a probable origin of domestication, two samples from the Middle East; Dubai (Arabian Peninsula) and Iran (Near Asia) were used. We also included a sample from Morocco since it is the sole representative from North African dromedary population in the NCBI database.

Table 12. Reference and accession number of the different mtDNA genomes used for genetic comparison and phylogeny

Sample	**Source**	**GenBank Accession number**	**Reference**
Reference genome	Dubai	NC_009849.1	Huang *et al.*, 2007
Arabian camel	Pakistan	KU605080	Mohandesan *et al.*, 2016
Arabian camel	Morocco	JN632608.1	Hassanin *et al.*, 2012
Arabian camel	Iran	KX554934	Tahmoorespur *et al.*, 2016
Arabian camel	Kenya	KU605078	Mohandesan *et al.*, 2016

Sequences of CB from the dromedary samples and from the downloaded references were aligned by Clustal O, Phylip 3.697 (Felsenstein, 2005). Then, they were used to infer phylogenetic trees by three methods: maximum likelihood, neighbor joining and parsimony using bootstrapping. Figure 9 goes over all the sequence analysis steps.

In another approach, ".phy" files from CB and CR alignments were visualized and concatenated into one sequence, using MEGA 7.0.26 (Tamura, 2007). Subsequently, phylogenetic tree from the whole sequenced parts through the samples was also constructed. As well, median joining method was applied via Network 5 (Bandelt *et al.*, 1999), inferring a schematic presentation of the relations between the different found haplotypes.

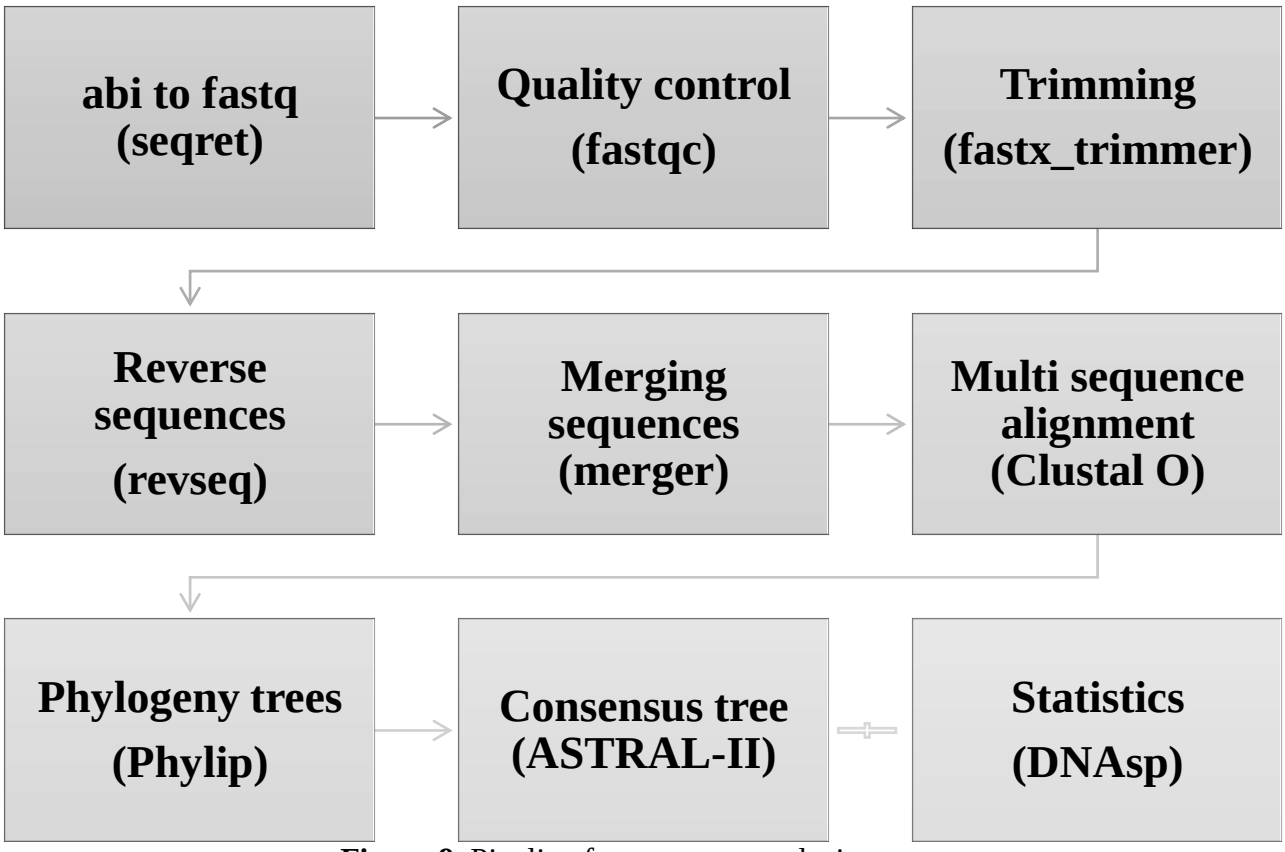


Figure 9. Pipeline for sequence analysis

III.3. Parentage test

III.3.1. Animal sampling

From the IRA's experimental herd, and along three reproducing seasons (2014 to 2016), 17 trios (17 offspring, 17 distinct mothers and 2 fathers) were sampled, totalizing 36 DNA samples.

Blood samples were drawn as described in the section III.1.1.2, then DNA was extracted as described in III.1.2.

III.3.2. Choice of the microsatellite markers

After being used and tested on the whole Tunisian dromedary camel population, the 20 microsatellite markers previously described in table 5 were checked to fulfill the criteria hereafter.

III.3.2.1. Polymorphism

For obvious reasons, the microsatellite marker should be polymorphic, at a threshold p = 0.05, i.e. the major allele frequency should not exceed 95%.

III.3.2.2. Linkage disequilibrium (LD) test

This parameter quantifies the probability that two genetic loci are inherited in a linked pattern. This pairwise probability was calculated through all the used microsatellites in order to eliminate those that are linked. This linkage may not be physical but only aParent due to the population differentiation effect. The usage of both markers is futile as the second does not carry any new information to the test.

III.3.2.3. Hardy-Weinberg test

This parameter was performed to estimate the impact of any mechanism of evolutionary change that can influence the microsatellite frequencies, and thus the stability and efficiency of the parentage test.

III.3.3. Experimental scheme for microsatellite panels evaluation

Different combinations of the microsatellite markers were tested in order to obtain an efficient but economical final parentage test. Markers with higher information content (PIC) and mean number of alleles (MEA) were tested first.

The software CERVUS 3.0 (Kalinowski *et al.,* 2007) was used to estimate the exclusion power (LOD and error rate) and the assignment confidence for each tested pair of animals.

The statistical test design consisted in two scenarios described hereafter. The evaluation of the efficiency of each markers set was based on the LOD, assignment rate and confidence estimates.

III.3.3.1. Double masked test

For each offspring, both the father and the mother were masked. CERVUS estimated the probability of each offspring-mother pair as well as offspring-father pair following the highest LOD. For each tested set of markers, the assignment rate and error rate were estimated.

III.3.3.2. Inferring paternity with known maternal origin

For each offspring, we suppose the mother is declared. The father is one of the two males: CFA or 373. Considering the highest LOD, each of the 17 offsprings were assigned to a father. The accuracy of the marker set was evaluated by calculating the error rate.

The sets of markers yielding the best assignment rates (99% or higher) with the minimal error rate were then compared on an economical basis. To minimize the test cost, the markers' number should be as low as possible and preferably, they can be run in one multiplex, i.e. have the same annealing temperature.

RESULTS

CHAPTER IV: Population genetics based on microsatellites

IV. Population genetics based on microsatellite markers

In this section, we try to explore the genetic variability at the nuclear level, using microsatellite markers.

IV.1. Preliminary study

This analysis aimed to offer a first insight on the level of variability and to gauge the utility of a larger study on the different ecotypes in Tunisia.

IV. 1.1. Statistics on the whole population

From the 7 amplified loci, only 4 successfully gave usable genotyping data. The number of alleles per locus varied from 6 to 7. No linkage disequilibrium was detected at p = 0.05. The characteristics of the microsatellites are shown in table 13. All genotyped samples were heterozygotes for all loci. Observed heterozygosity was significantly higher than the expected value (0.78). All loci showed significant ($p < 0.05$) departures from the Hardy-Weinberg proportions in the whole subpopulation.

Table 13. Number of alleles (NA), effective number of alleles (N_e), and size range of the microsatellites loci

Locus	Size range (bp)	NA	N_e
VOLP10	245-273	6	4.37
VOLP08	136-158	7	5.82
VOLP32	245-282	6	4.04
CVRL05	151-191	7	4.13
Mean	**-**	**6.5**	**4.61**

IV.1.3. Statistics on the ecotypes

Three groups were defined by their respective ecotypes, based on geographic origin and morphologic descriptions; Ardhaoui Medenine, Ardhaoui Tataouine, and Merzougui. The number of samples per group was balanced. Most alleles were equally present in all groups (table 14). Yet, two alleles of CVRL05 were private to the Merzougui group.

Table 14. Number of samples (N), number of alleles (NA), observed (H_0) and expected (H_e) heterozygosity for the main ecotypes of dromedary population in Tunisia

Locus	N	NA	H_0
Ardhaoui Medenine	21	23	1
Ardhaoui Tataouine	21	21	1
Merzougui	20	25	1

Since the heterozygotes were in excess, F_{IT} was negative (-0.226). F_{IS} was estimated to -0.295 ± 0.030 (table 15). Negative values of F_{IS} and F_{IT} values may be explained by random mating increasing heterozygosity. Within subpopulation inbreeding estimates, per group across loci, were of -0.369 (Ardhaoui Tataouine), -0.258 (Ardhaoui Medenine) and -0.265 (Merzougui). These negative values confirm the random mating of individuals in these groups.

Mean F_{ST} over all loci reached 0.052 (SD = 0.02), and showed a low to moderate differentiation between the ecotypes. The most informative marker was VOLP32, which indicated a 10 % differentiation between ecotypes. Accordingly, 95 % of the genetic variability was accorded to differentiation within each subpopulation. Moreover, mean

frequency of private alleles was low (0.087) and did not contribute greatly to the groups differentiation. The number of migrants (N_m = 4.2) was relatively high which may have prevented genetic drift from causing local genetic differentiation.

Table 15. F statistics at each locus across the three groups

Locus	F_{IS}	F_{IT}	F_{ST}
VOLP10	-0.304	-0.273	0.024
VOLP08	-0.208	-0.188	0.016
VOLP32	-0.340	-0.206	0.100
CVRL05	-0.327	-0.239	0.066
Mean	**-0.295 ± 0.030**	**-0.226 ± 0.019**	**0.052 ± 0.020**

The genetic differentiation between Ardhaoui Medenine and Merzougui groups was rather low (0.015). This could be mainly due to the cross breeding between the two subpopulations which share the same grazing area (Dhahar rangelands) in the transhumance period. F_{ST} was the highest (0.061) for Merzougui compared to Ardhaoui Tataouine group, reflecting the geographical isolation between these two regions. All Nei's genetic distance values were feeble (table 16), but pointed to a genetic resemblance between Ardhaoui Medenine and Merzougui groups, while ME and AT groups were genetically more distant. Neighbor-Joining tree (Figure 10) shows these genetic distances between the groups.

Table 16. Estimates of pairwise F_{ST} distances between the analyzed groups (above diagonal) and Nei's standard genetic distances (below diagonal)

	Ardhaoui Medenine	Ardhaoui Tataouine	Merzougui
Ardhaoui Medenine	-	0.041	0.015
Ardhaoui Tataouine	0.326	-	0.061
Merzougui	0.123	0.535	-

Figure 10. Neighbour-Joining phylogenetic tree based on Nei's standard genetic distance. ME: Merzougui, AB: Ardhaoui Medenine, AT: Ardhaoui Tataouine

IV.1.4. Discussion and Conclusions

As a preliminary work, this study gave a first insight on the variability of the dromedary population in Tunisia, using microsatellite markers.

Overall the population, the heterozygotes rate (100 %) was higher than that (60 %) reported by Ould Ahmed *et al.* (2010). This excess in heterozygotes can be explained by an isolate-breaking resulting from herd belonging to three different tribal regions. The population exhibited a significant Hardy-Weinberg disequilibrium. When associated to the heterozygotes excess, it may indicate either the occurrence of out-breeding or the presence of an over-dominant selection favoring heterozygote individuals. These results should be further

investigated in the main research work, involving a larger number of samples as well as that of the microsatellite markers.

Genetic differentiation between the considered groups; Ardhaoui Medenine, Ardhaoui Tataouine, and Merzougui (5 %) indicated a low to moderate level of 5 % and was comparable to that (8.3 %) previously reported for Tunisian camel population (Ould Ahmed *et al.*, 2010). It was also comparable to other camel populations such as the Indian (8.3 %) (Vijh *et al.*, 2007), and the Moroccan population (7 %) (Piro *et al.*, 2011). In other studies, F_{ST} between populations from different countries, as distant as Pakistani, Somali, and Saudi Arabian populations, were reported to range between 7 and 10 % (Shulz *et al.*, 2010).

On the other hand, Nei's genetic distance between the groups ranged between 0.1 and 0.5. While Ould Ahmed and his collaborators (2010) considered the geographic regions rather than the ecotypes in their study, they found Nei's genetic distance values varying from 0.1 to 0.3. They found Medenine and Tataouine regions to be the closest. Kebili region was at similar distance from Tataouine and Medenine. While the ecotypes do not faithfully represent the geographic regions in Tunisia, the genetic distances were comparable to those in the current study.

In conclusion, Ould Ahmed *et al.*, 2010 reported high variability of camel population in Tunisia. However, results from the current study were not consistent with results from other research works (Ould Ahmed *et al.*, 2010; Cherifi *et al.*, 2016), especially in terms of the excess of heterozygotes. This may be accorded to the limited number of samples and particularly of microsatellite markers. Furthermore, in this work did not include representatives of all known ecotypes. The reduced amount of data did not permit to inspect its clustering. Consequently, additional investigation is required to define if the Tunisian camel population presents separate genetic entities.

IV.2. Population variability based on 20 microsatellite markers

In the current study, we were interested in different aspects of the genetic results; quality and potential of the markers, and population description and structure.

IV.2.1. Experimental results

On the whole, DNA extraction yielded good quality and high concentration of the genomic DNA (figure). The DNA concentrations were estimated using an optical spectrophotometer. Overall, DNA concentrations ranged between 1000 and 3000 ng/µl. When the purity was poor (OD260/OD280 < 1.7), an additional phenol/chloroform purification was aPlied. The final purity of all samples ranged between 1.7 and 2.

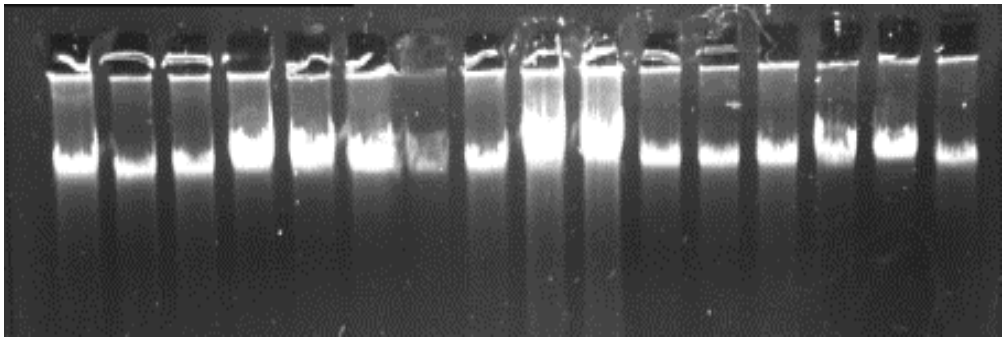


Figure 11. Electrophoresis of genomic DNA on agarose gel (0.8 %)

IV.2.2. Descriptive parameters of the markers

IV.2.2.1. Polymorphism and allele count

The CMS17 locus was monomorphic on all analyzed samples and thus was eliminated from all downstream analysis. The remaining 19 microsatellites were polymorphic at 99 % threshold and the maximum major allele frequency, through all loci, was of 80 %. With the exception of the locus VOLP32 that was bi-allelic (table 17). Maximum likelihood estimation of null alleles was noticeably high for two loci; CVRL6D and VOLP32, and six loci had no null alleles (likelihood ≤ 1 %).

More than half of the loci showed high informative potential as genetic markers, with PIC ≥ 70 % (table 17). Yet, three microsatellite markers had low informative content (PIC < 0.4) CMS18, CVRL6D and VOLP32.

Table 17. Parameters relative to the microsatellite loci in the studied population

Locus	N_a	P(N)	PIC
CMS9	7	0.204	0.707
CMS13	8	0.05	0.773
CMS15	8	0.149	0.748
CMS18	3	0.123	0.347
CMS25	5	0.147	0.633
CMS32	4	0.299	0.63
CMS50	8	0.173	0.829
CMS121	11	0.109	0.749
CVRL1D	17	0.018	0.808
CVRL4D	5	0	0.619
CVRL5D	10	0.808	0.701

CVRL6D	3	0.306	0.371
CVRL7	10	0.317	0.79
CVRL8	5	0	0.401
LCA66	8	0.097	0.734
VOLP10	7	0.236	0.73
VOLP32	2	0.893	0.211
YWLL44	11	0.092	0.596
YWLL59	5	0.181	0.538
Mean	7.368	-	0.627
SD	5.31	-	0.176

Na: Number of alleles. PIC: Polymorphic Information Content. SD: Standard Deviation.

IV.2.2.2. Linkage Disequilibrium (LD)

Out of 361 possible LD, Fisher's test for statistical significance gave only four significant linkage disequilibria (p = 0.005); CVRL4D and CVRL6D, YWLL44 and CVRL4D, CMS9 and VOLP10, and finally LCA66 and CMS50. At p = 0.001, no significant LD was detected. Sixteen pairwise correlations (Weir, 1979) were high (> 75 %), of which 4 were very high (90 %). CVRL6D was highly correlated to 5 loci. These correlations could be explained by the restraint number of alleles of this locus (3) and the high frequency (0.73) of its major allele which leads to a low number of possible genotypes and generates false correlation rates.

IV.2.2.3. Hardy Weinberg Equilibrium (HW)

The test for Hardy Weinberg equilibrium by locus showed that twelve loci were in departure from this equilibrium with high significance (p = 0.01), and that another locus was in departure at a threshold p = 0.05. Only 6 loci verified the HW equilibrium: CVRL4D, CVRL5D, CMS121, CMS9, CMS18, and VOLP32.

IV.2.2.4. Discussion

From the 20 used microsatellites, the only marker that was monomorphic in this study, and thus useless, was CMS17. In fact, this marker was developed for *C. bactrianus* and aPears to be monomorphic in *C. dromedarius*, as Almathen *et al.* (2016) found it to have a sole allele in 1083 samples from 20 countries.

In the present study, all the remaining 19 microsatellite loci were highly informative; the mean PIC was 0.62 and the mean number of alleles was 7.3. In general, the standard error of genetic distance estimates is reduced when each locus displays four or more alleles (Li *et al.*, 2002). Most loci (13) exhibited departure from HW equilibrium that most likely resulted from the effect of the studied population (presence of population structure and/or non random mating). No LD between the loci was significant at p = 0.01, and roughly 1 % of the LD tests were significant at p = 0.05. The physical linkage between these microsatellites cannot be completely discarded due to the absence of the mapping of these markers on the camel's genome. Nevertheless, this moderate linkage is most likely resulting from the properties of the loci; a restraint number of alleles or an unbalanced distribution of the allele frequencies.

To conclude, the microsatellites proved an adequate suitability for usage in the population characterization and structure. The markers have differential potentials for usage in parentage assay, which should further be examined.

IV.2.3. Variability in the whole population

The main purpose of this section is to evaluate (i) the genetic diversity of *Camelus dromedarius* population in Tunisia, (ii) the differentiation of the phenotypically distinguishable ecotypes or of local sub-populations, and (iii) relationships among them.

IV.2.3.1. Description of the population

IV.2.3.1.1. Allele count and frequencies

The number of alleles by locus was highly variable between the loci; from 18 alleles for the most polymorphic (CVRL1D) to merely two alleles for VOLP32. Yet, the average effective number of alleles (3.23) is largely inferior to the average number of alleles, which showed that an important number of low frequency alleles were present in the population (table 18). These alleles contribute poorly to the effective number and thus to the observed heterozygosity. Details of all allele frequencies are available in annex 4.

In their majority, allele frequencies of each locus were L-shaped (figure12), showing a distribution where few alleles were highly frequent, with multiple alleles sharing a small proportion in the population. This generally occurs in the case of a balance of genetic drift and mutation rate.

Table 18. Number of alleles (Na) and effective number of alleles (Ne) for all loci

Locus	**N_a**	**N_e**
CMS9	7	3.41
CMS13	8	4.41
CMS15	8	4
CMS18	3	1.52
CMS25	5	2.69
CMS32	4	2.7
CMS50	8	5.98
CMS121	11	3.98
CVRL1D	17	5.27
CVRL4D	5	2.6
CVRL5D	10	3.38
CVRL6D	3	1.56
CVRL7	10	4.82
CVRL8	5	1.68
LCA66	8	3.75
VOLP10	7	3.72
VOLP32	2	1.26
YWLL44	11	2.53
YWLL59	5	2.16
Mean	7.368	3.23

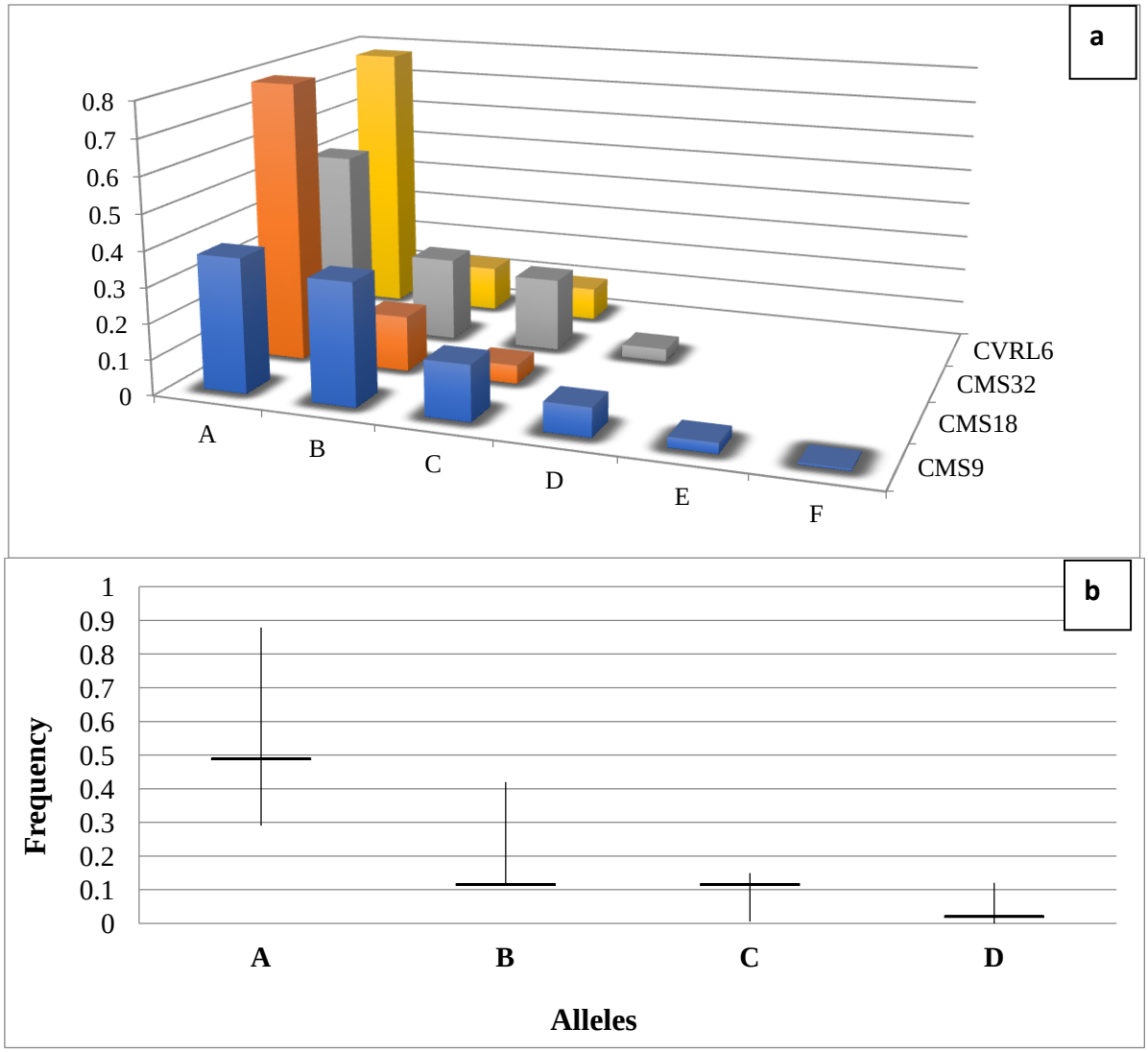


Figure 12. Normal distribution of the loci allele frequencies. (a). L-shaped distribution of allele frequencies for four different loci. (b). Mean and range of allele frequencies of the major allele (A), minor allele (D) and intermediate alleles (B and C) throughout all loci.

IV.2.3.1.2. Heterozygosity

The observed heterozygosity was largely variable across the loci; only 16 % of heterozygotes were observed at VOLP32, and the maximum proportion of heterozygotes (83 %) was observed at CMS50. Table 19 indicates the observed and the expected heterozygosity at all loci and for the whole samples. It also contains the unbiased heterozygosity (Nei, 1978), that corrects for low number of samples.

Table 19. Observed (H_0), expected (H_e) and unbiased (Hnb) heterozygosity estimates by locus

Locus	**H_0**	**H_e**	**H_{nb}**
CMS9	0.659	0.703	0.707
CMS13	0.715	0.771	0.776
CMS15	0.506	0.75	0.755
CMS18	0.375	0.362	0.364
CMS25	0.511	0.626	0.629
CMS32	0.535	0.623	0.627
CMS50	0.631	0.837	0.842
CMS121	0.755	0.751	0.754
CVRL1D	0.638	0.803	0.808
CVRL4D	0.586	0.616	0.619
CVRL5D	0.77	0.7	0.704
CVRL6D	0.287	0.379	0.382
CVRL7	0.344	0.803	0.808
CVRL8	0.282	0.409	0.411
LCA66	0.6	0.73	0.734
VOLP10	0.447	0.727	0.731
VOLP32	0.164	0.207	0.208
YWLL44	0.47	0.589	0.592
YWLL59	0.28	0.54	0.54
Mean	0.503	0.628	0.631

Compared to the expected heterozygosity, observed heterozygosity was in deficit in all loci with the exception of three; CVRL5D, CMS18 and CMS121. Figure 15 illustrates the deficit of observed heterozygosity rates in each locus plotted to its expected heterozygosity.
At two loci; CVRL7 and VOLP10 the deficit was severe. A large number of alleles was observed at each of these loci (10 alleles), but the heterozygotes proportions were of only 30 - 40 %, giving the highest departure from expected heterozygosity. The three markers VOLP32, CMS18, and CVRL6D presented the lowest heterozygosity (0.1 to 0.3). The small number of alleles (2 or 3) and the abundance of the major allele (allele frequency of 0.75) might explain this observation.

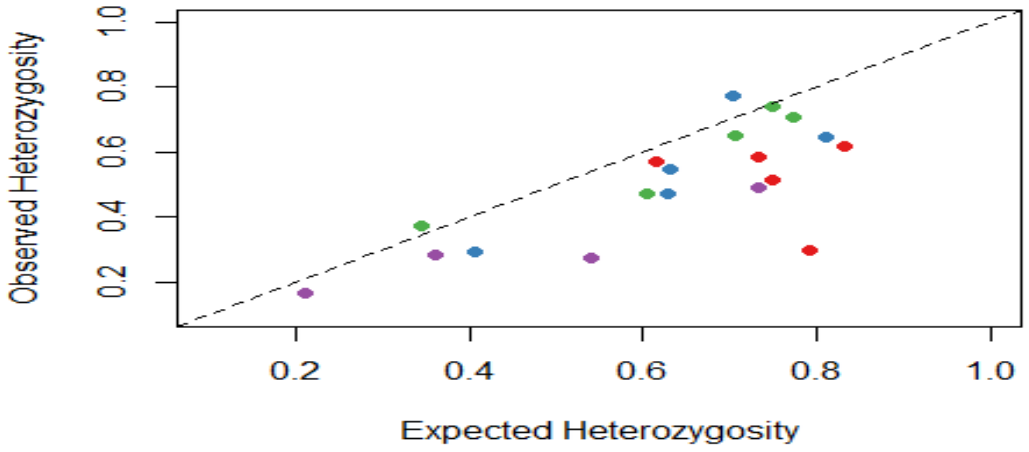


Figure 13. Observed heterozygosity plotted to the expected heterozygosity across the loci

Overall, heterozygotes in the population were present at 50 %, exhibiting a slight shortage compared to the expected 62 % of heterozygotes. In general, deficit in heterozygosity might ensue either from population structuring or inbreeding. As a natural consequence of this deficit, the population showed a significant (p = 0.05) departure from Hardy Weinberg equilibrium, over all loci.

IV.2.3.1.3. F statistics

All per locus F_{ST} values were very low (table 20). The marker YWLL44 was the most informative one and showed a moderate genetic differentiation among the population (F_{ST} = 0.04). Across all loci, the mean F_{ST} was of 0.017 with a standard deviation of 0.013, which indicated a lack of significant genetic structuring and a faint genetic divergence in the population.

On the other hand, mean F_{IS} (0.2) showed a moderate level of inbreeding. This value was within the confidence interval at 95 % after 10000 bootstraps (table 20). F_{IS} was highly variable between the different loci. Except at three loci, all F_{IS} values were positive, indicating more related individuals than expected under a random mating hypothesis. The exact test of statistical significance showed F_{IS} (Weir and Cockerham, 1987) to be significant at a threshold p = 0.05 at all loci, except for VOLP32, and it was highly significant at two loci (VOLP10 and CVRL7).

Finally, the level of variability of individuals compared to the total population was estimated using the average $F_{IT.}$ This index reached 0.209, and the 95 % Confidence Interval was 0.119 to 0.286. Thus, the variability between individuals was higher than that between sub-populations.

Table 20. F statistics estimates by locus in the whole population

Locus	**F_{IS} (W&C)**	**F_{IS} (R&H)**	**P**	**F_{ST}**
CMS9	0.069	0.049	0.04	0.017
CMS13	0.078	0.042	0.01	0.014
CMS15	0.331	0.208	0.01	0.004
CMS18	-0.029	-0.015	0.01	0.004
CMS25	0.189	0.183	0.01	0.026
CMS32	0.146	0.144	0.01	0.008
CMS50	0.252	0.153	0.01	0.019
CMS121	-0.001	-0.009	0.01	0.008
CVRL1D	0.207	0.109	0.01	0.015
CVRL4D	0.054	0.016	0.02	0.005
CVRL5D	-0.094	-0.022	0.01	0.005
CVRL6D	0.249	0.253	0.02	0.034
CVRL7	0.141	0.385	0.00	0.036
CVRL8	0.316	0.331	0.01	0.025
LCA66	0.184	0.065	0.01	0.034
VOLP10	0.39	0.431	0.00	0.006
VOLP32	0.212	0.214	0.09	0.036
YWLL44	0.228	0.126	0.02	0.048
YWLL59	0.487	0.322	0.03	0.003
Mean	0.204	-	0.02	0.017
95 % CI	0.118 – 0.288	-		-0.009 – 0.007

W&C: F_{IS} based on Weir & Cockerham (1984). R&H: F_{IS} based on Robertson & Hill (1984). P: exact test of significance. CI: Confidence Interval. R&H estimation gives more weigh to rare alleles than W&C estimation.

IV.2.3.1.4. Inbreeding

In pairwise comparisons by TrioML (Wang, 2007), 0.12 % of the comparisons showed relatedness (r) equal or greater than 0.5 and 1.7 % were equal or higher than 0.25 which is the genetic similarity between half siblings. Figure 16 shows the distribution of pairwise relatedness calculated by Dyadic maximum likelihood (Wang, 2007) and Triodic maximum likelihood (TrioML). Most comparisons showed relatedness under 0.2. Over all samples, the average inbreeding (Ritland, 1996) was 0.16 (SD = 0.046).

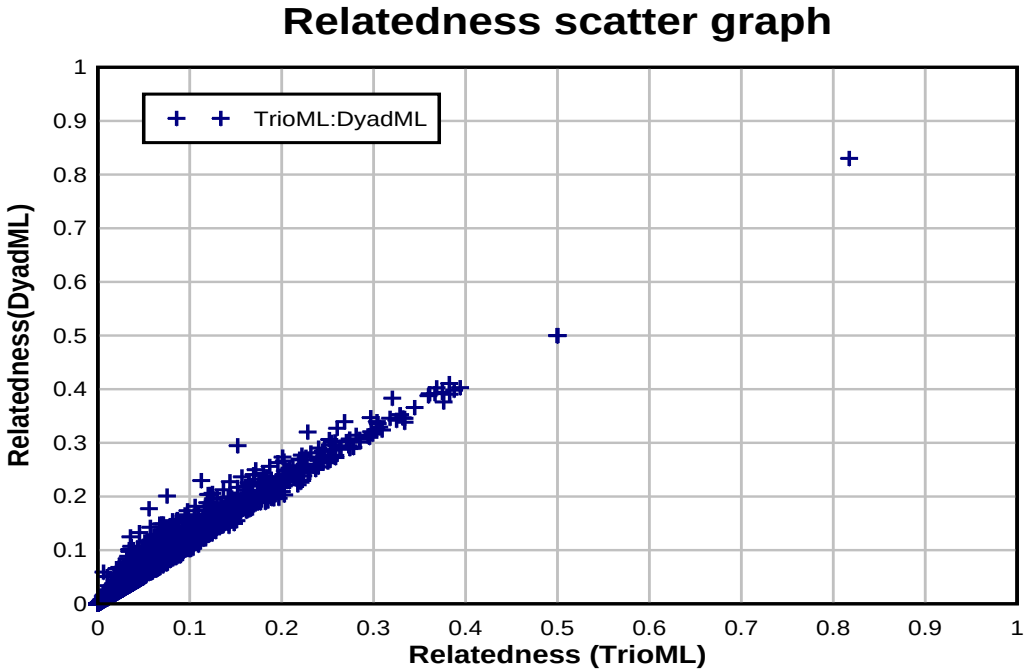


Figure 14. Pairwise relatedness of all samples by Dyadic and Triodic Maximum Likelihood (DyadML and TrioML)

IV.2.3.2. Population structure

IV.2.3.2.1. Descriptive parameters by location

The statistical parameters considered for the whole population were also estimated for each sampled region (Gabes was eliminated for low number of samples: 6), to evaluate the differences between them. Regarding the allele count, some alleles were absent in some regions (table 21). The average frequency of private alleles was similar in all regions and reached 2 %. Mean allelic richness was 5.4 with low difference between the geographic areas (SD = 0.04).

The exact genic and genotypic test proved that two loci were differentiated with high significance (p = 0.01) between animals from the three governorates; CMS32 (p = 0.0031, Standard Error = 0.0008) and CVRL7 (p = 0.010, Standard Error = 0.0017).

Table 21. Number of alleles (N_a), observed (H_0) and expected (He) heterozygosity in the different geographic regions

Locus	**Kebili**			**Tataouine**			**Medenine**		
	N_a	**H_0**	**H_e**	**N_a**	**H_0**	**H_e**	**N_a**	**H_0**	**H_e**
CMS9	5	0.48	0.69	5	0.72	0.70	6	0.51	0.70
CMS13	6	0.72	0.74	5	0.72	0.76	7	0.7	0.66
CMS15	7	0.37	0.75	7	0.5	0.75	7	0.71	0.62
CMS18	3	0.44	0.37	3	0.22	0.28	3	0.48	0.40
CMS25	4	0.55	0.60	4	0.27	0.55	4	0.42	0.62
CMS32	4	0.46	0.74	2	0.52	0.76	4	0.59	0.66
CMS50	8	0.65	0.82	7	0.75	0.84	7	0.53	0.82
CVRL1D	12	0.64	0.82	9	0.6	0.84	11	0.59	0.77
CVRL4D	3	0.48	0.58	3	0.54	0.60	4	0.71	0.69
CVRL5D	5	0.89	0.67	7	0.72	0.72	7	0.73	0.67
CVRL6D	3	0.42	0.38	3	0.22	0.28	3	0.27	0.33
CVRL7	7	0.27	0.76	6	0.45	0.74	8	0.32	0.80
CVRL8	3	0.35	0.35	4	0.23	0.36	4	0.23	0.40
CMS121	9	0.82	0.79	7	0.72	0.73	8	0.43	0.75
LCA66	8	0.71	0.73	5	0.52	0.76	5	0.42	0.53
VOLP10	7	0.41	0.77	6	0.68	0.72	6	0.44	0.79
VOLP32	2	0.17	0.16	2	0.3	0.32	2	0.05	0.12
YWLL44	8	0.51	0.59	6	0.42	0.71	8	0.52	0.65
YWLL59	3	0.22	0.53	3	0.27	0.56	5	0.27	0.45
Mean		**0.5**	**0.62**		**0.49**	**0.63**		**0.47**	**0.6**
SD		**0.2**	**0.19**		**0.19**	**0.19**		**0.19**	**0.19**

SD: Standard Deviation

Concerning the observed heterozygosity, values were strongly homogeneous throughout the locations and homozygotes excess was detected in all regions. For some loci (CMS9, CVRL1D, CVRL4D, CMS121 and LCA66), differential rates of heterozygosity were observed between the geographic areas.

Mean F_{IS} in Medenine and Tataouine were similar to mean F_{IS} in the whole population (0.207). Kebili showed F_{IS} slightly inferior, suggesting less effect of inbreeding in this area. F_{IS} values were highly significant in all regions (table 22). F_{IS} by locus showed differential values between the sampled areas. For eight loci, individuals were inbred in some regions, exhibiting moderate positive rate, but negative in others.

Table 22. F_{IS} values by locus and the exact fisher test (P) by location

Locus	**Kebili**		**Tataouine**		**Medenine**	
	F_{IS}	**P**	**F_{IS}**	**P**	**F_{IS}**	**P**
CMS9	0.29	N.S	-0.03	N.S	-0.04	N.S
CMS13	0.02	N.S	0.03	N.S	0.13	N.S
CMS15	0.49	**	0.32	*	0.12	N.S
CMS18	-0.19	N.S	0.19	N.S	-0.2	N.S
CMS25	0.08	*	0.5	*	0.3	N.S
CMS32	0.29	**	0.13	N.S	-0.03	N.S
CMS50	0.2	*	0.1	N.S	0.31	**
CVRL1D	0.21	**	0.28	**	0.17	N.S
CVRL4D	0.16	N.S	0.08	N.S	-0.1	N.S
CVRL5D	-0.34	N.S	-0.01	N.S	-0.21	N.S
CVRL6D	-0.11	N.S	0.19	N.S	0.33	N.S
CVRL7	0.63	**	0.38	**	0.73	**
CVRL8	-0.02	N.S	0.33	N.S	0.55	**
CMS121	-0.04	N.S	-0.007	N.S	0.08	N.S
LCA66	0.23	*	0.3	N.S	0.19	**
VOLP10	0.46	**	0.045	N.S	0.41	*
VOLP32	-0.08	N.S	0.07	N.S	0.61	N.S
YWLL44	0.13	N.S	0.39	**	0.14	N.S
YWLL59	0.58	**	0.5	**	0.49	**
Over all	**0.16**	**	**0.2**	**	**0.19**	**

N.S: Non Significant. *: significant at p = 0.05. **: significant at p = 0.01.

The mean inbreeding F (Ritland, 1996) was differential between the governorates. Individuals in Tataouine exhibited an average F = 0.208, while less inbreeding was observed in Kebili (0.164) and in Medenine (0.121). The number of migrants after correction for size was 7.264 (Barton et Slatkin, 1986).

Ewens-Watterson test for Neutrality (1985) was performed and details of the loci where proportions of homozygotes were lower than those expected in neutral conditions are presented in table 23. A deficit in homozygosity was observed in Medenine region at CMS50, and in Tataouine at two loci; CMS50 and LCA66. These loci give evidence for effects of either balancing selections or recent bottleneck. The YWLL44 locus test for neutrality was higher than the upper limit of neutrality (U95) in Medenine which might result either from the pressure of a directional selection or a population growth. All loci were neutral in Kebili region.

Table 23. Ewens-Watterson test for Neutrality (Only non neutral loci are shown)

		Observed F	**L*95**	**U*95**
CMS50	Medenine	0.183	**0.192**	0.589
	Tataouine	0.176	**0.187**	0.560
YWLL44	Medenine	0.525	0.140	**0.425**
LCA66	Tataouine	0.250	**0.255**	0.761

*: Based on 10000 bootstraps

Nei's genetic distance between the different defined geographic areas was estimated (table 24). In discordance to the geographic proximity of Medenine and Tataouine, these two

regions displayed the highest genetic distance. However, all genetic distances were closely low. In accordance, genetic identity between the different regions was very high. The genetic identity exceeding 95 % meant that low difference of allele frequencies was observed between the defined areas. After permutations, only 40 % of the Nei distances were superior to the real data, which puts high confidence in the results.

Table 24. Nei's genetic identity (above diagonal) and genetic distance (below diagonal)

	Kebili	**Medenine**	**Tataouine**
Kebili	****	0.958	0.963
Medenine	0.042	****	0.948
Tataouine	0.038	0.052	****

Pairwise F_{ST} values between the regions showed the populations to be genetically close. These values were negative between Medenine and both Kebili and Tataouine samples, indicating no genetic subdivision between the considered populations. Kebili and Tataouine exhibited a feeble pairwise F_{ST} of 0.009.

When two sub-populations were defined: South-Eastern area and South-Western area, Nei's distance (1978) dropped to 0.005. Other statistics on these regions are in the annex 5.

IV.2.3.2.2. Descriptive parameters by ecotype

In another approach, we tried to test genetic differentiation between the main ecotypes in the population: Ardhaoui, Merzougui and Targui. Pairwise F_{ST} values were very low between Ardhaoui and the two other ecotypes (table 25). The highest F_{ST} was observed between Merzougui and Targui but was still low (inferior to 0.05). Nei distances, on the other hand, were more important than those between geographic areas. Merzougui and Ardhaoui were the closest ecotypes. Targui ecotype showed moderate genetic distance from Ardhaoui and Merzougui that reached 0.08 and 0.132, respectively. Fisher test for genic differentiation overall loci proved the Merzougui samples to be relatively differentiated from Ardhaoui ($p = 0.004$) and Targui samples ($p = 0.002$), while Targui and Ardhaoui differentiation was not significant. In this test, p value over all loci and for all populations was highly significant (0.003). The genotypic test gave no significant differentiation for any pairwise comparison. F_{IS} values were moderate and similar for the three ecotypes: Ardhaoui 0.21, Merzougui 0.19, and Targui 0.27.

Table 25. Nei's distance (below diagonal) and pairwise F_{ST} (above diagonal) for three ecotypes

	Ardhaoui	**Merzougui**	**Targui**
Ardhaoui	****	0.001	0.010
Merzougui	0.037	****	0.025*
Targui	0.080	0.132	****

* Statistically significant at $p = 0.01$

IV.2.3.2.3. Data clustering

To explore the data and its clustering into genetic groups, different methods were aPlied; admixture using Structure and multivariate analysis (PCA, DCPA and FCA).

IV.2.3.2.3.1. Principal Component Analysis (PCA)

The Eigen values of the first two Principal Components (PC) were of 0.431 and 0.413, respectively. The three following PCs had Eigen values of 0.35. As shown in figure 15, the cumulative explanatory power of the PCs attained 75 % only when over 30 PCs are used.

Therefore, we could not define a low number of PCs to explain most of the observed variation in the data. As a consequence, plotting the data using PCs showed only feeble clustering (figure 15).

Plotting the data following the three first Principal Components (PC1, PC2 and PC3), with permutations, showed no clear grouping of the data (figure 16). A large proportion of the data was situated in intersections between the clusters. The main PCs were not able to explain the observed variability.

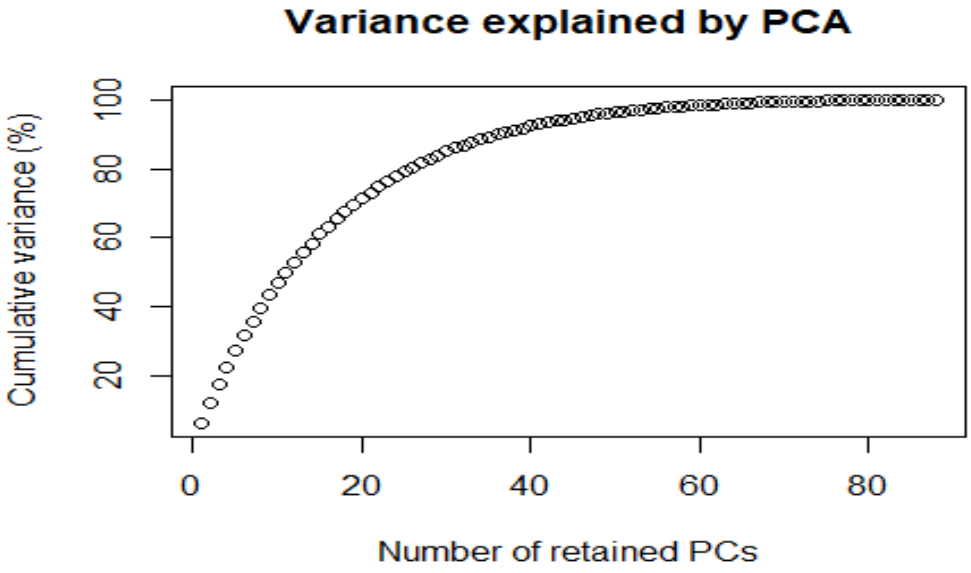
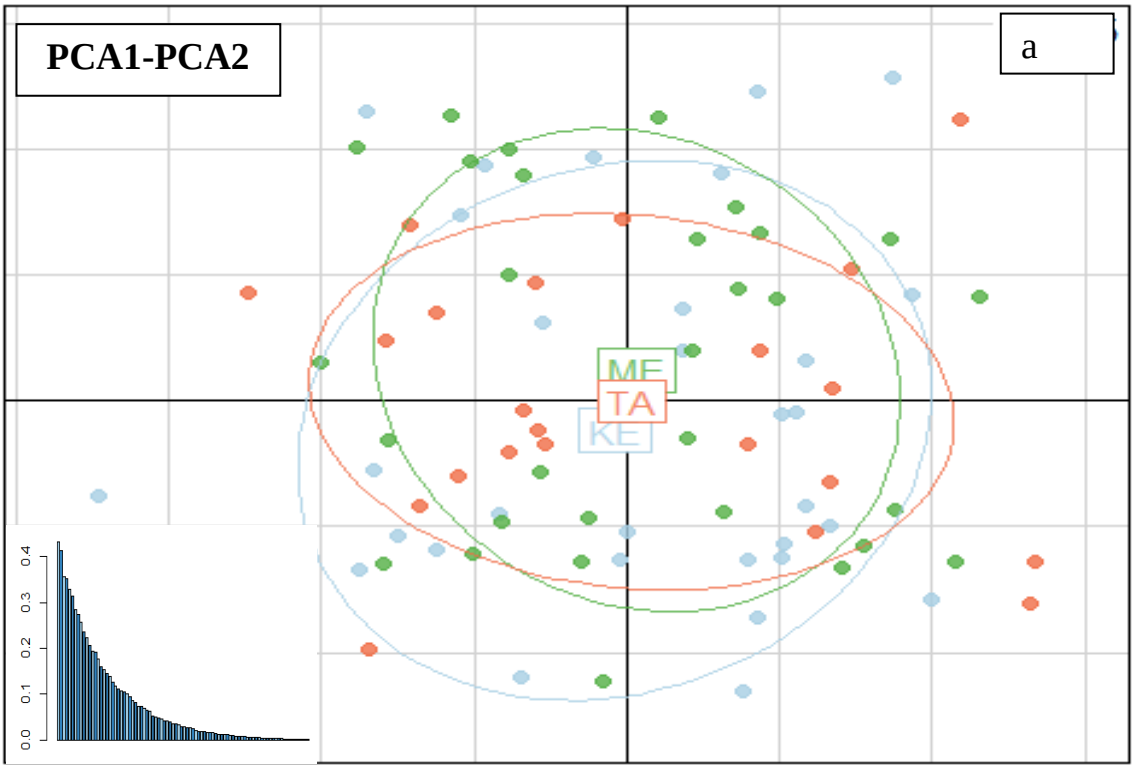


Figure 15. Cumulative explanatory power of PCs



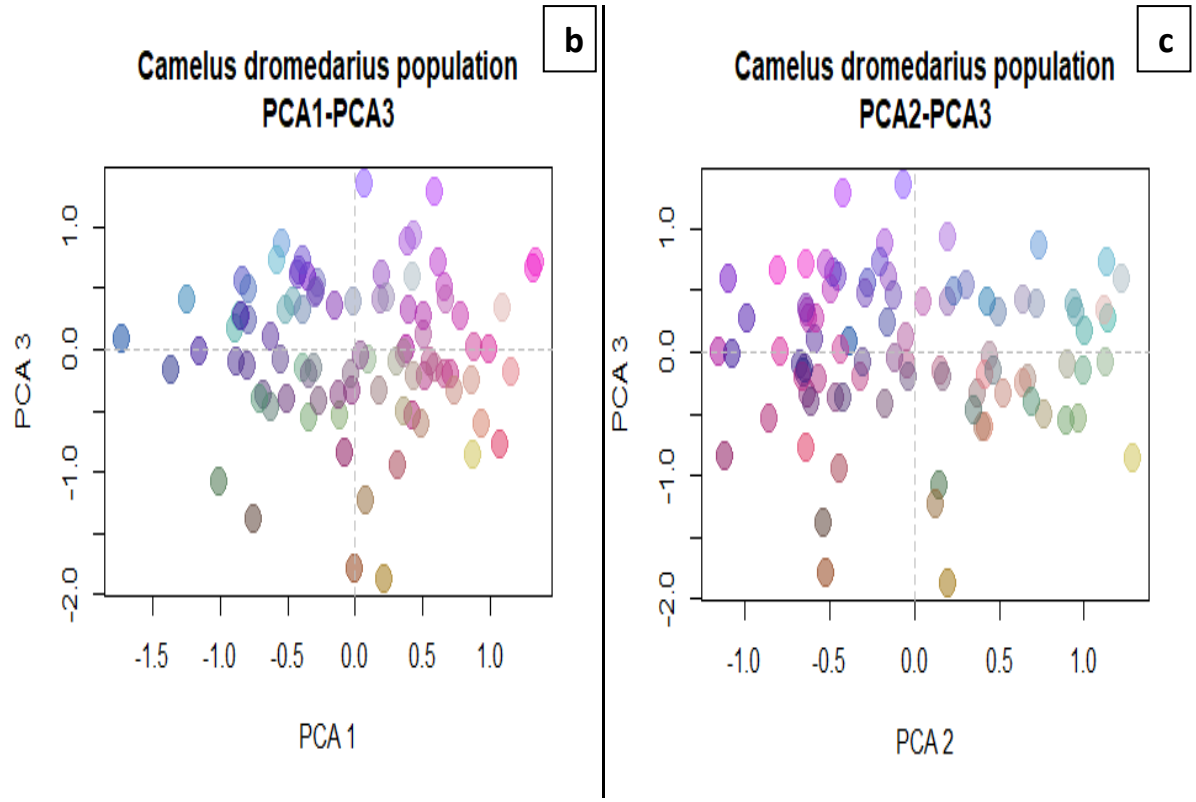


Figure 16. Principal component analysis projections. (a) Individuals distribution on PCA1 and PCA2, with Eigen values bar plot (lower left). Origin of samples; TA: Tataouine (in red), ME: Medenine (in green), and KE: Kebili (in blue). (b) Scatter plot diagram showing the distribution of individuals on PCA1 and PCA3 dimensions. (c) Scatter plot diagram showing the distribution of individuals on PCA2 and PCA3 dimensions.

IV.2.3.2.3.2. Discriminative Analysis of Principal Components (DAPC)

From DAPC analysis, 60 PCs were retained, which correspond to more than 99 % of explained cumulative variance. The plot of Bayesian Information Criteria (BIC) to the number of clusters (figure 17) permitted to determine the best number of clusters for the data (3). Two discriminate functions were retained for the clustering.

The resulting scatter plot of the data (figure 18), following the previously detailed parameters, showed three distinctive clusters, well defined and evidently distant. When only 10 PCs are used, the clustering is still clear and the groups are distinct but the distance between the groups diminished. The defined groups, under Discriminative Analysis, did not reflect the location origin of the samples (figure 19). Similarly, we observed that samples from different ecotypes were represented in each group. Hence, no complete discrimination between the ecotypes was observed.

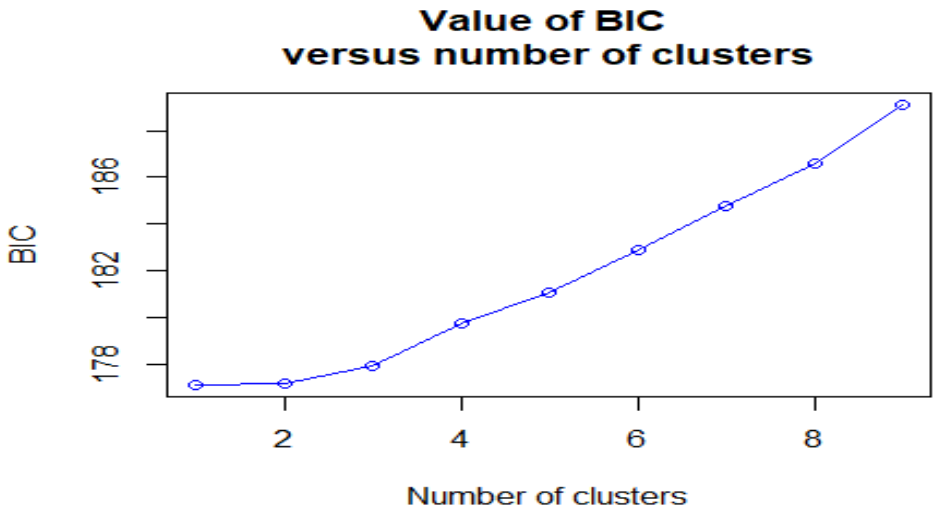


Figure 17. Bayesian information criterion and corresponding number of clusters

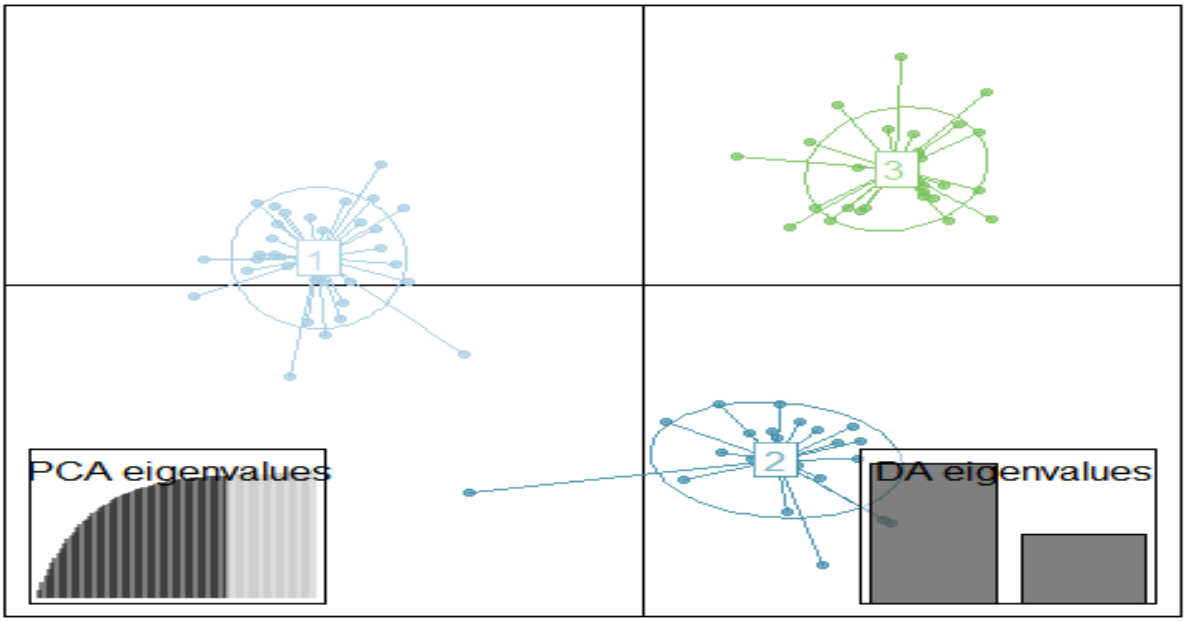


Figure 18. Scatter plot of the samples into 3 clusters using Discriminative Analysis (60 PCs and two discriminative functions were used)

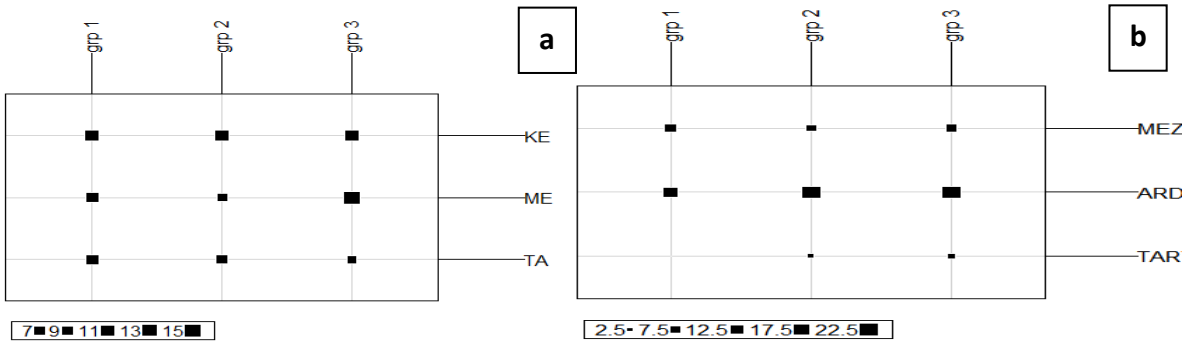


Figure 19. Distribution of samples according to their location (a) and to their ecotypes (b) into 3 groups (b) following DCPA clustering. KE: Kebili, ME: Medenine, TA: Tataouine, MEZ: Merzougui, ARD: Ardhaoui, and TAR: Targui

IV.2.3.2.3.3. Analysis of Factorial Component (AFC)

In another type of multivariate analysis, the Factorial Component (FC) analysis gave similar results to PCA. The cumulative power of the first four FC can only explain 13% of the observed variability in the data (figure 20). On a three dimensional space of FCs, a weak structure in the data and the clustering was observed. A visualization of the samples divided into two regions: Kebili and Tataouine-Medenine and plotted on the main FCA is presented in figure 21.

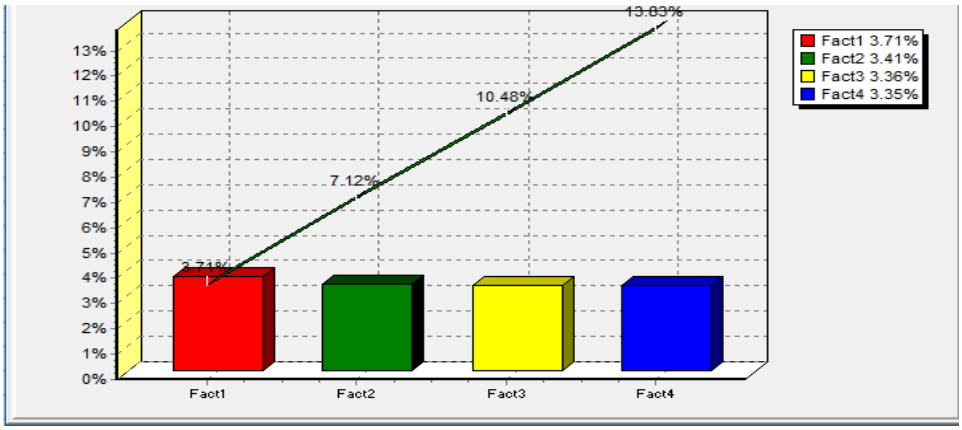


Figure 20. Explanatory power of the first four principal factors in FC Analysis

When arranged into three main ecotypes (Ardhaoui, Targui and Merzougui), the AFC gave a stronger clustering (figure 22). In fact, the ecotypes do not necessarily reproduce the geographical boundaries, and thus labeling the data by ecotype or location may give different results. The figure showed that Medenine and Tataouine samples, assembling Ardhaoui ecotype, were grouped (towards the right). Samples from Kebili and Gabes -that were identified as Merzougui ecotype- were grouped together (towards the left) but were not distant from the first group. A large proportion of the samples could not cluster distinctively. The Targui ecotype, presented in white, clustered together in a distinct group (upper right). Nevertheless, the limited number of samples (8) did not permit any conclusive assertion about this ecotype.

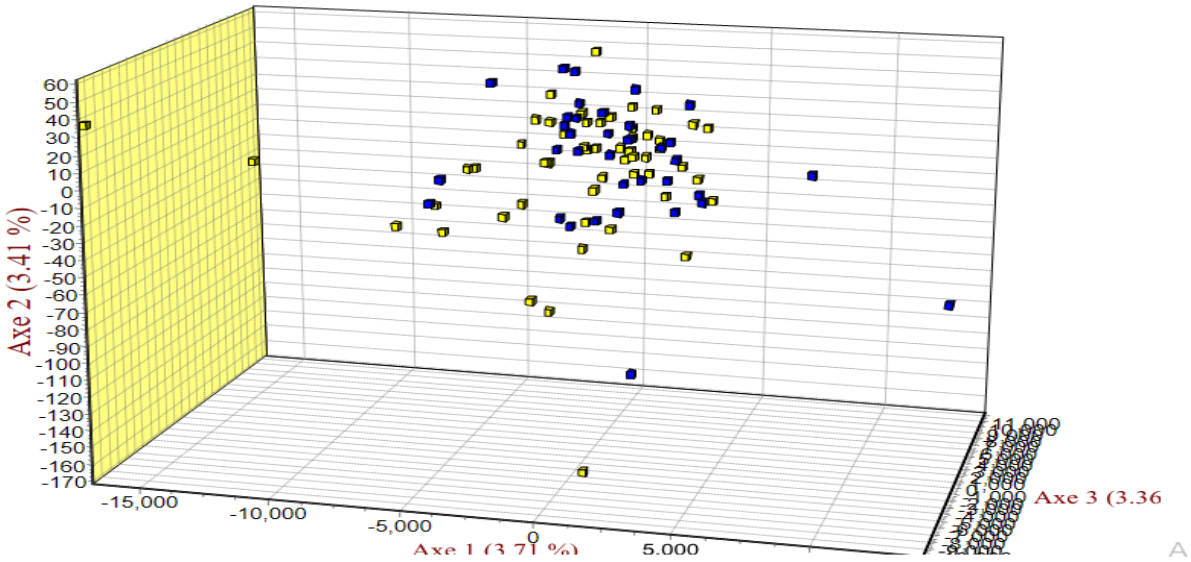


Figure 21. Three dimensional visualization of the samples on the factorial space issued from FC Analysis (Medenine and Tataouine in yellow, Kebili in blue)

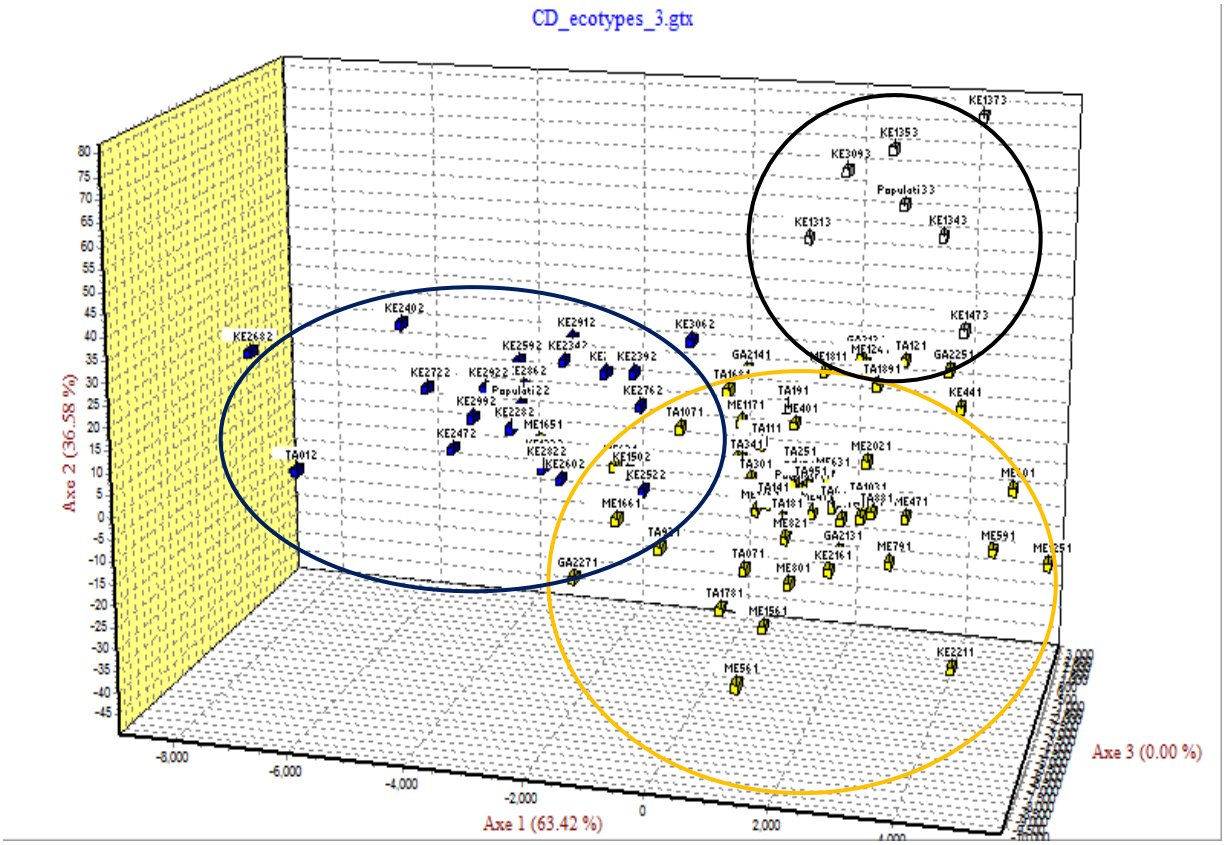


Figure 22. Scatterplot of the individuals on three main FC axes, according to their ecotypes (Merzougui in blue, Ardhaoui in yellow, and Targui in white).

IV.2.3.2.3.4. Bayesian clustering using STRUCTURE

Clustering using STRUCTURE software with and without the use of a prior (sample location) gave weak structure of the data (figure 23).

For K = 2, the presentation of the population structure demonstrated a nearly total admixture in the genetic material of the samples. When three ancestors are assumed, all individuals illustrated mixed genetic material from at least two ancestors, with differential contribution. Clusters of individuals with more genetic resemblance to one of the ancestors were visible, but this grouping was not consistent with the sample locations.

For K = 4, the wide admixture was clearly visible. Clusters could be discerned, but these clusters were not completely separate and did not form a unique genetic entity. The first group, displayed in blue and red, consisted in Merzougui ecotype. This group inherited its genetic material mostly from two ancestors (blue and red), with less contribution from other ancestors. The second, less obvious cluster, was the Targui ecotype, which exhibited a predominant genetic ancestor (in green), with fewer contributions from blue, red and even less from yellow ancestors. All other samples, from Ardhaoui Medenine, Ardhaoui Tataouine and also some of Merzougui ecotypes, displayed highly mixed individuals. This clustering pattern might originate from different initial ancestors (probably two or three) that were crossbred, resulting in generations of mixed individuals.

For K = 5, the same hybrid clusters are perceptible. An ancestor (purple) appears in all individuals, with the exception of the Merzougui group. The whole population displayed again large admixture.

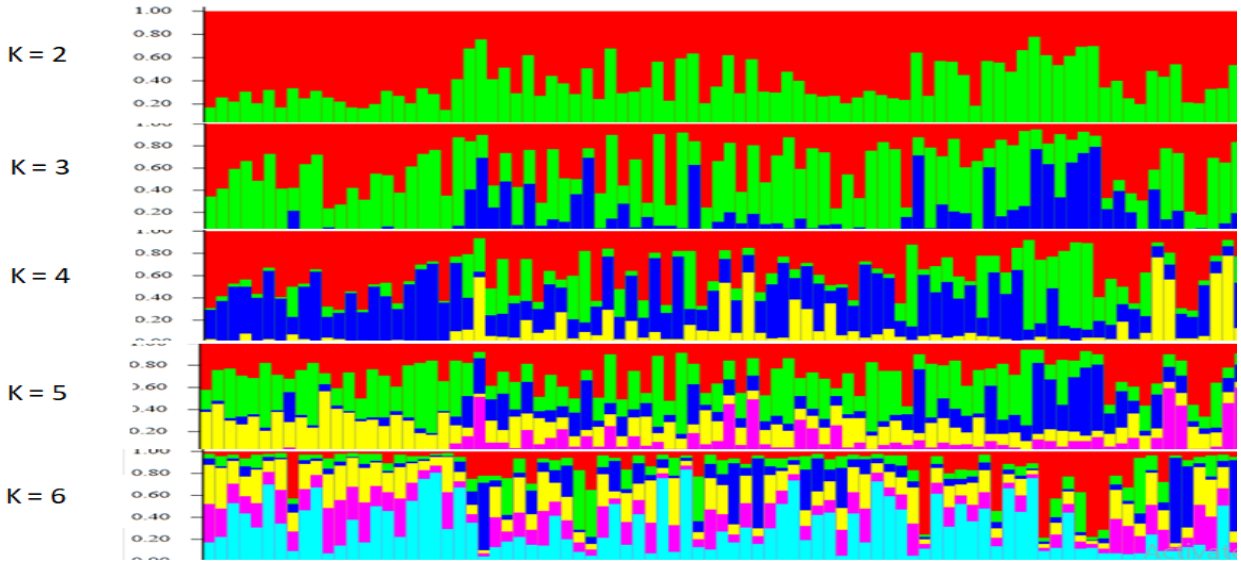


Figure 23. Bayesian analysis of the population structure and individual assignment for K = 2 to K = 5

IV.2.3.3. Bottleneck test

All four simulations of population expansion and decline resulted in comparable time, current N_e and ancient N_e values, proving a decline of the population from N_e = 20,000 (5-95 % quantiles: 5,000-75,000) to N_e = 700 (5-95 % quantiles: 200-27,000). The data also showed that the bottleneck most likely occurred around 6000 years ago, with 5-95 % quantiles 1,400-27,700 years (table 26). Plotting the MSVAR simulation results (figure 24) showed that the marginal posterior distributions of the demographic parameters were narrow and the estimated 95 % HPD (Highest Posterior Density) limits ranged from 1 to 4 in $\log_{10}$ scale. All posterior distributions (colored lines) were similar to each other but very different from the prior distributions (dashed, faint lines). Posterior distribution for ancestral N_e and current N_e showed very limited overlap, confirming the population collapse. In annex 6, we present all results from the performed runs; one with a generation time of 4 years, and one with 5 years, under the same four scenarios of population declination and growth. All the tests converged to the results presented in table 26 and figure 24.

Table 26. Time estimation for genetic bottleneck from MSVAR tests under two population decline and two population expansion scenarios

	Decline 10^3-10^5	**Decline 10^2-10^4**	**Expansion 10^5-10^3**	**Expansion 10^4-10^2**
N_0	704	738	767	756
0.025 HPD	187	202	211	194
0.975 HPD	2608	2749	2738	2890
N_1	19866	21765	19247	21028
0.025 HPD	5265	5612	4828	5637
0.975 HPD	75263	81907	71909	78158
t	6024	6152	7131	6381

0.025 HPD	1352	1359	1734	1422
0.975 HPD	26411	26361	30508	27758

N_0 = $\log_{10}$ current effective population size (N_e); N_1= $\log_{10}$ ancestral N_e; t = $\log_{10}$ time in years since demographic event started; HPD = 95 % highest probability density interval. Generation period was set to 4 years. The Gelman-Rubin's diagnostic test indicated convergence after $2.5*10^9$ MCMC iterations for the parameters N_0, N_1 and t.

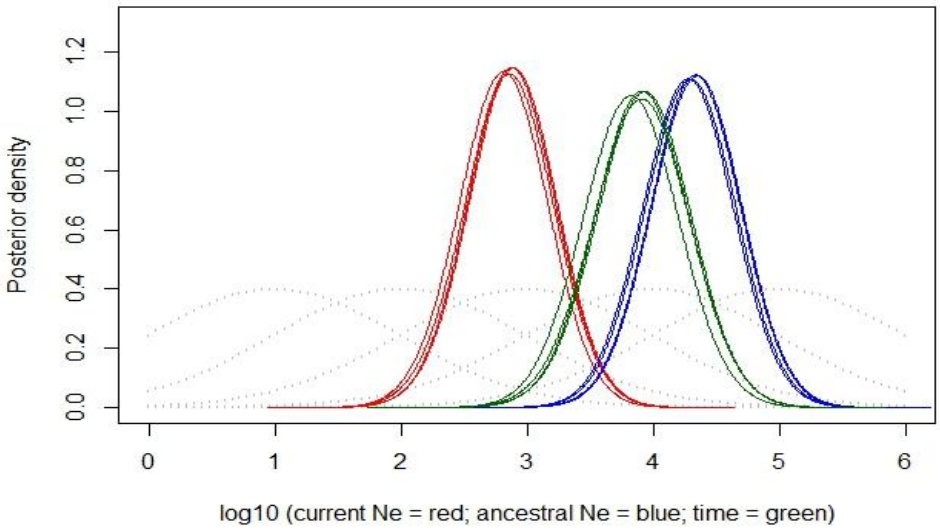


Figure 24. Estimated posterior distributions of highest posterior density intervals for current and ancient effective population size (N_e) and time since population change, after four MSVAR1.3 runs. All densities are represented in a $\log_{10}$ scale

IV.2.3.4. Phylogeny

The Neighbor Joining (NJ) tree showed similar genetic distance between all individuals and the grouping reflected neither the geographic origin nor the ecotypes (figure 25). The edges lengths varied between 0.04 (most branches) and 0.32 (figure 26).

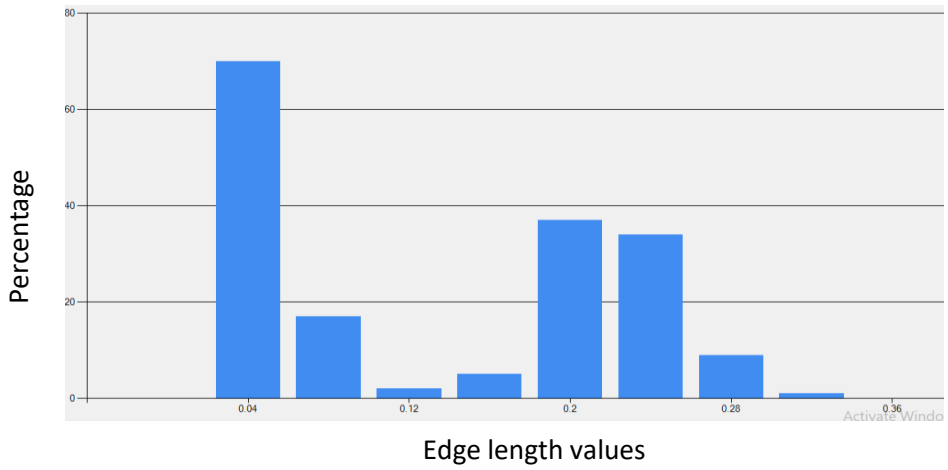


Figure 25. Distribution of branch lengths in the Neighbor Joining tree of all sampled individuals

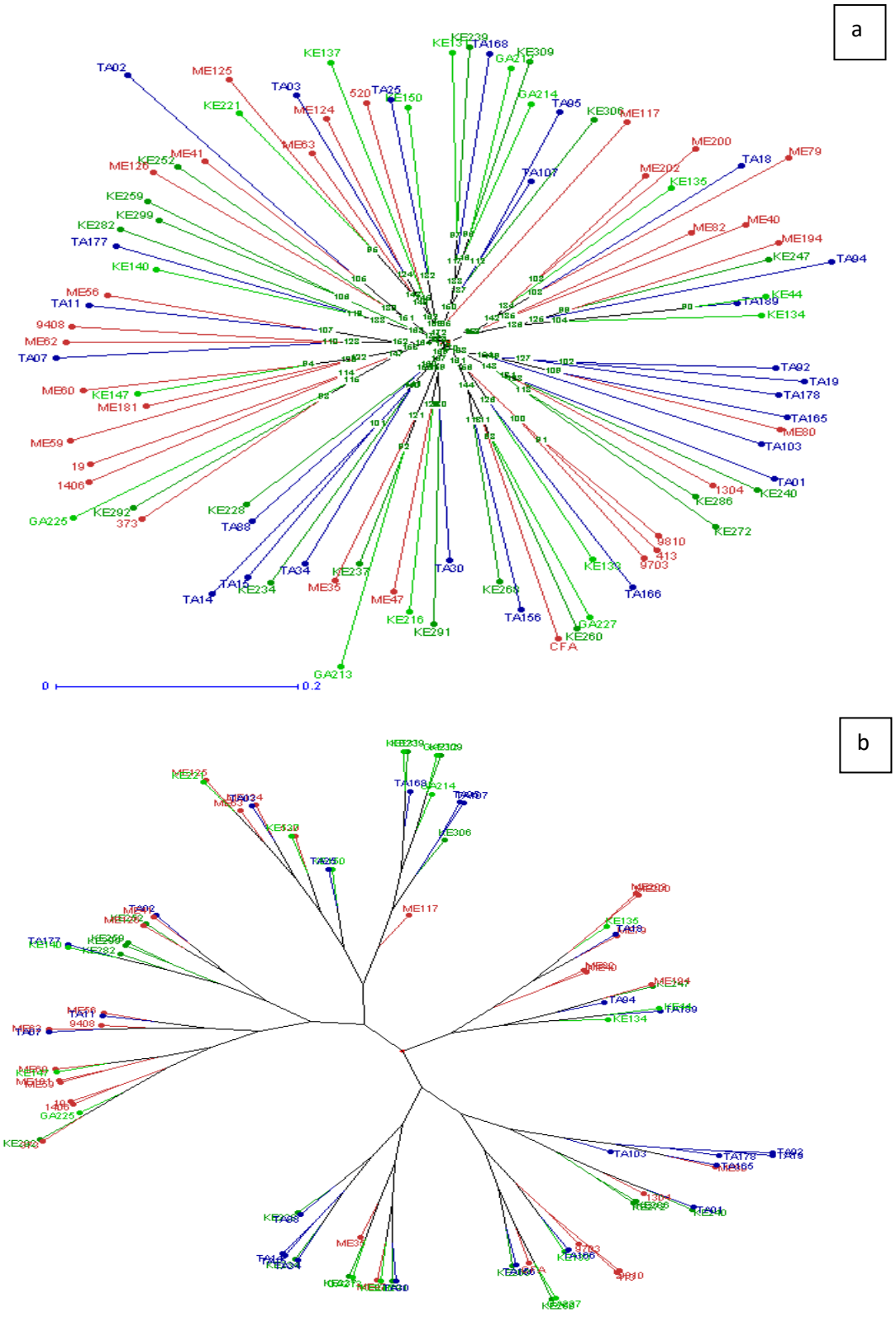


Figure 26. Phylogenetic tree (a) of all sampled dromedary camels in the Tunisian population (Neighbor joining method) and (b) branches grouping. Kebili samples displayed in green, Tataouine in blue and Medenine in red.

IV.2.4. Discussion

Despite of the recent and important decrease in the population head count, the dromedary camel population in Tunisia exhibited an important genetic variability confirmed by its high average of allele count (Na = 7.3), an allelic richness of 5, and an average PIC of 0.62.

The population showed a significant (p = 0.05, Wilcoxon test) deficit in heterozygotes that was confirmed by a highly significant (p = 0.01, Fisher test) departure from HW equilibrium. A comparable deficit was previously reported in a study by Ould Ahmed *et al.* (2010), where the observed heterozygosity in the Tunisian population did not exceed 0.46, while the expected heterozygosity was equal to that estimated in the current study (0.6). Likewise, populations of dromedary camel in different African and Asian countries exhibited similar homozygosity excess (Banerjee *et al.*, 2012, Vijh *et al*,. 2007). In the current study, this deficit is most probably related to non-random mating reflected in a moderate levels of both inbreeding (F_{IS} = 0.204) and individual inbreeding (F = 0.16) rates, rather than the existence of population structure, which proved to be feeble (F_{ST} = 0.018). This high ratio of heterozygotes can affect the high variability observed in the population. Thus, there is a necessity for developing a management plan of the genetic resources in order to (i) prevent anarchic mating and (ii) conserve the available genetic variability.

Inconsistently, in the preliminary study we found higher differentiation level of the population and F_{ST} reached 0.05. Ould Ahmed *et al.* (2010) also reported higher F_{ST} that reached 0.09. On the other hand, in a more recent study, Cherifi *et al.* (2017) found pairwise F_{ST} to range from 0.001 to 0.03 between geographical regions as distant as Egypt and Algeria and showed feeble structure among these populations.

A weak population structure was detected in the analysis of genetic admixture. The population showed a large common genetic ground shared by all individuals. In Bayesian clustering, the best fitting number of clusters (K) was equal to 4. However, the grouping was not evident and the clusters did not conform to the geographic origin of the samples. When the information about ecotypes was used, Merzougui ecotype showed a distinguishable group displaying an admixed genetic material from mostly two ancestors, while other individuals displayed differential contribution of all assumed 4 ancestors.

The multivariate methods (PCA, FCA and DAPC) yielded faint clustering of the individuals. A large proportion of the population was part of more than a cluster, showing large genetic characters shared between the different groups. These clusters were not completely in accordance neither with the geographic origin of the samples nor with their ecotypes. Furthermore, Nei distances were feeble when considering geographic areas as well as when considering the ecotypes. In fact, determining the ecotype of an individual is a highly subjective matter; while some morphologic description of the ecotypes is available (Chniter *et al.*, 2013), most sampled camels had intermediate phenotypes and their assignment to an ecotype or another was not obvious.

In general, poor structure was observed among the population in this study. Other previous works laid to similar conclusions. On a larger scale, Almathen and his collaborators (2016) reported little population structure in modern dromedaries, after examining samples from 21 countries, including Tunisia.

However, in a number of analyses, a faint differentiation of Merzougui was detected in some genetic parameters. When examining the ecotypes, significant (p = 0.003) genic differentiation was observed, overall loci. Specifically, Merzougui samples were relatively differentiated from Ardhaoui (p = 0.004) and Targui samples (p = 0.002), while Targui and Ardhaoui differentiation was not significant. The genotypic test gave no significant

differentiation for all pairwise comparisons. The difference between the groups was then more related to private alleles and allele frequencies than genotype frequencies.

The probable explanation is a high genetic flow between the considered areas and ecotypes, as shown by the estimation of the number of migrants. In fact, N_m between the regions and ecotypes was equivalent and was high (7.3). Actually, in several grazing areas, herds from different southern regions of the country gather, mostly in the spring, to take advantage of the available food resources. During these gatherings, male reproducers may be exchanged in order to avoid inbreeding inside the herds. The main largest grazing areas are situated between Tataouine and Medenine governorates (Dhaher and Ouaera) where animals from these regions get mixed. In the late years and due to successive drought years, the vegetation went scarcer and it became common for some camel herds to be transported, in their transhumance, to farther areas such as the Sebkha of Chott Jerid in the middle of the country, and to Behayer and Chareb regions in the South Eastern of the country. Consequently, exchange between breeders from distant regions has become more facilitated. This exchange of animals and their mating occurs –in general- with no regard to their ecotype.

As a final point, the population showed a sharp decline of 95% of the population size, about 6000 years ago. This decline precedes the domestication of dromedary camel and was observed in different dromedary populations and occurred between 5.1 to 8.6 thousand years ago (Almathen *et al.*, 2016).

In conclusion, despite of the weak structure and ecotype differentiation, the population exhibited high variability on the individual level. Consequently, the genetic potential can be exploited by selection of highly productive individuals. On the other hand, the genetically outlying and exceptional individuals can be exploited to generate distinctive groups that may lead to separate and interesting ecotypes, after generations.

RESULTS

CHAPTER V: Population genetics based on Mitochondrial DNA

V. Population genetics based on Mitochondrial DNA

The present work is the first insight on a maternal lineage specific genetic variability of dromedary camels in Tunisia. The absence of recombination in mtDNA gives it several unique properties that allow tracing back the phylogenetic origin of the treated samples.

V. 1. Experimental results

Using the extracted DNA, mitochondrial DNA loci were amplified and visualized on 1.5 % agarose gel to confirm the amplification and the size of amplicons. Figure 27 shows amplification check of the mtDNA (*cytochorme b*) PCR products.

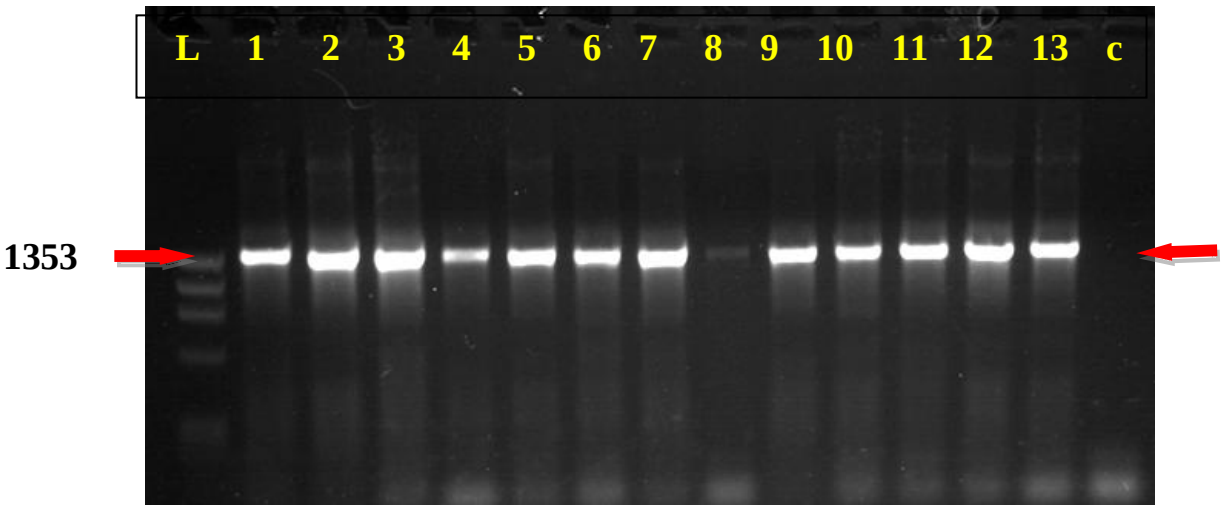


Figure 27. Amplification of the *Cytochorme b* gene in *Camelus dromedarius* mitochondrial DNA on agarose gel (1.5 %). L: Size ladder Φ X 174/*Hinf* I. 1-13: PCR products. c: Negative control.

Sequences from mitochondrial DNA amplification were of good quality (figure 29). For the *cytochorme b* gene, the length of the amplicons ranged between 398 and 932 bp. After eliminating 6 non-sequenced samples, more than half of the sequences had average quality score (Q) of 50, and all of them had average Q score ≥ 20. Figure 28 shows the distribution of Q scores throughout all sequences of *cytochorme b* gene. Likewise, amplicons from *D-loop* region were of good quality and ranged from 241 bp to 995 bp. Only 4 non exploitable sequences were removed and all remaining sequences had quality scores ranging from 20 to 55.

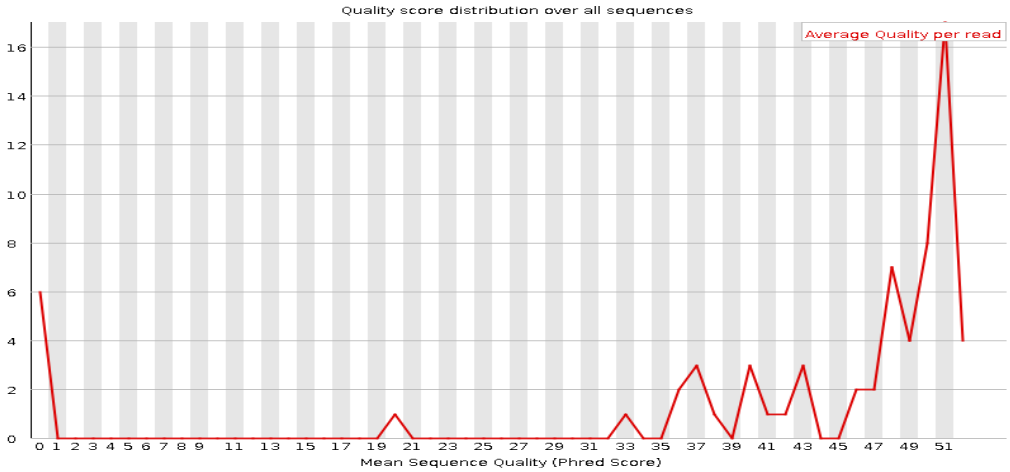


Figure 28. Per sequence quality scores of *cytochorme b* sequences

V.2. Sequence analysis

After trimming, merging forward and reverse sequences, and discarding sequences of less than 25 nucleotides, we obtained 119 sequences of good quality and length. Being paired end sequenced, 77 % of cytochorme b sequences were of 900 bp or longer. More than half of the *D-loop* sequences were of 680 bp or longer. Table 28 summarizes some information about the final sequences. Figure 29 shows the high quality scores of the trimmed and merged sequences.

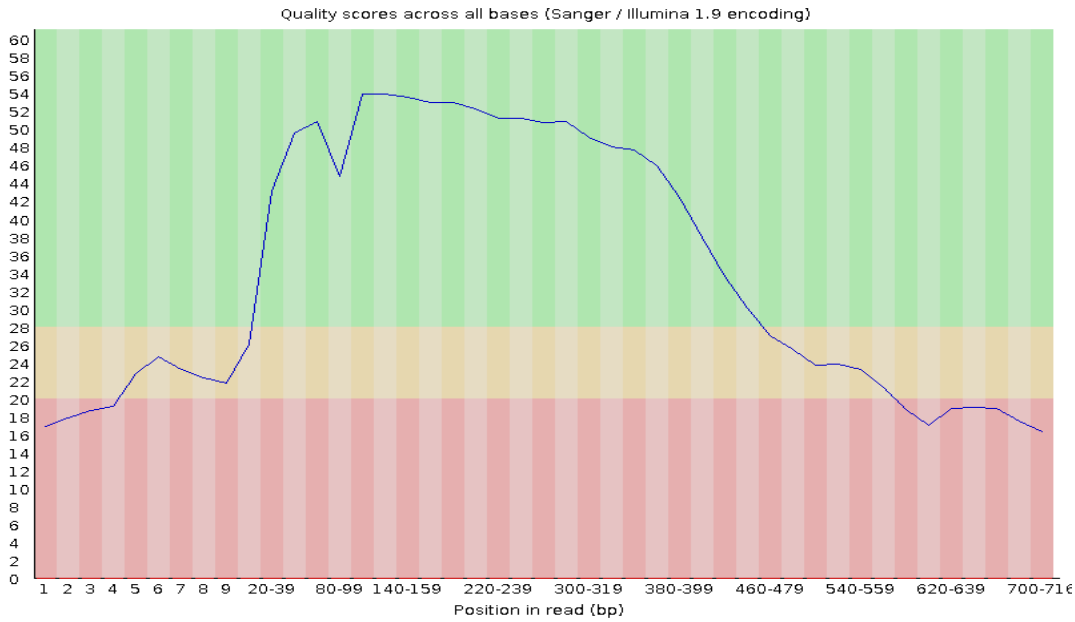


Figure 29. Quality score by base through all *D-loop* amplified sequences

Table 27. Sequence quality of trimmed *cytochorme b* and *D-loop* amplicons

	Cytochorme b	***D-loop***
Sequence length	364 - 946 bp	207 – 716 bp
Mean Quality score	45	51
N content	< 8 %	< 10 %
GC content	43 %	48 %

The amplified sequences were aligned to *Camelus dromedarius* mitochondrion reference genome (Genbank Accession number: NC_009849.1). On the reference, *cytochorme b* gene and *D-loop* region are of 1124 bp and 1140 bp, respectively. The Control region alignment was of 870 bp from 15,475 bp to 16,607 bp, while the *cytochorme b* alignment was of 561 bp from 14,635 bp to 15,195 bp. The use of an intermediate primer for the *D-loop* allowed overlaPing sequences and thus a larger common amplified region.

V.3. Variability by sequenced region

V.3.1. Overview

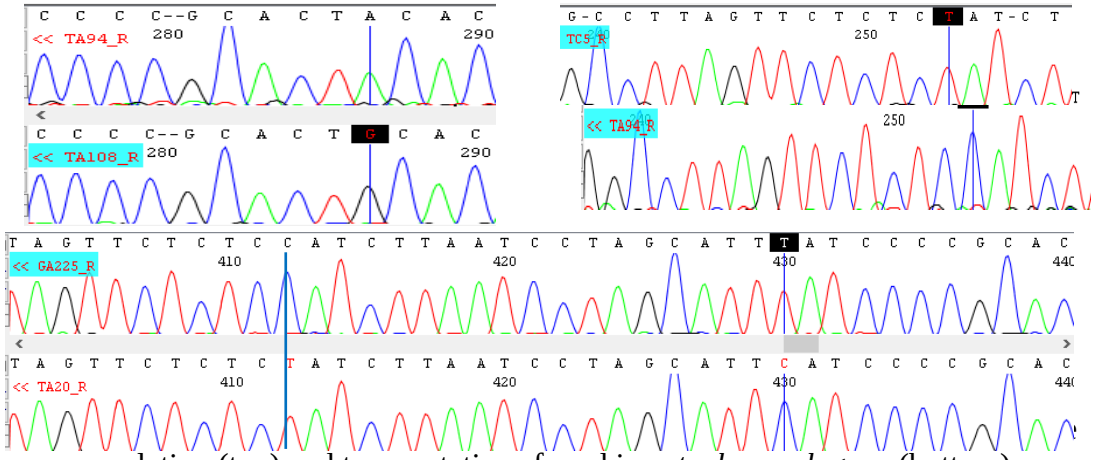
All detected polymorphisms were verified by eye to validate the sequencing quality. *Cytochorme b* sequences were more polymorphic than those of *D-loop*, with 12 segregating sites, resulting in 8 haplotypes (Table 28). In *D-loop* region, 4 polymorphic sites were observed, from which 2 were observed more than once, and yielded 5 haplotypes. All detected polymorphisms were point mutations (Annex 7).

Table 28. Nunber of sequences (N), number of polymorphic sites (S), number of singletons (s), number of parsimony informative sites (p), number of haplotypes (h) and invariable sites in the alignment to the reference (GenBank Accession Number: NC_009849.1)

	N	S	s	p	h	Invariable sites
Cytochorme b	64	12	5	7	8	549
D-loop region	62	4	2	2	5	870

Statistics on the haplotypes detected in the samples are summarized in table 29. A *cytochorme b* haplotype, distant by 6 mutations from the reference, was present in 3 samples. One of these samples had an additional SNP. All other haplotypes, in both sequenced regions, were distant by a sole mutation from the reference.

By comparing the haplotypes to the National Center for Biotechnology Information (NCBI) database, one haplotype of *D-loop* and two of *cytochorme b*, were previously detected in other dromedary camel populations. Three haplotypes from the *D-loop* region and five from *cytochorme b* gene were completely new. The following figure shows the chromatogram of some of the newly found mutations, compared to the abundant sequence.



population (top) and two mutations found in *cytochrome b* gene (bottom)

All mutations were present at less than 5 % of the samples. Most of these substitutions were transitions (87 %). Only one transversion was observed in each region of the mtDNA. Therefore, transition to transversion ratio for *cytochorme b* reached 12 and that of *D-loop* was of 4. Duchêne *et al.* (2011) reported these regions to have a high ratio of up to 19 and 15, respectively.

Table 29. Statistics on the haplotypes by analyzed region of the mtDNA

Region	Haplotype(s)	Mutations	Number of samples	Frequency
D-loop	A	0	56	0.903
	B* and C	1	2	0.032
	D* and E*	1	1	0.016
Cytochorme b	A	0	56	0.875
	B	7	2	0.031
	C*	6	1	0.015
	D*, E*, F*, G*, H	1	1	0.015

Haplotypes with asterisk (*) are new haplotypes.

Low nucleotide diversity (table 30) was observed for both sequenced regions, *D-loop* region more so than *cytochorme b*. Therefore, the sequences were –in average- very similar, as expected in the conserved mitochondrial DNA. The number of polymorphic sites is proportionally high (12 and 5).

Tajima's D quantifies the discordance between the estimate of theta from number of segregating sites and theta from average pair-wise sequence divergence. Fu's Fs quantifies the discordance in theta observed from the number of haplotypes and that from the average pairwise sequence divergence. Negative Fu's Fs values proved an excess of the number of alleles. Negative Tijama's D values were evidence for excess of low frequency polymorphisms than expected. This may result from a recent population expansion or from genetic hitchhiking. Since mtDNA, especially the coding *cytochorme b* gene, is generally highly conserved, the latter assumption is unlikely. Tajima's D values were negative and significant, confirming the possibility of a population growth. The controversy between results from Fu's Fs and Tajima's D resulted from the distant haplotypes (6 and 7 mutations of difference).

For *D-loop* region, the observed Fu's Fs and Tajima's D were negative but not significant. After 1000 simulations under neutral model demographic model, the probabilities of obtaining Fu's Fs and Tajima's D values equal or lower than the observed ones were significantly inferior to the observed ones ($p = 0.005$ and $p = 0.011$, respectively).

Under population growth model, the probability of simulation inferior or equal to observed was highly significant ($p = 0$) for both Tajima's and F (0.027) and the confidence interval at 99 % was -5.429 to 4.66. These results further outweigh the demographic growth assumption.

Table 30. Nucleotide diversity (π), haplotypes diversity (Hd) in the analyzed regions

	π	**Theta**	**K**	**Hd (SD)**	**Tajima's D**	**Fu's Fs**
Cytochorme b	0.00136	0.0045	0.762	0.236 (0.071)	-1.985*	-3.789
D-loop region	0.00026	0.0007	0.281	0.184 (0.065)	-1.453	-3.515

: $p < 0.1$; *: $p < 0.05$; SD: Standard Deviation

V.3.2. Phylogeny

Most haplotypes (4) in the *D-loop* region differed from the reference by a sole segregating site, and only one mutation was parsimonious. Thus, although two groups were discernible (figure 31), the genetic distances were very low. Three haplotypes were new and had the same distance to haplotypes from the Arabian Peninsula, Near Asia, Morocco and Africa. One haplotype had a single identical hit in blast, KU605080, which was reported in Pakistan.

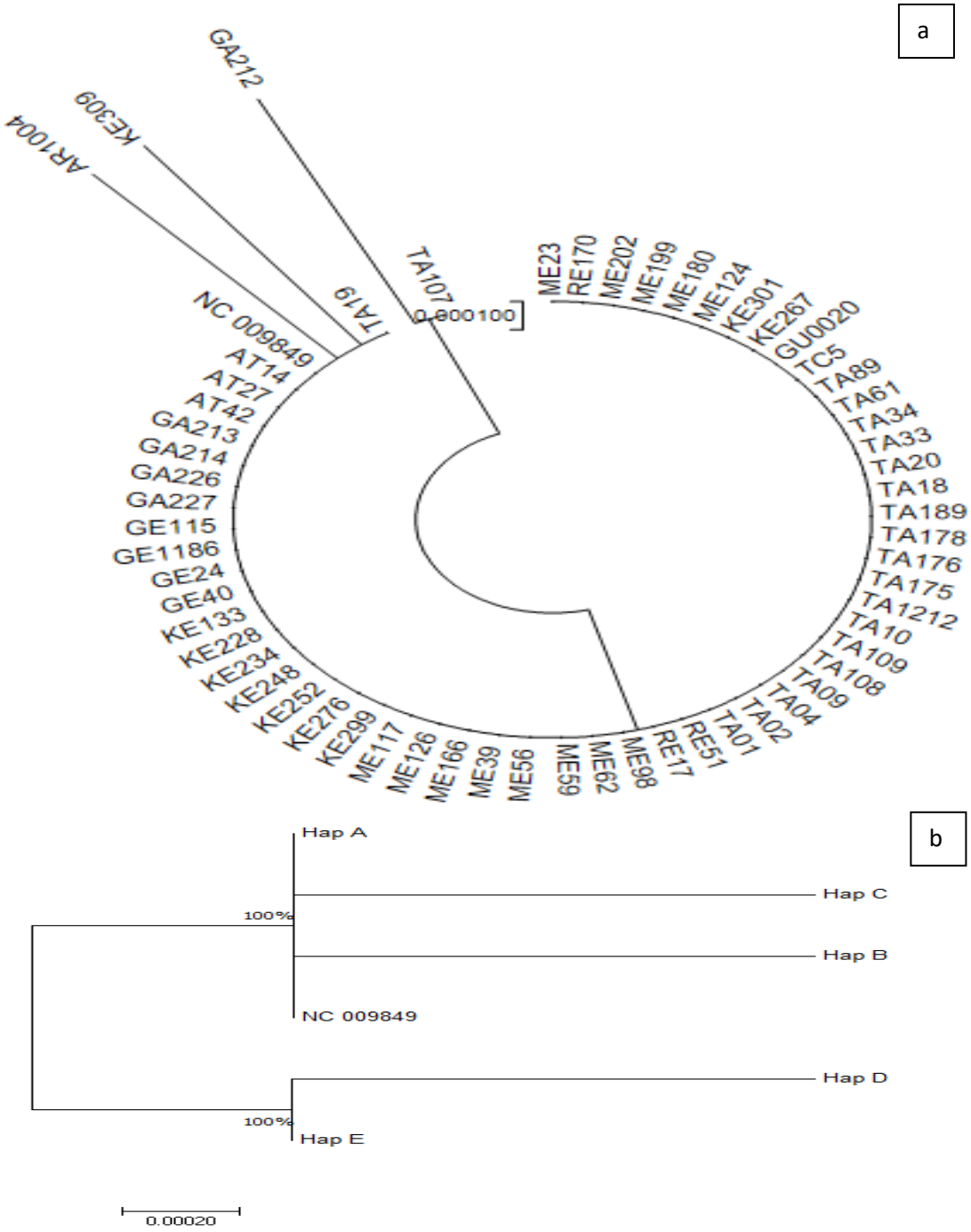


Figure 31. (a) Phylogenetic tree representing all the *D-loop* sequences, the reference (NC_009849) inferred by maximum likelihood. (b) Phylogenetic tree representing the five haplotypes and the reference mitochondrion genome (NC_009849). Length of short branches: 0.0004. Length of long branches: 0.0009.

The *cytochorme b* was more informative and yielded two groups of haplotypes (figure 32). The largest cluster was nearly identical to the reference and had 5 polymorphic sites; haplotypes D, E, F, G, and H. When sequences from different populations were included, the first group (the larger one) was genetically close to haplotypes from Arabian Peninsula (EU159113 and NC_009849), Near Asia (KU605080 and KX554934), and Morocco (JN632608). This group of haplotypes is predominating in the Tunisian population. The second group of haplotypes was less abundant and showed 100 % genetic identity with

previously reported haplotypes (KU605078 and others), from Ethiopia and Kenya. Genetic distance in terms of number of substitutions per site was of 0.0018 inside the two groups and 0.0054 between the two groups.

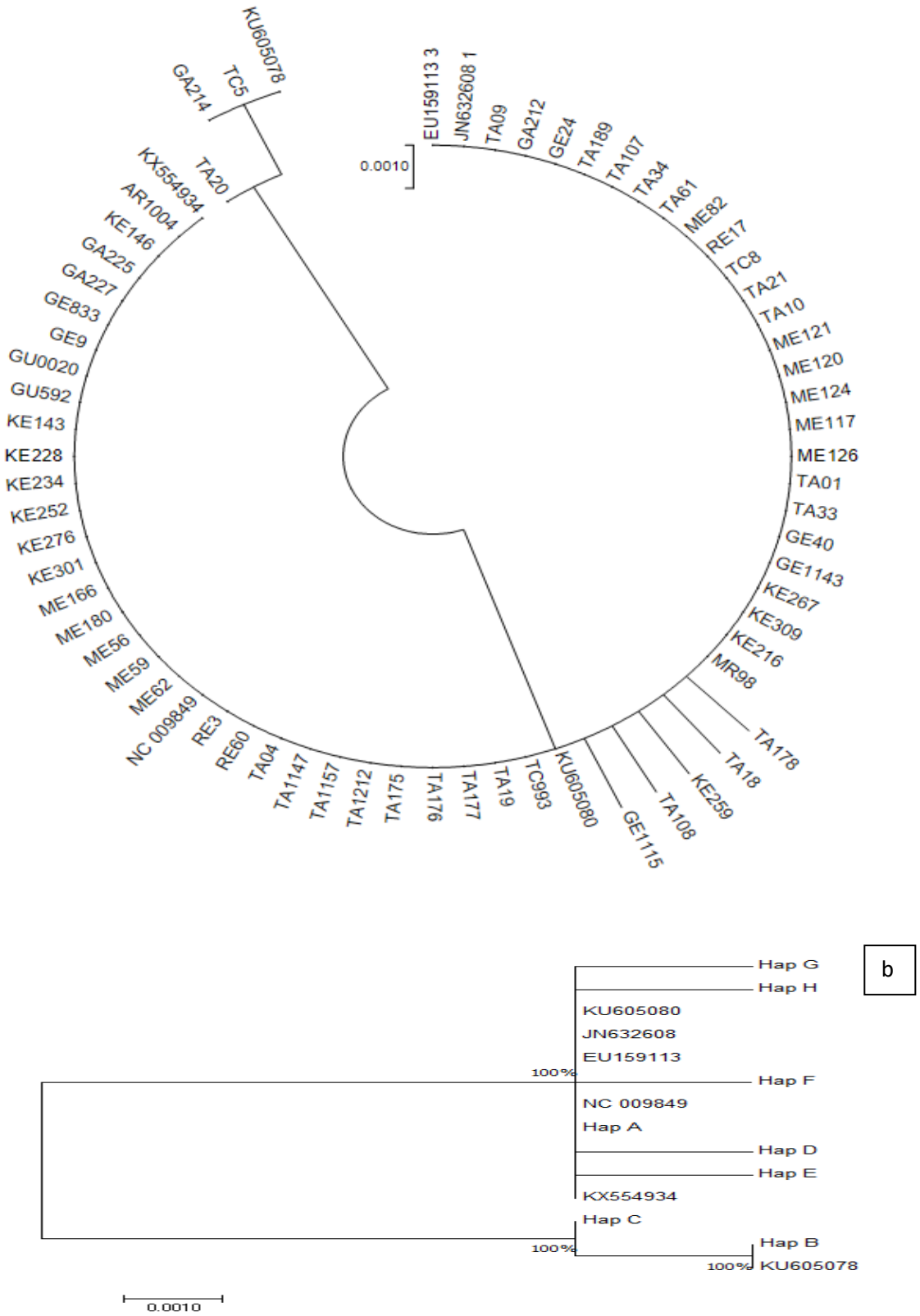


Figure 32. (a) Phylogenetic tree representing the seven haplotypes of *cytochorme b* gene in the Tunisian population, the reference mitochondrion genome (NC_009849), and the most

close sequences in GenBank. Length of the short branches: 0.0018 (1 mutation). Length of the long branches: 0.0054. (b) Phylogenetic tree representing all the *cytochorme b* sequences, the reference (NC_009849) and the five first best hits on Blast, inferred by maximum likelihood.

V.4. Variability in the consensus sequence

V.4.1. Overview

The longest fragments from *cytochorme b* gene and *D-loop* region, from the same animal, were concatenated into consensus sequences. The final sequences were of 813 bp from position 14,635 to 15,195 and from 16,319 to 16,607 (numbering according to GenBank Accession number NC_009849.1), with no gaps. These sequences spanned 561 bp of the *cytochorme b* gene and 264 bp of the *D-loop* region. The segregating sites were 8, yielding 10 haplotypes.

Most of the variable sites were singletons. Thus the nucleotide diversity was low (table 31) and up to 98 % of the sequence was conserved (figure 33). Eight haplotypes were scored only once, resulting in an important haplotypes diversity (0.43) and low pairwise nucleotide difference (1.12). These haplotypes could be explained by single, recent mutations.

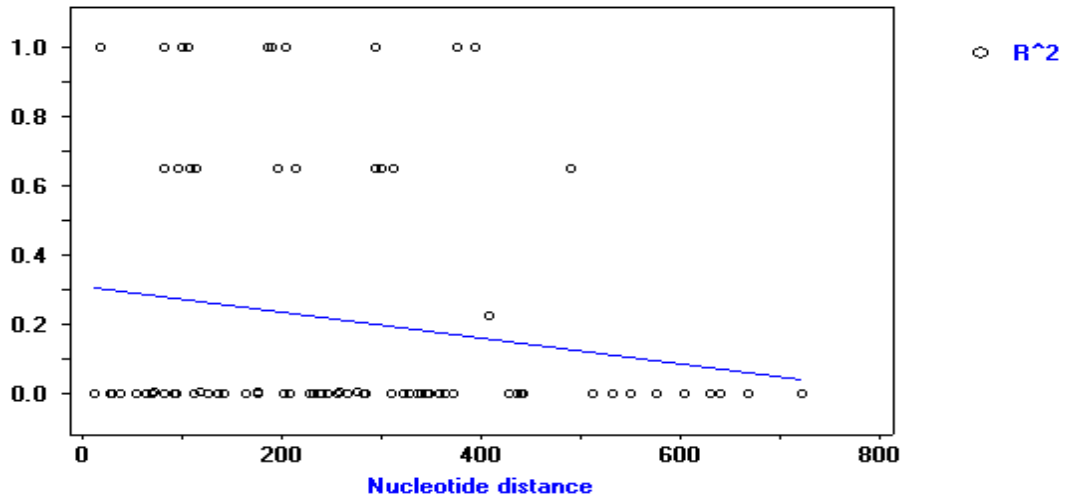


Figure 33. Nucleotide difference of all consensus sequences

Table 31. Statistics on the consensus sequence

	π	Theta	K	Hd (SD)	Tajima's D	Fu's Fs
Consensus	0.00138	1.779	1.126	0.436 (0.089)	-1.858*	-4.838*

*: $p < 0.05$

While the sequences were highly similar, with low π value, two haplotypes were proportionally highly different, with 5 mutations. This explains well the negative Tajima's D, and its statistical significance. This may result from either the action of a purifying selection or a demographic growth.

Generally, Fu's Fs test has robust statistical power to detect population expansions. Significant and negative Fu's Fs outweighs the hypothesis of population expansion assumption that was previously noticed in section III.4.2. Figure 34 visualizes the discrepancy between the observed pairwise difference and the expected difference for a population in

expansion. In its whole, the plot shows an observed curve very similar to the expected one. The observed curve presents a sole peak, with low frequency, resulting from the observed distant haplotypes.

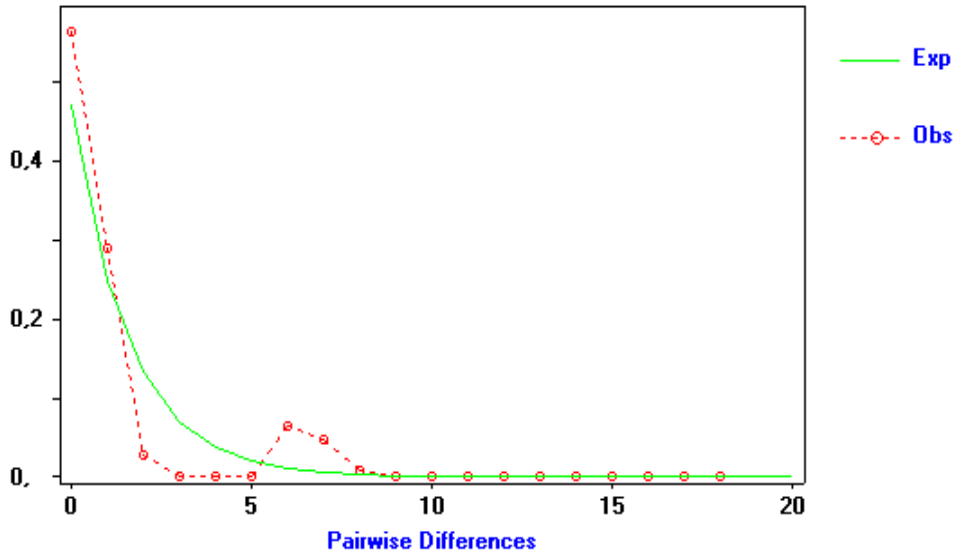


Figure 34. Mismatch distribution observed pairwise difference plotted to the expected pairwise difference

To test the hypothesis of a population growth, Fu's Fs and Tajima's D were estimated in 10000 simulations, under the different demographic models; Standard Neutral Model (SNM), population growth, population decline, bottleneck, and population split/admixture models. All simulations converged towards negative and highly significant Tajima's D ($p = 0$) and Fs ($p < 0.05$).

Based on the consensus sequence, two groups were defined. The first group was the most abundant and was congruent to the reference (Huang *et al.,* 2007). Five haplotypes were included in this group, as each differed only by one segregating site from the abundant haplotype. The second group included three samples. Interestingly, these samples were from three different geographic areas; Gabes, Tataouine and Medenine. These two groups were distant by 6 point mutations from the reference mitochondrion genome. Statistics on these groups are summarized in table 32.

Table 32. Number of sequences (N), nucleotide diversity (π), number of haplotypes (nh), haplotypes diversity (DH), number of polymorphic sites (S) in the analyzed regions, and average number of differences (K)

	N	**S**	**nh**	**Hd (SD)**	**K**	**π**
Group 1	45	6	5	0.357 (0.09)	0.387	0.00048
Group 2	3	2	3	1	1.333	0.00164
Overall	48	13	10	0.436 (0.09)	1.125	0.00138

Between populations, the fixed differences were of 6 distinct mutations. The two populations shared no mutations. All segregating sites were private for each population. The average number of nucleotide differences between the populations was 6.5 and the average number of nucleotide substitutions per site between the populations Dxy = 0.008. Number of net nucleotide substitutions per site between populations Da = 0.007.

P value for Chi2 test of genetic differentiation were highly significant (p = 0.000). Gene flow estimates (Nei, 1973) reached 1.5, and giving an important differentiation with G_{ST} = 0.14, and high F_{ST} reaching 0.87. The estimates of the Divergence Time was T = 3,761 years ago.

Statistics about the genetic difference between samples from the geographic areas showed no particular grouping or differentiation. These statistics are presented in table 33. Pairwise comparison parameters of these areas are presented in table 34.

Table 33. Number of individuals (N), number of segregating sites (S), number of haplotypes (h), average number of differences (K), nucleotide diversity (π), and haplotype diversity (Hd) per geographic area

	N	**S**	**h**	**K**	**π**	**Hd**
Kebili	14	7	2	1	0.0018	0.142
Tataouine	20	9	5	0.818	0.0014	0.337
Medenine	14	0	1	0	0	0

Table 34. Shared mutations between the populations (M), average number of nucleotide substitution per site between the populations (D_{XY}), number of net nucleotide substitutions per site between populations (D_a), average number of nucleotide difference (K_{XY}), Nei's G_{ST}, and Fst

	M	**D_{XY}**	**D_a**	**K_{XY}**	**G_{ST}**	**F_{ST}**
Kebili –Tataouine	6	0.0015	$-7\ 10^{-5}$	0.87	0.001	-0.044
Kebili – Medenine	0	0.0008	0	0.5	0	0
Tataouine – Medenine	0	0.0007	0	0.409	0.03	0

V.4.2. Phylogeny

The phylogenetic trees, based on Neighbor Joining and on Maximum likelihood, were computed based on each sequenced region, alone. The concatenation of these trees into one tree, based on common branches of reference sequences, is presented in figure 36. This method gave similar results to that using the consensus sequence.

The median joining network (figure 35) of the consensus sequences gave similar results to maximum likelihood phylogenetic trees. Two main groups are clearly defined. The large node included most of the samples. Its star shape is usually observed when a group of individuals come from the same ancestral lineage, with a number of new mutations. The haplotypes are then recent and did not cumulate larger difference from the abundant haplotypes and did not spread in the population.

The median joining network visualized the important distance (6 mutations) between the two groups of haplotypes. The second group of three individuals had two segregating sites, which made it not clear which haplotype mutated from the other. In a large study, Almathen *et al.* (2016) studied mtDNA from wide geographic area spanning 21 countries from African, Asian and Australian continents. The results showed two maternal lineages; (i) a dominant haplotype presenting few mutated haplotypes, and (ii) a second less abundant lineage, distant by 6 to 8 mutations from the first one. Estimates of mutation age and divergence years based on rho statistics gave estimations of 3,272 years (SD = 1,335 years) between the two groups.

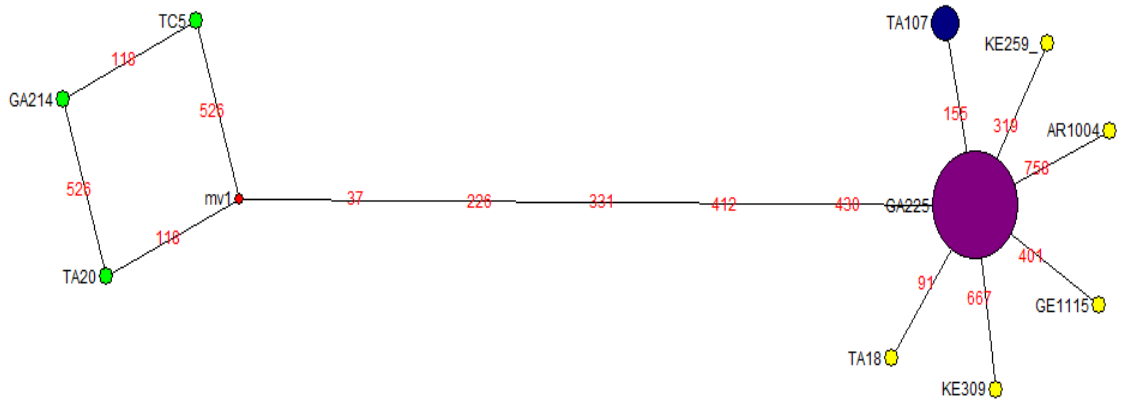


Figure 35. Median joining network presentation of haplotypes groups based on consensus sequence. The red node represents a missing haplotype or homoplasy.

V.5. Conclusion

Results from mtDNA analysis confirmed the high variability of Camel population in Tunisia. This variability was not correlated to the geographic areas or the known ecotypes. Since there was one abundant haplotype, no correlation was observed between the haplotypes distribution and geographic localization.

Two main maternal lineages were observed. In the first lineage, the abundant haplotype was genetically identical to the mitochondrion genome reference (GenBank Accession number: NC_09849). This haplotype was reported to be the most abundant in all dromedaries' populations so far studied (Almathen *et al.*, 2016). This haplotype is considered to be related to the origin of domestication, in the Arabian Peninsula, indicating an intensive trading and human movements from this area to North Africa. Other haplotypes were less frequent and genetically very close to the reference (one mutation). Most of these haplotypes were new (3 in *D-loop* region and 4 in *cytochrome b* gene). In addition, one haplotype had a single 100 % identical hit in GenBank with a sample from Pakistan. Although two independent mutations can lead to the same haplotype, the possibility of an exchange with Near Asia is also plausible.

The second maternal lineage, genetically distant from the first one, was previously reported in samples from Kenya and Ethiopia. The Eastern African dromedary population had been previously reported to be genetically distinctive from the widely panmixed dromedary populations (Almathen *et al.*, 2016). Historically, caravan routes through camels, known as Incense and Silk roads, followed a trans-Saharian trading road linking Port Sudan and the Horn of Africa (Eastern Africa) to Adrar in Southern Algeria (Bulliet, 1975; Heiss, 2012; Knoll and Burger, 2012). The Eastern African ancestral lineage probably reached Tunisia through this route or lately through exchange with the Algerian population. Recent introduction of camels from nearby regions (Libya) probably strengthened the abundance of the Arabian Peninsula haplotype, compared to those from the enclosed region of Eastern Africa.

This pattern of one common haplotype associated to minor low frequency or private haplotypes is suggestive of recent expansion of populations (Slatkin and Hudson, 1991; Rogers and Harpending, 1992). Neutrality tests applied in search for additional signs of demographic growth were in concordance. Measures of Fu's Fs statistic and Tajima's D test

from *Cytochorme b* region sequences and from consensus sequences were significantly different from zero. Both Tajima's D and Fu's Fs tests were negative and significant, confirming the demographic expansion history of the population. Combined to results from 10000 simulations, under different demographic evolution models, the population expansion assumption was confirmed.

Fu and Li's D* (-1.38) and F* (-1.88) tests were not significant at p = 0.1. When combined to significant Fs, these results prove the absence of background selection (Fu, 1997).

All this information indicates that the current population of dromedaries in Tunisia to result from different genetic flows, in concordance to the history and the largely panmictic genetic aspect of Asian and African dromedary populations (Almathen *et al.*, 2016; Burger, 2016; Cherifi *et al.*, 2017).

RESULTS

CHAPTER VI: Parentage test

7

VI. Parentage test

After evaluating the 20 microsatellite markers and estimating their statistic parameters on 130 samples of *C. dromedarius* from Tunisia (section **IV.2**), they were filtered and their power in parentage testing was evaluated. Hereafter, we present the process of the development of this test.

VI. 1. Marker filtering

Knowing the estimation of the statistic parameters for each marker, the microsatellites were filtered for basic criteria:

i. The polymorphism at a threshold p = 0.005: only one marker (CMS17) was monomorphic and thus was eliminated. All the remaining 19 microsatellites were polymorphic at 99 % threshold.
ii. Pairwise linkage disequilibrium: Locus CVRL6D appeared to be transmitted in linkage with CMS52 (p = 0.067) and was highly correlated to 4 other loci with a correlation exceeding 75 % (Weir, 1979). It was discarded to eliminate redundant information. All the remaining markers were in no significant linkage (p = 0.05).
iii. Number of alleles: With the exception of the locus VOLP32 that was bi-allelic, all the remaining loci were multi-allelic. CMS121, CVRL5D, CVRL1D, YWLL44 and CVRL7 were highly multiallelic with more than 10 alleles. They were privileged in determining the markers combination for the test. On the other hand, VOLP32, CMS18 and CMS32 had the lowest number of alleles (≤ 5).
iv. Polymorphic Information Content (PIC): This parameter is relative to the number of alleles and their frequencies. PIC of most loci was higher than 70 % (CVRL1D, CMS50, CMS15, CVRL7, VOLP10, LCA66, CVRL5D, CMS121, CMS9 and CMS13). Five loci had relatively low PIC; VOLP32 (0.2), CVRL6D, and CMS18 (0.3).

After filtering, 18 markers were suitable for parentage assay, with differential informative power. The 12 markers with the highest PIC and number of alleles (CMS121, CVRL5D, CVRL1D, YWLL44, CVRL7, CMS50, CMS15, VOLP10, LCA66, CMS9 and CMS13) were selected for the sets of parentage assay, and then tested for their exclusion power. For economic reasons, we tried to restrain the number of markers to an optimal minimum. Finally, we tried to choose microsatellites with the same annealing temperature. Amplifying all the markers in a single multiplex PCR and sequencing reaction optimizes both the cost and the running time.

VI.2. Exclusion power for parentage test

The exclusion power of each combination of markers was tested -using CERVUS- on a set of 17 trios with known parental links, following the two scenarios of one or two unknown parents. Table 35 details the exclusion power, in term of LOD score of the assigned parents, of different tested combinations of markers. Usage of 14 or more markers gave a nearly 100 % exclusion power for both scenarios. The main utilization of this type of parentage test is to detect one unknown parent (male parent). Therefore, the exclusion power for two unknown parents could be acceptable at a lower rate.

A minimum of 13 microsatellites were needed in the double unknown parents test to yield a likelihood (LOD) of 99 %. Only 10 markers were necessary to reach a power exclusion of 99 % if one parent is known. However, the exclusion power of 10 microsatellites when two parents are unknown is rather low and not decisive.

Then, we compared different sets of 12 markers yielding the best assignment rates: LOD scores should be 99 % or higher for one unknown parent and 90 % or higher for two unknown parents, and error rates should be less than 10^{-3}. This comparison aimed to choose the best combination from an economical perspective.

Table 35. Minimal exclusion power of some of the tested combinations of markers

Number of markers	**Exclusion power** ***two unknown parents***	**Exclusion power** ***one unknown parent***
14 markers or more	99.999 %	99.999 %
13 markers	98.896 %	99.999 %
12 markers	98.618 %	99.999 %
12 markers (other set of markers)	97.210 %	99.999 %
10 markers	88.667 %	99.009 %
8 markers	72.828 %	93.436 %

A final parentage assay consisting of 12 microsatellite markers was defined. Table 36 summarizes the likelihood and error rates over all tested trios. The error rates were estimated by CERVUS for each assignment; we calculated the mean error rates of all these assignments, and their standard deviation. The error rate given by comparing the assigned parent to the real data was also calculated (observed error rate).

With a declared dam, the assignment rate was of 100 %, with a likelihood of 99.99 % and an estimated error rate of 10^{-9} (table 36). All tested trios gave the same LOD score and error rate, which gave a high confidence to the parentage assay. Compared to the real data, all young animals were rightly assigned to their sires with no error. When the sire is the declared parent, all offspring were assigned rightly with no error.

Table 36. Error rates and LOD score estimates, and observed error rates of the assigned parent(s) using the proposed parentage test, for different scenarios

Test	**Dam known**	**Sire known**	**Double unknown**
LOD score Dam	-	99.02 %	98.61 %
Error rate	-	10^{-7}	10^{-3}
LOD score Sire	99.99 %	-	99.99 %
Error rate	10^{-9}	-	10^{-7}
Observed error rate	0	0	0

In the double unknown parents test, all the sires were correctly assigned, but the error rate increased to 10^{-7} (SD = 10^{-9}). This error rate is still acceptable, and the second possible male had an LOD score of less than 98 %. The dams, coming from the same herd and being related (an overall 34 % of shared genetic material), were more difficult to assign. The correct dam had the highest LOD score (≥ 98 %) for all the tested offspring. But, in less than 20 % of the cases (3 individuals), the second highest LOD score was higher than 97 %. When compared to the real data, there was no error in the assigned offspring but the mean estimated error rate based on CERVUS calculations was 10^{-3}.

Finally, the cost of the parentage test was estimated by calculating the DNA extraction, PCR reagents, capillary sequencing costs and adding a manipulation fee. The final charge was set to 40 DNT per individual.

VI.3. Discussion and conclusion

The proposed parentage test uses 12 of the most informative microsatellite markers with an exclusion power of 99 % when one parent is known and 98 % if both parents are unknown.

In general, the Tunisian camel population displays a mean of 15% of inbreeding (Ould Ahmed *et al.*, 2010). The animals used in this study exhibited an inbreeding rate of 34 %.

Therefore this parentage test should be efficient in an average inbred herd in Tunisia, and even better when investigating individuals from different herds.

The assay is based on techniques that are customary in most research laboratories. It costs less than 2 % of the average price of an adult female.

Given the limited spread of individual identification and mating recordings of dromedaries in Tunisia, this assay will allow the determination of parental relationships for different aims. In particular, it will provide a mean to choose the best sire to use for mating, in order to minimize the inbreeding risk, and it will offer the possibility of verifying pedigrees of animals, which is the foundation for productive selection and genetic improvement.

GENERAL CONCLUSION

1. High variability: Nineteen microsatellites showed that the population of dromedaries in Tunisia exhibited significant variability, demonstrated by high allelic richness (5) and average PIC (0.6), which was confirmed by the mtDNA analysis, demonstrated in the total number of haplotypes (12) and the haplotype diversity (0.7 in the *cytochorme b* gene). The traditional breeding mode of this species in Tunisia permitted better genetic diversity preservation of this livestock species.

2. Inbreeding risk: A heterozygote deficit associated to a moderate rate of inbreeding was observed. It was observed that with normal breeding practices, populations in small areas can become inbred. Especially in the absence of constant trade with surrounding geographic areas that reduces the population's effective number.

3. Weak population structure: Microsatellite genotyping data showed feeble structure (F_{ST} = 0.05) and shallow clustering of the dromedaries according to their geographic origin and their ecotypes. Both the large common genetic ground and the significant gene flow explained these results. The observed genetic diversity is more related to individual variability than to sub-population differentiation.

4. Parentage assay: A parentage test was proposed for Maghrebi camels, using 12 microsatellite markers. It offers a methodical approach for avoiding inbreeding risk, by choosing suitable mating sires. Furthermore, it offers a practical tool for checking the reliability o the pedigree data information and trace back genealogy of animals.

5. Ancestral origins: The dromedary population in Tunisia shows two maternal lineages related to Middle East and Eastern Africa. The most abundant lineage (more than 80 %) is present in most North African, Arabian Peninsula and Near Asian populations. This lineage most probably came in different waves of trading and human movements towards North Africa, from Near Asia, as it was also identified as the focal of domestication. The second maternal lineage (5 %) is more abundant in dromedary samples from Eastern Africa (Ethiopia and Kenya). This population was reported to be genetically distant and in relative isolation from other panmixed populations across North Africa and Near Asia. A probable exchange with this population brought this maternal lineage to Tunisia through trading.

The combined results from microsatellite and mtDNA analyses allowed outlining congruent recommendations:

1. Inbreeding risk: The major threat for the genetic variability and the productivity would be inbreeding. As much as the inbreeding level is moderate and not alarming, it is important to keep it at manageable levels. The anarchic mating between different ecotypes should also be controlled in order to reduce the observed high heterozygosity. A practical step would consist in sensitizing dromedary owners towards better husbandry practices; avoiding random mating of different ecotypes, using different male and female breeding lines, and introducing an outcross (primarily rams).

2. Genetic improvement: Defining the selection priorities and implementing automatic practice of identification and performance and mating recordings would be the first steps for genetic improvement. A database of these performances will allow defining the national

average and setting the statistical parameters for selection plans. Genetic improvement would rely more on the individual genetic variability than on specific ecotypes.

In addition, the population offers a high potential for the development of distinct genetic groups using the available individual variability.

The parentage assay developed in this work offers an important tool to check the mating records.

PERSPECTIVES

The promising genetic potential of dromedaries in Tunisia provides the fundament for genetic improvement. Combining this latter to other breeding and production factors (ex.: intensification, nutrition, milk automation, etc.) will enhance camel productivity and largely integrate camel livestock in the country's economy.

In practice, a realistic approach to genetic improvement would rely on putting up some nucleus herds where performance recording, progeny testing and then selection can be performed. One objective of such structure is to realize genetic improvement of the whole population performance after generations of oriented selection. In addition, it can also provide selected breeding males to the outside population, as the ordinary structure of camel herds allows genetic progress via male introductions. Such practice can both disseminate genetic progress and prevent inbreeding in the targeted herds.

On the other hand, as the construction of an 80K SNP-Chip for dromedaries is being developed (GeneSeek company), a larger genetic characterization study relaying on Single Nucleotide Polymorphisms (SNPs) may be more informative about the structure of dromedary population in Tunisia. It also may offer the possibility of detecting associations of SNPs to phenotypic traits and of marker assisted selection.

The late worldwide developments in camel genetics and genomics must be exploited on the country's level to develop and promote dromedary genetic potential, in Tunisia. This livestock species presents strategic resources for food security, in arid lands.

REFERENCES

-A-

Agapito J, Rodriguez J, Herrera-Velit P (2008). Parentage testing in alpacas (Vicugna pacos) using semi-automated fluores- cent multiplex PCRs with 10 microsatellite markers. Animal Genetics. 39: 201-203.

Alexander M, Ho SY, Molak M, Barnett R, Carlborg Ö, Dorshorst B, Honaker C, Besnier F, Wahlberg P, Dobney K, Siegel P, Andersson L, Larson G (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.

Almathen F, Charruau P, Mohandesan E, Mwacharo J. M, Orozco-terWengel P, Pitt D, Abdussamad A. M, Uerpmann H, Uerpmann M, De Cupere B, Magee P, Alnaqeeb M. A, Salim B, Raziq A, Dessie T, Abdelhadi O. M, Banabazi M. H, Al-Eknah M, Walzer C, Faye B, Hofreiter M, Peters J, Hanotte O, Burger PA (2016). Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. Proceedings of the National Academy of Sciences. 113(24): 6707-6712.

Allali I, Arnold JW, Roach J, Cadenas MB, Butz N, Hassan HM, Koci M, Ballou A, Mendoza M, Ali R, Azcarate-Peril AM (2017). A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. BMC Microbiology. 2017 17:194.

Al-Atiyat RM (2015). The power of 28 microsatellite markers for parentage testing in sheep. Electronic Journal of Biotechnology. 18(2): 116-121.

Al-swailem AM, Shehata MM, Abu-Duhier FM, Al-Yamani EJ, Al-Busadah KA, Al-Muhaimeed AN, Al-Qahtani FH, Manee MM, Al-Shomrani BM, Al-Qhtani SM, Al-Harthi AS, Inan MS, Otu HH (2010). Sequencing, Analysis, and Annotation of Expressed Sequence Tags for Camelus dromedarius. Plos one. 5(5): e10720.

Atigui M, Marnet PG, Harrabi H, Bessalah S, Khorchani T, & Hammadi M (2016). Relationship between external and internal udder and teat measurements of machine milked dromedary camels. Tropical animal health and production. 48(5): 935-942.

-B-

Bandelt HJ, Forster P, Röhl A (1999). Median-joining networks for inferring intraspecific phylogenies. Mathematisches Seminar, Universität Hamburg, Germany. Molecular Biology and Evolution. 16(1): 37-48.

Banerjee, P, Joshi J, Sharma R. Vijh K (2012). Population differentiation in dromedarian camel: A comparative study of camel inhabiting extremes of geographical distribution. International Journal of Animal Veterinary Advance. 4(2): 84-92.

Barton NH and Slatkin M (1986). A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. Heredity. 56: 409-415.

Bayes T, Price R (1763). An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S. Philosophical Transactions of the Royal Society of London. 53: 370-418.

Beaumont MA (1999). Detecting population expansion ad decline using microsatellites. Genetics. 153(4): 2013-2029.

Bedhiaf-Romdhani S, Chahbani I, Djemali M (2015). Caractérisation phénotypique et moléculaire de l'espèce cameline au sud Tunisien par des marqueurs AFLP. Proceedings of the 1st International Seminar on Livestock and Wild life in Arid and Desertic lands (SIEFAD).

Belgacem AO, Sghaier M, Ouessar M (2011). Vulnérabilité de l'écosystème pastoral face au changement climatique dans le Gouvernorat de Médenine. Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH. Ministère de l'Equipement, de l'Aménagement du Territoire et du DéveloPement Durable (Tunisie). P 276.

Bensasson D, Zhang D, Hartl DL, Hewitt GM (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology and Evolution. 16(6): 314-321.

Bulliet RW (1975). The camel and the wheel. Cambridge, Harvard University press. P 327.

Burger P (2016). The history of Old World camelids in the light of molecular genetics. Tropical Animal Health and Production. 48: 905-913.

Buschmann H (1965). Blood group studies in pigs. In: Maousek J. (eds) Blood groups of animals. Springer. Dordrecht. ISBN: 978-94-017-5834-5.

-C-

Cherifi Y, Gaouar SBS, Guastamacchia R, El-Bahrawy KA, Abushady AMA, Sharaf AA, Harek D, Lacalandra GM, Saïdi-Mehtar N, Ciani E (2017). Weak genetic structure in Northern African dromedary camels reflects their unique evolutionary history. Plos one. 12(1): e0168672.

Chniter M, Hammadi M, Khorchani T, Krit R, Ben wahada A, Ben Hamouda M (2013). Classification of Maghrebi camels (camelus dromedarus) according to their tribal affiliation and body traits in southern Tunisia. Emirates journal of food and Agriculture. 25(8): 625-634.

Ciofi C, Funk SM, Coote T, Cheesman DJ, Hammond RL, Saccheri IJ, Bruford MK (1998). Genotyping with microsatellite markers. In: Karp A, Isaac PG, Ingram DS (eds) Molecular tools for screening Biodiversity, Springer, Dordrecht. P 195-201.

Coates BS, Sumerford DV, Miller NJ, Kim KS, SaPingtonBlair TW, Leslie DS, Lewis C (2009). Comparative Performance of Single Nucleotide Polymorphism and Microsatellite Markers for Population Genetic Analysis. Journal of Heredity. 100(5): 556-564.

Cornuet JM, Luikart G (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics. 144(4): 2001-2014.

Crow JF and Kimura M (1970). An introduction to population genetics theory. Harper and Row, New York. P 591.

Csilléry K (2009). Statistical inference in population genetics using microsatellites. Ph.D. The University of Edinburgh. P 172.

Cummings MP, Handley SA, Myers DS, Reed DL, Rokas A, Winka K (2003). Comparing bootstrap and posterior probability values in the four-taxon case. System Biology. 52(4): 477-87.

-D-

da Silva EC, McManus MC, Guimarães MP, Gouveia MG, Facó O, Pimentel DM, Caetano AR, Paiva SR (2014). Validation of a microsatellite panel for parentage testing of locally adapted and commercial goats in Brazil. Genetics and Molecular Biology. 37(1): 54-60.

DeFaveri J, Viitaniemi H, Leder E, Merilä J (2013). Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. Molecular ecology resources. 13: 377-392.

Di Rocco F, Zambelli A, Mate L, Vidal-Rioja L (2010). The complete mitochondrial DNA sequence of the guanaco (Lama guanicoe): comparative analysis with the vicuna (Vicugna vicugna) genome. Genetica. 138(8): 813-818.

Dirienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M (1994). Mutational processes of simple-sequence repeat loci in human populations. Proceedings of National Academy of Science USA. 91: 3166-3170.

Djemali M (2005). Camel genetic ressources in North Africa. FAO-ICAR Seminar on Camelids: Cuerrent Status of Genetic Ressouces, Recording and Production Systems in African, Asian and American Camelids. Sousse, Tunisia, 2004. ICAR: Technical series. 11: 51-60.

Douady CJ, Delsuc F, Yan Boucher W, Doolittle F, Douzery EJP (2003). Comparison of Bayesian and Maximum Likelihood Bootstrap Measures of Phylogenetic Reliability. Molecular Biology and Evolution. 20(2): 248-254.

Douris R (1922). APlication medico-legale des groupes sanguins humains. Discussion de la patemite. Bulletains de Science; Pharmacologie. 30: 90-107.

Duchêne S, Archer F, Vilstrup J, Caballero S, Morin PA (2011). Mitogenome phylogenetics: The impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. Plos one. 6(11): e27138.

-E-

Erixon P, Svennblad B, Britton T, Oxelma B (2003). Reliability of Bayesian Posterior Probabilities and Bootstrap Frequencies in Phylogenetics. System Biology. 52(5): 665–673.

Estoup A, Jarne P, Cornuet JM (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Molecular Ecology. 11: 1591-1604.

Evdotchenko D, Han Y, Preuss S, Geldermann H (2003). New polymorphic microsatellite loci for different camel species. Molecular Ecology Notes. 3: 431-434.

Ewens WJ (1979). Mathematical population genetics 1: Theoretical introduction. Springer-Verlag, Berlin. P: 418.

-F-

FAO (2008). L'état des resources zoogénétiques pour l'alimentation et l'agriculture dans le monde, édité par B. Rischkowski et D. Pilling. Rome. P 551.

FAO-ISAG (2011). Molecular genetic characterization of animal genetic resources. FAO Animal Production and Health Guidelines. No. 9. Rome. P 85.

FAO, UNEP (2000). World Watch List for domestic animal diversity. Beate D. Scherf. 3rd Edition. Food and Agriculture Organization of the United Nations. Rome. P 746.

Faye B, Grech S, Korchani T (2004). Le dromadaire, entre féralisation et intensification. Anthropozoologica. 39(2): 7-14.

Fitak RR, Mohandesan E, Codander J, Burger P (2016). The de novo genome assembly and annotation of a female domestic dromedary of North African origin. Molecular Ecology Resources. 16: 314-324.

Fu YX (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 147(2): 915-925.

-G-

Galtier N, Nabjolz B, Glémin S, Hurst GDD. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Molecular ecology. 18: 4541-4550.

Garcia-Sandoval R (2014). Why some clades have low bootstrap frequencies and high Bayesian posterior probabilities. Israel Journal of Ecology & Evolution. 60(1): 41-44.

Gärke C, Ytournel F, Bed'hom B, Gut I, Lathrop M, Weigend S, Simianer H (2011). Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. Animal Genetics. 43(4): 419-28.

Garza JC, Slatkin M, Freimer NB (1995). Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. Molecular Biology and Evolution. 12: 594-603.

Gissi C, Iannelli F, Pesole G (2008). Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity. 101: 301-320.

-H-

Hammadi M, Khorchani T, Khaldi K, Majdoub A, Abdouli H, Slimane N, Portetelle D, Renaville R (2001). Effect of diet suPlementation on growth and reproduction in camels under arid range conditions. Biotechnology Agronomy Society and Environment. 5(2): 69–72.

Hara K, Watabe H, Sasazaki S, Mukai F, Mannen H (2010). Development of SNP markers for individual identification and parentage test in a Japanese Black cattle population. Animal Science Journal. 81(2): 152-157.

Hassanin A, Delsuc F, Ropiquet A, Hammer C, Jansen van Vuuren B, Matthee C, Ruiz-Garcia M, Catzeflis F, Areskoug V, Nguyen TT, Couloux A (2012). Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. Comptes Rendus Biologies. 335(1): 32-50.

Hayan L, Gurtowski J, Yoo L, Nattestad M, Marcus S, Goodwin S, McCombie WR, Schatz MC (2016). Third-generation sequencing and the future of genomics. BioRxiv (preprint). doi:10.1101/048603.

Heide M (2010). The domestication of the camel: biological, archaeological and inscriptional evidence from Mesopotamia, Egyypt, Israel and Arabia, and traditional evidence from the Hebrew Bible. Ugarit-Forschungen. 42: 82-100.

Heintzman P, Zazula G, Cahill J, Reyes A, MacPhee R, Shapiro B (2015). Genomic data from extinct North American Camelops revise Camel evolutionary history. Molecular Biology Evolution. 32(9): 2433-2440.

Heiss J (2012). Caravans from South Arabia: Roads and organization. Camels in Asia and North Africa. Interdisciplinary Perspectives on Their Significance in Past and Present, eds Knoll EM, Burger PA (Austrian Academy of Science Press, Vienna). P 131–140.

Herraez DL, Schäfer H, Mosner J, Fries HR, Wink M (2005). Comparison of microsatellite and single nucleotide polymorphism markers for the genetic analysis of a Galloway cattle population. Journal of biosciences. 60(7-8): 637-643..

Heyen DW, Beever JE, Da Y (1997). Exclusion probabilities of 22 bovine microsatellite markers in fluorescent multiplexes for semiautomated semiautomated parentage testing. Animal Genetics. 28: 21-7.

Hooker SB and Boyd WC (1934). Blood-Grouping as a Test of Non-Paternity. American Institute of Criminal Law and Criminology. 187: 1934-1935.

Huang Q, Xu FY, Shen H, Deng HY, Liu YJ (2002). Mutation patterns at dinucleotide microsatellite loci in humans. American Journal of Human Genetics. 70: 625-634.

Hudson RR, Slatkin M, Maddison WP (1992). Estimation of levels of gene flow from DNA sequence data. Genetics. 132(2): 583-589.

Huelsenbeck JP, Larget B, Miller RE, Ronquist F (2002). Potential APlications and Pitfalls of Bayesian Inference of Phylogeny. Systematic Biology. 51(5): 673-688.

-J-

Jinliang W (2007). Triadic IBD coefficients and aPlications to estimating pairwise relatedness. Genetic Resources. 89: 135-153.

Jombart T, Devillard S, Balloux F (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics. 2010 11 :94.

Jones AG, Small CM, Paczolt KA, Ratterman NL (2010). Technical review A practical guide to methods of parentage analysis Molecular Ecology Resources. 10: 6-30.

-K-

Kalinowski ST, Taper ML, Marshall TC (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology. 16: 1099-1106.

Kamoun M. (1995). Le lait du dromadaire: production, aspects qualitative et aptitude la transformation. Options Méditerranéennes. 13: 81-102.

Kamoun M, Rekik B, Bouzari M, Tayechi L (2009). Quality of camel meat marketed by butchers in Tunisia. Second conference of the iternatioal society of camelid research and developmet. Djerba, Tunisia 12-14th, 2009.

Karp A, Isaa PG, Ingram DS (1998). Molecular Tools for Screening Biodiversity. Chapman and Hall, London.

Khorchani, T., M. Ismail, M. Hammadi, M. Moslah et M. Chemmem (1996). Sauvegarde du dromadaire et amélioration de sa productivité : Bilan de principales recherches menées l'institut des régions arides de Médenine (Tunisie). Revue des régions arides N° spécial. 1996: 368-376.

Kimura M (1983). The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, MA. P 380.

Kimura M and Crow JF (1964). The number of alleles that can be maintained in a finite population. Genetics. 49: 725-38.

Knoll EM and Burger PA (2012). Camels in Asia and North Africa: Interdisciplinary perspectives on their past and present significance. Austrian Academy of Sciences Press. Veröffentlichungen zur Sozialanthropologie volume 18.

Koch M, Hadfield JD, Sefc KM, Sturmbauer C (2008). Pedigree reconstruction in wild fish populations. Molecular Ecology. 17: 4500-4511.

Kristensen TN, Hoffmann A, Pertoldi C, Stronen AV (2015). What can livestock breeders learn from conservation genetics and *vice versa*? Frontiers in Genetics. 6: 38.

Kristensen TN and Sørensen AC (2005). Inbreeding-Lessons from animal breeding, evolutionary biology and conservation genetics. Animal Science. 50: 121-133.

Kruglyak S, Durrett RT, Schug MD, Aquadro CF (1998). Equilibrium distributions of microsatellite repeat length resulting from a balance between sliPage events and point mutations. Proceedings National Academy Sciences USA. 95: 10774-10778.

-L-

Lang K, Wang Y, Plante Y (1996). Fifteen polymorphic dinucleotide microsatellites in llamas and alpacas. Animal Genetics. 27: 285-294.

Lee CL (1975). Current status of paternity testing. Family law quarterly, American Bar Association. 9(4): 615-633.

Lee C, Abdool A, Huang CH (2009). PCA-based population structure inference with generic clustering algorithms. BMC Bioinformatics. 10(S1): S73.

Lee H, Gurtowski J, Yoo S, Nattestad M, Marcus S, Goodwin S, MrCombie WR, Schatz M (2016). Third generation sequencing and the future of genomics. BioRxiv (preprint). doi: 10.1101/048603.

Li WH (1976). Effect of migration on genetic distance. American Naturalist. 110: 841-847.

Li M, Zhao S, Bian C, Wang H, Wei H, Liu B, Yu M, Fan B, Chen S, Zhu M, Li S, Xiong T, Li K (2002). Genetic relationships among twelve Chinese indigenous goat populations based on microsatellite analysis. Genetic Selection and Evolution. 34: 729-744.

Luikart G, Biju-Duval MP, Ertugrul O (1999). Power of 22 microsatellite markers in fluorescent multiplexes for parentage testing in goats (Capra hircus). Animal Genetics. 30: 431-438.

-M-

Mahmoud AH, Alshaikh MA, Aljumaah RS, Moammed OB (2012). Genetic variability of camel (Camelus dromedarius) populations in Saudi Arabia based on microsatellite markers. African Journal of Biotechnology. 51(11): 11173-11180.

Mahrous KF, Ramadan HA, Abdel-Aziem SH, Abdel Mordy M, Hemdan DM (2011). Genetic variations between camel breeds using microsatellite markers and RAPD techniques. Journal of APlied Biosciences. 39: 2626-2634.

Manly FJ (1985). Tests of the theory of natural selection: An overview. Journal of the Royal Society of New Zealand. 15(4): 411-432.

Mariasegaram M, Pullenayegum S, Ali M, Shah RS, Penedo MC, Wernery U, Sasse J (2002). Isolation and characterization of eight microsatellite markers in *Camelus dromedarius*and cross-species amplification in *C.bactrianus* and *Lama pacos.* Animal Genetics. 33(5): 385-387.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998). Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology. **7**: 639-655.

Mburu, D, Ocheing W, Kuria G, Jianlin H, Kaufmann B, Rege JE, Hanotte O (2003). Genetic diversity and relationships of indigenous Kenzan camel (*Camelus dromadarius*) opulation. Animal Genetics. 34: 26-32.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH (1953). Equation of State Calculations by Fast Computing Machines. Journal of Chemical Physics. 21(6): 1087.

Miller SA, Dykes DD, Polesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research 16(3): 1215.

Miyamoto M, Motooka D, Gotoh K, Imai T, Yoshitake K, Goto N, Iida T, Yasunaga T, Horii T, Arakawa K, Kasahara M, Nakamura S (2014). Performance comparison of second- and third-generationsequencers using a bacterial genome with two chromosomes. BMC Genomics. 21(15): 699.

Mohandesan E, Speller CF, Peters J, Uerpmann HP, Uerpman M, De Cupere B, Hofreier M, Burger P (2017). Combined hybridization capture and shotgun sequencing for ancient DNA analysis of extinct wild and domestic dromedary camel. Molecular Ecologic Resources. 17(2): 300-313.

-N-

Nasr N, Ben Salem M, Rachidi Y. L, Benissad J, Medouni Y (2000). Mutation des systems d'élevage et de gestion des parcours en zones arides (sud est tunisien). Revue des régions arides 8: 57-77.

Nei M (1972). Genetic distance between populations. The American naturalist. 106(949): 283-292.

Nei M and Roychoudhury AK (1974). Sampling variances of heterozygosity and genetic distance. Genetics 76:379-390. Bayesian posterior probabilities. Israel Journal of Ecology & Evolution. 60(1): 41-44.

Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 89(3): 583-590.

Nei M (1982). Evolution of human races at the gene level, P. 167-181. In Bonne-Tamir,B., Cohen,T. and Goodman,R.M. (eds.), Human genetics, part A: The unfolding genome. Alan R. Liss, New York, NY.

Nei M (1987). Molecular Evolutionary Genetics. Columbia University Press, New York, NY

Nielsen R, Mattila DK, Clapham PJ, Palsbøll PJ (2001). Statistical approaches to paternity analysis in natural populations and aPlications to Northern Atlantic humpback whale. Genetics. 157: 1673-1682.

Nolte M, Kotze A, Van der Bank FH, Grobler JP (2005). Microsatellite markers reveal low genetic differentiation among South American Camelus dromedrius populations. South African Journal of Animal Sciences. 35: 152–161.

Nouairia G, Kdidi S, Ben Salah R, Mohammed H, Khorchani T, Yahyaoui MH (2015). Assessing genetic diversity of three Tunisian dromedary camel (*Camelus dromedarius*) sub populations using microsatellite markers. Emirates Journal of Food and Agriculture. 27(4): 362-366.

-O-

Obreque, V., L. Coogle, P. J. Henney, E. Bailey, R. Mancilla, J. Garcia-Huidobro, P. Hinrichsen and E. G. Cothran. 1998. Characterisation of 10 polymorphic alpaca dinucleotide microsatellites. Animal Genetics. 29: 461-462.

OEP (2016). Plan national de développement du secteur camelin à l'horison 2020. Ministère de l'agriculture, des ressources d'eau et de la pêche (unpublished).

Ohta T and Kimura M (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genome Research. 22: 201-204.

Ould Ahmed M, Ben Salem F, Bedhiaf S, Djemali M. (2007). Caractérisation des pratiques de l'élevage du dromadaire (Camelus dromedarius) en Tunisie. Revue de l'Institut National Agronomique de Tunisie. 22: 127-138.

Ould Ahmed M. and Djemali M (2011). Caractérisation de la population des dromadaires : (*Camelus dromedarius*) en Tunisie. Editions universitaires europeennes EUE. 2011:172.

Ould Ahmed M, Ben Salem F, Bedhiaf S, Rekik B, Djemali M. (2010). Genetic diversity in Tunisian dromedary (Camelus dromedarius) populations using microsatellite markers. Livestock Science. 132: 182-185.

-P-

Panetto CJ, Machado M, da Silva GB, Barbosab R, Gouveiados G, Leitec S, Peixotoa CD (2017). Parentage assignment using SNP markers, inbreeding and population size for the Brazilian Red Sindhi cattle. Livestock Science. 213: e08.008.

Parker J, Helmstetter AJ, Devey D, Wilkinson T, Papadopulos AST (2017). Field-based species identification of closely-related plants using real-time nanopore sequencing. Scientific reports. 7: 8345.

Penedo M, Caetano A, Cordova K (1999). Eight microsatellite markers for South American camelids. Animal Genetics. 30: 161-168.

Perrier X, Flori A, Bonnot F (2003). Data analysis methods. In: Hamon, P.; Seguin, M.; Perrier, X. & Glaszmann, J.C. (Ed.). Genetic diversity of cultivated tropical plants. Enfield Science Publishers, Montpellier. P 43-76.

Pickford M, Morales J, Soria D (1995). Fossil camels from the UPer Miocene of Europe: implications for biogeography and faunal change. Geobiosystems. 28: 641–650.

Piro M, Bouazzati O, Bengoumi M, Allali K, Achaaban MR, Benjouad A, Nabich A, Ouragh L (2011). Genetic characterization of Moroccan camel populations using microsatellites markers. Journal of Camel Practice and Research. 18(2): 167-172.

Pritchard JK and Feldman MW (1996). Statistics for microsatellite variation based on coalescence. Theoretical Population Biology. 50: 325-344.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics. 155(2): 945-959.

Priyanka PR and Barve SS (2011). Standardization and Optimization of mtDNA Isolation and Molecular Genetic Analysis of *D-loop* Region in Animal Natural Fibres. International Journal of Zoological Research. 7: 190-200.

-R-

Rendel J (1957). Studies of Cattle blood groups. Acta Agriculturae Scandinavia. 8(3): 191-215.

Rendel J and Gahne B (1961). Parentage tests in cattle using erythrocyte antigens and serum transferrins. Animal Science. 3(3): 307-314.

Richard S (2003). Near Eastern Archeology: A reader. Eisenbrauns p120. ISBN-13: 978-1575062341.

Ritland K (1996). Estimators for pairwise relatedness and inbreeding coefficients. Genetic Resources. 67: 175-186.

Robertson A, Hill WG (1984). Deviations from Hardy-Weinberg proportions: Sampling variances and use in estimation of inbreeding coefficients. Genetics. 107: 703-718.

Rogers AR, Harpending H (1992). Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution. 9(3): 552-569.

-S-

Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 4(4): 406-425.

Samman MA, Al-Saleh AA, Sheth K (1993). The Karyotype of the Arabian Camel, *Camelus dromedarius*. Journal of King Saud University. 5: 57-64.

Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of National Academy of Science of the United States of America. 74(2): 5463-5467.

Sasse, J, Mariasegaram M, Balu R, Kinne J, Wernery U (2000). South American camelid microsatellite amplification in Camelus dromedaries. Animal Genetics. 31: 75-76.

Sawilowsky, Shlomo S, Fahoome, Gail C (2003). Statistics via Monte Carlo Simulation with Fortran. Journal of Modern APlied Statistical Methods, Michigan. ISBN 0-9740236-0-4.

Schulz U, Tupac-Yupanqui I, Martinez A, Mendez S, Delgado C, Gomez M, Dunner S, Canon J (2010). Diversity. 2(4): 561-571.

Sghaier M (2003). Camel Marketing in Tunisie. The Camel APlied Research and Development Network (CARDN). ACSAD. CARDN/ACSAD/Camel: p108.

Sghaier M., 2004. Camel production systems in Africa. FAO-ICAR Seminar on Camelids: Cuerrent Status of Genetic Ressouces, Recording and Production Systems in African, Asian and American Camelids. Sousse, Tunisia, 2004. ICAR: Technical series. 11: 19-30.

Shah MG, Qureshi AS, Reissmann M (2014). Phenotypic and Genetic differentiation of Pakistani camel breed. Lambert Academic publishing. P 230.

Simianer, H., Reist-Marti, S.B., Gibson, J., Hanotte, O., Rege, J.E.O., 2003. An approach to the optimal allocation of conservation funds to minimize loss of genetic diversity between livestock breeds. Ecological Economics. 45: 377–392.

Slatkin M and Hudson RR (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics. 129(2): 555-562.

Spencer PBS, Wilson KJ (2010). "Parentage testing of racing camels (Camelus dromedarius) using microsatellite DNA typing." Animal Genetics. 41: 121-126.

Stormont C (1967). Contribution of blood typing to dairy science progres. Journal of Dairy Science 50(2): 253-260.

Suzuki Y, Glazko GV, Nei M (2002). Over credibility of molecular phylogenies obtained by Bayesian phylogenetics. Proceedings of the National Academy of Sciences USA. 99: 16138–16143.

-T-

Tajima F (1983). Evolutionary relationship of DNA sequences in finite populations. Genetics. 105: 437-460.

Tajima F (1989). "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism". Genetics. 123 (3): 585–95.

Theilmann JL, Skow LC, Baker JF, Womack JE (1989). Restriction fragment length polymorphisms for growth hormone, prolactin, osteonectin, α crystallin, γ crystallin, fibronectin and 21-steroid hydroxylase in cattle. Animal Genetics. 20(3): 257-266.

Tuomey M, Leidy J (1854). September 26th; Description of some new fossils, from the cretaceous rocks of the southern states; description of a fossil aParently indicting an extinct species of the camel tribe. Proceedings of the academy of natural sciences of Philadelphia. 7 (1854-1855): 167-173.

-V-

Valdes AM, Slatkin M, Freimer NB (1993). Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133: 737-749.

Vijh RK, Tantia MS, Mishra B, Kumar B (2007). Genetic diversity and differentiation of dromedarian camel of India. Animal Biotechnology. 18: 81- 90.

-W-

Wan QH, Wu H, Fujihara T, Fang SG (2004). Which genetic maker for which conservation genetics issue? Electrophoresis 25: 2165-2176.

Wang J (2007). Triadic IBD coefficients and aPlications to estimating pairwise relatedness. Genetic resources 89(3): 135-53.

Watterson GA (1975). On the number of segregating sites in genetical models without recombination. Theoretical Population Biology **7** (2): 256-276.

Weir BS (1979). Inferences about linkage disequilibrium. Biometrics 35(1): 235-254.

Weir BS and Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. Evolution 38(6): 1358-1370

Werner FAO, Durstewitz G, Haermann FA, Thaller G, Krämer W, Kollers S, Buitkamp J, Georges M, Brem G, Mosner J, Fries JR (2003). Detection and characterization of SNPs useful for identity control and parentage testing in major European dairy breeds. Animal Genetics. 35: 44-49.

Williamson G, Payne WJA (1978). An introduction to animal husbandry in the tropics. Third Edition. London, UK: Longman. P485. ISBN 978-0-582-46813-9

Wright P (1931). Evolution in 91endelian populations. Genetics 16: 97–159.

Wright S (1951). The genetic structure of populations. Annals of Eugenics 15: 323-354.

Wright P (1987). A refinement of Porter's strategies. Strategic Management Journal. 8(1): 93-101.

Wu H, Guang X, Al-Fageeh MB, Cao J, Pn S, Zhou H, Zhang L, Abutarboush MH, Xie Z, ALshanqeeti AS, Zhang Y, Al-Shomrani BM, Zhang D, Li J, Manee MM, Yang , Liu Y, Zhang J, Altammami MA, Wang S, Yu L, Liu S, Ba L, Liu C, Meng F, Wang S, Li L, Li X, Wu K, Zhang S, Wang J, Yin Y, Yang H, AL-swailem AM, Wang J (2014). Camelid genomes reveal evolution and adaptation to desert environments. Nature communications. 5:5188. ISSN: 2041-1723.

-Y-

Yang D, Liang H, Gao Y, Lin S, He Z, Gao J, Su H, Li Q, Ma X, Ou X (2017). Noninvasive fetal genotyping of pternally inherited alleles using trageted massively parallel sequencing in parentage testing cases. Transfusion 57(6):1505-1514.

-Z-

Zaho X, Chu M, Li N, Wu C (2001). Paternal inheritance of mitochondrial DNA in the sheep (Ovine aries). Science in China Series C: Life Sciences 44:321-326.

Zangemeister W (1929). A new paternity test (serology test for paternity). Medico-legal Journal 5.

Zhang DX, Hewitt GM (2003). Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Molecular Ecology 12:563–584

DATABASE AND WEBSITES

FAO (2018). Statistics Division (FAOSTAT). http://faostat4.fao.org. [Last consultation : February, 2018]

NIH (2017). National Human Genome Research Institute. https://www.genome.gov [Last consultation: November, 2017]

NCBI (2018). National Center for Biotechnology Information. https//www.ncbi.nlm.nih.gov [Last consultation: March, 2018]

OEC (2018). Observatory of Economic Complexity, database. http://atlas.media.mit.edu/. [Last consultation : March, 2018].

OEP (2017). Office de l'ELevage et de la Faune Sauvage. www.oep.nat.tn. [Last consulation : December, 2017]

SOFTWARE

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Belkhir K., Borsa P., Chikhi L., Raufaste N. and Bonhomme F., 2004. GENETIX 4.05.2, Logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions ; CNRS UMR 5000, Université de MontpellierII, France. http://www.univ-montp2.fr/~genetix/genetix/constr.htm#download

Bryant D, Moulton V (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Molecular biology and evolution 21: 255–265.

Dieringer D and Schlötterer C (2003). Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3(1):167-169.

Earl, Dent A and vonHoldt BM (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4(2):359-361. doi:10.1007/s12686-011-9548-7.

Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14(8):2611-2620.

FASTX Toolkit (2010). by **Hannon Lab**. http://hannonlab.cshl.edu/fastx_toolkit/index.html Key: citeulike:9103573

Felsenstein, J (2005). PHYLIP (phylogeny inference package) version 3.6. Distributed by Author. Department of Genome Sciences, University of Washington, Seattle. Cladistics 5:164-166.

Glaubitz JC (2004). CONVERT: a user friendly program to reformat diploid genotypic data for commonly used population genetic software packages. Molecular Ecology Notes 4:309-310.

Jombart T (2008). Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405.

Krolzig HM (1998). Econometric modelling of markov-switching vector autoregressions using MSVAR for Ox. Institute of economics and statistics and Nuffield college, Oxford.

Langella O (1999). POPULATIONS v1.2.32 (2002): a population genetic software. CNRS UPR9034. http://bioinformaticsorg/~tryphon/populations

Peakall R and Smouse PE (2006). GENALEX v6.501 genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara RB, Simpson GL, olymos P, Stevens MH, Wagner H, Oksanen MJ (2013). Package Vegan fro R. Community ecology package 2 :9

Rice P, Longden I, Bleasby A (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics 16(6):276-277.

RStudio Team (2015). RStudio: Integrated development for R. RStudio, Inc., Boston. URL: http://www.rstudio.com

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007). ClustalW and ClustalX version 2.0. Bioinformatics 23:2947-2948.

Paradis E (2010). Pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26(3):419-420.

Perrier X, Jacquemoud-Collet JP (2006). DARwin software. http://darwin.cirad.fr/Darwin

Rousset F, Raymond M, Rousset F (1995). GENEPOP Version 1.2: population genetics softwarefor exact tests and ecumenicism. Journal of Heredity 86:248-249.

Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Librado P, Guirao-Rico S, Ramos-Onsins SE, Sanchez-Garcia A (2017). DNA Sequence Polymorphism Version 6.10.04 December 13, 2017. http://www.ub.es/dnasp

Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:539. doi:10.1038/msb.2011.75

Tamura K, Dudley J, Nei M & Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596-1599.

Wall L (2010). Perl Language Reference Manual - for Perl version 5.12.1. Network Theory Ltd. ISBN: 1906966028-9781906966027.

Yeh FC, Boyle T, Rongcai Y, Ye Z, Xiyan JM (1999). Popgene: Microsoft Window-based Freeware for Population Genetic Analysis. https://sites.ualberta.ca/

SEQUENCES

Huang, X., Shah, R.S. and Khazanehdari, K.A. (2009). Camelus dromedarius mitochondrion, complete genome. GenBank NCBI Reference Sequence: NC_009849

Tahmoorespur M, Javadmanesh A, Azghandi M, Abbassi-Daloii T, Yousefi S, Khabiri AA, Shahabi A, Torabi A (Unpublished). Sequencing and bioinformatic analysis of mitogenome in Iranian Bacterianus and Dromedaries camels. GenBank Accession number: KX554934

Mohandesan E, Fitak RR, Yadamsuren A, Chuluunbat B, Abdelhadi O, Raziq A, Juhasz J, Nagy P, Burger PA (Unpublished). Whole mitogenome scans reveal signatures of purifying selection in the genus Camelus. GenBank Accession number: KU605080 and KU605078

Ji R, Cui P, Ding F, Geng J, Gao H, Zhang H, Yu J, Hu S, Meng H (2009). Monophyletic origin of domestic bactrian camel (Camelus bactrianus) and its evolutionary relationship with the extant wild camel (Camelus bactrianus ferus). Animal Genetics 40(4):377-382. GenBank Accession number: NC_009628.2

Hassanin A, Delsuc F, Ropiquet A, Hammer C, Van Vuuren JTT, Couloux A (2012). Pattern and timing of diversification of Cetartiodactyla (Mammalia Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. Comptes Rendus Biologies 355(1):32-50. Accession Number: JN632608.1

ANNEXES

Annex 1. Primer sequences, annealing temperatures and Gene Bank accession numbers of the used microsatellites

Microsatellite	Primer sequences F R	Annealing temperatures*	Ta**	Gene Bank Accession number
VOLP10	CTTTCTCCTTTCCTCCCTACT CGTCCACTTCCTTCATTTC	55	60	AF305231
VOLP32	GTGATCGGAATGGCTTGAAA CAGCGAGCACCTGAAAGAA	55	58	AF305234
YWLL44	CTCAACAATGCTAGACCTTGG GAGAACACAGGCTGGTGAATA	55-60	60	-
YWLL59	TGTGCAGGAGTTAGGTGTA CCATGTCTCTGAAGCTCTGGA	50-58	60	-
CVRL08	AATTCCTGTGATTTTATACACA CATGTCATGAAAGCTACAGTA	58	60	AF217608
CVRL07	GTGCAGCGTCCAAATAGTCA CCAGCATCGTCCAGTATTCA	50-58	60	AF091125
CVRL6D	TTTTAAAAATTCTGACCAGGAGTCTG CATAATAGCCAAAACATGGAAACAAC	55	60	AF217606
CVRL5D	CCTTGGACCTCCTTGCTCTG GCCACTGGTCCCTGTCATT	60	60	AF217602
CVRL4D	CCCTACCTCTGGACTTTG CCTTTTTGGGTATTTTCAG	60	60	AF217604
CVRL1D	GAAGAGGTTGGGGCACTAC CAGGCAGATATCCATTGAA	55	60	AF217601
CMS50	TTTATAGTCAGAGAGAGTGCTG TGTAGGGTTCATTGTAACA	55	60	AF329149
CMS32	ACGGACAAGAACTGCTCATA ACAACCAATAAATCCCCATT	58	58	AF329146
CMS25	GATCCTCCTGCGTTCTTATT CTAGCCTTTGATTGGAGCAT	58	58	AF380345
CMS18	GAACGACCCTTGAAGACGAA AGCAGCTGGTTTTAGGTCCA	58	58	AF329148

CMS17	TATAAAGGATCACTGCCTTC AAAATGAACCTCCATAAAGTTAG	55	58	AF329147
CMS15	AAATACTTAAAGGTTCCCAGA TTGTAAACTAAAGCCAGAAAG		58	AF329151
CMS13	TAGCCTGACTCTATCCATTTCTC ATTATTTGGAATTCAACTGTAAGG	55	58	AF329158
CMS121	CAAGAGAACTGGTGAGGATTTTC AGTTGATAAAAATACAGCTGGAAAG		60	AF329159
CMS9	TGCTTTAGACGACTTTTACTTTAC ATTTCACTTTCTTCATACTTGTGAT	55	60	AF329160
LCA66	GTGCAGCGTCCAAATAGTCA CCAGCATCGTCCAGTATTCA		60	AF091125

* FAO-ISAG recommendations. ** Annealing temperature used in the current study

Annex 2. Composition of the Multiplexes

Multiplex	**Markers**
Multiplex I	CMS121
	CVRL1D
	CVRL4D
	CVRL5D
	CVRL6D
	LCA66
Multiplex II	CMS50
	CMS9
	CVRL07
	CVRL08
	YWLL44
	YWLL59
	VOLP10
Multiplex III	CMS13
	CMS15
	CMS17
	CMS18
	CMS25
	CMS32
	VOLP32

Annex 3. Concentration of the different reagents in the mtDNA PCR

Reagent	**Final concentration**
Buffer	1X
$MgSO_4$	1.5 mM
dNTPs	0.2 mM
Primers (F and R)	0.5 µM
Taq polymerase (Sigma)	1 U
DNA	50 ng

Annex 4. Allelic frequencies for all studied loci

LCA66	CMS121	CVRL8	CVRL7	CVRL6D	CVRL5D	CVRL4	CVRL1D	CMS50	CMS32	CMS25	CMS18	CMS15	CMS13	CMS9	
0.345	0.366	0.737	0.294	0.776	0.395	0.488	0.355	0.243	0.518	0.453	0.79	0.368	0.29	0.377	**A**
0.327	0.302	0.231	0.252	0.129	0.348	0.354	0.192	0.213	0.237	0.383	0.156	0.275	0.273	0.343	**B**
0.16	0.122	0.012	0.188	0.094	0.122	0.122	0.102	0.158	0.207	0.104	0.052	0.175	0.22	0.156	**C**
0.113	0.075	0.012	0.141		0.069	0.029	0.084	0.152	0.036	0.046		0.075	0.104	0.081	**D**
0.023	0.052	0.006	0.035		0.029	0.005	0.066	0.097		0.011		0.056	0.087	0.029	**E**
0.011	0.023		0.035		0.011		0.06	0.067				0.018	0.011	0.005	**F**
0.011	0.017		0.029		0.005		0.018	0.054				0.018	0.005	0.005	**G**
0.005	0.017		0.011		0.005		0.018	0.012				0.012	0.005		**H**
	0.011		0.005		0.005		0.018								**I**
	0.005		0.005		0.005		0.018								**J**
	0.005						0.018								**K**
							0.012								**L**
							0.012								**M**
							0.006								**N**
															O
							0.006								**P**
							0.006								**Q**

YWLL59	YWLL44	VOLP32	VOLP10
0.526	0.61	0.879	0.451
0.427	0.133	0.12	0.138
0.032	0.075		0.132
0.006	0.058		0.126
0.006	0.052		0.108
	0.023		0.036
	0.005		0.006
	0.005		
	0.005		
	0.005		

Annex 5. Estimation of observed (H_0), non biased (Hnb) and expected heterozygosity, and fixation index (F_{IS}) by locus, arranged into two geographic regions; South East and South West of Tunisia

Locus	**South-East**				**South-West**			
	H_0	**Hnb**	**He**	**F_{IS}**	**H_0**	**Hnb**	**He**	**F_{IS}**
CMS9	0.735	0.706	0.699	-0.014	0.542	0.695	0.685	0.134
CMS13	0.717	0.772	0.764	0.076	0.714	0.78	0.769	0.052
CMS15	0.571	0.755	0.748	0.126	0.411	0.76	0.749	0.32
CMS18	0.32	0.333	0.329	0.027	0.457	0.414	0.408	0.229
CMS25	0.471	0.635	0.629	0.13	0.571	0.616	0.607	0.229
CMS32	0.568	0.597	0.591	0.008	0.484	0.669	0.659	0.237
CMS50	0.612	0.847	0.839	0.177	0.657	0.836	0.824	0.131
CVRL1D	0.64	0.796	0.788	0.141	0.636	0.821	0.809	0.253
CVRL4D	0.634	0.633	0.627	-0.026	0.514	0.607	0.598	0.117
CVRL5D	0.73	0.719	0.712	-0.002	0.828	0.685	0.675	-0.081
CVRL6D	0.226	0.32	0.317	0.192	0.382	0.461	0.454	0.189
CVRL7	0.384	0.827	0.819	0.382	0.285	0.765	0.754	0.531
CVRL8	0.215	0.431	0.427	0.386	0.382	0.387	0.381	0.013
CMS121	0.711	0.733	0.726	-0.014	0.823	0.788	0.776	-0.015
LCA66	0.576	0.731	0.724	0.126	0.636	0.719	0.708	0.021
VOLP10	0.52	0.744	0.736	0.336	0.342	0.718	0.708	0.46
VOLP32	0.176	0.253	0.25	0.31	0.147	0.138	0.136	-0.065
YWLL44	0.47	0.609	0.603	0.174	0.47	0.568	0.560	0.009
YWLL59	0.288	0.555	0.549	0.295	0.266	0.535	0.526	0.751
Mean	0.503	0.631	0.625	0.20	0.503	0.63	0.62	0.18
S. D	0.185	0.178	0.176	0.06	0.184	0.178	0.176	0.08

S.D: Standard Deviation

Annex 6. Results from two runs of simulations of population size evolution, following 4 different scenarios; population decline (Dec) and expansion (Exp).

Run 1: Gen=5	**Dec 3-5 (log)**	**Dec 3-5**	**Dec 2-4 (log)**	**Dec 2-4**	**Exp 5-3 (log)**	**Exp 5-3**	**Exp 4-2 (log)**	**Exp 4-2**
N_0	2.884	766	2.852	712	2.875	750	2.808	643
0.025 HPD	2.311	205	2.260	182	2.305	202	2.227	169
0.975 HPD	3.433	2711	3.425	2665	3.435	2723	3.369	2341
N_1	4.346	22202	4.293	19671	4.334	21610	4.272	18744
0.025 HPD	3.769	5881	3.705	5080	3.754	5687	3.683	4820
0.975 HPD	4.925	84173	4.868	73812	4.905	80398	4.848	70475
Time	3.922	8366	3.900	7951	3.911	8157	3.821	6626
0.025 HPD	2.734	543	3.239	1734	3.284	1924	3.187	1538.4
0.975 HPD	4.606	40384	4.536	34417	4.533	34153	4.451	28299

Run 2: Gen=4	**Dec 3-5 (log)**	**Dec 3-5**	**Dec 2-4 (log)**	**Dec 2-4**	**Exp 5-3 (log)**	**Exp 5-3**	**Exp 4-2 (log)**	**Exp 4-2**
N_0	2.847	704	2.867	738	2.884	767	2.878	756
0.025 HPD	2.272	187	2.304	202	2.323	211	2.287	194
0.975 HPD	3.416	2608	3.439	2749	3.437	2738	3.460	2890
N_1	4.298	19866	4.337	21765	4.284	19247	4.322	21028
0.025 HPD	3.721	5265	3.749	5612	3.683	4828	3.751	5637
0.975 HPD	4.876	75263	4.913	81907	4.856	71909	4.892	78158
Time	3.779	6024	3.789	6152	3.853	7131	3.804	6381
0.025 HPD	3.131	1352	3.133	1359	3.239	1734	3.152	1422
0.975 HPD	4.421	26411	4.420	26361	4.484	30508	4.443	27758

Gen : time of generation (in years), N_0: Final population size, N_1: Initial population size, HPD: Highest Probability Densit

Annex 7. Mutations detected in the Control Region sequences

Region	Mutation	Number of samples	Frequency
D-loop	G ↔ A	2	0.032
	G ↔ C	2	0.032
	G ↔ A	1	0.016
	G ↔ A	1	0.016
Cytochrome b	A ↔ G	3	0.045
	G ↔ A	1	0.014
	T ↔ C	3	0.045
	T ↔ C	1	0.014
	G ↔ A	3	0.045
	A ↔ T	1	0.014
	G ↔ A	3	0.045
	T ↔ C	1	0.014
	C ↔ T	3	0.045
	T ↔ C	3	0.045
	A ↔ G	1	0.014
	G ↔ A	2	0.031