e-ISSN 1734-9168
http://www.isez.pan.krakow.pl/en/folia-biologica.html

Folia Biologica (Kraków), vol. 72 (2024), No 2
https://doi.org/10.3409/fb_72-2.07

Insight into the genetic diversity of Selle Français horse populations based on whole D-loop sequencing of mitochondrial DNA

Katarzyna Ropka-Molik*, Abdelhanine Ayad*, Adriana D. Musiał, Omar Besseboua, Sofiane Aissanou, Monika Stefaniuk-Szmukier, and Katarzyna Piórkowska

Accepted June 07, 2024 *Published online June 25, 2024* *Issue online June 28, 2024*

Original article

Ropka-Molik K., Ayad A., Musiał A.D., Besseboua O., Aissanou S., Stefaniuk-Szmukier M., Piórkowska K. 2024. Insight into the genetic diversity of Selle Français horse populations based on whole D-loop sequencing of mitochondrial DNA Folia Biologica (Kraków) **72**: 65-73.

The Selle Français (SF) breed originated from native France horses crossed with Thoroughbreds, Arabs, Anglo-Arabs and French Trotters, which allowed for individuals to be obtained with excellent predispositions and a versatile usage for equestrian sports. The performance of the SF horse is widely accepted in the fields of show jumping and eventing; however, they have successfully started to compete in other areas such as dressage, vaulting, carriage driving and TREC competitions. To date, information about the genetic structure of the SF breed has been very limited. In the present report, Sanger sequencing was applied to analyse the whole D-loop of mitochondrial DNA of 98 samples representing Selle Français horses originating from two populations: Algerian (n=47) and French (n=51). The result obtained showed the presence of 19 SNPs, which allowed us to divide the samples into 41 haplotypes (OR909723-OR909773). A higher number of haplotypes was detected in the Algerian population (n=24) compared to the French population (n=14), but both populations did not represent unique clusters and showed a high degree of genetic similarities. The result showed that the identified haplotypes of the SF belonged to four clades: L, A, B and G. The present report gives strong evidence that the Selle Français breed has been established based on a great influence of the Thoroughbred, whose close connection is more visible in the French population. The haplogroup assignment to L clades proved that the SF breed has European origin as ancestral haplogroup L was assigned to West and South Europe while A to West Asia and G and B to Middle East. Our results show SF as an example of a contemporary breed of sport horses with European roots further influenced by local breeds.

Key words: mtDNA, SNPs, haplotypes, haplogroups, Sanger sequencing, horses

Katarzyna Ropka-Molik✉, Adriana D. Musiał, Monika Stefaniuk-Szmukier, Katarzyna Piórkowska, Department of Animal Molecular Biology, National Research Institute of Animal Production, Balice, Poland. E-mail: katarzyna.ropka@iz.edu.pl
Abdelhanine Ayad, Sofiane Aissanou, Department of Environment and Biological Sciences, Faculty of Nature and Life Sciences, University of Bejaia, 06000 Bejaia, Algeria.
Omar Besseboua, Department of Agronomic and Biotechnological Sciences, Faculty of Nature and Life Sciences, University H. Benbouali, 02000, Chlef, Algeria.

*These authors have contributed equally to this work.

Warmblood horses are bred mostly for competition in Olympic disciplines such as dressage, showjumping and eventing. The breeding goal for major modern warmblood horse organisations is to produce healthy high-performance animals that are able to compete at the Grand Prix level. Several locally-raised breeds have been refined by a strict selection of performance characteristics, in response to the demands of sports competition.

The first mention of the Selle Français as a breed origin come from the 19th century when in northern France, some native mares were bred with the Norfolk Trotter or to an advantage with Thoroughbred stallions (Collectif 2006). Until the middle of the 20th century, these half-blood French horses were used for light fieldwork and as saddle horses, and were maintained throughout the country. An important moment in SF breeding – the establishment of the

French Saddle Horse studbook in 1958 – began the period of selection towards refining the breed's sport abilities. The breed's performance is now widely accepted in show jumping and eventing ; however, these horses are successfully starting to compete in other events such as dressage, vaulting, carriage driving and TREC competitions (Techniques de Randonnée Équestre de Compétition; www.sellefraincaise.fr). The crossing with Thoroughbreds, Arabs, Anglo-Arabs and French Trotters allows for individuals to be obtained with excellent predispositions and a versatile usage in equestrian sports (L'Agriculture 1966). Since the 1995 Studbook, limitations on the entry of individuals with at least one SF parent have been introduced (Dubois & Ricard 2007). The SF horses originated from native France horses, but this breed is also valued and maintained in other countries including Brazil, Argentina, Morocco and Algeria.

Mitochondrial DNA (mtDNA) has become the marker of choice for improving the phylogeny robustness in genetic characterisation studies (Achilli *et al.* 2012). The mtDNA has a strictly maternal inheritance (Hutchison *et al.* 1974), which means that mtDNA haplotypes should be shared by all individuals within a maternal family line. The mtDNA is useful for studying the evolution of closely-related breeds, and many studies have focused on the mitochondrial D-loop region – the most variable part of mtDNA (Ishida *et al.* 1994) – due to a higher substitution rate than in the rest of the mtDNA genome (Cann *et al.* 1984). Furthermore, molecular studies have used the mtDNA nucleotide sequence for the assessment of horse diversity (Kusza *et al*. 2013; Sziszkosz *et al*. 2016; Yang *et al.* 2018; Csizmár *et al.* 2018; Effa *et al.* 2021; Atsenova *et al.* 2022) and for conservation purposes (Dell *et al.* 2020; Liu *et al.* 2021).

To date, very few studies have been carried out on the genetic diversity within the SF horse breed. The genetic relationships among SF horse populations have not been investigated using the mtDNA analysis method. Moreover, most of the studies in this field have used a small part of the hypervariable region; however, increasing the length of the investigated part affects the robustness the analysis (Achilli *et al.* 2012; Cardinali *et al.* 2016; Hudson 2017). In this report, we present the first molecular characterisation of SF horses based on a mtDNA sequence, to show the genetic diversity and origin of this breed and to verify the obtained results in terms of the historical origin of the breed. Based on sequences of the entire mtDNA control region, the presented study indicates the differences in the genetic diversity structure between two populations – French and Algerian. The presence of new mitochondrial haplotypes described in the SF horses compared to breeds involved in the founding of the SF breed can confirm the involvement of individual breeds (taking into account the geographical region). Moreover, the use of DNA variability can shed new light on the breeding practices used since the breed was recognised and their compliance with the breeding documentation.

Materials and Methods

This research was approved by the Scientific Council of the Faculty of Nature and Life Sciences (Report of the Faculty's Scientific Council #05 dated 11 November 2020, University of Bejaia, Algeria). Concerning the ethical aspects, the experimental procedure was performed according to good veterinary practice under farm conditions.

Sampling

A total of 98 samples from SF horses were collected in Europe (French population) (n=51) and in Algeria (n=47). Thirty hairs approx. 7 cm long were collected from the root of the neck or the tail skin. The obtained materials were preserved in an individually labelled paper envelopes and stored in dry and dark conditions at 20°C until the further analysis.

DNA isolation and genotyping using Sanger sequencing

The Sanger Sequencing method was applied to identify all SNPs within the whole hypervariable mtDNA region. mtDNA sequencing is a powerful and sensitive method that allows for the recognition of variability in terms of a dam line analysis.

Genomic DNA was extracted from all samples using the Sherlock AX kit (A&A Biotechnology, Poland), according to the protocol. The quality and concentration of the DNA was measured with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and about 150 ng of DNA was used for the PCR reaction.

For the detection of variants, primers covering the hypervariable regions (HVR1A; HVR1B; HVR2) of the D loop were designed using the Primer3Plus program (primer3plus.com), based on the mitochondrial reference genome GenBank NC_001640.1 as a reference (X79547.1; Xiufeng and Árnason 1994): the first amplicon for HVR1A F: AACGTTTCCTCCCAAGGACT and R: GTAGTTGGGAGGGTTGCTGA (397bp); the second amplion for HVR1B F: ACCCCATCCAAGTCAAATCA and R: CAGGTGCACTTGTTTCCTATG (461bp); and the third

amplicon for HVR2 F: ACCTACCCGCGCAGTAAGCAA and R: ACGGGGGAAGAAGGGTTGACA (304bp). The amplicons overlapped to fully cover the entire sequence of interest.

Furthermore, the primers were designed to avoid amplification of the NUMTs (Nuclear Mitochondrial DNA) (Nergadze *et al.* 2010). The PCR reaction was performed with an AmpliTaq Gold 360 Master Mix (Life Technologies California, USA) and an annealing temperature of 56°C, using the Veriti Thermal Cycler (Life Technologies California, USA). The PCR reactions were performed according to the protocol and the PCR profile was as follows: 95°C – 10 min; (95°C–30 s, 56°C–30 s; 72°C–30 s) for 35 cycles; and 72°C – 7 min. The PCR mixture included: 5 µl of 2x AmpliTaq Gold 360 Master Mix; 1.2 µl of 360 GC Enhance; 0.5 µl of 10 pmol primers mix; 3.3 µl nuclease-free water; and about 150 ng of isolated DNA. The obtained PCR product was purified from free nucleotides and primers using EPPiC (A&A Biotechnology) and was then prepared for sequencing using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Life Technologies California, USA). The reactions were performed according to the protocol and the PCR profile was as follows: 96°C – 1 min; (96°C for 10 s, 50°C for 5 s; 60°C for 4 min) for 25 cycles (Veriti Thermal Cycler; Life Technologies California, USA). The reaction mixture included 4 µl of BigDye™ Terminator 3.1 Ready Reaction Mix; 1 µl of 10 pmol primer; 4 µl of nuclease-free water; and 1 µl of purified PCR product. Next, the products were purified using the BigDye XTerminator Purification Kit (ThermoFisher Scientific). According to the protocol, sequencing was performed using a Genetic Analyser 3500x1 capillary sequencer and was analysed using GeneMapper® Software (Applied Biosystems).

To check the quality of the sequencing, the obtained reads were compared using the BLAST algorithm (https://blast.ncbi.nlm.nih.gov). The sequences were read using FinchTV 1.3.0 (Geospiza Inc. 2004-2005) and were compared with a horse mitochondrial reference genome (GenBank NC_001640.1; X79547.1) using MEGA software.

Statistical analysis of the genetic diversity and phylogenetic reconstruction

As a background of the breeds potentially related to SF, the following sequences from GenBank were used: Thoroughbreds (KT221835.1; KT221840.1; KT221841.1; KT221833.1; KT221843.1; KT221830.1) and Arabian (JN398448.1). Moreover, the samples of different haplotypes of Arabian horses from Algeria (n=5; PP795368-PP795372) and Europe (n=5; PP795373-PP795377), and Barb horses (n=5; PP795363-PP795367) from Algeria were sequenced according to the procedure described in the section above. These samples were also used during clustering to show the breeds potentially involved in the SF breed's establishment.

The SNPs found in two populations of SF horses were analysed using phylogenetic methods. A neighbour-joining tree demonstrating the diversity of the studied mitochondrial DNA segments was created with Mega 11 software (Tamura *et al.* 2021) and iTOL 6.8.1 (Letunic & Bork 2021), with an implemented Maximum Likelihood. To visualise the connections between all the SF mitochondrial DNA haplotypes, the median-joining network (Bandelt *et al.* 1999) was constructed in the PopART phylogenetic software (http://www.popart.otago.ac.nz, accessed January 2024).

Results and Discussion

In the present report, the Sanger sequencing of all the 98 samples representing Selle Français horses showed the presence of 19 SNPs, which allowed us to divide the samples into 41 haplotypes (Table 1). The new SF haplotypes were submitted to GenBank and received the accession numbers OR909723-OR909773 (Table 1). The distribution of the number of horses in individual haplotypes indicates the overrepresentation of some of them (from 4 to 6 horses per haplotype) (Table 1). A higher number of haplotypes was detected in the Algerian population (n=24) compared to the French population (n=14)(Table 1) (Figure 1). What is interesting is that only 3 haplotypes (OR909730; OR909749; OR909763) were common for both populations, and represented a total of 11 individuals. The neighbour-joining clustering method did not show a unique block for any of the populations. We observed partial clustering within the populations in separated blocks, as well as the mixed clustering for the rest of the horse's origins in both Algeria and France (Figure 1).

In the nineteenth century, with the increased popularity in the breeding of Thoroughbred (TB) horses, the first breed selected only for racing, and the use of the TB in cross-breeding, local mares were the basis of most saddle horse breeding in Europe, with the Selle Français (SF) being no exception (Langlois and Hernu 2003). In the stages of creating the breed, half-blood horses played a key role, in particular: Norman half-bloods (area of Caen), Vendée half-bloods (area of La Roche sur Yon) and Centre half-bloods (area of Cluny). In 1958, the three

Table 1

SNPs and haplotypes identified in both the Algerian and French populations (HG – haplogroups according to Achilli *et al*. 2021)

Haplotype (GenBank acc. number)	HG	No. of horses	Population Alg./Fr.	15495	15496	15534	15585	15602	15603	15604	15649	15650	15666	15703	15720	15870	15871	15960	15975	16543	16557_16558in	16629
Ref: C_001640.1				T	A	C	G	C	T	G	A	A	G	T	G	C	C	C	T	T	–	A
OR909724	A	6	6 / 0	C	G	T	·	T	C	·	G	·	·	·	A	·	·	·	·	·	·	·
OR909728	L	6	6 / 0	C	G	T	·	T	C	·	G	·	·	·	A	T	T	T	·	·	C	G
OR909755	L	5	0 / 5	C	G	T	A	T	C	A	G	·	·	·	A	T	T	T	·	·	C	G
OR909756	A	5	0 / 5	C	·	·	A	T	·	·	·	G	·	·	A	T	T	T	·	·	·	·
OR909763	A/B	5	2 / 3	C	·	·	·	·	·	·	·	G	A	·	A	·	·	·	C	·	·	·
OR909725	A	4	4 / 0	C	G	T	A	T	C	A	G	·	·	·	A	·	·	·	·	·	·	·
OR909731	A/L	4	4 / 0	C	G	T	A	T	C	A	G	·	·	·	A	·	·	·	·	·	C	·
OR909734	A	4	4 / 0	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	C	·	C	·
OR909754	A	4	0 / 4	C	·	·	A	T	·	·	·	G	·	·	A	·	·	·	C	·	·	·
OR909760	L	4	0 / 4	C	G	T	A	·	C	·	G	·	·	·	A	T	T	T	·	·	C	G
OR909730	A/B	3	2 / 1	C	·	·	·	·	·	·	·	G	A	·	A	·	·	·	C	·	C	·
OR909749	A	3	2 / 1	C	·	·	A	·	·	·	·	G	A	·	A	·	·	·	C	·	·	·
OR909765	A/L	3	0 / 3	C	G	T	·	T	C	·	G	·	·	·	A	·	T	T	·	·	C	G
OR909723	A	2	2 / 0	C	·	·	A	T	·	·	·	G	·	·	A	T	·	·	C	·	·	·
OR909726	A	2	2 / 0	C	G	T	A	T	C	·	G	·	·	·	A	·	·	·	·	·	·	·
OR909727	A	2	2 / 0	C	·	·	A	T	·	·	·	G	·	·	A	·	·	·	·	·	·	·
OR909729	L	2	2 / 0	C	G	T	A	T	C	·	G	·	·	·	A	T	T	T	·	·	C	G
OR909732	A	2	2 / 0	C	G	T	·	T	C	·	G	·	·	·	A	·	·	·	·	·	C	·
OR909733	A	2	2 / 0	C	·	·	·	·	·	·	·	·	·	·	·	·	·	·	C	·	C	·
OR909735	A	2	2 / 0	C	·	·	·	T	·	A	·	·	·	·	A	·	·	·	·	·	C	·
OR909736	A	2	2 / 0	C	G	T	A	·	C	·	G	·	·	·	A	·	·	·	·	·	C	·
OR909739	A	2	2 / 0	C	·	·	·	·	·	·	·	G	·	·	·	·	·	·	C	·	C	·
OR909747	A	2	0 / 2	·	·	·	·	·	·	·	·	R	·	·	·	·	·	·	C	·	·	·
OR909753	L	2	0 / 2	C	G	T	·	T	C	A	G	·	·	·	A	T	T	T	·	·	C	G
OR909757	G	2	0 / 2	C	·	·	·	T	·	·	·	G	A	C	A	T	T	·	C	·	·	·
OR909766	A	2	0 / 2	C	·	·	A	T	·	·	·	G	A	·	A	T	·	·	·	·	·	·
OR909767	A	2	0 / 2	C	·	·	A	T	·	·	·	G	·	·	A	·	T	T	·	·	·	·
OR909746	A	1	0 / 1	·	·	·	·	·	·	·	·	G	·	·	·	·	·	·	C	·	·	·
OR909750	A	1	0 / 1	C	·	·	A	T	·	·	·	G	·	·	A	·	·	·	C	A	·	G
OR909752	A	1	0 / 1	C	·	·	·	T	·	·	·	G	·	·	A	·	·	·	C	·	·	·
OR909758	A	1	0 / 1	C	·	·	·	T	·	·	·	G	·	·	A	T	·	T	C	·	·	·
OR909759	A	1	0 / 1	C	·	·	·	·	·	·	·	R	·	·	·	·	·	·	C	·	·	·
OR909761	A	1	0 / 1	C	·	·	·	T	·	·	·	G	·	·	A	T	T	T	·	·	·	·
OR909762	A	1	0 / 1	C	·	·	A	T	·	·	·	R	·	·	A	T	T	T	·	A	·	G
OR909764	A	1	0 / 1	C	·	·	A	T	·	·	·	G	A	·	A	·	·	·	C	·	·	·
OR909768	A	1	0 / 1	C	·	·	·	T	·	·	·	·	·	·	A	·	·	·	C	A	·	G
OR909769	A	1	0 / 1	C	·	·	·	T	·	·	·	·	A	C	A	T	·	·	C	·	·	·
OR909770	A	1	0 / 1	C	·	·	·	T	·	·	·	·	·	·	A	T	·	T	C	A	·	G
OR909771	A	1	0 / 1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	C	·	·	·
OR909772	A	1	0 / 1	C	·	·	·	T	·	·	·	G	A	C	A	T	·	·	C	·	·	·
OR909773	A	1	0 / 1	C	·	·	A	T	·	·	·	G	A	C	A	T	·	·	C	·	·	·

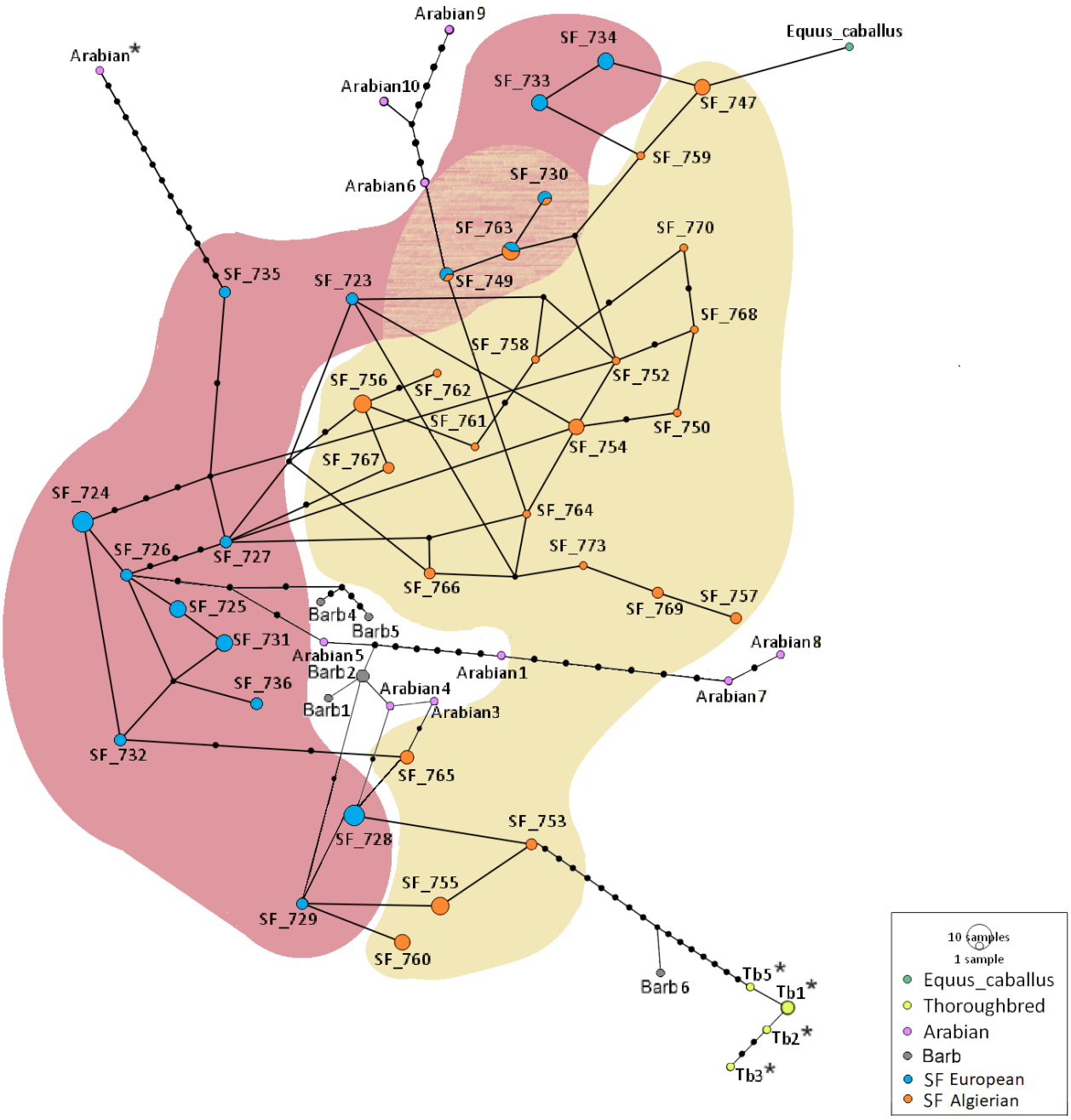

Fig. 1. Median-joining network of the identified SF haplotypes with the Thoroughbred (KT221835.1; KT221840.1; KT221841.1; KT221833.1; KT221843.1; KT221830.1), Arabian (JN398448.1), Arabian horses from Algeria (2-6) and Europe (1, 7-10) and Barb horses (1-6), and *Equus caballus* (NC_001640). The yellow background includes haplotypes represented by the French population, while the red background includes haplotypes represented by the Algerian population. The size of the nodes corresponds to the number of samples, while the strokes on the branches correspond to the number of polymorphisms. The network was constructed with the PopART phylogenetic software.

half-bred clades were grouped, together with the half-bred Anglo-Arabs of the South-West, under the name Selle-Français. This led to the classic French breed's entry into the highest level of sports competition as the Selle Français (SF), also known as the Cheval de Selle Français, meaning 'French Saddle Horse' (www.fei.org). The studbook of SF horses established in 1958 is still open and allows for horses with at least one SF parent to be admitted (Dubois & Ricard 2007).

To date, information about the genetic structure of the SF breed has been very limited. Due to the open stud book, there is a requirement to monitor the genetic diversity and the involvement of other breeds, as well as to identify the founders' origins. In turn, the usage of mtDNA variability can help to identify

the dam line structure in Selle Français populations coming from different geographical regions, which was the aim of the present study. A greater diversity was observed in the Algerian population for which 27 haplotypes were detected (24 were unique only to this population). As we show in Figure 1, some of the SF horses from the Algerian population were clustered closely to the Arabian mtDNA sequence, which may indicate the predominance of Arabian horse crosses in this region. Moreover, the analysis of the additional samples representing Barb horses (Algerian origin) and Arabian horses from Algeria and Europe shows that the SF horses from the Algeria region clustered in the vicinity of the Arabian and Barb horses from this exact region (Arabian 2-6) (Figure 1). Part of the Selle Français horses from two populations were grouped close to the Arabian horses from Europe (Arabian horses 7-10), which can confirm the common origin and the proximity of both populations. In turn, part of the French population (as well as two haplotypes from Algeria – SDF_728 and SF_729) were clustered near the Thoroughbred, which follows the historic origin of the SF breed and can be explained by general historical trends in the TB's expansion within Europe (Yoon *et al.* 2018) and a pedigree data analysis performed for the French population (Pirault *et al.* 2013). The obtained results also confirm the share of the same TB crossing with the Algerian population. In a study concerning the genetic structure of horses representing French populations based on genealogical data, a predominant share of TB horses was indicated in the Selle Français origins. The authors estimated that the TB constituted 46.7% of the SF founder origins, while in total, French founders for the SF breed accounted for 92.9% (Pirault *et al.* 2013). Moreover, French Trotters, Arabs and Anglo-Arabs accounted for 4.1%, 2.5% and 2.7%, respectively of the founder origins for the French population of SF horses (Pirault *et al.* 2013). The participation of other breeds and saddle horses in the SF origins was estimated to be 11.2% (Pirault *et al.* 2013).

The clustering of the French population of Selle Français horses in the vicinity of TB horses based on the mtDNA sequence can be supported by previous reports using microsatellite loci (Leroy *et al.* 2009). STR profiling was applied to investigate 34 populations of French horse breeds and showed a 98% similarity between SF and French TB horses. Moreover, the SF was clustered within the warm-blooded breed group, and after pair testing, this breed showed no significant differences between TB and Anglo-Arab breeds (Leroy *et al.* 2009). The study by Kusza *et al.* (2013) indicated that mtDNA results combined with genome variability estimated by STR loci polymorphisms can be an excellent tool for showing the genetic structure of horse breeds. At the next stage of research, the SF mtDNA data obtained in the present study should be used together with STR profiling as a comprehensive approach, to give a full view of the genetic diversity of geographically-separated populations as well as the whole breed.

The profiling of the whole D-loop of mtDNA of the Selle Français horses indicated a strong relationship between both of the analysed populations. The Algerian and French populations did not represent unique clusters and showed a high degree of genetic similarities, which may have resulted from the common geographical origin of both populations. As one of the analyses performed, the obtained mtDNA SF haplotypes were compared to the generally known and recognised haplotypes according to the Achilli *et al.* (2012) nomenclature and classification. The result indicated that the identified haplotypes of the SF breed belong to four clades: L (JN398422.1, JN398423.1, JN398424.1, JN398425.1, JN398427.1); A (JN398377.2, JN398378.1, JN398384.1, JN398385.1); B (JN398391.1); and G (JN398412.1) (Figure 1). A majority of the identified horses represented haplogroups A and L (Figure 2), with no apparent significant difference in this share between populations (Table 1). The haplogroup G was detected in only two horses representing the French population.

Due to the geographic distribution of mtDNA haplogroups established by (Ma *et al.* 2018), the L haplogroup is mainly characteristic of European horses, B describes the horses originating from West Asia, while the A and G haplogroups occur in the highest frequencies in the East Asia region. In turn, according to Achilli *et al.* (2012), both the G and B haplogroups – representing the exact JN398412.1 and JN398391.1 reference sequences, respectively – were assigned to Arabian horses and to an unspecified Syrian Breed from the Middle East. Moreover, the identified haplotypes assigned to the L haplogroups (JN398425.1, JN398427.1) were detected previously in some horses from South Europe, such as Italian the Breed and Maremmano (Achilli *et al.* 2012), which can indicate the involvement of the above in the SF breed creation. The comprehensive study of the dam line origins of the TB breed showed that the founder haplotypes belong to, among others, the L, A and G haplogroups (Hill *et al.* 2002). Moreover, an analogous approach used for an Arabian maternal phylogenetic analysis showed the presence of all the haplogroups observed in SF horses – A, L and G (Khanshour & Cothran 2013). Such an assignment can also confirm the predominant role of both breeds in the Selle Français formation and indicates multiple origins of the founder mares.

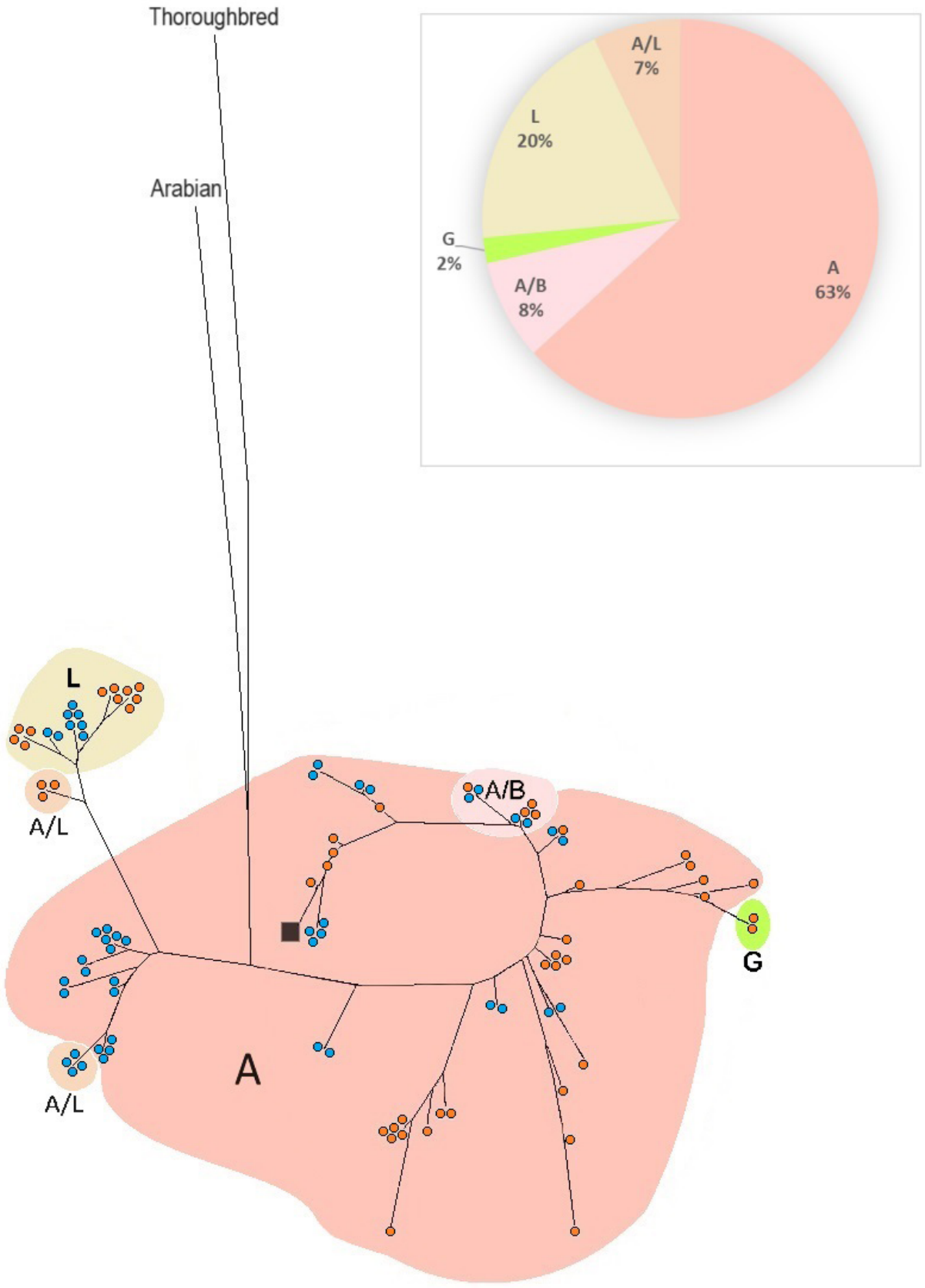

Fig. 2. The neighbour-joining radiation tree of 98 SF mtDNA sequences belonging to two populations – French (orange dots) and Algerian (blue dots) – with reference sequences of the Arabian (JN398448.1), Thoroughbred (KT221835.1) and *Equus caballus* (NC_001640; marked as a black square). The yellow background includes individuals belonging to the L haplotype, according to Achilli *et al*. (2012); the red background includes individuals belonging to the A haplotype; the green – G; the orange – A/L; and pink – A/B. The tree was created with Mega 11 software.

Conclusion

The present report provides strong evidence that the Selle Français breed has been established based on a great influence of the Thoroughbred, whose close connection is more visible in the French population. A greater involvement of Arabian horses and Barb horses originating from Algeria was confirmed for the Algerian SF population. The haplogroup assignment L clades proved that the SF breed has a European origin, as the ancestral haplogroup L was assigned to West and South Europe while A was assigned to Western Asia and G and B to the Middle East (Ning *et al.* 2019). Our results show that SF is an example of a contemporary breed of sports horse with European roots further influenced by local breeds. This information on the genetic diversity of Selle Français populations may be essential to optimise strategies for the development of horse breeding aimed at improving desirable phenotypic traits, maintaining the desired genetic diversity and monitoring the involvement of other breeds.

A limitation of the study is the lack of a Trotter horse mtDNA sequence, as this breed was used as one of the Selle Français breed founders. In the future, the results concerning mtDNA should be extended with microsatellite data to give a full view of the genetic diversity at the genomic DNA level.

Conflict of Interest

The authors declare no Conflict of Interest in this article.

Funding

The present study was supported by the statutory activity of the National Research Institute of Animal Production, Poland (No. 01-18-08-11).

Author Contributions

Research concept and design: K.R.-M., A.A.; Collection and/or assembly of data: A.D.M., O.B., S.A., M.S.-S., K.P.; Data analysis and interpretation: K.R.-M., A.A.; Writing the article: K.R.-M., A.A.; Critical revision of the article: K.R.-M., A.A., A.D.M., O.B., S.A., M.S.-S., K.P.; Final approval of article: A.D.M., O.B., S.A., M.S.-S., K.P.

References

Achilli A., Olivieri A., Soares P., Lancioni H., Kashani B.H., Perego U.A., Nergadze S.G., Carossa V., Santagostino M., Capomaccio S., Felicetti M., Al-Achkar W., Penedo M.C.T., Verini-Supplizi A., Houshmand M., Woodward S.R., Semino O., Silvestrelli M., Giulotto E., Pereira L., Bandelt H.J., Torroni A. 2012. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. Proc. Natl. Acad. Sci. U.S.A. **109**: 2449-2454. https://doi.org/10.1073/pnas.1111637109

Atsenova N., Palova N., Mehandjyiski I., Neov B., Radoslavov G., Hristov P. 2022. The Sequence Analysis of Mitochondrial DNA Revealed Some Major Centers of Horse Domestications: The Archaeologist's Cut. J. Equine Vet. Sci. **109**: 103830. https://doi.org/10.1016/j.jevs.2021.103830

Bandelt H.J., Forster P., Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. **16**: 37-48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Cann R.L., Brown W.M., Wilson A.C. 1984. Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics **106**: 479-499. https://doi.org/10.1093/genetics/106.3.479

Cardinali I., Lancioni H., Giontella A., Capodiferro M.R., Capomaccio S., Buttazzoni L., Biggio G.P., Cherchi R., Albertini E., Olivieri A., Cappelli K., Achilli A., Silvestrelli M. 2016. An Overview of Ten Italian Horse Breeds through Mitochondrial DNA. PLoS One **11**: e0153004. https://doi.org/10.1371/journal.pone.0153004

Collectif. 2006. Les races de chevaux et poneys. Éditions Artémis, 127 pp.

Csizmár N., Mihók S., Jávor A., Kusza S. 2018. Genetic analysis of the Hungarian draft horse population using partial mitochondrial DNA D-loop sequencing. PeerJ **6**: e4198. https://doi.org/10.7717/peerj.4198

Dell A.C., Curry M.C., Yarnell K.M., Starbuck G.R., Wilson P.B. 2020. Mitochondrial D-loop sequence variation and maternal lineage in the endangered Cleveland Bay horse. PLoS One **15**: e0243247. https://doi.org/10.1371/journal.pone.0243247

Dubois C., Ricard A. 2007. Efficiency of past selection of the French Sport Horse: Selle Français breed and suggestions for the future. Livest. Sci. **112**: 161-171. https://doi.org/10.1016/j.livsci.2007.02.008

Effa K., Rosenbom S., Han J., Dessie T., Beja-Pereira A. 2021. Genetic Diversities and Historical Dynamics of Native Ethiopian Horse Populations (*Equus caballus*) Inferred from Mitochondrial DNA Polymorphisms. Genes **12**: 155. https://doi.org/10.3390/genes12020155

Hill E.W., Bradley D.G., Al-Barody M., Ertugrul O., Splan R.K., Zakharov I., Cunningham E. P. 2002. History and integrity of thoroughbred dam lines revealed in equine mtDNA variation. Anim. Genet. **33**: 287-294. https://doi.org/10.1046/j.1365-2052.2002.00870.x

Hudson W. 2017. Whole-loop mitochondrial DNA D-loop sequence variability in Egyptian Arabian equine matrilines. PLoS One **12**: e0184309. https://doi.org/10.1371/journal.pone.0184309

Hutchison C.A., Newbold J.E., Potter S.S., Edgell M.H. 1974. Maternal inheritance of mammalian mitochondrial DNA. Nature **251**: 536-538. https://doi.org/10.1038/251536a0

Ishida N., Hasegawa T., Takeda K., Sakagami M., Onishi A., Inumaru S., Komatsu M., Mukoyama H. 1994. Polymorphic sequence in the D-loop region of equine mitochondrial DNA. Anim. Genet. **25**: 215-221. https://doi.org/10.1111/j.1365-2052.1994.tb00196.x

Khanshour A.M., Cothran E.G. 2013. Maternal phylogenetic relationships and genetic variation among Arabian horse populations using whole mitochondrial DNA D-loop sequencing. BMC Genet. **14**: 83. https://doi.org/10.1186/1471-2156-14-83

Kusza S., Priskin K., Ivankovic A., Jedrzejewska B., Podgorski T., Jávor A., Mihók S. 2013. Genetic characterization and population bottleneck in the Hucul horse based on microsatellite and mitochondrial data. Biol. J. Linn. Soc. **109**: 54-65. https://doi.org/10.1111/bij.12023

L'Agriculture M. de 1966. France,. Ministère de l'agriculture, Service de l'information et de la documentation generale. Rev. française l'agriculture.

Langlois B., Hernu V. 2003. An attempt to predict the earning status of a thoroughbred in France by genealogical data. Anim. Res. **52**: 79-85. https://doi.org/10.1051/animres:2003006

Leroy G., Callède L., Verrier E., Mériaux J. C., Ricard A., Danchin-Burge C., Rognon X. 2009. Erratum to: Genetic diversity of a large set of horse breeds raised in France assessed by microsatellite polymorphism. Genet. Sel. Evol. **41**: 31. https://doi.org/10.1186/1297-9686-41-31

Letunic I., Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. **49**: W293-W296. https://doi.org/10.1093/nar/gkab301

Liu S., Fu C., Yang Y., Zhang Y., Ma H., Xiong Z., Ling Y., Zhao C. 2021. Current genetic conservation of Chinese indigenous horses revealed with Y-chromosomal and mitochondrial DNA polymorphisms. G3 (Bethesda). **11**: jkab008. https://doi.org/10.1093/g3journal/jkab008

Ma H., Wu Y., Xiang H., Yang Y., Wang M., Zhao C., Wu C. 2018. Some maternal lineages of domestic horses may have origins in East Asia revealed with further evidence of mitochondrial genomes and HVR-1 sequences. PeerJ **6**: e4896. https://doi.org/10.7717/peerj.4896

Nergadze S.G., Lupotto M., Pellanda P., Santagostino M., Vitelli V., Giulotto E. 2010. Mitochondrial DNA insertions in the nuclear horse genome. Anim. Genet. **41**: 176-185. https://doi.org/10.1111/j.1365-2052.2010.02130.x

Ning T., Ling Y., Hu S., Ardalan A., Li J., Mitra B., Chaudhuri T. K., Guan W., Zhao Q., Ma Y., Savolainen P., Zhang Y. 2019. Local origin or external input: Modern horse origin in East Asia. BMC Evol. Biol. **19**: 217. https://doi.org/10.1186/s12862-019-1532-y

Pirault P., Danvy S., Verrier E., Leroy G. 2013. Genetic Structure and Gene Flows within Horses: A Genealogical Study at the French Population Scale. PLoS One **8**: e61544. https://doi.org/10.1371/journal.pone.0061544

Sziszkosz N., Mihók S., Jávor A., Kusza S. 2016. Genetic diversity of the Hungarian Gidran horse in two mitochondrial DNA markers. PeerJ **4**: e1894 https://doi.org/10.7717/peerj.1894

Tamura K., Stecher G., Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol. Biol. Evol. **38**: 3022-3027. https://doi.org/10.1093/molbev/msab120

Xiufeng X., Árnason Ú. 1994. The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. Gene **148**: 357-362. https://doi.org/10.1016/0378-1119(94)90713-7

Yang L., Kong X., Yang S., Dong X., Yang J., Gou X., Zhang H. 2018. Haplotype diversity in mitochondrial DNA reveals the multiple origins of Tibetan horse. PLoS One **13**: e0201564. https://doi.org/10.1371/journal.pone.0201564

Yoon S.H., Lee W., Ahn H., Caetano-Anolles K., Park K. Do, Kim H. 2018. Origin and spread of Thoroughbred racehorses inferred from complete mitochondrial genome sequences: Phylogenomic and Bayesian coalescent perspectives. PLoS One **13**: e0203917. https://doi.org/10.1371/journal.pone.0203917