ORIGINAL ARTICLE

New Insight into the human genetic diversity in North African populations by genotyping of SNPs in *DRD3*, *CSMD1* and *NRG1* genes

Souhir Mestiri[1,2] | **Sami Boussetta**[3] | **Andrew J. Pakstis**[4] | **Sarra El Kamel**[3] | **Amel Ben Ammar El Gaaied**[3] | **Kenneth K. Kidd**[4] | **Lotfi Cherni**[2,3]

[1]Laboratory of Genetics, Biodiversity and Bioresource Valorization (LR11ES41), University of Monastir, Monastir, Tunisia

[2]Higher Institute of Biotechnology of Monastir, Monastir University, Monastir, Tunisia

[3]Laboratory of Genetics, Immunology and Human Pathologies, Faculty of Sciences of Tunis, University of Tunis El Manar, Tunis, Tunisia

[4]Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA

Correspondence
Souhir Mestiri, Laboratory of Genetics, Biodiversity and Bioresource Valorization (LR11ES41), University of Monastir, Monastir 5000, Tunisia.
Email: mestiriss@hotmail.fr

Abstract

Background: The single nucleotide polymorphisms (SNPs) of the dopamine D3 receptor (*DRD3*), the CUB and sushi multiple domains 1 (*CSMD1*) and the neuregulin 1 (*NRG1*) genes were used to study the genetic diversity and affinity among North African populations and to examine their genetic relationships in worldwide populations.

Methods: The rs3773678, rs3732783 and rs6280 SNPs of the *DRD3* gene located on chromosome 3, the rs10108270 SNP of the *CSMD1* gene and the rs383632, rs385396 and rs1462906 SNPs of the *NRG1* gene located on chromosome 8 were analysed in 366 individuals from seven North African populations (Libya, Kairouan, Mehdia, Sousse, Kesra, Smar and Kerkennah).

Results: The low values of F_{ST} indicated that only 0.27%–1.65% of the genetic variability was due to the differences between the populations. The Kairouan population has the lowest average heterozygosity among the North African populations. Haplotypes composed of the ancestral alleles ACC and ACAT were more frequent in the Kairouan population than in other North African populations. The PCA and the haplotypic analysis showed that the genetic structure of populations in North Africa was closer to that of Europeans, Admixed Americans, South Asians and East Asians. However, analysis of the rs3732783 and rs6280 SNPs revealed that the CT microhaplotype was specific to the North African population.

Conclusions: The Kairouan population exhibited a relatively low rate of genetic variability. The North African population has undergone significant gene flow but also evolutionary forces that have made it genetically distinct from other populations.

KEYWORDS
CSMD1, *DRD3*, Haplotypic analysis, North Africa, *NRG1*

1 | INTRODUCTION

The dopamine D3 receptor (*DRD3*) is a member of the dopamine receptor G protein-coupled receptor family. It is located in the limbic areas of the brain (Sokoloff et al., 1992) and plays a role in congnitive and emotional functions (Bombin et al., 2008). The *DRD3* gene (OMIM accession number: *126451) encodes the DRD3 receptor, which is located on chromosome 3q13.3 (Le Coniat et al., 1991). The rs6280 and rs3732783 polymorphisms of the *DRD3* gene have been extensively studied to demonstrate genetic associations with neuropsychiatric disorders (Gassó et al., 2020; Yang et al., 2016; Zhang et al., 2011). The rs6280 (Gly9Ser) is a functional single nucleotide polymorphism (SNP), which corresponds to a cytosine (C) to thymine (T) substitution leading to a glycine (Gly) to serine (Ser) substitution in the extracellular N-terminal domain of the receptor (Utsunomiya et al., 2012). It has been shown that the presence of the glycine leads to a significant increase in dopamine-binding affinity and amplifies dopamine intracellular signalling (Jeanneteau et al., 2006). The rs6280 and rs3732783 polymorphisms are genetically close (27 basepairs). These two SNPs constitute a microhaplotype, which has been studied by Kidd et al. (2014). The rs3732783 (Ala17Ala) is a synonymous variant that has no effect on the resulting protein sequence product. Zainal Abidin et al. (2015) reported an association of rs3732783 polymorphism with the development of impulse control behaviour among Malaysian Parkinson's disease patients. The rs3773678 polymorphism of *DRD3* gene is an intronic polymorphism that has been associated with nicotine dependence in Chinese population (Wei et al., 2012).

The CUB and sushi multiple domains 1 (CSMD1) protein plays a very important role in the regulation of complement activation and inflammation in the developing central nervous system (Kraus et al., 2006). The CSMD1 protein is encoded by the *CSMD1* gene (OMIM accession number: *608397), which is located on chromosome 8p23 and contains 69 exons (Sun et al., 2001). Recently, it has been shown that the expression level of the *CSMD1* gene increases after antipsychotic treatment in patients with schizophrenia (Liu et al., 2019). Several polymorphisms of the *CSMD1* gene-like rs10503253 and rs2616984 have been associated with the onset of schizophrenia and multiple neurodevelopmental disorders (Athanasiu et al., 2017; Bocharova et al., 2017; Donohoe et al., 2013). However, no studies have been performed with the polymorphism rs10108270 of *CSMD1* gene. Neuregulin 1 (NRG1) is a multifunctional protein. It plays a fundamental role in the development of the peripheral nervous system and in the process of nerve repair (Bare et al., 2011; Ronchi et al., 2016).

The *NRG1* gene (OMIM accession number: *142445) is made up of 21 alternately spliced exons. Most of the exons are located in a 200-kb region at the 3′ end of the gene (Steinthorsdottir et al., 2004). Recently, it has been shown that the expression of the *NRG1* gene is associated with the clinical risk of psychosis (Jagannath et al., 2018). In 2002, Stefansson et al. (2002) identified for the first time *NRG1* as a susceptibility gene for schizophrenia in the Icelandic population. Several association studies have confirmed this finding in other populations (Fukui et al., 2006; Jagannath et al., 2017; Mostaid et al., 2016). However, other investigations have failed to show the association of *NRG1* with schizophrenia (Allen et al., 2008; Loh et al., 2013). These conflicting results may be due to the difference in allele frequencies between populations for *NRG1* and other genetic risk factors. Indeed, Gardner et al. (2006) analysed 13 SNPs of the *NRG1* gene, including rs385396 and rs1462906 in 39 populations worldwide. They suggested that the allele and haplotypic frequencies of these SNPs are extremely different between populations. They also showed a significant difference in the structure of the linkage disequilibrium between Europe and other continental groups such as America, Central/South Asia, East Asia, Middle East/North Africa, Oceanic and Sub-Saharan Africa. These results may be due to events in the history of populations from different geographic regions (Gardner et al., 2007).

Several other studies have studied the genetic diversity of the North African populations (Boussetta et al., 2019; Serra-Vidal et al., 2019). Recently, we have used the markers of the *DRD2/ANKK1* locus to show that the North African population structure is the result of a mixture of European, Asian and Sub-Saharan lineages (Mestiri et al., 2021). To confirm this result, we analysed the SNPs rs6280, rs3732783, rs3773678 of the *DRD3* gene, the SNP rs10108270 of the *CSMD1* gene and the SNPs rs385396, rs383632 and rs1462906 of the *NRG1* gene in seven North African populations. The allele and haplotypic frequencies of the North African populations were compared with those of other worldwide populations in order to place it in the global context and to better understand its evolutionary history.

2 | MATERIALS AND METHODS

2.1 | Study populations

The study was performed on seven North African populations (Figure 1) consisting of 362 healthy and unrelated individuals who gave their informed written consent for DNA genotyping.

FIGURE 1 Geographical location of the seven populations analysed in this study

The Smar sample includes 62 individuals. Smar is a Berber village in the south-east of Tunisia. The Berbers can be descendants of Capsian and Neolithic peoples (Murdock, 1959). The Berbers of the region of Smar have been living in Tunisia since the end of the 16th century.

The Kesra sample includes 44 individuals. Kesra is a Berber village in north-west of Tunisia. During the Arab invasion, the Berber population took refuge in this mountainous region (Ibn Khaldoun, 1968).

The Kerkennah sample is made up of 46 individuals. Kerkennah is a group of islands off the coast of southern Tunisia, where several past civilizations like the Amazigh (Berbers), Phoenicians and Romans have settled. The Islamic history of Kerkennah began with the Ottomans (Fehri, 2009).

The Sousse sample is made up of 48 individuals. Sousse, a port city in east-central Tunisia, was founded by the Phoenicians in the 11th before the common era. After the fall of the Roman Empire, the city came under the control of the Vandals and then the Byzantines. In the seventh century, the Arabs conquered the city and introduced the Islamic religion and the Arabic language (Djelloul, 2006). The Normans of Sicily (Norwich, 1992) then the Spaniards invaded Sousse. It received contributions from the Ottoman Turks in the 16th century, then it came under the control of the French in 1881 (Pétridès, 1910).

The Mahdia sample comprises 47 individuals. Mahdia is located in the center-east of Tunisia. It was the capital of the Fatimids (921–973) thanks to its proximity to the sea and to its promontory, on which a military colony had existed since the time of the Phoenicians (Favreau, 1995).

The Kairouan sample includes 45 individuals. Kairouan is a city in central Tunisia. It was the first Muslim city in the Maghreb built around 670. It was a center of learning attracting Muslims from other countries (McEvedy & Jones, 1978).

The Libyan sample comprises 70 individuals collected from six different communities in Libya. Libya is a country in North Africa that is part of the Maghreb. Several civilizations have succeeded in Libya, the Phoenicians, the Romans, the Vandals, the Byzantines, the Arabs, the Spaniards, the Ottomans and the Italians. The majority of Libyans are of mixed Berber-Arab culture (Camps, 1981).

Data from the 26 global populations of the 1000 Genomes (1KGP) project (Table 1) were added to those obtained from the seven North African populations of this study to allow a comparison between populations (Consortium, 2015). The populations of the Americas (Puerto Ricans, Colombians, Peruvians and Mexican Americans) are mixtures of West African and European origin. Peruvians have the highest portion of Native American ancestry.

2.2 | DNA genotyping

The extraction of genomic DNA from all individuals was carried out by the phenol-chloroform method. We analysed three SNPs of the *DRD3* gene, one SNP of the *CSMD1* gene and three SNPs of the *NRG1* gene (Table 2). Genotyping of all SNPs was performed using the TaqMan® assay according to the manufacturer's protocol. An AB7900 thermal cycler using SDS software read the 384-well plates.

2.3 | Statistical analysis

The PLINK 1.09 software (Purcell et al., 2007) determined the calculation of the allele and genotypic frequencies. The Hardy–Weinberg equilibrium was estimated by a chi-square test. Haplotypes formed with the seven SNPs in all populations were obtained from the LDlink website (Machiela & Chanock, 2015). The calculation of the distance matrix F_{ST} was determined by the software ARLEQUIN v3.11 (Excoffier & Lischer, 2010) and the principal component analysis was carried out with the software PAST (Hammer et al., 2001). The analysis of the haplotypes obtained with the seven SNPs was performed using the PHASE v2.1.1 software (Stephens et al., 2001; Stephens & Scheet, 2005). The degree of linkage disequilibrium (LD) was assessed by Haploview software.

TABLE 1 The size of the global populations of the 1000 Genomes (1KGP) project analysed in this study

Population	Sample size
African (AFR)	
Yoruba in Ibadan, Nigeria (YRI)	108
Luhya in Webuye, Kenya (LWK)	99
Gambian in Western Gambia (GWD)	113
Mende in Sierra Leone (MSL)	85
Esan in Nigera (ESN)	99
Americans of African Ancestry in SW USA (ASW)	61
African Carribbeans in Barbados (ACB)	96
Ad Mixed American (AMR)	
Mexican Ancestry from Los Angeles, USA (MXL)	64
Puerto Ricans from Puerto Rico (PUR)	104
Colombians from Medellin, Colombia (CLM)	94
Peruvians from Lima, Peru (PEL)	85
East Asian (EAS)	
Han Chinese in Bejing, China (CHB)	103
Japanese in Tokyo, Japan (JPT)	104
Southern Han Chinese (CHS)	105
Chinese Dai in Xishuangbanna, China (CDX)	93
Kinh in Ho Chi Minh City, Vietnam (KHV)	99
European (EUR)	
Utah Residents from North and West Europe (CEU)	99
Toscani in Italia (TSI)	107
Finnish in Finland (FIN)	99
British in England and Scotland (GBR)	91
Iberian population in Spain (IBS)	107
South Asian (SAS)	
Gujarati Indian from Houston, Texas (GIH)	103
Punjabi from Lahore, Pakistan (PJL)	96
Bengali from Bangladesh (BEB)	86
Sri Lankan Tamil from the UK (STU)	102
Indian Telugu from the UK (ITU)	102

3 | RESULTS

3.1 | Allele frequencies and heterozygosity

Single nucleotide polymorphism sites were polymorphic in all populations. For all SNPs studied the Hardy–Weinberg equilibrium (HWE) were respected at a significance level of 1% (Table S1). Table 3 displays the allele frequencies of all SNPs in all populations of North Africa. The results did not reveal any differences in the allele frequencies of the five SNPs rs6280, rs3732783, rs385396, rs383632 and rs1462906 in North African populations. However, a significant difference at the 0.05 level in the allele frequencies of rs3773678 between the Sousse and Libya populations was demonstrated ($p = 0.014$). In addition, the allele frequencies of rs10108270 from Kairouan were also different at the 0.05 level from those of Kerkennah ($p = 0.006$), Smar ($p = 0.018$) and Libya ($p = 0.02$). The ancestral C allele of rs3732783 was the least frequent allele in the North African populations studied; its frequency varied from 2.3% in Kerkennah to 11.5% in Sousse. Nonetheless, the ancestral alleles rs385396 A and rs383632 C were the most frequent alleles in these populations with frequencies ranging from 88.5% in Sousse to 97.1% in Libya for rs385396 and from 88.5% in Sousse to 97.8% in Kairouan for rs383632 (Table 3). Table 4 shows the heterozygosity for each locus and for each North African population. The average heterozygosity for each population varied from 0.283 in Kairouan to 0.333 in Smar. In general, the heterozygosity of each locus was close to the maximum value of heterozygosity (0.5) except those of loci rs3732783, rs385396 and rs383632. Indeed, the average heterozygosity varied from 0.127 for rs3732783 and rs383632 to 0.492 for rs6280.

3.2 | Gene diversity analysis

Total genomic diversity through all loci among populations (H_T), the genetic diversity between individuals on populations (H_S) and the diversity between populations (F_{ST}) are presented in Table 5. The findings propose that the genomic diversity of all the loci was relatively high (mean $H_T = 0.3063$), this could be owing to the diversity between individuals inside the same population (mean $H_S = 0.3031$). The low F_{ST} values (0.0027–0.0165) indicate that a high proportion (98.35%–99.73%) of the total genetic variability is the result of within-population variation while 0.27%–1.65% of this variability is due to the differences among the seven North African populations. The frequencies of the ancestral alleles of the SNPs analysed in the North African population and the global populations were shown in Figure 2. The frequencies of the ancestral alleles rs383632 (C) and rs385396 (A) of the *NRG1* gene in the North African populations are similar to those of the African, Admixed American, European and South Asian populations but are higher than those of the East Asian population. The rs3773678 (A), rs6280 (C), rs10108270 (A) and rs1462906 (T) ancestral alleles frequencies in the North African population are close to those of the Admixed American, East Asian, European and South Asian populations but are lower than those of the African population. The rs3732783 (C) ancestral allele frequency in the North African population is similar to those of other world populations.

TABLE 2 Description of the seven single nucleotide polymorphisms in the *DRD3*, *CSMD1* and *NRG1* genes

SNP rs-number	Chromosome nucleotide position GRCh38.p12	Chromosome	Gene	Localization on the gene	Forward strand nucleotide alleles (ancestral /derived)	Protein change
rs6280	114,171,968	3q13.31	*DRD3*	Exon 1, 5′UTR	C/T	Gly9Ser
rs3732783	114,171,942	3q13.31	*DRD3*	Exon 2, 5′UTR	C/T	Ala17Ala
rs3773678	114,151,231	3q13.31	*DRD3*	Intron	A/G	
rs10108270	4,333,271	8p23	*CSMD1*	Intron	A/C	
rs385396	31,950,578	8p12	*NRG1*	Intron	A/C	
rs383632	31,948,532	8p12	*NRG1*	Intron	C/T	
rs1462906	32,039,076	8p12	*NRG1*	Intron	T/C	

Abbreviations: SNP, single nucleotide polymorphism; UTR, untranslated region.

TABLE 3 Allele frequencies in North African populations at seven SNPs (rs6280, rs3732783 and rs3773678 of the *DRD3* gene; rs10108270 of the *CSMD1* gene and rs385396, rs383632 and rs1462906 of the *NRG1* gene

	Kairouan	Kerkennah	Kesra	Mehdia	Smar	Sousse	Libya
rs6280							
C	42 (0.477)	40 (0.444)	35 (0.438)	35 (0.398)	58 (0.475)	34 (0.425)	60 (0.435)
T	46 (0.523)	50 (0.556)	45 (0.562)	53 (0.602)	64 (0.525)	46 (0.575)	78 (0.565)
rs3732783							
C	6 (0.067)	2 (0.023)	5 (0.058)	8 (0.093)	8 (0.065)	11 (0.115)	9 (0.064)
T	84 (0.933)	84 (0.977)	81 (0.942)	78 (0.907)	116 (0.935)	85 (0.885)	131 (0.936)
rs3773678							
A	26 (0.361)	35 (0.398)	28 (0.326)	29 (0.322)	43 (0.358)	27 (0.281)	60 (0.435)
G	46 (0.639)	53 (0.602)	58 (0.674)	61 (0.678)	77 (0.642)	69 (0.719)	78 (0.565)
rs10108270							
A	30 (0.333)	49 (0.533)	35 (0.398)	39 (0.443)	61 (0.492)	40 (0.426)	67 (0.486)
C	60 (0.667)	43 (0.467)	53 (0.602)	49 (0.557)	63 (0.508)	54 (0.574)	71 (0.514)
rs385396							
A	87 (0.967)	85 (0.924)	79 (0.94)	77 (0.917)	111 (0.895)	85 (0.885)	134 (0.971)
C	3 (0.033)	7 (0.076)	5 (0.06)	7 (0.083)	13 (0.105)	11 (0.115)	4 (0.029)
rs383632							
C	88 (0.978)	84 (0.933)	79 (0.94)	85 (0.904)	110 (0.902)	85 (0.885)	128 (0.97)
T	2 (0.022)	6 (0.067)	5 (0.06)	9 (0.096)	12 (0.098)	11 (0.115)	4 (0.03)
rs1462906							
T	20 (0.222)	16 (0.178)	18 (0.209)	9 (0.122)	32 (0.262)	14 (0.149)	26 (0.191)
C	70 (0.778)	74 (0.822)	68 (0.791)	65 (0.878)	90 (0.738)	80 (0.851)	110 (0.809)

3.3 | Principal component analysis (PCA)

The PCA plot depicted in Figure 3 shown the genetic relationship between North African and world populations. PC1 (79.65%) and PC2 (16.46%) clearly separate the North African population from the East Asian and African populations, respectively. The North African population was clustered with the European, South Asian and Admixed American populations in the lower left region of the PCA plot.

3.4 | Analysis of haplotype frequencies

Haplotypic analysis of the rs10108270, rs383632, rs385396 and rs1462906 polymorphisms of chromosome 8 in all the populations studied showed the presence of 10 haplotypes among the 16 possible haplotypes (Figure 4). The most common haplotype, CCAC, represents 39.5% of all North African chromosomes. The ACAT haplotype with ancestral allele at all four SNPs is more frequent in the Kairouan population (11.7%) than in other North African

TABLE 4 Heterozygosities and average heterozygosity of seven SNPs in the *DRD3*, *CSMD1* and *NRG1* genes in North African populations

	rs6280	rs3732783	rs3773678	rs10108270	rs385396	rs383632	rs1462906	Average
Kairouan	0.499	0.125	0.461	0.444	0.064	0.043	0.345	0.283
Kerkennah	0.494	0.045	0.479	0.498	0.140	0.125	0.293	0.296
Kesra	0.493	0.109	0.439	0.479	0.113	0.113	0.331	0.297
Mehdia	0.479	0.169	0.437	0.494	0.152	0.174	0.214	0.303
Smar	0.499	0.122	0.460	0.500	0.188	0.177	0.387	0.333
Sousse	0.489	0.204	0.404	0.489	0.204	0.204	0.254	0.321
Libya	0.492	0.120	0.492	0.500	0.056	0.058	0.309	0.289
Average	0.492	0.127	0.453	0.486	0.131	0.127	0.304	0.303

TABLE 5 Genetic diversity analysis for each locus and all loci in the North African population

Locus	Heterozygosiy within populations (H_S)	Total heterozygosity (H_T)	Gene diversity (F_{ST})
rs6280	0.4920	0.4933	0.0027
rs3732783	0.1275	0.1290	0.0110
rs3773678	0.4531	0.4576	0.0098
rs10108270	0.4862	0.4938	0.0155
rs385396	0.1310	0.1329	0.0142
rs383632	0.1276	0.1297	0.0165
rs1462906	0.3046	0.3083	0.0121
All loci	0.3031	0.3063	0.0116
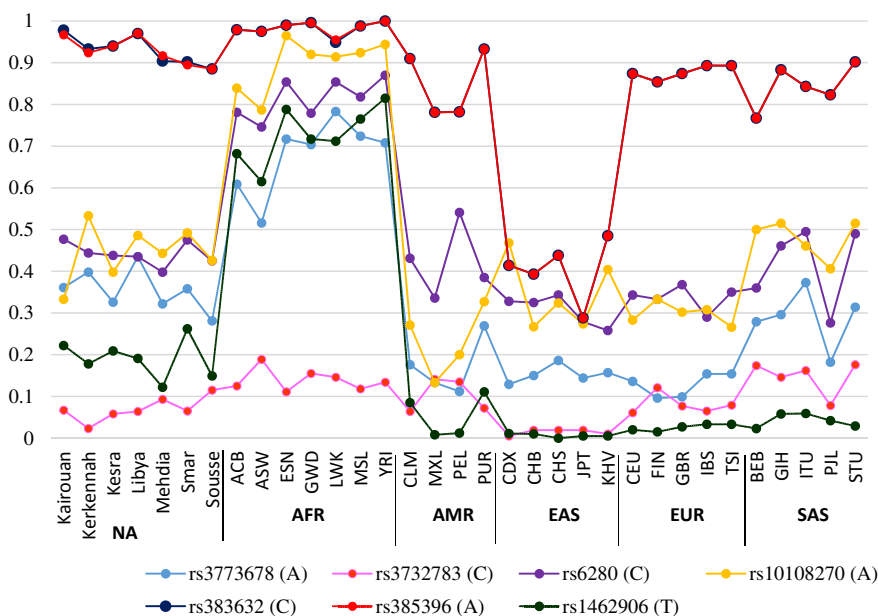
FIGURE 2 Frequency of the ancestral allele of SNPs in the North African and the world populations. AFR, African; AMR, Admixed American; EAS, East Asian; EUR, European; NA, North African; SAS, South Asian. Abbreviations for other populations are shown in Table 1

populations (1.1%–6.7%). The CTCT haplotype with ancestral allele (T) of rs1462906 is only detected in the Sousse population. Since the rs10108270 polymorphism of the *CSMD1* gene is very distant from the other SNPs of the *NRG1* gene on chromosome 8 (~27 megabytes), we have estimated the frequencies of the haplotypes

consisting only of the SNPs rs383632, rs385396 and rs1462906 of the *NRG1* gene in Table S2. We noticed that the results obtained are very close to those determined with the combination of the four SNPs of chromosome 8. In fact, the TCT haplotype is only identified in the Sousse population.

FIGURE 3 Principal component analysis (PCA) of North African and world populations based on a dataset
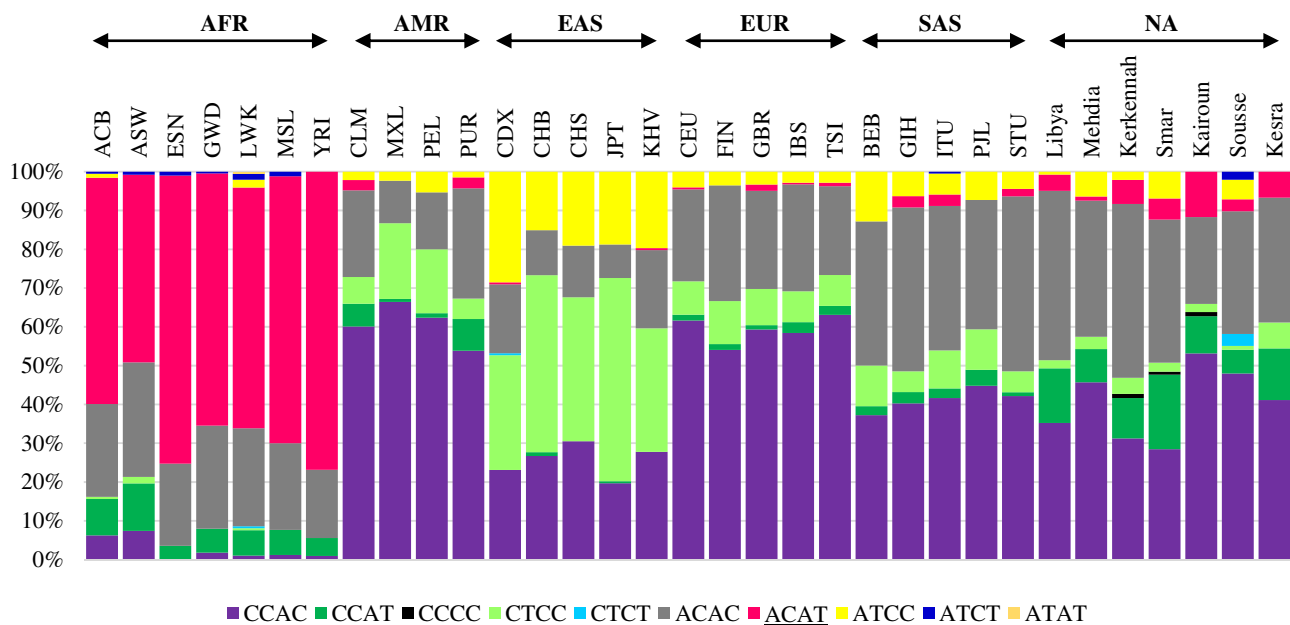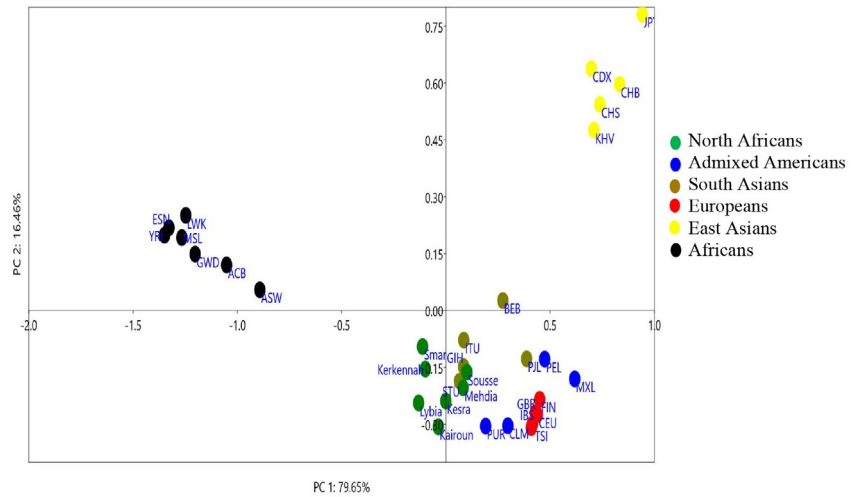

FIGURE 4 Haplotypic distribution of the chromosome 8 single nucleotide polymorphisms (rs10108270, rs383632, rs385396, rs1462906) in the North African and world populations. AFR, African; AMR, Admixed American; EAS, East Asian; EUR, European; NA, North African; SAS, South Asian. Abbreviations for other populations are shown in Table 1. The ancestral haplotype is underlined

Haplotypic analysis of the rs3773678, rs3732783 and rs6280 SNPs of *DRD3* gene on chromosome 3 in all the populations studied showed the presence of seven haplotypes among the eight possible haplotypes (Figure 5). The frequencies of the haplotypes composed of these three SNPs in the North African population are presented in Tablr S3. The ACC haplotype composed of the ancestral alleles of the three SNPs of chromosome 3 is more frequent in the Kairouan population (6.4%) than in other North African populations (0.8%–2.8%). The GCT haplotype with ancestral allele (C) of rs3732783 is specific to the Mehdia population of the North African region. No other population has this haplotype.

The comparison of the haplotypic profile of the North African population with those of the 26 populations of the 1000 Genomes project was carried out with the SNPs of chromosome 8 (Figure 4) and those of chromosome 3 (Figure 5). The frequency of ACAT haplotype with ancestral alleles is very high in the African population (65.89%) but low in the North African (5.38%), European (0.7%), Admixed American (1.59%), East Asian (0.2%) and South Asian (1.64%) populations (Figure 4). The frequency of the ACC haplotype, composed of ancestral alleles, in the North African population (1.75%), is close to that of the Admixed American (0.72%), East Asian (0.5%) and European (0%) populations but it is different from that of the African
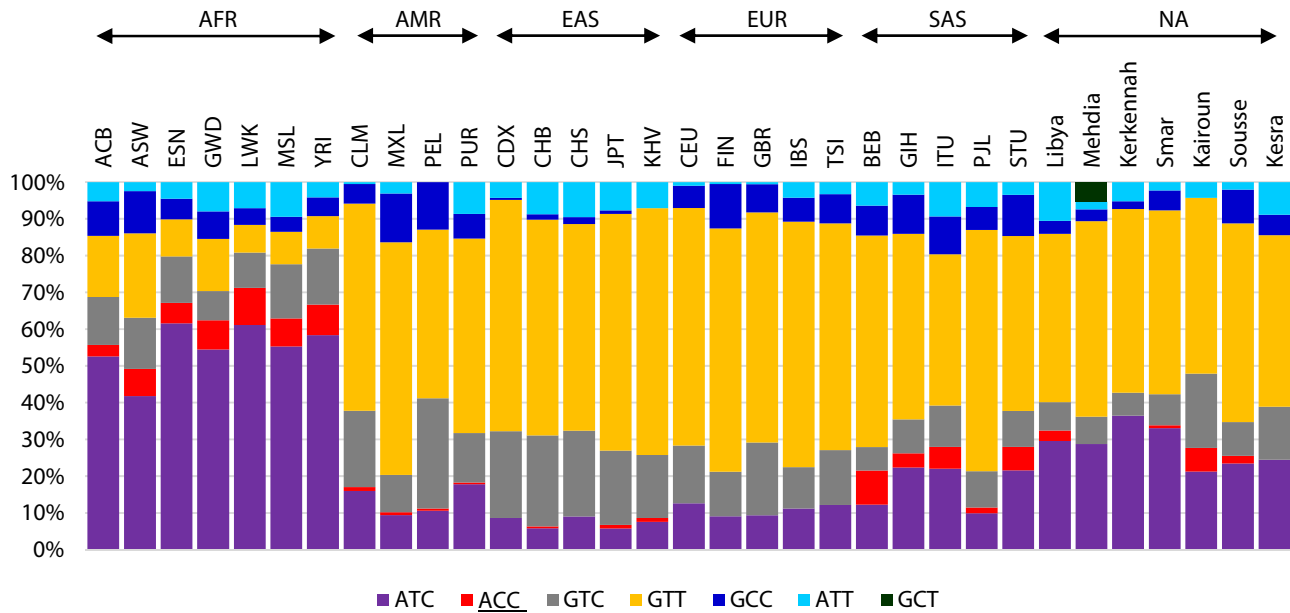
FIGURE 5 Haplotypic distribution of the chromosome 3 single nucleotide polymorphisms (rs3773678, rs3732783, rs6280) in the North African and world populations. AFR: African; AMR: Ad Mixed American; EAS: East Asian; EUR: European; NA: North African; SAS: South Asian. Abbreviations for other populations are shown in Table 1. The ancestral haplotype is underlined
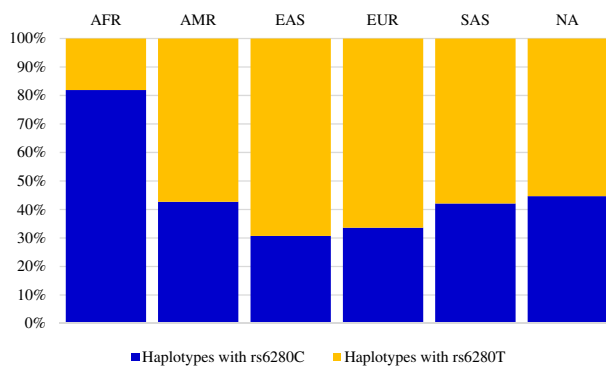

FIGURE 6 Frequency of haplotypes with or without the rs6280T-derived allele in the North African and world populations. AFR, African; AMR, Ad Mixed American; EAS, East Asian; EUR, European; NA, North African; SAS, South Asian
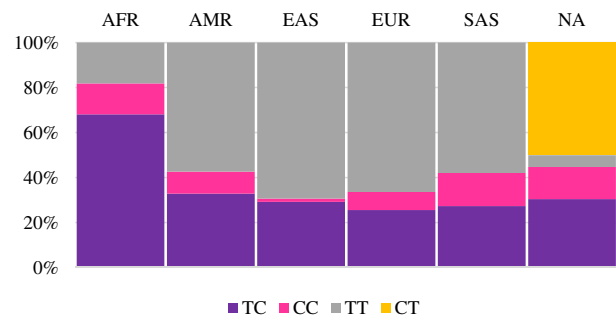

FIGURE 7 Haplotypic distribution of the chromosome 3 single nucleotide polymorphisms (rs3732783, rs6280) in the North African and world populations. AFR, African; AMR, Ad Mixed American; EAS, East Asian; EUR, European; NA, North African; SAS, South Asian

(7.19%) and South Asian (5.32%) populations. The GTT haplotype composed of the derived alleles of the three SNPs of chromosome 3 is more frequent in the North African (49.46%), South Asian (52.15%), European (64.41%), East Asian (61.81%) and Admixed American (54.03%) populations than in the African population (12.18%) (Figure 5).

Figure 6 showed the frequencies of haplotypes carrying the T-derived allele of SNP rs6280. The frequency of these haplotypes was 18.1% in the African population while it varied between 55.3% in the North African population and 69.3% in the East Asian population. Analysis of microhaplotypes––haplotypes of less than 300 nucleotides

(Kidd et al., 2014)—made up of exonic polymorphisms of the *DRD3* gene in the North African population and global populations revealed that the rs3732783C-rs6280T microhaplotype is present only in the North African population. It represents 50% of the microhaplotypes determined in this population (Figure 7). The rs3732783T-rs6280T microhaplotype consisting of derived alleles has a low frequency (5.2%) in the North African population while its frequency varies between 18% and 69% in the other populations. In the North African population, the frequency of the microhaplotype presenting the ancestral alleles rs3732783C and rs6280C (14.3%) is close to that of the South Asian (14.7%)

and African (13.7%) populations but it is higher than that of the European (8%), American (9.8%) and East Asian (1.5%) populations (Figure 7).

3.5 | Linkage disequilibrium

Linkage disequilibrium (LD) between rs10108270, rs383632, rs385396 and rs1462906 SNPs of chromosome 8 and between rs3773678, rs3732783 and rs6280 SNPs of chromosome 3 is measured by the values of D′, r^2 and LOD (Logarithm of odds) for the North African populations and the world populations. The values of r^2 are presented in Figures 8 and 9.

Significant LD between the two SNPs rs383632 and rs385396 of the *NRG1* gene in chromosome 8 was observed for all North African populations. The LD between rs3773678 and rs6280 of the *DRD3* gene in chromosome 3 was significant in all North African populations except in those of Libya and Kesra. However, the exonic SNPs rs6280 and rs3732783 of the *DRD3* gene were not in linkage disequilibrium in all North African populations (Figure 8).

The pairwise LD between the SNPs of chromosome 8 and those of chromosome 3 for the North African population and the other global populations are presented in Figure 9. LD analysis of chromosome 8 SNPs identified a haplotypic block on 2 kb consisting of rs383632 and rs385396 in the North African population and in all global populations except African population. For chromosome 3 SNPs, we observed significant linkage disequilibrium between the polymorphisms rs3732783 and rs6280 of

DRD3 gene in the North African population and global populations.

4 | DISCUSSION

In this study, the rs3773678, rs3732783 and rs6280 SNPs of *DRD3* gene, the rs10108270 SNPs of *CSMD1* gene and the rs383632, rs385396 and rs1462906 SNPs of *NRG1* gene were explored in seven population groups of North Africa. A regional analysis assessed the extent of genetic variation and affinity between those groups and a worldwide analysis aimed at genetically locating the North African population in relation to the global populations.

Regional analysis based on comparison of allele frequencies, PCA and LD analysis revealed a general similarity among the populations of the North African region. The low F_{ST} values for each SNP (mean F_{ST} 0.0116) showed that the North African populations studied are not very different from one another at these neurological polymorphisms. This result was in agreement with that of Gardner et al. (2007) who also presented low values of F_{ST} for the SNPs rs383632, rs385396 and rs1462906 of the *NRG1* gene in the populations of the Middle East/North Africa region. While Kidd et al. (2004) showed a high F_{ST} values for rs1462906 (0.457) and rs385396 (0.251) in global population. These different results could be due to the variability of the allele frequencies of the *NRG1* gene according to the geographic regions. However, the heterozygosity evaluation on the seven SNPs studied suggested that the Kairouan population has the lowest average
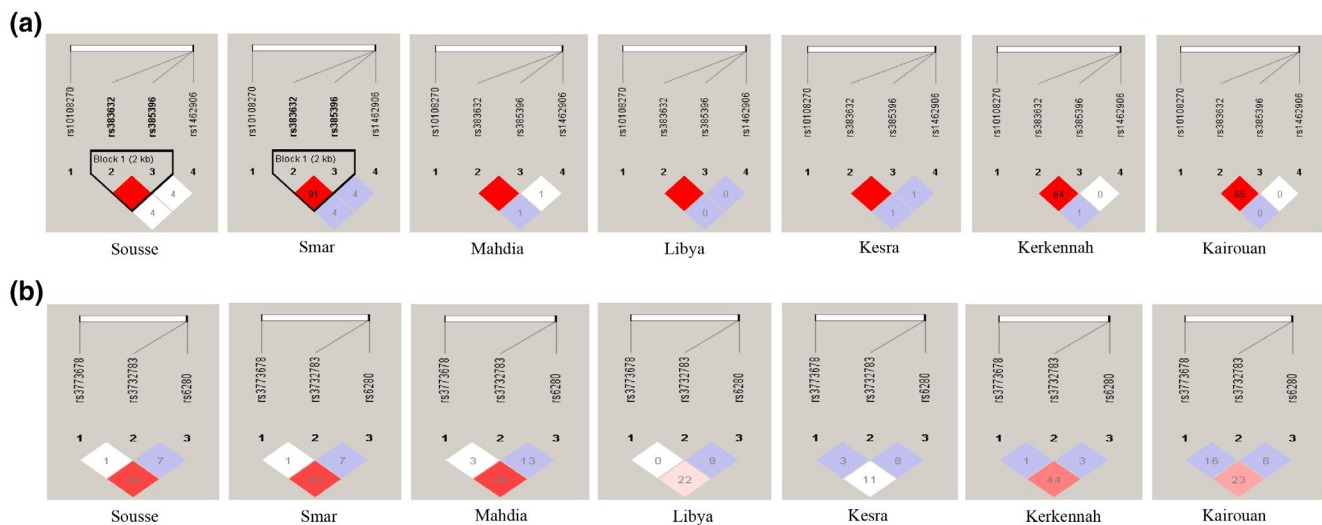

FIGURE 8 Linkage disequilibrium patterns between rs10108270, rs383632, rs385396 and rs1462906 SNPs of chromosome 8 (a) and rs3773678, rs3732783 and rs6280 SNPs of chromosome 3 (b) in North African populations. The r^2 values are presented in each square. Red squares indicate a statistically significant LD between the pair of SNPs. Shaded red squares indicate that D′ < 1 and LOD ≥ 2, bright red squares without a number indicate that D′ = 1 and LOD ≥ 2, white squares indicate that D′ < 1 and LOD < 2 and blue squares indicate that D′ = 1 and LOD < 2
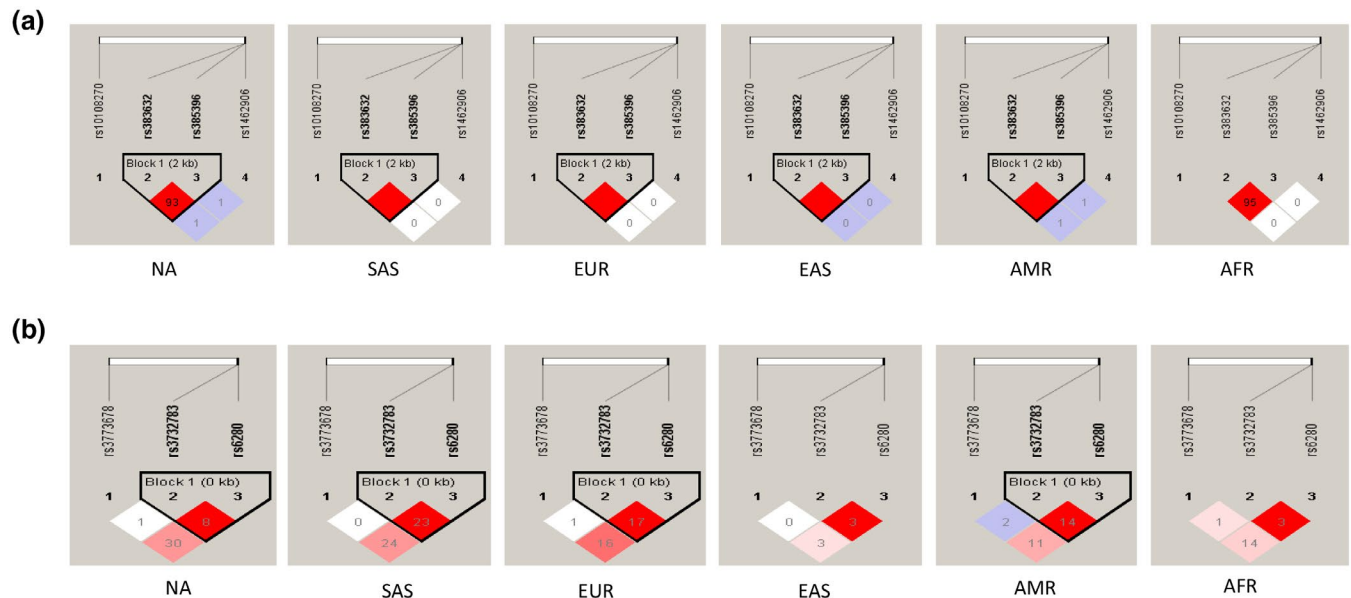
FIGURE 9 Linkage disequilibrium patterns between rs10108270, rs383632, rs385396 and rs1462906 SNPs of chromosome 8 (a) and rs3773678, rs3732783 and rs6280 SNPs of chromosome 3 (b) in the North African and world populations. AFR, African; AMR, Ad Mixed American; EAS, East Asian; EUR, European; NA: North African; SAS, South Asian. The r^2 values are presented in each square. Red squares indicate a statistically significant LD between the pair of SNPs. Shaded red squares indicate that D′ < 1 and LOD ≥ 2, bright red squares without a number indicate that D′ = 1 and LOD ≥ 2, white squares indicate that D′ < 1 and LOD < 2 and blue squares indicate that D′ = 1 and LOD < 2

heterozygosity (0.283) among the North African populations. Therefore, within the North African population, the Kairouan population group has a low rate of genetic variability. In addition, our result of the haplotypic analysis of the North African population provided relevant information. Indeed, haplotypes composed of the ancestral alleles ACC of chromosome 3 and ACAT of chromosome 8 were more frequent in the Kairouan population than in other North African populations. We also showed that the Mahdia and Sousse populations were distinguished from other North African populations by the presence of the GCT and CTCT haplotypes, respectively. The presence of these specific haplotypes could be the result of recombination and mutations or a new flow of genes, which play a major role in genetic diversity and represent an important evolutionary force (Scally, 2016). Our result is comparable to those obtained by the analysis of other genetic markers. Effectively, by analysing SNPs of the *COMT* gene, Boussetta et al. (2019) suggested that the Sousse population is distinguished from other North African populations. In addition, by analysing seven Alu insertion polymorphisms, Frigi et al. (2014) highlighted a differentiation between Mahdia and the majority of North African populations. This genetic diversity observed in the populations of Sousse and Mahdia could be due to their close geographical location and their common history. In fact, the city of Sousse is located 28 kilometres from that of Mahdia. The two cities occupy a strategic geographical location

in the center-east of Tunisia on the Mediterranean Sea. These two cities have seen the passage of several civilizations such as Phoenicians, Romans, Vandals, Byzantines, Arabs, Spanish, Ottomans, Andalusians, and French who participated in the genetic diversity of North African populations (Favreau, 1995; Norwich, 1992; Pétridès, 1910).

The comparison of the North African population to the whole of the world populations revealed a differentiation of the sub-Saharan African populations from the North African, European, South Asian, East Asian and Admixed American populations. Indeed, the frequency of the ancestral allele of certain SNPs like rs6280, rs3773678, rs1462906 and rs10108270 in the sub-Saharan African populations was different from that of other populations. Our result confirmed that of Gardner et al. (2006) who demonstrated that the allele frequency of rs1462906 in the sub-Saharan African populations was different from that of the world populations. PCA and haplotypic profile analysis also showed genetic differentiation of the sub-Saharan African populations from other global populations. This differentiation could be explained by the existence of the Sahara, which constitutes a geographical barrier to migration and against gene flow (Harich et al., 2010). Indeed, 10,000 years ago during the optimum climatic Holocene, the drying up of the Sahara began and has led for about 5,000 years to the current climate and to a rupture between the north and the south of Africa (Brooks et al., 2005).

Despite the geographic, socio-cultural and linguistic distances that separate the North African and sub-Saharan African populations, the existence of an ancestral genetic component between the two groups of populations has been shown. Indeed, we have demonstrated that the ancestral ACC haplotypes of chromosome 3 and ACAT of chromosome 8 were present in the North African and sub-Saharan African populations. We have already demonstrated in a previous study that the genes of the North African populations were subjected to a Sub-Saharan influence (Mestiri et al., 2021). Our results were consistent with the idea of 'out of Africa' dispersal for *Homo sapiens* presented by Henn et al. (2012). More recently, D'Atanasio et al. (2018) proved that the period of Humid Africa favoured the expansion of Neolithic civilization and contact between populations, which contributed to the distribution of Y lineages in North and Sub-Saharan Africa. All these events could explain the presence of common haplotypes between the North African and African populations.

On the other hand, PCA and the haplotypic analysis of *DRD3*, *CSMD1* and *NRG1* polymorphisms in North African and worldwide populations showed that the genetic structure of populations in North Africa is closer to that of Europeans, Admixed Americans, South Asians and East Asians. This result proved the existence of a considerable Eurasian genetic component for the North African populations since it was shown that recent European contact had an influence on the Admixed American population (Gravel et al., 2013). Our findings were in agreement with other observations made using different genetic markers. In recent study, using SNPs from the *DRD2/ANKK1* locus, we demonstrated by PCA the genetic affinity of the North African population to the South Asian and Admixted American populations (Mestiri et al., 2021). In 2017, Rodríguez-Varela et al. (2017) performed a PCA on the Guanche individuals (originated from Canary Islands) and populations from North Africa, Europe and the Middle East using genome-wide autosomal SNPs. They revealed a strong genetic affinity of the Guanches to modern Northwest African populations like Tunisians and Algerians. Further research confirmed our result by analysing the HLA genes. Indeed, Hajjej et al. (2017) showed that the Tunisian population is genetically close to the Iberians and Western Mediterranean but it was different from the Sub-Saharan population. Other studies based on the analysis of mitochondrial DNA have shown that the haplogroups detected in the North African population are mostly of Eurasian origin (Cherni et al., 2005; Cherni et al., 2009; Fadhlaoui-Zid et al., 2011; Frigi et al., 2017; Harich et al., 2010). Indeed, the U5, R0a, V, U3 and J1b haplogroups of Middle East and European origin as well as the M1b and U6 haplogroups of Southwest Asian

origin are specific to the North African populations (González-Andrade et al., 2007; Olivieri et al., 2006; Pennarun et al., 2012). Arauna et al. (2019) recently studied the genetic history of populations in the coastal area of North Africa. They demonstrated a flow of genes from the Mediterranean coasts of North Africa to Europe (Tuscany and the Iberian Peninsula) probably in the fourth century during the movements of peoples following the fall of the Roman Empire and during the expansion of Arabs in the seventh century and the Christian kingdoms later.

We then focused on the analysis of exonic polymorphisms of the *DRD3* gene, the synonymous SNP rs3732783 and the non-synonymous SNP rs6280 in the North African population and other global populations. Comparison of the frequencies of haplotypes carrying the rs6280T-derived allele and haplotypes carrying the rs6280C ancestral allele proved that all populations except the sub-Saharan African populations showed an increase in the frequency of haplotypes carrying the rs6280T-derived allele. This result could be explained by a recent positive selection, which affected the *DRD3* gene only in North African, European, South Asian, East Asian and Admixed American populations. It is very likely that the rs6280 SNP was the subject of this positive selection since it has an effect on the function of the gene. Indeed, it has been shown that the ancestral allele rs6280C has a greater affinity to dopamine and increased signalling responses (Jeanneteau et al., 2006). The most striking result of our analysis was the specificity of the microhaplotype presenting the ancestral C allele of rs3732783 and the T-derived allele of rs6280 to the North African population. This microhaplotype represented 50% of all haplotypes in the North African populations. However, the frequency of the microhaplotype presenting the ancestral alleles rs3732783C and rs6280C in the North African populations was very low compared to those of other populations worldwide. Our study showed that the rs3732783 and rs6280 polymorphisms were not in LD in all North African populations. The generation of new haplotypes specific to the North African region could be due to different evolutionary forces occurred over many human generations and which can act at the genomic level by mutations and recombination, at the demographic level by migrations and at the selective level following climate or nutritional change. All these events have allowed the North African population to be genetically distinct from other populations.

An important limitation of this study was the limited number of SNPs analysed. Indeed, to regroup individuals into distinct groups according to their different origins, it is necessary to consider a very large number of markers

making it possible to have a panel with sufficient discriminating power. In addition, the analysis of a large number of markers will avoid bias in the estimation of inter- and intrapopulation variation (Willing et al., 2012). On the other hand, the study of the genetic diversity and the structure of the North African population could be carried out using a new approach, which analyses transcriptomic markers specific to the population.

In conclusion, our study suggested that the North African genome lies in an intermediate position between Europe and Asia. The antiquity and richness of the North Africa colonisation history allowed the emergence of a particular genetic structure, which made North African populations genetically distinct from other populations.

ETHICAL COMPLIANCE

Ethics Committee approves this work for Research in Life Sciences and Health of the ISBM (CER-SVS/ISBM).

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors have declared no conflicts of interest.

AUTHOR CONTRIBUTIONS

Souhir Mestiri: Data interpretation; writing-original draft. **Sami Boussetta and Sarra Elkamel:** Methodology; data acquisition and analysis. **Andrew J. Pakstis, Kenneth K. Kidd and Amel Ben Ammar Elgaaied:** Writing-review and editing. **Lotfi Cherni:** Project administration; supervision.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Souhir Mestiri https://orcid.org/0000-0003-3199-9040
Sarra El Kamel https://orcid.org/0000-0002-1110-5225
Lotfi Cherni https://orcid.org/0000-0002-7050-6539

REFERENCES

Allen, N. C., Bagade, S., McQueen, M. B., Ioannidis, J. P., Kavvoura, F. K., Khoury, M. J., Tanzi, R. E., & Bertram, L. (2008). Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: The SzGene database. *Nature Genetics*, *40*(7), 827–834. https://doi.org/10.1038/ng.171

Arauna, L. R., Hellenthal, G., & Comas, D. (2019). Dissecting human North African gene-flow into its western coastal surroundings. *Proceedings of the Royal Society B*, *286*(1902), 20190471. https://doi.org/10.1098/rspb.2019.0471

Athanasiu, L., Giddaluru, S., Fernandes, C., Christoforou, A., Reinvang, I., Lundervold, A. J., Nilsson, L. G., Kauppi, K., Adolfsson, R., Eriksson, E., Sundet, K., Djurovic, S., Espeseth, T., Nyberg, L., Steen, V. M., Andreassen, O. A., & Le Hellard, S. (2017). A genetic association study of CSMD1 and CSMD2 with cognitive function. *Brain, Behavior, and Immunity*, *61*, 209–216. https://doi.org/10.1016/j.bbi.2016.11.026

Bare, D. J., Becker-Catania, S. G., & DeVries, G. H. (2011). Differential localization of neuregulin-1 type III in the central and peripheral nervous system. *Brain Research*, *1369*, 10–20. https://doi.org/10.1016/j.brainres.2010.10.092

Bocharova, A., Stepanov, V., Marusin, A., Kharkov, V., Vagaitseva, K., Fedorenko, O. Y., Bokhan, N., Semke, A., & Ivanova, S. (2017). Association study of genetic markers of schizophrenia and its cognitive endophenotypes. *Russian Journal of Genetics*, *53*(1), 139–146.

Bombin, I., Arango, C., Mayoral, M., Castro-Fornieles, J., Gonzalez-Pinto, A., Gonzalez-Gomez, C., Moreno, D., Parellada, M., Baeza, I., Graell, M., Otero, S., Saiz, P. A., & Patino-Garcia, A. (2008). DRD3, but not COMT or DRD2, genotype affects executive functions in healthy and first-episode psychosis adolescents. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *147*(6), 873–879. https://doi.org/10.1002/ajmg.b.30710

Boussetta, S., Cherni, L., Pakstis, A. J., Ben Salem, N., Elkamel, S., Khodjet-el-Khil, H., Kidd, K. K., & Elgaaied, A. B. A. (2019). Usefulness of COMT gene polymorphisms in North African populations. *Gene*, *696*, 186–196. https://doi.org/10.1016/j.gene.2019.02.021

Brooks, N., Chiapello, I., Lernia, S. D., Drake, N., Legrand, M., Moulin, C., & Prospero, J. (2005). The climate-environment-society nexus in the Sahara from prehistoric times to the present day. *The Journal of North African Studies*, *10*(3–4), 253–292. https://doi.org/10.1080/13629380500336680

Camps, G. (1981). L'origine des Berbères. Islam, société et communauté. *Anthropologie du Maghreb. FRA*, *12*, 9–33.

Cherni, L., Fernandes, V., Pereira, J. B., Costa, M. D., Goios, A., Frigi, S., Yacoubi-Loueslati, B., Ben Amor, M., Abdelhakim, S., Amorim, A., Ben Ammar El Gaaied, A., & Pereira, L. (2009). Post-last glacial maximum expansion from Iberia to North Africa revealed by fine characterization of mtDNA H haplogroup in Tunisia. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *139*(2), 253–260. https://doi.org/10.1002/ajpa.20979

Cherni, L., Loueslati, B. Y., Pereira, L., Ennafaa, H., Amorim, A., & El Gaaied, A. B. A. (2005). Female gene pools of Berber and Arab neighboring communities in central Tunisia: Microstructure of mtDNA variation in North Africa. *Human Biology*, *77*, 61–70. https://doi.org/10.1353/hub.2005.0028

Consortium, G. P. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

D'Atanasio, E., Trombetta, B., Bonito, M., Finocchio, A., Di Vito, G., Seghizzi, M., Romano, R., Russo, G., Paganotti, G. M., Watson,

E., Coppa, A., Anagnostou, P., Dugoujon, J. M., Moral, P., Sellitto, D., Novelletto, A., & Cruciani, F. (2018). The peopling of the last Green Sahara revealed by high-coverage resequencing of trans-Saharan patrilineages. *Genome Biology*, *19*(1), 1–15. https://doi.org/10.1186/s13059-018-1393-5

Djelloul, N. (2006). *Sousse ancient Hadrumetum* (Contraste ed.). Oloron-Sainte Marie.

Donohoe, G., Walters, J., Hargreaves, A., Rose, E. J., Morris, D., Fahey, C., Bellini, S., Cummins, E., Giegling, I., Hartmann, A. M., Moller, H. J., Muglia, P., Owen, M. J., Gill, M., O'Donovan, M. C., Tropea, D., Rujescu, D., & Corvin, A. (2013). Neuropsychological effects of the CSMD1 genome-wide associated schizophrenia risk variant rs10503253. *Genes, Brain and Behavior*, *12*(2), 203–209. https://doi.org/10.1111/gbb.12016

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fadhlaoui-Zid, K., Rodríguez-Botigué, L., Naoui, N., Benammar-Elgaaied, A., Calafell, F., & Comas, D. (2011). Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the Nile Valley. *American Journal of Physical Anthropology*, *145*(1), 107–117. https://doi.org/10.1002/ajpa.21472

Favreau, R. (1995). *Études d'épigraphie médiévale* (Vol. *1*). Presses Univ.

Fehri, A. (2009). Kerkena in the course of history: research unit acculturation of Mediterranean Tunisia, Faculté des lettres et des Sciences humaines de Sfax, Cercina Center for Research in Mediterranean Islands. Kerkena. Série: Rive Méditerranéenne, 9.

Frigi, S., El Gaaied, A. B. A., & Cherni, L. (2014). An investigation of the genetic diversity of the Kerkennah islands and Mahdia (Tunisia) using biparental markers. *Annals of Human Biology*, *41*(1), 53–60. https://doi.org/10.3109/03014460.2013.824025

Frigi, S., Mota-Vieira, L., Cherni, L., van Oven, M., Pires, R., Boussetta, S., & El-Gaaied, A. B. A. (2017). Mitochondrial DNA analysis of Tunisians reveals a mosaic genetic structure with recent population expansion. *Homo*, *68*(4), 298–315. https://doi.org/10.1016/j.jchb.2017.05.003

Fukui, N., Muratake, T., Kaneko, N., Amagane, H., & Someya, T. (2006). Supportive evidence for neuregulin 1 as a susceptibility gene for schizophrenia in a Japanese population. *Neuroscience Letters*, *396*(2), 117–120. https://doi.org/10.1016/j.neulet.2005.11.015

Gardner, M., González-Neira, A., Lao, O., Calafell, F., Bertranpetit, J., & Comas, D. (2006). Extreme population differences across Neuregulin 1 gene, with implications for association studies. *Molecular Psychiatry*, *11*(1), 66–75. https://doi.org/10.1038/sj.mp.4001749

Gardner, M., Williamson, S., Casals, F., Bosch, E., Navarro, A., Calafell, F., Bertranpetit, J., & Comas, D. (2007). Extreme individual marker FST values do not imply population-specific selection in humans: The NRG1 example. *Human Genetics*, *121*(6), 759–762. https://doi.org/10.1007/s00439-007-0364-9

Gassó, P., Arnaiz, J. A., Mas, S., Lafuente, A., Bioque, M., Cuesta, M. J., Diaz-Caneja, C. M., Garcia, C., Lobo, A., Gonzalez-Pinto, A., Parellada, M., Corripio, I., Vieta, E., Castro-Fornieles, J., Mané, A., Rodriguez, N., Boloc, D., Saiz-Ruiz, J., Bernardo, M., & Group, P. E. P. S. (2020). Association study of candidate genes with obesity and metabolic traits in antipsychotic-treated patients with first-episode psychosis over a 2-year period.

Journal of Psychopharmacology, *34*(5), 514–523. https://doi.org/10.1177/0269881120903462

González-Andrade, F., Sánchez, D., González-Solórzano, J., Gascón, S., & Martínez-Jarreta, B. (2007). Sex-specific genetic admixture of mestizos, amerindian kichwas, and afro-ecuadorans from Ecuador. *Human Biology*, *79*(1), 51–77. https://doi.org/10.1353/hub.2007.0024

Gravel, S., Zakharia, F., Moreno-Estrada, A., Byrnes, J. K., Muzzio, M., Rodriguez-Flores, J. L., Kenny, E. E., Gignoux, C. R., Maples, B. K., Guiblet, W., Dutil, J., Via, M., Sandoval, K., Bedoya, G., 1000 Genomes Project, Oleksyk, T. K., Ruiz-Linares, A., Burchard, E. G., Martinez-Cruzado, J. C., & Bustamante, C. D. C. (2013). Reconstructing Native American migrations from whole-genome and whole-exome data. *PLoS Genetics*, *9*(12), e1004023. https://doi.org/10.1371/journal.pgen.1004023

Hajjej, A., Almawi, W. Y., Hattab, L., El-Gaaied, A., & Hmida, S. (2017). The investigation of the origin of Southern Tunisians using HLA genes. *Journal of Human Genetics*, *62*(3), 419–429. https://doi.org/10.1038/jhg.2016.146

Hammer, O., Harper, D. A., & Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, *4*(1), 1–9.

Harich, N., Costa, M. D., Fernandes, V., Kandil, M., Pereira, J. B., Silva, N. M., & Pereira, L. (2010). The trans-Saharan slave trade-clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages. *BMC Evolutionary Biology*, *10*(138), 1–18. https://doi.org/10.1186/1471-2148-10-138

Henn, B. M., Botigué, L. R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J. K., Fadhlaoui-Zid, K., Zalloua, P. A., Moreno-Estrada, A., Bertranpetit, J., Bustamante, C. D., & Comas, D. (2012). Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genetics*, *8*(1), e1002397. https://doi.org/10.1371/journal.pgen.1002397

Ibn Khaldoun, A. (1968). *Histoire des Berbères et des dynasties musulmanes de l'Afrique septentrionale. Traduite de l'arabe par le baron de Slane* (Vol. *1*). P. Geuthner.

Jagannath, V., Gerstenberg, M., Walitza, S., Franscini, M., Heekeren, K., Rössler, W., Theodoridou, A., & Grünblatt, E. (2018). Neuregulin 1 (NRG1) gene expression predicts functional outcomes in individuals at clinical high-risk for psychosis. *Psychiatry Research*, *266*, 143–146. https://doi.org/10.1016/j.psychres.2018.05.025

Jagannath, V., Theodoridou, A., Gerstenberg, M., Franscini, M., Heekeren, K., Correll, C. U., Rossler, W., Grunblatt, E., & Walitza, S. (2017). Prediction analysis for transition to schizophrenia in individuals at clinical high risk for psychosis: The relationship of DAO, DAOA, and NRG1 variants with negative symptoms and cognitive deficits. *Frontiers in Psychiatry*, *8*, 292. https://doi.org/10.3389/fpsyt.2017.00292

Jeanneteau, F., Funalot, B., Jankovic, J., Deng, H., Lagarde, J.-P., Lucotte, G., & Sokoloff, P. (2006). A functional variant of the dopamine D3 receptor is associated with risk and age-at-onset of essential tremor. *Proceedings of the National Academy of Sciences*, *103*(28), 10753–10758. https://doi.org/10.1073/pnas.0508189103

Kidd, K. K., Pakstis, A. J., Speed, W. C., & Kidd, J. R. (2004). Understanding human DNA sequence variation. *The Journal of Heredity*, *95*(5), 406–420. https://doi.org/10.1093/jhered/esh060

Kidd, K. K., Pakstis, A. J., Speed, W. C., Lagacé, R., Chang, J., Wootton, S., Haigh, E., & Kidd, J. R. (2014). Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics. *Forensic Science International: Genetics*, *12*, 215–224. https://doi.org/10.1016/j.fsigen.2014.06.014

Kraus, D. M., Elliott, G. S., Chute, H., Horan, T., Pfenninger, K. H., Sanford, S. D., Foster, S., Scilly, S., Welcher, A. A., & Holers, V. M. (2006). CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. *The Journal of Immunology*, *176*(7), 4419–4430. https://doi.org/10.4049/jimmunol.176.7.4419

Le Coniat, M., Sokoloff, P., Hillion, J., Martres, M.-P., Giros, B., Pilon, C., Schwartz, J. C., & Berger, R. (1991). Chromosomal localization of the human D 3 dopamine receptor gene. *Human Genetics*, *87*(5), 618–620. https://doi.org/10.1007/BF00209024

Liu, Y., Fu, X., Tang, Z., Li, C., Xu, Y., Zhang, F., Zhou, D., & Zhu, C. (2019). Altered expression of the CSMD1 gene in the peripheral blood of schizophrenia patients. *BMC Psychiatry*, *19*(1), 1–5. https://doi.org/10.1186/s12888-019-2089-4

Loh, H. C., Tang, P. Y., Tee, S. F., Chow, T. J., Choong, C. Y., Lim, S. Y., & Yong, H. S. (2013). Neuregulin-1 (NRG-1) and its susceptibility to schizophrenia: A case–control study and meta-analysis. *Psychiatry Research*, *208*(2), 186–188. https://doi.org/10.1016/j.psychres.2013.01.022

Machiela, M. J., & Chanock, S. J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*, *31*(21), 3555–3557. https://doi.org/10.1093/bioinformatics/btv402

McEvedy, C., & Jones, R. (1978). *Atlas of world population history*. Penguin Books Ltd.

Mestiri, S., Boussetta, S., Pakstis, A. J., Elkamel, S., Elgaaied, A. B. A., Kidd, K. K., & Cherni, L. (2021). Genetic diversity of the North African population revealed by the typing of SNPs in the DRD2/ANKK1 genomic region. *Gene*, *777*, 145466. https://doi.org/10.1016/j.gene.2021.145466

Mostaid, M. S., Lloyd, D., Liberg, B., Sundram, S., Pereira, A., Pantelis, C., Karl, T., Weickert, C. S., Everall, I. P., & Bousman, C. A. (2016). Neuregulin-1 and schizophrenia in the genome-wide association study era. *Neuroscience & Biobehavioral Reviews*, *68*, 387–409. https://doi.org/10.1016/j.neubiorev.2016.06.001

Murdock, G. P. (1959). *Africa its peoples and their culture history*. McGrawHill.

Norwich, J. J. (1992). *The Normans in Sicily*. Penguin Global.

Olivieri, A., Achilli, A., Pala, M., Battaglia, V., Fornarino, S., Al-Zahery, N., Scozzari, R., Cruciani, F., Behar, D. M., Dugoujon, J. M., Coudray, C., Santachiara-Benerecetti, A. S., Semino, O., Bandelt, H. J., & Torroni, A. (2006). The mtDNA legacy of the levantine early upper palaeolithic in Africa. *Science*, *314*(5806), 1767–1770. https://doi.org/10.1126/science.1135566

Pennarun, E., Kivisild, T., Metspalu, E., Metspalu, M., Reisberg, T., Moisan, J.-P., Behar, D. M., Jones, S. C., & Villems, R. (2012). Divorcing the Late Upper Palaeolithic demographic histories of mtDNA haplogroups M1 and U6 in Africa. *BMC Evolutionary Biology*, *12*(1), 1–12. https://doi.org/10.1186/1471-2148-12-234

Pétridès, S. (1910). *Hadrumetum'the Catholic encyclopedia*. Robert Appleton Company.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi, https://doi.org/10.1086/519795

Rodríguez-Varela, R., Günther, T., Krzewińska, M., Storå, J., Gillingwater, T. H., MacCallum, M., Arsuaga, J. L., Dobney, K., Valdiosera, C., Jakobsson, M., Gotherstrom, A., & Girdland-Flink, L. (2017). Genomic analyses of pre-European conquest human remains from the Canary Islands reveal close affinity to modern North Africans. *Current Biology*, *27*(21), 3396–3402. e3395. https://doi.org/10.1016/j.cub.2017.09.059

Ronchi, G., Haastert-Talini, K., Fornasari, B. E., Perroteau, I., Geuna, S., & Gambarotta, G. (2016). The Neuregulin1/ErbB system is selectively regulated during peripheral nerve degeneration and regeneration. *European Journal of Neuroscience*, *43*(3), 351–364. https://doi.org/10.1111/ejn.12974

Scally, A. (2016). The mutation rate in human evolution and demographic inference. *Current Opinion in Genetics & Development*, *41*, 36–43. https://doi.org/10.1016/j.gde.2016.07.008

Serra-Vidal, G., Lucas-Sanchez, M., Fadhlaoui-Zid, K., Bekada, A., Zalloua, P., & Comas, D. (2019). Heterogeneity in Palaeolithic population continuity and Neolithic expansion in North Africa. *Current Biology*, *29*(22), 3953–3959.e3954. https://doi.org/10.1016/j.cub.2019.09.050

Sokoloff, P., Giros, B., Martres, M., Andrieux, M., Besancon, R., Pilon, C., Bouthenet, M. L., Souil, E., & Schwartz, J. (1992). Localization and function of the D3 dopamine receptor. *Arzneimittel-Forschung*, *42*(2A), 224–230.

Stefansson, H., Sigurdsson, E., Steinthorsdottir, V., Bjornsdottir, S., Sigmundsson, T., Ghosh, S., Brynjolfsson, J., Gunnarsdottir, S., Ivarsson, O., Chou, T. T., Hjaltason, O., Birgisdottir, B., Jonsson, H., Gudnadottir, V. G., Gudmundsdottir, E., Bjornsson, A., Ingvarsson, B., Ingason, A., Sigfusson, S., ... Stefansson, K. (2002). Neuregulin 1 and susceptibility to schizophrenia. *The American Journal of Human Genetics*, *71*(4), 877–892. https://doi.org/10.1086/342734

Steinthorsdottir, V., Stefansson, H., Ghosh, S., Birgisdottir, B., Bjornsdottir, S., Fasquel, A. C., Olafsson, O., Stefansson, K., & Gulcher, J. R. (2004). Multiple novel transcription initiation sites for NRG1. *Gene*, *342*(1), 97–105. https://doi.org/10.1016/j.gene.2004.07.029

Stephens, M., & Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *The American Journal of Human Genetics*, *76*(3), 449–462. https://doi.org/10.1086/428594

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics*, *68*(4), 978–989. https://doi.org/10.1086/319501

Sun, P. C., Uppaluri, R., Schmidt, A. P., Pashia, M. E., Quant, E. C., Sunwoo, J. B., Gollin, S. M., & Scholnick, S. B. (2001). Transcript map of the 8p23 putative tumor suppressor region. *Genomics*, *75*(1–3), 17–25. https://doi.org/10.1006/geno.2001.6587

Utsunomiya, K., Shinkai, T., Sakata, S., Yamada, K., Chen, H. I., De Luca, V., Hwang, R., Ohmori, O., & Nakamura, J. (2012). Genetic association between the dopamine D3 receptor gene polymorphism (Ser9Gly) and tardive dyskinesia in patients with schizophrenia: a reevaluation in East Asian populations. *Neuroscience Letters*, *507*(1), 52–56. https://doi.org/10.1016/j.neulet.2011.11.050

Wei, J., Chu, C., Wang, Y., Yang, Y., Wang, Q., Li, T., Zhang, L., & Ma, X. (2012). Association study of 45 candidate genes in nicotine dependence in Han Chinese. *Addictive Behaviors*, *37*(5), 622–626. https://doi.org/10.1016/j.addbeh.2012.01.009

Willing, E. M., Dreyer, C., & Van Oosterhout, C. (2012). Estimates of genetic differentiation measured by F ST do not necessarily require large sample sizes when using many SNP markers. *PLoS One*, *7*(8), e42649. https://doi.org/10.1371/journal.pone.0042649

Yang, B., Niu, W., Chen, S., Xu, F., Li, X., Wu, X., Cao, Y., Zhang, R., Yang, F., Wang, L., Li, W., Xu, Y., He, L., & He, G. (2016). Association study of dopamine receptor genes polymorphisms with the risk of schizophrenia in the Han Chinese population. *Psychiatry Research*, *245*, 361–364. https://doi.org/10.1016/j.psychres.2016.08.052

Zainal Abidin, S., Tan, E. L., Chan, S. C., Jaafar, A., Lee, A. X., Abd Hamid, M. H. N., Abdul Murad, N. A., Pakarul Razy, N. F., Azmin, S., Ahmad Annuar, A., Lim, S. Y., Cheah, P. S., Ling, K. H., & Ibrahim, N. M. (2015). DRD and GRIN2B polymorphisms and their association with the development of impulse control behaviour among Malaysian Parkinson's disease patients. *BMC Neurology*, *15*(1), 1–10. https://doi.org/10.1186/s12883-015-0316-2

Zhang, F., Fan, H., Xu, Y., Zhang, K., Huang, X., Zhu, Y., Sui, M., Sun, G., Feng, K., Xu, B., Zhang, X., Su, Z., Peng, C., & Liu, P. (2011). Converging evidence implicates the dopamine D3 receptor gene in vulnerability to schizophrenia. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *156*(5), 613–619. https://doi.org/10.1002/ajmg.b.31203

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.