



The radial expansion of the Diego blood group system polymorphisms in Asia: mark of co-migration with the Mongol conquests

Florence Petit¹ · Francesca Minnai^{1,2} · Jacques Chiaroni^{1,3} · Peter A. Underhill⁴ · Pascal Bailly^{1,3} · Stéphane Mazières¹ · Caroline Costedoat¹

Received: 24 February 2018 / Revised: 9 June 2018 / Accepted: 18 July 2018 / Published online: 24 August 2018

© The Author(s) 2018. This article is published with open access

Abstract

Red cell polymorphisms can provide evidence of human migration and adaptation patterns. In Eurasia, the distribution of Diego blood group system polymorphisms remains unaddressed. To shed light on the dispersal of the Di^a antigen, we performed analyses of correlations between the frequencies of DI^*01 allele, C2-M217 and C2-M401 Y-chromosome haplotypes ascribed as being of Mongolian-origin and language affiliations, in 75 Eurasian populations including DI^*01 frequency data from the HGDP-CEPH panel. We revealed that DI^*01 reaches its highest frequency in Mongolia, Turkmenistan and Kyrgyzstan, expanding southward and westward across Asia with Altaic-speaking nomadic carriers of C2-M217, and even more precisely C2-M401, from their homeland presumably in Mongolia, between the third century BCE and the thirteenth century CE. The present study has highlighted the gene-culture co-migration with the demographic movements that occurred during the past two millennia in Central and East Asia. Additionally, this work contributes to a better understanding of the distribution of immunogenic erythrocyte polymorphisms with a view to improve transfusion safety.

Introduction

Central and East Asia underwent key human expansions since Paleolithic that have contributed to the present-day repartition of many cultural and biological features in Asia and beyond [1–5] (Fig. 1). Notably, archeological and historical records pointed out a reinforcement of population displacements since the Bronze Age (about the second millennium BCE), implying several main Steppe nomadic populations [6]. This period of expansion and exploration is

first exemplified by the Indo-European Andronovo culture, which appeared throughout the South Russian steppe and Kazakhstan during the second millennium BCE then diffused eastwards to the Upper Yenisei in the Altai mountains, westwards in the Ural mountains, and southwards until the Amu-Darya basin [7].

The Iron Age Scythians (about 700 to 300 BCE) is another outstanding example of Eurasian expansion. Originating from the Andronovo culture near the Volga river, they occupied the Pontic of the Black Sea (611 BCE), dominated Mesopotamia and Judea, reached Egypt and penetrated several times into Central Europe. When they reigned over the greater part of Central Asia steppes, the Scythians had an important part in the establishment of transcontinental trade, notably the Silk Road [1].

Following the decline of the Scythians, nomadic horsemen peoples flourished in the Altai, near Lake Baikal and in Selenga valley, later migrating to Central Europe where they mingled with the Franco-Germanic populations. Noteworthy were the European Huns led by Attila, who reached Europe at the fourth century CE [8], and the Mongol khans (emperors), who conquered most of Eurasia between the third century BCE and thirteenth century CE.

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41431-018-0245-9>) contains supplementary material, which is available to authorized users.

✉ Caroline Costedoat
caroline.costedoat@univ-amu.fr

¹ Aix Marseille Univ, CNRS, EFS, ADES, Marseille, France

² University of Sassari, Località Piandanna, 07100 Sassari, Italy

³ Etablissement Français du Sang PACA Corse, Marseille, France

⁴ Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

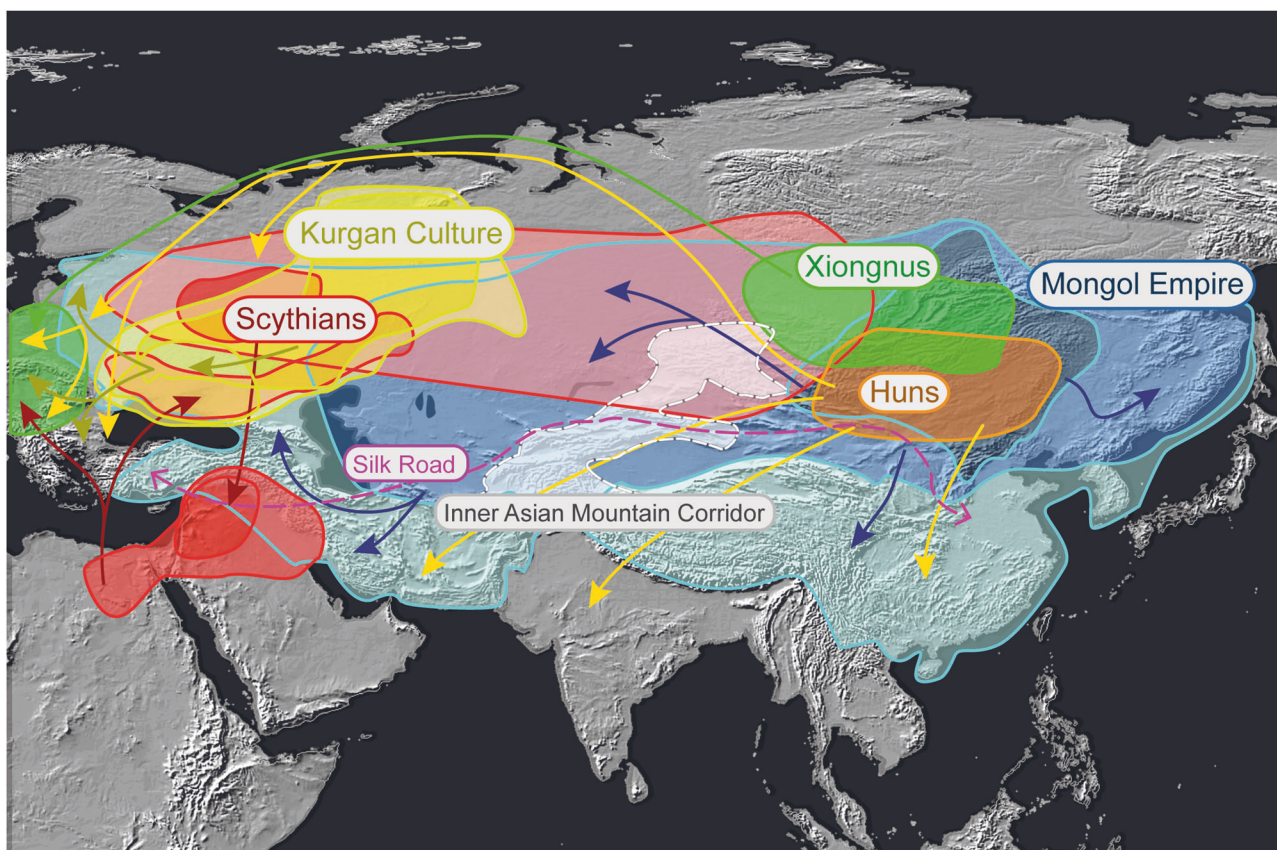


Fig. 1 Main migratory and trade movements in Eurasia since the Neolithic. The nomadic mongoloid Xiongnu people established their first empire in the Northern regions of Mongolia between 209 BCE and 93 CE to migrate thereafter to Central Europe, where they mingled

with the Franco-Germanic populations. The plotting of arrows indicates general trajectories and do not represent the actual course taken during the movements of populations

The khans enjoyed strong social prestige, so did their relatives and descendants, resulting in selection pressure by culture for many generations [6].

The different Steppe nomads were successively Indo-European, Finno-Ugric and Altaic-speakers [6]. Altaic languages include at least Turkic, Mongolic and Tungusic families and are spoken from Turkey and Moldova to Russian Far East [9] (Fig. 2). At genetic level, the Indo-European language has been related to the major Y-chromosome R1-M17 lineage assigned to the centrifugal expansion of the Yamna culture [10]. Altaic-speaking pastoral nomadic populations are mostly carriers of the pan-Eurasian C2-M217 Y-lineage [11, 12]. C2-M217(xM48) patrilineage (embedding C2-M401 and its derivative C2*-ST) is nowadays amplified in Mongols, populations bordering on Mongolia [11, 13–15] and in north Eurasians [5, 16, 17]. Recent refinements of the geographical patterns and age of C2-M217 and its sub-lineages pointed out C2*-ST as part of the founder paternal lineages of all Mongolic-speaking populations, rather than Genghis Khan himself or his relatives [18]. The phylogeographical pattern of the mtDNA diversity in this area is plural and notably witnesses for the same period the admixture of Eastern and Western Asian lineages [19].

Similarly to the uniparental genetic markers, red cell genetic polymorphisms can also trace back past population expansions [20–22] and additionally, be sensitive to natural selection [23, 24]. Notably, the Diego blood group system is a key anthropological marker, since it has evidenced the peopling of the Americas from Asia about 15,000 years ago (reviewed in ref. [20]). Functionally, the Diego blood group system totalizes 22 antigens dispersed at 16 antigenic sites on the erythrocyte membrane glycoprotein band 3. A single substitution (rs2285644, hg19 chr17:g.42328621G>A) defines two alleles, *DI*02* and *DI*01* responsible for the Di^b and Di^a antigens, respectively. This transition occurs in the *SLC4A1* gene (17q21.31) which codes for band 3 polypeptide. Noticeably, band 3 is the place of several symptomatic patterns. Some *SLC4A1* substitutions results in preeclampsia, renal tubular acidosis, and Southeast Asian Ovalocytosis (SAO) [25], and as far as the present genetic marker is concerned, the Di^b and Di^a antigens are highly involved in cases of hemolytic disease of the newborn (HDNB). Hence, such implication suggests a putative contribution of natural selection in the present-day repartition of the Di^a antigen amongst populations. But given that its occurrence in Eurasia roughly coincides with the

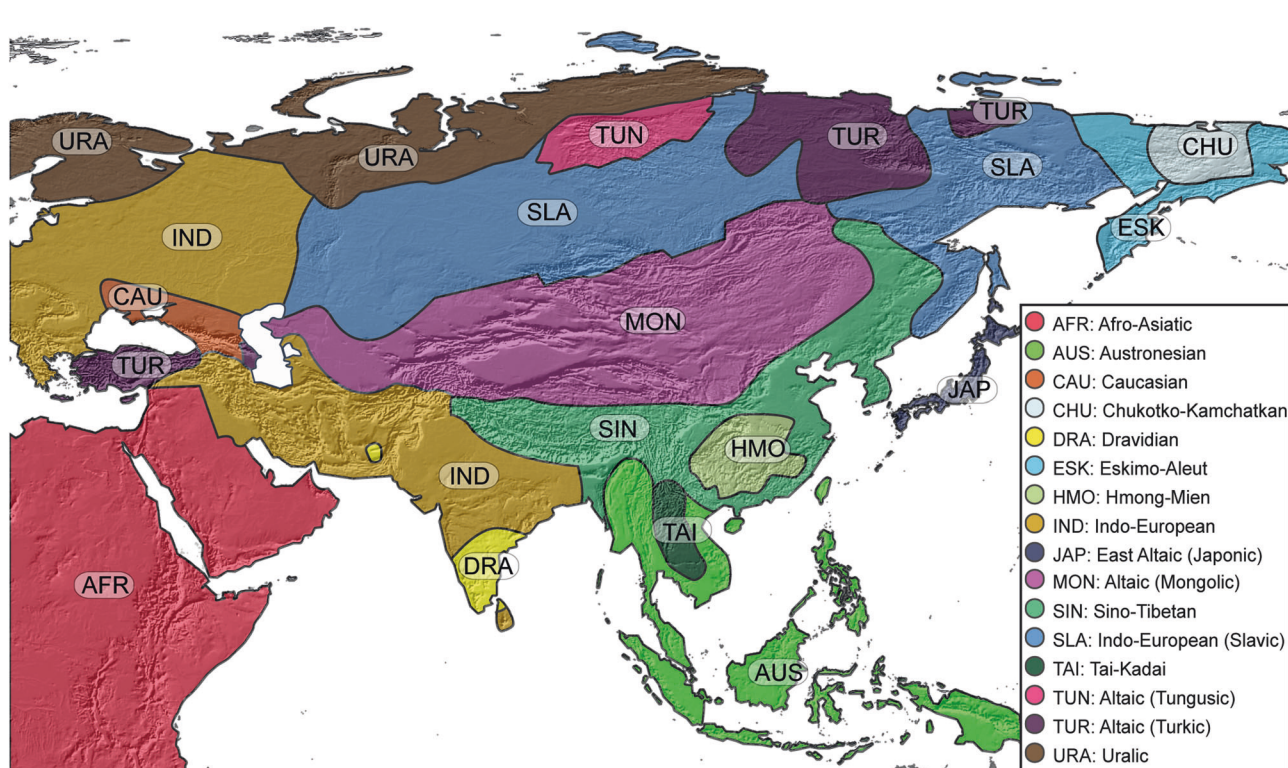


Fig. 2 Main pan-Eurasian linguistic families with wide geographical coverage. All isolates and small centers are not represented

boundaries of the Mongol empires, the actual explanation of the present-day geographical repartition of Di^a in Eurasia remains unaddressed.

To tackle this question, we generated original $DI*01$ data and compared them with previously published data for 75 Eurasian populations screened for the Diego blood group and the Y-chromosome. The effect of the geographical coordinates, linguistics and amount of Y haplogroups and variance on $DI*01$ distribution received statistical support to discard the role of natural selection and to assume that the Di^a antigen has co-migrated with the Mongol expansions.

Materials and methods

Allele frequency compilation

Twenty-one Eurasian populations of the HGDP-CEPH panel, representing original data in the present study, were genotyped for rs2285644G>A (Table S1), in addition to four Kyrgyz, four Afghan and eight Pakistani populations previously screened in refs. [15] and [26]. Primers, probes, amplification and capillary conditions were taken from the multiplex SNaPshot array described

in ref. [26]. SLC4A1 gene variants have been submitted in public database at <https://databases.lovd.nl/shared/transcripts/00019010> (Individual IDs from: 00164740 to 00164758).

We completed with $DI*01$ allelic frequency data from 38 Eurasian populations from previous studies (Table S1). If only the number of Di^a positive phenotypes was mentioned so that it was not possible to distinguish homozygous from heterozygous, the frequency of the $DI*01$ allele was estimated using Bernstein's gene-counting method with the formula $p(DI*01) = 1 - \sqrt{p[1 - f(Di(a+))]}$, where $f(Di(a+))$ is the proportion of Di^a antigen carriers in the population [27]. By assuming Hardy-Weinberg equilibrium, this method allows calculation of gene frequencies from phenotype frequencies even in the presence of ambiguous cases (e.g., hidden and recessive alleles) but prevents any further selection tests from those frequencies. Data were then plotted onto a single map using the Kriging algorithm of SURFER software version 12.0 (Golden Software, LLC).

For all these same 75 populations, we then merged the $DI*01$ allele frequency data with Y-chromosome C2-M217 and derivatives frequency data from previous studies (Table S1, ISOGG v. 12.90, March 2017, https://isogg.org/tree/ISOGG_HapgrpC.html).

Depicting the spread of *DI*01* allele across Eurasia

Detecting signal of co-migration

To characterize the geographical dispersal of *DI*01* across Eurasia, the non-parametric Spearman's rank correlation coefficient was estimated between *DI*01* frequency, latitude, longitude, and frequencies of C2-M217 and C2-M401.

In addition, simple logistic regression models using R software version 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria) and XLSTAT software version 2017.01 were adjusted to explain the presence of *DI*01* by geographical coordinates.

Analysis of variance influenced by language categories

In addition to allele frequencies, we collected the linguistic affiliation rallied in 7 well-represented categories for accurate statistical analyzes (Table S1) using Ethnologue (<https://www.ethnologue.com/>), Glottolog v. 3.0 (<http://glottolog.org/>) and the World Atlas of Language Structures online (<http://wals.info/>).

One-way ANOVA model was evaluated using the Fisher *F* test to determine whether the amount of information provided by the language factor was significant enough to explain the variance of *DI*01* and C2-M217 frequencies.

Testing the polarity of genetic diversity

With the intention of evidencing a radial expansion from Mongolia to the rest of Asiatic continent, we tested whether the polarity of Y-chromosome genetic diversity agrees with past Mongolian expansions. Indeed, it is accepted that diversity is expected to be the highest at the source population and would decrease as a function of the distance from the source, a.k.a. the founder effect [28–30] or the down-the-line exchange model for archeology records [31]. To this aim, we gathered 234 Y-SNP-STR (Short Tandem Repeats: DYS389I, DYS390, DYS391, DYS392, DYS393, DYS439) haplotypes from 15 Asian populations from Japan to Afghanistan (Table S2).

For each population, we estimated the mean variance encompassed in C2-M217 and C2-M401 individuals, and plotted onto a kriging map. We then estimated the non-parametric Spearman's rank correlation coefficient between mean C2-M217 and C2-M401 variances with latitude and longitude.

Testing natural selection

In order to measure the contribution of natural selection in the repartition of the Di^a antigen, and thus *DI*01*, we ran two Hardy–Weinberg (HW) equilibrium tests on rs2285644

genotypes from the 1000 Genome Project database (1KGP) [32] and HGDP-CEPH panel samples including 26 and 31 Eurasian populations respectively, thanks to PLINK version 1.9 [33].

Results

Spread directional of *DI*01* allele across Eurasia

Figure 3 presents the *DI*01* allele frequency in Eurasia. Highest frequencies were observed in Mongolia (frequency = 0.098) and Turkmenistan (0.121), followed by Kyrgyzstan (mean frequency = 0.081) and Japan (mean frequency = 0.058). Lower mean frequencies occurred in Cambodia (0.050), China (0.049), North eastern China (Daur and Mongol populations), sporadically in Central (Tu) and western (Uyгур) China, and a few pocket areas in South eastern China (Zhuang, Yizu, She and Han), followed by Korea (0.042), Russia and Thailand (0.033). However, the *DI*01* allele frequency was null in some populations of southern (Dai, Lahu, Miao, Naxi and Tujia) and northern (Hezhen and Oroqen) China. The *DI*01* allele was completely absent in Indonesia (Sumatra), Israel (Hebrew), Russia Caucasus (Adygei), and South India (Tamil Nadu, Irula, and Kurumba).

Table 1 presents the correlation test between *DI*01* allele frequency, geography, and frequencies of C2-M217 and C2-M401. Results indicated that the amount of *DI*01* correlated positively with latitude ($\rho = 0.291$, $p = 0.011$), longitude ($\rho = 0.297$, $p = 0.010$) and amount of Mongolian-origin Y haplogroups ($\rho = 0.490$, $p = 0.018$; $\rho = 0.428$, $p = 0.042$). Similarly, C2-M217 frequency varied as latitude and longitude did ($\rho > 0.520$, $p < 0.011$). Alternatively said, frequencies of *DI*01* decreased along with amounts of C2-M217 and C2-M401 toward South and West Asia.

In addition, adjusted simple logistic regression model showed that the presence of *DI*01* could be explained by longitude ($p = 0.098$), in case significance threshold taken was 0.1 (Table S3).

Link with cultural linguistic traits

The *DI*01* allele distribution portrayed in relation to language showed highest frequencies in Altaic-speakers (mean frequency = 0.071), followed by Eskimo-Aleut, Uralic, Chukchi Kamchatkan and North Caucasian speakers (0.029) from North Eurasia and the Hmong/Tai-Kadai speakers (0.027) from South eastern China. The *DI*01* allele is almost complete absent (< 0.015) in the Austronesian/Asiatic speakers from Austronesia and the Indo-European family from Middle East and West Asia.

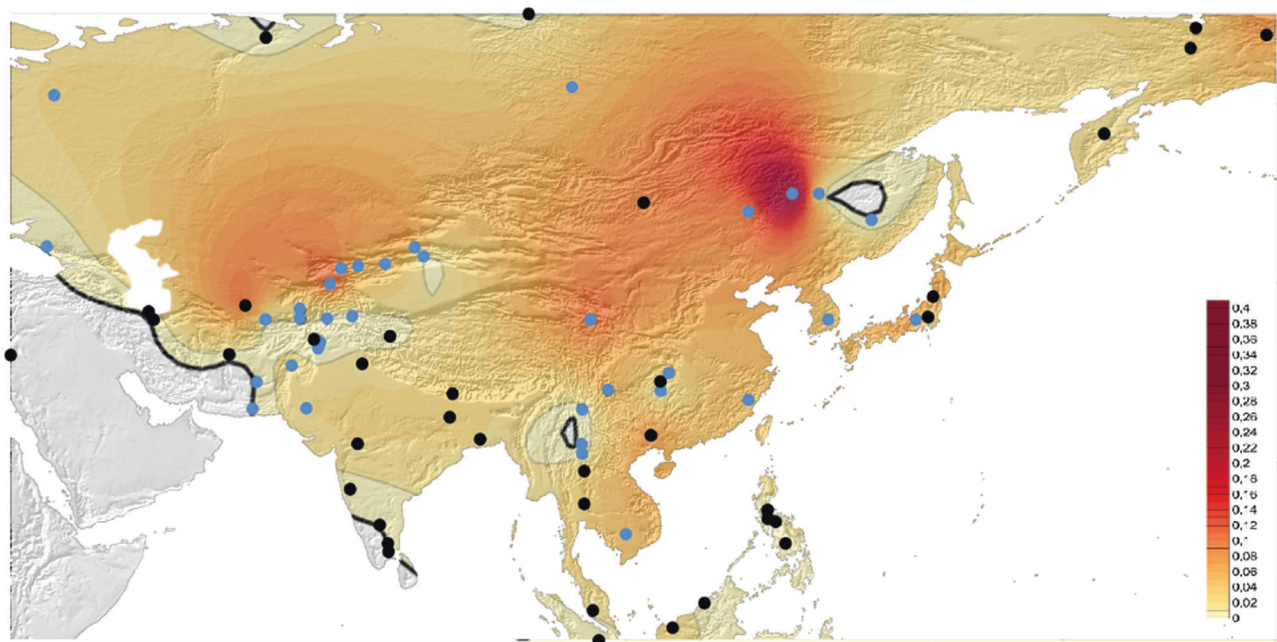


Fig. 3 Mapping of *DI*01* allele frequency amongst 75 Eurasian populations. Black dots: populations from previous surveys. Blue dots: 21 populations of the HGDP-CEPH panel genotyped in this present

study, and in ref. [26] (Table S1). The frequency range 0–0.01 is indicated by marked black lines (in bold for 0, simple for 0.01)

Table 1 Spearman’s rank correlation ρ coefficient between *DI*01* allele frequency in 75 Eurasian populations, geography, C2-M217, and C2-M401 haplogroup frequencies

	Latitude (°N)	Longitude (°E)	C2-M217 freq.	C2-M401 freq.
	ρ (<i>p</i> -value)	ρ (<i>p</i> -value)	ρ (<i>p</i> -value)	ρ (<i>p</i> -value)
<i>DI*01</i> allele frequency	0.291 (0.011)	0.297 (0.010)	0.490 (0.018)	0.428 (0.042)
C2-M217 frequency	0.520 (0.011)	0.623 (0.002)	—	—

ρ estimated value of non-parametric Spearman’s rank correlation coefficient, *p*-value pairwise two-sided *p*-value, *freq.* frequency

Table 2 Fisher’s *F* test of the amount of information provided by the language factor in comparison with the mean allele frequency of *DI*01*, C2-M217, and C2-M401 in 75 Eurasian populations

	D° of freedom	<i>DI*01</i> allele freq.	C2-M217 freq.	C2-M401 freq.
		<i>F</i> (<i>p</i> -value)	<i>F</i> (<i>p</i> -value)	<i>F</i> (<i>p</i> -value)
Language category	6	3.926 (0.002)	4.155 (0.001)	2.080 (0.070)

freq. frequency

Table 2 presents the effects of language factor in the dispersal of *DI*01*, C2-M217, and C2-M401. One-way ANOVA shows that the variances of *DI*01* allele (Fisher’s $p = 0.002$) and C2-M217 (Fisher’s $p = 0.001$) frequencies were significantly explained by the linguistic background.

Polarity of genetic diversity with Y-STRs

Table 3 presents the Spearman’s rank correlation results between the mean variance of the Y-STR markers for C2-M217, C2-M401 groups (Table S4) and geographical coordinates. A significant correlation ($p = 0.010$; $p = 0.042$) was

obtained only with latitude, so that as variance increased so did latitude. These results were concordant with [12] with significant correlation coefficient with latitude, but our data did not allow to provide the same finding with longitude.

Then, we mapped the Y-STRs mean variance Mongolian-origin for C2-M217 and C2-M401 (Fig. 4). Y-STRs variance for C2-M217 was higher in northern than in southern populations, ranging from 0.431 (Northern Hans in China), to less than 0.100 in the Hazaras from Pakistan (0.096) and even 0.000 in Cambodia and for the Burusho population in Pakistan (Fig. 4a). Figure 4b shows a hotspot of mean variance of C2-M401 in Mongolia (0.105), decreasing towards Pakistan.

Testing natural selection

Runs of HW equilibrium test allowed to discard natural selection at the rs2285644 locus in all populations from the 1KG project and HGDP-CEPH panel ($p_{\chi^2} = 1$) but 2: Han ($p_{\chi^2} = 0.002$) and Tu ($p_{\chi^2} = 0.027$) from China (Table S5).

Discussion

Historical marks of co-migration from Mongolia and the Diego-Y-chromosome duet in Central Asia

The C2-M217xM48 male-lineage, which amplified in Altaic-speaking pastoral nomadic groups [12, 18] signposts significantly the East-to-West Mongolian expansions which may have also towed the Diego blood group polymorphism across Eurasia. Additional support of our assumption is

Table 3 Spearman's rank correlation ρ coefficient (p -value) of the Y-STRs mean variance for the two C2-M217 and C2-M401 population groups

Y-STRs variance	Latitude (°N)	Longitude (°E)
C2-M217 group (15 populations)	0.618 (0.010)	0.104 (0.713)
C2-M401 group (6 populations)	0.829 (0.042)	0.257 (0.623)

given by the contrasting pattern observed in India, where *DI*01* has been only detected in the northern populations (Bhil Madhya, Rajbanshi Bengali, Bihar, Oraon, and Punjab) and further North of India (Nosherpa, Nepal, Sindh, Balochi and Hazara, Pakistan) populations, with null *DI*01* frequencies in the South (Tamil Nadu, Irula, and Kurumba populations) and South-East. This pattern could be potentially explained by the conquests in India following the Mongolian expansion. The Mughal—or Mogul—Empire (1526–1857) was founded in northern India in 1526 by Babur, descendant of Tamerlane (1336–1405), first ruler in the Timurid dynasty, commonly considered nomadic Turco-Mongol conqueror of Eurasian Steppe [34] and of Genghis Khan. At the end of Akbar's reign in 1605, the Empire stretched from Punjab, Bengal to Gujarat. Some territories were notwithstanding conquered in a southerly direction in the mid-seventeenth century [35] leaving no visible genetic traces using our data.

Several independent genetic features, such as the *EDAR* and *ADH1B*47His* alleles whose fixation would have started through positive selection at the earliest stages of Neolithic, also showed a clear geographical cline amongst East Asians due to human expansion [36, 37], but preceding the Mongols.

Worldwide, very few non-Mongol-origin individuals have been identified as carrying Di^a antigen [38–40]. Sporadic occurrence of *DI*01* in population bordering on

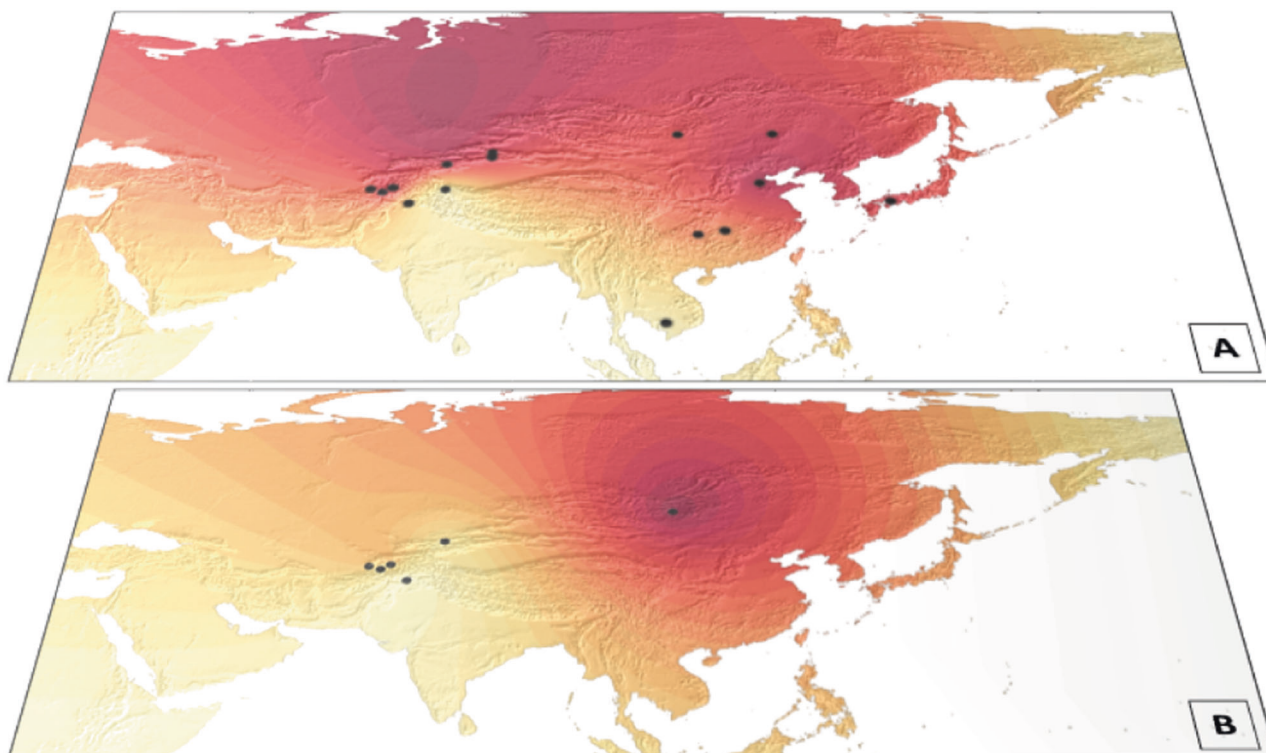


Fig. 4 a, b Mapping of the mean variance of Y-STRs amongst 15 Eurasian populations. **a** Mean variance of C2-M217 carriers from

15 populations (see further details in Table S4). **b** Mean variance of C2-M401 carriers from 6 populations (included in the 15)

the Mongolian expansion, such as in Poland that has been invaded by the Mongolian Tartars during the thirteenth century and between the fifteenth and seventeenth centuries [41] and in Afghanistan and Pakistan populations close to the Hazara [26], might originate from admixture.

Thus, we relied on the sharp cut signal of the Y-chromosome to point out the polarity of the Di^a expansion. The pattern is consistent with an East Asian source that have expanded through the several Mongol expansions, strengthened by linguistic background and also in concordance with the diffusion of a package of East Asian mtDNAs as presented in ref. [19].

Could the presence of Di^a in continental Asia populations also be attributable to natural selection?

Given the role of band 3 onto which the Diego antigens are erected, natural selection is expected to have acted at some point. Noteworthy is the association of the Di^a antigen with severe or even fatal HDNB, that would have strongly disrupted the penetration of the *DI*01* allele in a Di^b world. As previously described in Native Amerindians [20], the role of natural selection at the rs2285644 locus in continental Eurasian populations do not receive support from the present data.

In-depth examination of the *DI*01* allele

To attempt to spot, within the *SLC4A1* gene, alleles usually associated with the *DI*01* allele, we have estimated by using PLINK version 1.9, the allele frequencies of SNPs belonging to this gene and available in the 1KGP. Positive and significant Spearman's rank correlation coefficients were found between the distribution of rs2285644 (derived nucleotide: A) and that of SNP mutations previously known to be responsible for preeclampsia, renal tubular acidosis, and mostly SAO [25]; the two latter being usually associated [42]. To go further, among the 12 mutations known to be responsible for the SAO [43], we have identified 4 SNPs: rs5036 (hg19 chr17:g.42338945T>C, ancestral nucleotide: C, responsible for band 3-Memphis non-synonymous polymorphism [44]), rs16940582 (hg19 chr17:g.42339745C>T, derived nucleotide: T), rs16940585 (hg19 chr17:g.42339762G>A, derived nucleotide: A) and rs2521602 (hg19 chr17:g.42336424G>A, derived nucleotide: A), whose distribution was always, worldly and only at Asiatic scale too, correlated with rs2285644 (A). These different results could be considered as a track of association between Diego polymorphism and SAO. Taking into account the sparse geographical coverage of the sample used (10 Asian among 26 populations), the results should be confirmed by additional studies.

To conclude with, the aim of this study was to disentangle the geographical distribution of the *DI*01* allele of the Diego blood group system in Eurasia. Our data demonstrated large variations in frequency ranges with a hotspot in Mongolia, and a significant positive correlation with geographical coordinates. Our findings suggested that *DI*01* carriers could have crossed into West Asia from Mongolia as indicative of a striking historical and linguistic event: the expansions of Altaic-speaking pastoral nomads from Mongolia with high reproductive success. In a broader context, the description and understanding of the present-day geographical distribution of red cell phenotypes come under a multidisciplinary approach as herein developed. Beyond the anthropological interest, the elucidation of well-circumscribed areas of immunologic polymorphisms is a crucial step to ensure in blood transfusion safety, and this is, in our opinion, what the present study contributes to.

Acknowledgements This work was partially supported by APR 2016-11-MAZIERES-AM. FM was supported by the Erasmus + program of the European Union.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Frachetti MD. Multiregional emergence of mobile pastoralism and nonuniform institutional complexity across Eurasia. *Curr Anthropol*. 2012;53:2–38.
2. Keyser C, Bouakaze C, Crubézy E, Nikolaev VG, Montagnon D, Reis T, et al. Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Hum Genet*. 2009;126:395–410.
3. Shi H, Qi X, Zhong H, Peng Y, Zhang X, Ma RZ, et al. Genetic evidence of an East Asian origin and Paleolithic Northward migration of Y-chromosome haplogroup N. *PLoS ONE*. 2013;8:e66102.
4. Yunusbayev B, Metspalu M, Metspalu E, Valeev A, Litvinov S, Valiev R, et al. The Genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genet*. 2015;11:e1005068.
5. Zhong H, Shi H, Qi X-B, Xiao C-J, Jin L, Ma RZ, et al. Global distribution of Y-chromosome haplogroup C reveals the

- prehistoric migration routes of African exodus and early settlement in East Asia. *J Hum Genet.* 2010;55:428–35.
6. Anthony DW. The horse, the wheel, and language: how Bronze-Age riders from the Eurasian Steppes shaped the modern world. Princeton: Princeton University Press; 2007.
 7. Koryakova L, Epimakhov AV. The Urals and Western Siberia in the Bronze and Iron Ages. Cambridge: Cambridge University Press; 2007.
 8. Thompson EA. The Huns. Oxford: Wiley-Blackwell; 1999.
 9. Ruhlen M. A guide to the world's languages. Volume 1. Classification. Stanford: Stanford University Press; 1987.
 10. Semino O. The genetic legacy of Paleolithic *Homo sapiens* in extant Europeans: a Y-chromosome perspective. *Science.* 2000;290:1155–9.
 11. Dulik MC, Osipova LP, Schurr TG. Y-chromosome variation in Altaian Kazakhs reveals a common paternal gene pool for Kazakhs and the influence of Mongolian expansions. *PLoS ONE.* 2011;6:e17548.
 12. Balaresque P, Poulet N, Cussat-Blanc S, Gerard P, Quintana-Murci L, Heyer E, et al. Y-chromosome descent clusters and male differential reproductive success: young lineage expansions dominate Asian pastoral nomadic populations. *Eur J Hum Genet.* 2015;23:1413–22.
 13. Abilev S, Malyarchuk B, Derenko M, Wozniak M, Grzybowski T, Zakharov I. The Y-chromosome C3* star-cluster attributed to Genghis Khan's descendants is present at high frequency in the Kerey clan from Kazakhstan. *Hum Biol.* 2012;84:79–89.
 14. Katoh T, Munkhbat B, Tounai K, Mano S, Ando H, Oyungere G, et al. Genetic features of Mongolian ethnic groups revealed by Y-chromosomal analysis. *Gene.* 2005;346:63–70. Feb.
 15. Di Cristofaro J, Pennarun E, Mazieres S, Myres NM, Lin AA, Temori SA, et al. Afghan Hindu Kush: where Eurasian subcontinent gene flows converge. *PLoS ONE.* 2013;8:e76748.
 16. Malyarchuk B, Derenko M, Denisova G, Wozniak M, Grzybowski T, Dambueva I, et al. Phylogeography of the Y-chromosome haplogroup C in northern Eurasia: Y-chromosome haplogroup C phylogeography. *Ann Hum Genet.* 2010;74:539–46.
 17. Zhong H, Shi H, Qi X-B, Duan Z-Y, Tan P-P, Jin L, et al. Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Mol Biol Evol.* 2011;28:717–27.
 18. Wei L-H, Yan S, Lu Y, Wen S-Q, Huang Y-Z, Wang L-X, et al. Whole-sequence analysis indicates that the Y chromosome C2*-Star Cluster traces back to ordinary Mongols, rather than Genghis Khan. *Eur J Hum Genet.* 2018. <https://www.nature.com/articles/s41431-017-0012-3>
 19. Cheng B, Tang W, He L, Dong Y, Lu J, Lei Y, et al. Genetic imprint of the Mongol: signal from phylogeographic analysis of mitochondrial DNA. *J Hum Genet.* 2008;53:905–13.
 20. Bégat C, Bailly P, Chiaroni J, Mazieres S. Revisiting the Diego blood group system in Amerindians: evidence for gene-culture comigration. *PLoS ONE.* 2015;10:e0132211.
 21. Dugoujon J-M, Hazout S, Loirat F, Mourrieras B, Crouau-Roy B, Sanchez-Mazas A. GM haplotype diversity of 82 populations over the world suggests a centrifugal model of human migrations. *Am J Phys Anthropol.* 2004;125:175–92.
 22. Villanea FA, Bolnick DA, Monroe C, Worl R, Cambra R, Leventhal A, et al. Brief communication: evolution of a specific O allele (O1vG542A) supports unique ancestry of Native Americans. *Am J Phys Anthropol.* 2013;151:649–57.
 23. Fumagalli M, Cagliani R, Pozzoli U, Riva S, Comi GP, Menozzi G, et al. Widespread balancing selection and pathogen-driven selection at blood group antigen genes. *Genome Res.* 2009;19:199–212.
 24. Baum J, Ward RH, Conway DJ. Natural selection on the erythrocyte surface. *Mol Biol Evol.* 2002;19:223–9.
 25. Jarolim P, Palek J, Amato D, Hassan K, Sapak P, Nurse GT, et al. Deletion in erythrocyte band 3 gene in malaria-resistant Southeast Asian ovalocytosis. *Proc Natl Acad Sci USA.* 1991;88:11022–6.
 26. Mazieres S, Temory SA, Vasseur H, Gallian P, Di Cristofaro J, Chiaroni J. Blood group typing in five Afghan populations in the North Hindu-Kush region: implications for blood transfusion practice. *Transfus Med.* 2013;23:167–74.
 27. Bernstein F. Zusammenfassende Betrachtungen über die erblichen Blutstrukturen des Menschen. *Z Für Indukt Abstamm-Vererb.* 1925;37:237–370.
 28. Chiaroni J, Underhill PA, Cavalli-Sforza LL. Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc Natl Acad Sci USA.* 2009;106:20174–9.
 29. Provine WB. Ernst Mayr: genetics and speciation. *Genetics.* 2004;167:1041–6.
 30. Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci USA.* 2005;102:15942–7.
 31. Renfrew LC. The emergence of civilization: Cyclades and the Aegean in the third millennium B.C. London: Methuen Publishing Ltd; 1972.
 32. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. *Nature.* 2015;526:68–74.
 33. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
 34. Manz BF. The rise and rule of Tamerlane. Cambridge: Cambridge University Press; 1989.
 35. Sharma SR. Mughal empire in India: a systematic study including source material. New Delhi: Atlantic Publishers & Dist; 1999.
 36. Bryk J, Hardouin E, Pugach I, Hughes D, Strotmann R, Stoneking M, et al. Positive Selection in East Asians for an EDAR Allele that Enhances NF- κ B Activation. *PLoS ONE.* 2008;3:e2209.
 37. Peng Y, Shi H, Qi X-B, Xiao C-J, Zhong H, Ma RZ, et al. The ADH1B Arg47His polymorphism in East Asian populations and expansion of rice domestication in history. *BMC Evol Biol.* 2010;10:15.
 38. Issitt PD, Wren MR, Rueda E, Maltz M. Red cell antigens in Hispanic blood donors. *Transfusion.* 1987;27:117.
 39. Levine P, Robinson EA, Layrisse M, Arends T, Domingues Sisco R. The Diego blood factor. *Nature.* 1956;177:40–1.
 40. Simmons RT. The apparent absence of the Diego (Dia) and the Wright (Wra) blood group antigens in Australian aborigines and in New Guineans. *Vox Sang.* 1970;19:533–6.
 41. Kuśnierz-Alejska G, Bochenek S. Haemolytic disease of the newborn due to anti-Dia and incidence of the Dia antigen in Poland. *Vox Sang.* 1992;62:124–6.
 42. Bruce LJ, Wrong O, Young MT, et al. Band 3 mutations, renal tubular acidosis and South-East Asian ovalocytosis in Malaysia and Papua New Guinea: loss of up to 95% band 3 transport in red cells. *Biochem J.* 2000;350:41–51.
 43. Paquette AM, Harahap A, Laosombat V, Patnode JM, Satyagraha A, Sudoyo H, et al. The evolutionary origins of Southeast Asian Ovalocytosis. *Infect Genet Evol.* 2015;34:153–9.
 44. Yannoukakos D, Vasseur C, Driancourt C, Blouquait Y, Delaunay J, Wajcman H, et al. Human erythrocyte band 3 polymorphism (band 3 Memphis): characterization of the structural modification (Lys 56-Glu) by protein chemistry methods. *Blood.* 1991;78:1117–20.