



**HAL**  
open science

# Rapid change in red cell blood group systems after the main Out of Africa of Homo sapiens

Stéphane Mazières, Silvana Condemi, Wassim El Nemer, Jacques Chiaroni

## ► To cite this version:

Stéphane Mazières, Silvana Condemi, Wassim El Nemer, Jacques Chiaroni. Rapid change in red cell blood group systems after the main Out of Africa of Homo sapiens. *Scientific Reports*, 2025, 15 (1), pp.1597. <10.1038/s41598-024-83023-0>. <hal-04878570>

**HAL Id: hal-04878570**

**<https://hal.science/hal-04878570v1>**

Submitted on 10 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License



## OPEN Rapid change in red cell blood group systems after the main Out of Africa of *Homo sapiens*

Stéphane Mazières<sup>1✉</sup>, Silvana Conde mi<sup>1</sup>, Wassim El Nemer<sup>1,2</sup> & Jacques Chiaroni<sup>1,2</sup>

Despite the advances in paleogenomics, red cell blood group systems in ancient human populations remain scarcely known. Pioneer attempts showed that Neandertal and Denisova, two archaic hominid populations inhabiting Eurasia, expressed blood groups currently found in sub-Saharan and a rare “rhesus”, part of which is found in Oceanians. Herein we fully pictured the blood group genetic diversity of 22 *Homo sapiens* and 14 Neandertals from Eurasia living between 120,000 and 20,000 years before present (yBP). From the ABO, Rh, Kell, Duffy, Kidd, MNS, Diego, H, secretor and Indian systems, we noted that the blood group allele diversity in the Neandertals remained unchanged since 120,000 yBP, while *H. sapiens* conquered Eurasia with blood group alleles presently exclusive to non-African populations, suggesting they may have differentiated right after the Out of Africa, between 70,000 and 45,000 yBP. Notably, Ust’Ishim possessed unknown alleles that may illustrate the lost genetic heritage of the early Eurasians. Lastly, Neandertals shared a unique Rh haplotype from which we updated the current *RHD* phylogeny. The contribution of this study is twofold. It enlightens the expansion patterns of *H sapiens* and recalls the anthropological effectiveness of genetic polymorphisms currently being surveyed for transfusion safety and pregnancy monitoring.

**Keywords** blood group systems, Upper Paleolithic, *Homo sapiens*, Neandertal, Out of Africa

The encounter of *Homo sapiens* with Neandertal and Denisova, two archaic populations inhabiting Eurasia since 400,000 yBP (years before present), has led to a major cultural and biological disruption<sup>1</sup>. Despite few forays outside of Africa before 200,000 yBP<sup>2,3</sup>, *Homo sapiens* recurrently met with Neandertal since 100,000 yBP in the Levant and Zagros mountains, and have mated with them almost immediately<sup>4,5</sup>. Material and paleoecological evidence also revealed that *Homo sapiens* successfully conquered the Arabian Peninsula as early as 125,000 yBP, indicating that *Homo sapiens* went East first and fast, then North<sup>6</sup>. In particular, *Homo sapiens* had already penetrated Australia 20,000 km away from the Levant since 60,000 yBP<sup>7</sup>, while it did not reach Northwestern Europe by 3,500 km until around 45,000 yBP<sup>8</sup>. Lastly, the expansion of *Homo sapiens* in Eurasia coincides with the disappearance of several archaic hominid populations, marking the end of the Middle Paleolithic ca. 45 – 35,000 yBP.

In the context of the biological and cultural transition that occurred during the Upper Paleolithic (UP), the contribution of paleogenomics has been outstanding. One remarkable achievement is the deduction of the social structure of a Neandertal group<sup>9,10</sup>. On a wider scale, paleogenomics highlighted a schism of gene diversity between the archaic and *Homo sapiens* populations. On the one hand, the heterozygosity of the late Neandertals and Denisova was up 5 times lower than the one of modern-day populations, reflecting inbreeding and declining demography<sup>11</sup>. On the other hand, the *H. sapiens* of the early stages of UP segregated into distinct genetic clusters depicting a higher gene diversity than the previous inhabitants of Eurasia<sup>12,13</sup>. Before extinction, Denisova and the Late Neandertals have passed a part of their gene pool down to *H. sapiens*, visible in phenotype traits, including red cell blood group systems<sup>14,15</sup>.

Red cell polymorphisms comprise the biological variation in three areas: the cell interior (hemoglobin, enzymes), membrane components (membrane skeleton, water and ion exchange channels) and membrane surface (blood group antigens). For several decades, they constituted the classical genetic markers that initiated the description of human variability in the mid-20th century<sup>16</sup>. However, the emergence of molecular biology in the 1980s and the screening of mitochondrial DNA, the Y chromosome and short and long interspersed repetitive elements relegated blood group systems to their main application, blood transfusion<sup>17</sup>. It was the advent of genomic sequencing that restored the anthropological and evolutionary interests in all red blood cell genetic polymorphisms.

<sup>1</sup>Aix Marseille Univ, CNRS, EFS, ADES, Marseille, France. <sup>2</sup>Etablissement Français du Sang PACA Corse, Marseille, France. ✉email: stephane.mazieres@univ-amu.fr

Indeed, the molecular approach of red blood cell polymorphisms allowed to exemplify all evolutionary forces behind human genetic diversity. Genetic drift and migration are explained by the allele distribution of the DI (Diego) system<sup>18,19</sup>. Selection, either positive and purifying, are represented by the FY (Duffy), MNS and ER systems<sup>20–22</sup>, and KX deficiency<sup>23</sup> respectively. Balanced selection is represented by two types of compromise. The first is the advantage of the heterozygote carrying a deleterious allele in an environment unfavorable to the healthy homozygote: hemoglobin S and band 3 deficiency<sup>24,25</sup>. The second is the maintenance of genetic polymorphism by fluctuating actions due to different pathogens, favoring one or another phenotype in turn: this is the case of the ABO system<sup>26</sup>. Lastly, admixture with archaic populations is illustrated by the RH (named “rhesus” by the broad audience) system<sup>15</sup>. Besides, genomics has made it possible to outline the evolutionary history of red blood cell polymorphisms. For instance, although the A and B antigens of the ABO system are common amongst Primates<sup>26</sup>, the A antigen in humans originates from an *ABO*\*A allele that would have re-emerged recently, in the early *H. sapiens* times ca. 260,000 yBP, from the fusion of *ABO*\*B with *ABO*\*O, themselves derived from an *ABO*\*A allele that probably vanished 2 Ma ago<sup>27</sup>. As far as the RH system is concerned, the ancestral haplotypes linked the *RHCE* alleles coding for the Rhc and Rhe antigens with *RHD* alleles coding for weak and incomplete D antigens, and are mostly found in Sub-Saharan populations and Neandertal<sup>15,28</sup>. The complete D antigen arose later to constitute the Dce haplotype, also called R0 and notably carried by Denisova, from which the seven other DCe, DcE, DCE, dce, dCe, dcE and dCE (or R1, R2, RZ, r, r', r'' and ry) haplotypes derive. Moreover, some of these which are exclusive to non-African populations<sup>15,29</sup>.

To date, there are 47 systems in the red blood cell, but around ten are responsible for transfusion accidents and severe hemolytic diseases of the fetus and the newborn (HDFN) (retrieved from the International Society of Blood Transfusion, ISBT, website <http://www.isbtweb.org>). Recently, these ten blood group systems have entered the debate on understanding the evolutionary history of Neanderthals and Denisovans and the present-day blood group system diversity<sup>14,15,21,22,30</sup>. These pioneer studies on archaic populations have revealed that the sub-Saharan gene diversity extended as far as Siberia 100,000yBP. Noticeable is the variation at the *ABO* locus, the existence of non-secretor alleles presently advantageous, a rare RH haplotype likely introgressed in the ancestors of modern Oceanians, and a *RHD*-*RHCE* haplotype causing today's severe HDFN<sup>14,15</sup>.

In the present study, we aimed to picture the oldest blood group system variability available so far, with two goals in mind: to test whether the Neandertal blood group diversity was specific to the archaic populations or shared with contemporaneous *H. sapiens* populations from the initial Upper Paleolithic and to sketch the evolutionary history of the archaic and *H. sapiens* lineages from blood group systems. For this, we completed the blood group panorama obtained previously for four high-quality archaic individuals<sup>15</sup> with 23 Eurasian *Homo sapiens* and 14 Neandertals living between 120,000 and 20,000 yBP.

## Results

### Description of the blood group systems in Upper Paleolithic individuals

Allele counts, genotypes and comparison with present-day YRI, CHB and CEU allele frequencies are presented in Supplementary Material 1, Tables S5 and S6. At the *ABO* locus, we detected polymorphism for both Neandertals and *H. sapiens*. The genetic background of the O01 blood group (inactive transferase due to c.261delG, *ABO*\*O.01 alleles) was the most frequently observed. Moreover, noteworthy is the occurrence of the non-deletional *ABO*\*O.02.01 allele (c.53G>T, c.220C>T, c.261G, c.297A>G, c.526C>G, c.802G>A) in the related Sunghir2 and Sunghir3, as well as two Neandertals, ChagyrskayaD and GoyetQ56-1. Presently this allele codes for the O03 blood group, an ambiguous O phenotype found in less than 5% of the populations from Western Europe<sup>31–35</sup>. Two *ABO*\*B alleles were found in Siberian and East Asian individuals, respectively Yana2, and AR33K. For the *FUT1* and *FUT2* loci that code for precursor structure of the ABO antigens onto red cells and their secretion in body fluids, two *FUT1*\*01.02 (c.35C>T) alleles were found in two *H. sapiens*, Ust'-Ishim and GoyetQ116-1. *FUT2*\*01 (c.390 C>T) was encountered in high frequency in *H. sapiens* and Ust'-Ishim carried an unknown *FUT1*\*01 c.564 C>T without functional consequence (p.His188=). Interestingly, three individuals, Ust'-Ishim, AR33K and Bacho Kiro BK1653 carry an unknown secretor allele, *FUT2*\*01 (c.390 C>T, c.896 C>T), for which the estimated function might be damaged (CADD C-score: 21.1)<sup>36,37</sup>. Remarkably, non-secretor alleles are exclusive to the Neandertals understudied. Amongst these, we once again found *FUT2*\*01N.04 (c.604 C>T, p.Arg202ter) in Vindija87, previously described in Vindija33.19, Chagyrskaya8, and Denisova3<sup>15</sup>, and exclusive to Asian and Oceanian populations<sup>34,35</sup>.

For the RH system, the Neandertals share the key substitutions of the peculiar haplotype formerly described in the Altai (Denisova5), Chagyrskaya8 and Vindija33.19 Neandertals at the homozygous state (Supplementary Material 2, File S1)<sup>15</sup>. This haplotype, made of *RHD*\*03.04 (c.186G>T, c.410C>T, c.455A>C) plus c.602C>G and c.733G>C in cis with *RHCE*\*01.05 (*RHCE*\*ceEK, c.48G>C, c.712A>G, c.787A>G, c.800T>A), encode for partial RhD, Rhc and Rhe antigens, involved in blood transfusion and pregnancy issues (HDFN). We also suspect one *RHCE*\*01.04 (a.k.a. *RHCE*\*ceAR: c.48G>C, c.712A>G, c.733C>G, c.787A>G, c.800T>A) in Vindija87. Nowadays, this allele is involved in delayed hemolytic transfusion reaction and mild HDFN and is found in Sub-Saharan-origin populations<sup>38</sup>. Amongst *H. sapiens*, all readable tracts match with *RHD*\*01, *RHCE*\*01 and *RHCE*\*03 in three of the most ancient Eurasia *H. sapiens*, Kostenki14, Sunghir IV and Ust'-Ishim. Today, these alleles encode for complete RhD, Rhc, Rhe and RhE antigens. Lastly, Ust'-Ishim has one *RHD*\*01 (c.165 C>T, p.Thr55=, without the c.1213 C>G transversion of the *RHD*\*01W.68 weak D allele), rare in modern-day populations (<0.08%)<sup>34,35,39</sup>.

About the KEL system, *KEL*\*02 is predominant. GoyetQ56-1, the sole Neandertal with readable alleles of the Jsa/Jsb antigens, carries a *KEL*\*02.06 like Altai Neandertal, Vindija33.19 and Chagyrskaya8<sup>15</sup>, while all ancient *H. sapiens* possess *KEL*\*02.07. Note that Ust'-Ishim and Sunghir2 are positive for c.1680 A>C (p.Pro560=), a substitution presently encountered with c.1339 C>T or c.2107G>C in two uncommon *KEL*\*01 alleles

responsible for a weak expression of KEL antigens in Eurasia, *KEL\*01M.03* and *KEL\*01M.06*<sup>35,40</sup>. The two latter substitutions are absent in our data.

About the FY, JK and MNS systems, while the ancestral *FY\*02* and *GYPB\*04* are detected in both *H. sapiens* and archaic individuals, the derived *FY\*01* and *GYPB\*03* would be restrictive to *H. sapiens*. Similarly, *JK\*01* is shared by Neandertal, Denisova and *H. sapiens* while *JK\*02* is found in *H. sapiens* only. The *GYPB\*04* allele was often observed with c.251 C>G, and once c.230T and c.251G in the Les Cottés Neandertal, but without c.208G>T and c.270+5 g>t. Presently, the MNS alleles combining c.208T, c.230T, c.251G and c.270+5 g>t are usually found in the null alleles coding for the S-U +<sup>var</sup> phenotype, exclusive of Sub-Saharan populations<sup>41</sup>. Note that *GYPB* in Kostenki14, with c.143 C>T, c.230 C>T, and c.251 C>G but without c.208G>T, may be on his way to becoming *GYPB\*03N*.

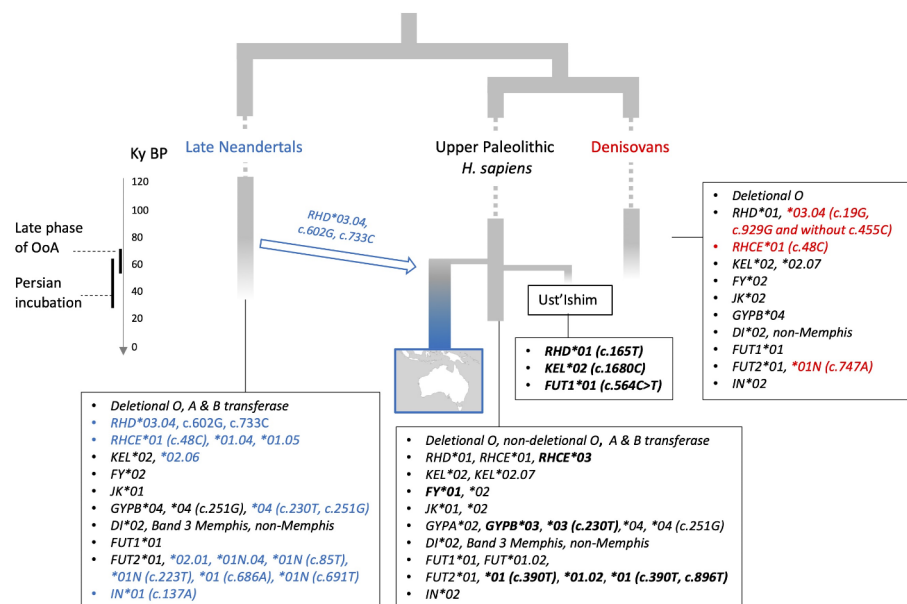
Lastly, two Band3 Memphis alleles (Diego blood group system) have been observed in one archaic and one *H. sapiens* from Siberia, Denisova11 and Yana1. Also, two astounding alleles of the Diego and Indian blood group systems may have been detected: *DI\*01* in AR33K (Supplementary Material 2, Fig S1) and *IN\*02* c.137G>A (p.Arg46Gln) in the Neandertal from the Caucasus, Mezmaiskaya2 (Table S5). The former, which did not receive statistical approval here, may have appeared in South Siberia before migrating to the Americas and Central Asia<sup>18,19</sup>. The latter is almost non-existent in modern humans (1/730,000)<sup>39</sup> but occurs at the same locus as the *IN\*01* allele (c.137G>C, p.Arg46Pro) with a different nucleotide substitution.

## Discussion

In the present study, we have supplemented the attempt of<sup>15</sup> with several *H. sapiens* and late Neanderthals to picture the red cell blood group gene diversity in Eurasia when three hominids inhabited almost simultaneously. The most striking pattern is that blood group gene diversity appears partitioned according to the hominid lineages (Fig. 1).

On one hand, amongst the Late Neanderthals, the present study showed the prevalence of ancestral alleles and that the blood group allele diversity remained unchanged during their last 80,000 years of existence. For two blood group systems, RH and KEL, the Neandertal alleles are almost exclusively found in modern populations from Sub-Saharan Africa. We also corroborated the richness of non-secretor alleles (*FUT2*). The non-secretor trait in Primate species, that is the absence of soluble ABO antigens in body fluids (saliva, breast milk...), is unique to humans and it may have been fixed since the Neolithic by positive selection after gut viral infestation<sup>42–44</sup>. However, the origin of *FUT2* alleles altering the protein function in pre-Neolithic times remains unresolved. As far as the Denisovans are considered, it is harsh to extrapolate but Denisova3 shares two derived and non-Neandertal blood group alleles with *H. sapiens*, *RHD\*01* and *KEL\*02.07*. Similarly, the Band3 Memphis variant was already polymorphic in the three *Homo* lineages.

On the other hand, the oldest genomes available (Ust'Ishim, Bacho Kiro, Zlaty Kun, Tianyan, Oase1, Kostenki14, 36–45 ka) show that *H. sapiens* entered in Eurasia with a set of functional alleles absent in Neanderthals and Denisova, and already derived from their ancestral state: *ABO\*O.01.02*, *FY\*01*, *RHD\*01* in *cis* with *RHCE\*03* forming the DcE (R2) haplotype, *KEL\*02.07*, *FUT1\*01.02* c.35C>T, *GYPB\*04* and *GYPB\*04* c.251C>G. Presently, some of these blood group alleles are only found in non-African populations<sup>31</sup>.



**Fig. 1.** Neighbor-Joining tree of red cell blood group alleles in Upper Paleolithic *H. sapiens*, Neanderthals and Denisova. Each branch received 100% of 1,000 bootstraps iterations. Bold black, *H. sapiens* characteristic alleles; blue: Neandertal characteristic alleles, red: Denisova characteristic alleles. Arrow: probable intergression event detected by<sup>15</sup>. OoA: Out of Africa. <https://d-maps.com>.

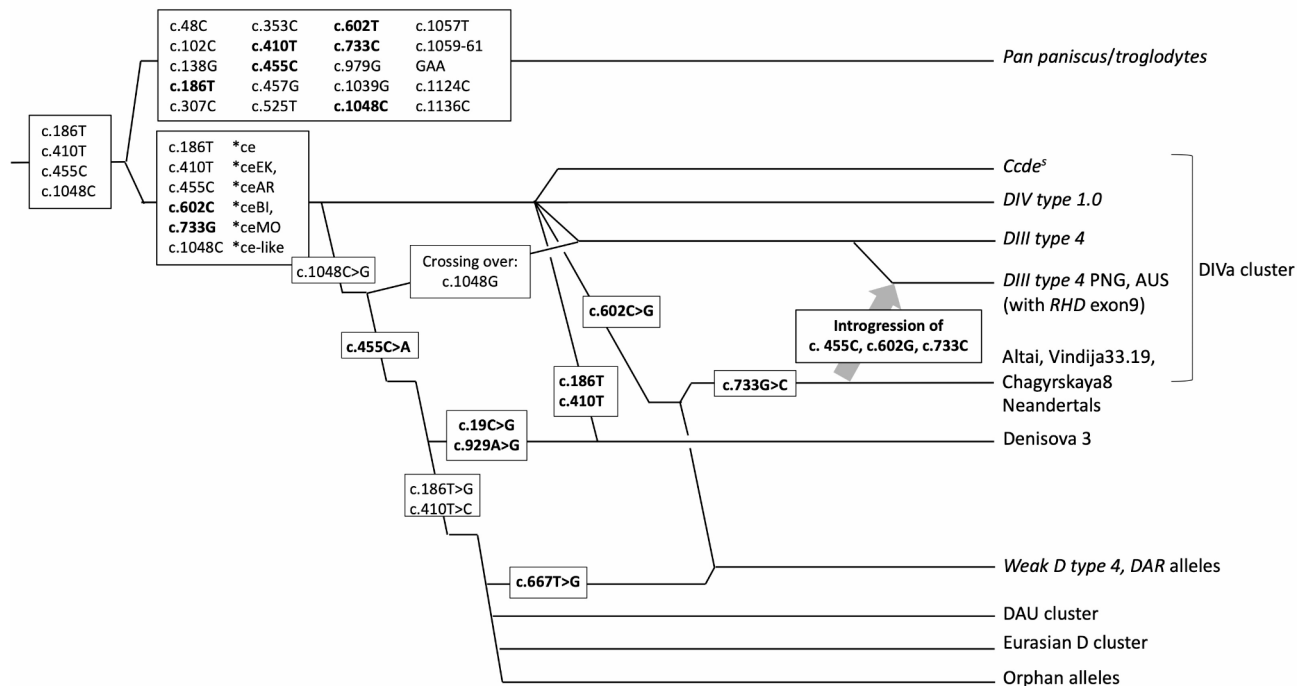
Additionally, we detected three unknown non-functional alleles (*RHD\*01* c.165C>T, *FUT1\*01* c.564C>T, and *KEL\*02.07* c.1680A>C) in Ust'-Ishim. Maybe these alleles were part of the dead-end *H. sapiens* lineage whose genetic ancestry did not contribute to the present-day Eurasian populations<sup>12,45,46</sup>.

Given the wide distribution of the derived alleles amongst the earliest *H. sapiens* of Eurasia coupled with their total absence in Africa, a first scenario could assume that these alleles have emerged in African *H. sapiens* before the Out of Africa, spread in Eurasia, then disappeared in Africa only. It would be tempting to point to natural selection, but the joint disappearance of independent alleles coding for phenotypes of a wide variety of susceptibility to natural selection<sup>47</sup> is unlikely. Moreover, given that the highest gene diversity is found in Africa and that the expansion of *H. sapiens* is characterized by an initial founder effect<sup>48</sup>, these alleles should be found in African populations. An alternative could be an outbreak of red cell blood group alleles outside Africa right before the conquest of Eurasia. Genetic, lithic and paleoecological evidence from the Initial Upper Paleolithic indicated that the late phase of *H. sapiens* expansion outside Africa circa 70–60,000 years ago was followed by a standstill of at least 15,000 yBP in the Persian Plateau<sup>49</sup>. Analogously to the Beringian occupation ca. 20,000 yBP during which the Beringian populations paused several millennia and differentiated before entering the Americas with genetic polymorphisms shared by Native Americans only<sup>50</sup>, *H. sapiens* could have resided long enough in the Persian Plateau to differentiate<sup>51</sup>. This view makes Western Asia a good candidate for the place of incubation of the functional *ABO\*O.01.02*, *FY\*01*, *RHCE\*03*, *KEL\*02.07*, *FUT1\*01.02* c.35 C>T, and *GYPB\*04* before expansion towards East (Tianyaun), North (Zlatý kůň, Ust'-Ishim) and Western Eurasia (Oase1, Bacho Kiro) more than 45ky ago. Interestingly, this time frame coincides with the arrival from Africa of two basal haplogroups: the FT haplogroup of the Y-chromosome and the L3 mtDNA lineage<sup>52–54</sup>. Later on, both FT and L3 eventually diversified outside of Africa around 47,000 and 55,000 yBP into respectively the male NO and female R Eurasian haplogroup before settling in Southern Siberia 45,000 years ago (Ust'-Ishim). Hence, the diversification of the FT Y-chromosomes and L3 mtDNAs could have occurred concomitantly with non-African blood group alleles, between the Levant and Southern Siberia. The emergence and fixation of independent genes after a long-term halt insinuates that demography has primarily accounted for the red cell genetic variation of the first Eurasian *H. sapiens*. However, this is restrictive and omits the evolutionary models for living beings that combine neutral and non-neutral events, heterogeneous mutation rates and temporal variation of selective forces<sup>55–60</sup>. Particularly, fluctuating selection may explicate that the non-secretor alleles (*FUT2*), present in Upper Paleolithic populations but advantageous from the Neolithic, were part of the standing genetic variation before the onset of a selection pressure. A second impulse of *H. sapiens* populations from the Persian Plateau would have expanded circa 38 ka to lead to Kostenki14, Yana, Goyet116-1 and Sunghir Eurasians<sup>51</sup>, but could not be detected from red cell blood group functional alleles. Note that the distribution of c.251 C>G in the *GYPB* gene (MNS blood group system) in Neandertal, Denisova and Paleolithic *H. sapiens* from Siberia and present-day Sub-Saharan populations suggests that the c.251 C>G allele could have diversified earlier, in Africa, where it has matured in *H. sapiens* to *GYPB\*03N.01* coding for the present-day S-U+<sup>var</sup> phenotype<sup>41,61</sup>.

In East Asia, AR33K indicates that before 30ky the Diego blood group might have already taken shape, before settling the Americas and comigrating with the nomads of the Mongolian steppes<sup>18,19</sup>. Unfortunately, we could not corroborate any other *DI\*01* in Siberian individuals dating from the initial colonization of the Americas (AfontovaGora2, USR1 in Alaska, 12ky, Supplementary material 2, Fig S1). So far, the earliest *DI\*01* carrier might be NEO239, a 7,500-year individual from the Yana Rhinoceros Horn Site (Supp Fig. 1)<sup>62</sup>. We also searched for putative *IN\*01* alleles (Indian blood group system) in 1236 ancient genomes from Caucasus and Middle Eastern<sup>63–66</sup> surrounding the area of Mezmaiskaya2, but the only individuals with readable data carried the reference allele, *IN\*02* (c.137G, p.46Arg). So far, both *IN\*01* and *DI\*01* alleles remain enigmatic as to when and where they first emerged.

Taking advantage of the well-defined *RHD* phylogeny and the accuracy of the *RHD* archaic sequences, we attempted to complement the current *RHD* phylogeny with the Neandertal and Denisova *RHD* haplotypes (Fig. 2, Supplementary Material 2, File S2). The *RHD* gene is one of the most screened loci of the red cell blood group systems, as it encodes for a clinically significant blood group antigen, the D antigen<sup>67</sup>. Four main clusters compose the phylogeny of *RHD*: *DIV*, *weak D type 4*, *DAU* and *Eurasian D* with *RHD\*01* as defined as a reference<sup>28,68,69</sup>. With 6 iterations from Siberia, Caucasus, Balkans and Meuse Basin, we confirmed the exclusiveness of the *RHD-RHCE* haplotype amongst the Late Neandertal<sup>15</sup>. The combined occurrence of c.186T, c.410T, and c.455 C in the Late Neandertals indicate that the Neandertal *RHD* belong to the *DIVa* cluster. We added an introgression event to lead to the Oceanian *DIII type 4* (*RHD\*03.04*)<sup>15,70</sup>. The Denisova allele, with the ancestral c.186C>T and c.410C>T but derived c.455C>A and c.1048C>G in *cis* is ideal to embody an intermediate branch right after the *DIVa* cluster, onto which we added the original c.19C>G and c.929A>G substitutions. Once the Neandertal and Denisova *RHD* incorporated, we could infer the sequential occurrence of the five substitutions leading to the non-DIVa cluster<sup>28</sup>, c.186T>G, c.410T>G, c.455C>A, and c.1048C>G. Excluding the *RHCE* exon7 hybrid, the c.1048C>G substitution distinguished the most ancestral branches *Pan* and *DIV type 1.0* from the rest of the tree. We hence assumed that c.1048C>G occurred first. To explain its presence in the *DIII type 4* and Neandertal haplotype, we used the crossing-over conveniently drawn between the *Eurasian D* cluster to the *DIII type 4* allele<sup>28,69</sup>. Denisova3, dated by 52–76ky BP and carrier of c.455C>A c.186T>G and c.410T>C, indicates that these three substitutions occurred secondly, before 76ky, to lead to the *weak D*, *DAU* and *Eurasian D* clusters to which the other Denisova allele and Ust'-Ishim ones, *RHD\*01* and *RHD\*01* (c.165T), belong.

To reinforce our assumption, we attempted to root the main functional blood group system substitutions and *RHD* branch nodes with the Human Genome Dating tool (<https://human.genome.dating><sup>71</sup>, Supplementary Material 1, Table S7). Nevertheless, three reasons prevented this approach. First, in numerous cases, the alternate nucleotide used for the date estimates is not the actual derived one. Second, numerous major substitutions are lacking, notably for the RH blood group system. Third and more importantly, we found numerous date



**Fig. 2.** Updated phylogenetic tree of the *RHD* gene.

inconsistencies with archaeological evidence. For instance, rs150073306 (c.733G>C, p.Val245Leu) which belongs to the peculiar Neandertal-Oceanian *RHD* haplotype, is thought to have emerged 8,3ky ago<sup>71</sup>.

Lastly, we proposed introgression to be the mechanism behind the Neandertal *RHD* haplotype in modern Oceanians, because we previously described this haplotype in one Papuan and one Aboriginal individual and demonstrated that a 15,700 base pairs (bp) segment of this haplotype was likely introgressed<sup>15</sup>. However, Incomplete Sorting Lineage (ILS) could also explain why modern Oceanians share an archaic haplotype with Neandertals<sup>72</sup>. A proper test to distinguish introgression from ILS for every blood group allele shared with Neandertal and Denisova is beyond the topic of the present paper. Nevertheless, a foretaste could be to calculate the probability that a 15,700 bp haplotype shared through ILS remains untruncated by recombination. For this, we replicated the Gamma calculations performed by<sup>73</sup>, using the genome-wide recombination rate obtained specifically for Papuans and Aborigines<sup>74</sup> (See Supplementary Material 2, File S2 for the detailed procedure). We found that the length of haplotype acquired through ILS in Papuans and Aborigines is estimated to be around 3,100 bp and 2,700 bp respectively. Hence, the likelihood of retention of a 15,700 bp archaic fragment in modern Oceanians since the divergence with Neandertal is low, estimated at  $P=0.037$  (Papuans) and  $P=0.019$  (Aborigines). But overall, similarly as<sup>73</sup> have noticed for *EPAS1*, if ILS had occurred, the Neandertal *RHD* haplotype would have been found in numerous other modern human populations besides Oceanians, including those from Africa. Today, the database of *RHD* genetic variations is densely nourished and yet, the Neandertal haplotype is limited to Oceanians. Hence, the length of the haplotype – more than 5x longer than expected under ILS – and its geographic repartition makes ILS unlikely.

To conclude, the present foray into the molecular polymorphisms of blood group systems mirrors the previous genetic inferences about the pattern of arrival of *H. sapiens* in Eurasia. The data showed that numerous derived alleles were unique to *H. sapiens*. Notably, the RH and Duffy systems show a clearcut repartition between in and out of Africa, allocated to the main Out of Africa expansion of *H. sapiens* circa 60-70ky. Indeed, whilst Neandertal stored since 120 ka ancestral and presently sub-Saharan functional blood group alleles, *H. sapiens* who had just left Africa underwent an allele diversification before expanding throughout Eurasia, which explains why some red cell blood group antigens are presently present outside of Africa. The lost *H. sapiens* lineage, notably embodied by Ust-Ishim, carried blood group alleles not found in any other *H. sapiens*. Lastly, the present study recalls the anthropological effectiveness of genetic polymorphisms currently being surveyed for transfusion safety and pregnancy monitoring. The present study highlighted numerous alleles presently found in modern humans, shared with Neandertal and Denisova. In the future, deciphering introgression versus ILS would considerably enlighten the evolutionary history of the red cell genetic polymorphisms.

## Materials and methods

### Sample choice and genotype readability

Initially, we selected 61 Eurasians dating from 120 to 16.5ky including 22 medium and low-coverage archaic hominids (Supplementary Material 1, Table S1). Low coverage will eventually lead to unavailable genotypes,

and a pertinent approach to prevent from is imputation<sup>75</sup>. However, imputation relies on a reference panel, which assumes that the imputed genotype will be of Neandertal or modern human ancestry, depending on the reference panel used. Moreover, imputation would prevent the description of unknown alleles whose MAF is close to zero, such as the authentic RH haplotype and non-secretor alleles of Altai (Denisova5), Chagyrskaya8 and Vindija33.19 Neandertals and Denisova3<sup>15</sup>. In the present study, where the Denisova, Neandertal and *H. sapiens* lineages are confronted based on the variation of blood group systems, we are concerned that imputation with a determined reference panel might bias the comparison. Hence, blood group alleles were annotated as they appeared, regardless of the ancestry of the carriers.

Furthermore, because red cell blood group loci are scarcely represented in the Allen Ancient DNA Resource 1240k dataset<sup>76</sup> (Supplementary Material 1, Table S2) and allele designation for the RH, H and secretor blood group systems requires the examination of the entire exons, we chose only whole-genome sequences. To evaluate the readability of the genotypes, we gathered the mean coverage, base and map qualities of the codons of interest (Supplementary Material 1, Tables S3 and S4). The individuals of the Adriatic shore, the three clusters defined by (Fu<sup>12,45</sup>) as Věstonice, Goyet, and El Miron, the Mal'ta individual, several low-coverage Neandertals of Western Europe and Chagyrskaya-Okladnikov caves could not be properly called for red cell blood group alleles. Hence, the present results rely on the genetic picture provided by 34 Upper Paleolithic individuals (Supplementary Material 1, Table S5 and S6). We ended up with 36 readable samples: 14 Neandertals, 22 *H. sapiens* and one offspring of Neandertal and Denisova. Their name, lineage, geographical location, age in ky, mean coverage, bibliographic reference and nuclear DNA European Nucleotide Archive access number are listed in Supplementary Material 1, Table S1.

### Exploration procedure for blood group alleles

We focused on ten blood group systems of anthropological and medical interest, that is a set of genetic markers that enable both the tracing of ancient human migration, natural selection and archaic admixture, together with the safety of blood transfusion, transplants and pregnancy monitoring<sup>67,77</sup>: ABO, MNS, Rh (RH), KEL, Duffy (FY), Kidd (JK), Diego (DI), H (*FUT1* gene), H secretor (*FUT2* gene) and Indian (IN). For these blood group systems, we searched the genotype at the functional single nucleotide substitutions listed by the ISBT.

For the ABO, MNS, KEL, FY, JK, DI and IN systems, we searched for the main functional single nucleotide substitutions in relation to the conventional transcripts (International Society of Blood Transfusion, ISBT, <https://www.isbtweb.org>; UCSC Genome<sup>78</sup>). For each, we called the genotype using the Integrative Genome Viewer (IGV)<sup>79</sup>, the samtools, vcftools and tabix commands<sup>80,81</sup>. For the ABO system, we focused on the five single-nucleotide substitutions responsible of the A and B enzymes bringing the A and B antigens onto the red cells, as well as their absence (O blood group)<sup>82</sup>. We added 18 substitutions to characterize the two other main O alleles, *ABO\*O.01.02*, *ABO\*O.02.01* as well as *ABO\*O.01.11*, predominant in Native American populations, that may have arisen right before the initial peopling of the Americas, in Beringia or Siberia<sup>83</sup>. For the RH and H systems, we rebuilt the *RHD*, *RHCE* and *FUT2* coding sequences and aligned to the conventional reference using Clustal Omega tool at EMBL-EBI<sup>37</sup>. We paid specific attention to polymorphisms responsible for partial, weak, and negative RH variants as well as stop codons in the *FUT2* gene.

### Warnings about the reference sequence

While calling genotypes, we paid four specific attentions to the hg19 reference allele. First, as shown previously<sup>15</sup>, the genomic sequence of six loci (*ABO*, *KEL*, *GYPB*, *RHCE*, *SLC4A1* and *FUT1*) need to be reverse-complemented to 5'-3'. Second, although genotype calling is mapped with hg19 (GRCh37), herein we referred to the conventional nomenclature sequence of the ISBT. Hence, caution has to be taken because the genomic sequence may differ from the mRNA<sup>84</sup>. Noteworthy are *ABO* and *RHD*<sup>85</sup>. The conventional coding alleles are respectively *ABO\*A1* (NM\_020469.3, uc004cda.1) and *RHD\*01* (NM\_016124.6, uc001bjz.3), while the genomic sequence of hg19 has 5 substitutions of *ABO\*O.01.02* (NC\_000009.11: c.106G>T, c.188G>A, c.189 C>T, c.220 C>T, c.261delG) and the c.1136 C>T of *RHD\*10.00 RHD\*DAU0* plus the c.1154-31 C>T of *RHD\*DEL37* (NC\_000001.10). Third, the ISBT stated recently that the initiation codon of *FUT2* is henceforth the first ATG at the beginning of exon 2 (GRCh37, 19: 49,206,214), which involves rectifying the previous amino acid and nucleotide numbering by +11 and +33 respectively (see the ISBT 018 H track of changes by v6.1 31-MAR-2023). Fourth, we defined the ancestral allele according to previous phylogenetic attempts (*ABO*, *ACKR1*, *FUT1*, *FUT2*, *SLC4A1*, *RHD*, *RHCE*<sup>21,27,28,55,86,87</sup>, comparison with *Pan* (*KEL*, *SLC14A1*, *CD44*, UCSC Genome Browser on Chimp Jan. 2018, Clint\_PTRv2/panTro6 and Bonobo May 2020, Mhudiblu\_PPA\_v0/panPan3<sup>88</sup>, or contrasted allele frequency between sub-Saharan and non-African populations (*GYPB*, *GYPB*; *KEL*)<sup>16,31,61,67</sup>.

### One-allele genotypes and false homozygous

Low-coverage sequences are prone to omit all available alleles which leads to false homozygous calls. Solutions for genotype imputation in low-coverage sequences exist but these assume the population to be strictly biallelic (e.g.<sup>89</sup>), yet for some blood group nucleotide substitutions more than 2 alleles exist (e.g. for *RHD*, c.697G>A or C). In addition to being restrictive, the use of genotype imputation from low-coverage sequences would prevent from the detection of novel haplotypes and alleles, like *RHD\*03.04*+c.602 C>G+c.733G>C and *FUT2\*01N* (c.747T>A) in Neandertals<sup>15</sup>. Hence, when one allele was encountered, no zygosity has been imposed. Nevertheless, although low coverage cannot fully establish the likelihood of a genotype at the individual level, the repeated observation of an allele among several individuals gives greater confidence in its existence.

### Genotypes in admixed *H. sapiens*

Previously, Neandertal-ancestry DNA segments have been finely mapped in Oase1, Ust'-Ishim, Kostenki14, Tianyuan and the four Bacho Kiro individuals (~35-45ky)<sup>45,90</sup>. For the loci successively covered, we wondered

if the genes of red cell blood group systems were authentically of *H. sapiens*-ancestry. For this, we superimposed the coordinates of the genes with the spatial distribution of Neanderthal DNA tracts >0.2 cM in length. Only ABO in BK\_BB7\_240 fall within a Neanderthal tract (Supplementary Material 1, Table S4).

### Genetic relationships between *H. sapiens*, Neandertal and Denisova lineages

Based on the genotypes listed in Supplementary Material 1, Tables S5 and S6, we calculated the blood group allele frequencies of in each lineage. Obviously, the Denisova lineage suffers from its sample size ( $N=2$ , Denisova3 and Denisova11, the Neandertal-Denisova offspring). Operating PHYLIP version 3.698<sup>91</sup>, we then generated 1,000 data sets, calculated the Nei's genetic distance<sup>92</sup> for each matrix and built a consensus Neighbor Joining tree of the 1,000 bootstrap reiterations.

### Data availability

The datasets analyzed during the current study are publicly available in the ancient genomes project of the Max Planck Institute for Evolutionary Anthropology (<https://www.eva.mpg.de/genetics/genome-projects/>) and the European Nucleotide Archive (<https://www.ebi.ac.uk/ena/browser/home>). Access codes for each sample are provided in this published article as well as in the present supplementary information files. The reference blood group alleles are listed in the Working Parties of the International Society of Blood Transfusion (<https://www.isbtweg.org>). Conventional cDNA transcripts are presented at the UCSC Genome browser (<https://genome.ucsc.edu>). Data is provided within the manuscript and supplementary information files.

Received: 28 May 2024; Accepted: 11 December 2024

### References

- Higham, T. et al. The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* **512**, 306–309 (2014).
- Harvati, K. et al. Apidima Cave fossils provide earliest evidence of Homo sapiens in Eurasia. *Nature* **571**, 500–504 (2019).
- Hershkovitz, I. et al. The earliest modern humans outside Africa. *Science* **359**, 456–459 (2018).
- Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
- Guran, S. H., Yousefi, M., Kafash, A. & Ghasidian, E. Reconstructing contact and a potential interbreeding geographical zone between Neanderthals and anatomically modern humans. *Sci. Rep.* **14**, 20475 (2024).
- Armitage, S. J. et al. The southern route 'out of Africa': evidence for an early expansion of modern humans into Arabia. *Science* **331**, 453–456 (2011).
- Clarkson, C. et al. Human occupation of northern Australia by 65,000 years ago. *Nature* **547**, 306–310 (7AD).
- Mylopotamitaki, D. et al. Homo sapiens reached the higher latitudes of Europe by 45,000 years ago. *Nature* **626**, 341–346 (2024).
- Lalueza-Fox, C. et al. Genetic evidence for patrilineal mating behavior among Neandertal groups. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 250–253 (2011).
- Skov, L. et al. Genetic insights into the social organization of Neanderthals. *Nature* **610**, 519–525 (2022).
- Prüfer, K. et al. The complete genome sequence of a Neandertal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
- Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
- Hajdinjak, M. et al. Reconstructing the genetic history of late Neandertals. *Nature* **555**, 652–656 (2018).
- Villanea, F. A., Huerta-Sanchez, E. & Fox, P. K. ABO genetic variation in Neandertals and Denisovans. *Mol. Biol. Evol.* **38**, 3373–3382 (2021).
- Condemi, S. et al. Blood groups of Neandertals and Denisova decrypted. *PLoS One*. **16**, e0254175 (2021).
- Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton University Press, 1994).
- Schlötterer, C. The evolution of molecular markers — just a matter of fashion? *Nat. Rev. Genet.* **5**, 63–69 (2004).
- Begat, C., Bailly, P., Chiaroni, J. & Mazieres, S. Revisiting the Diego Blood Group System in Amerindians: Evidence for Gene-Culture Comigration. *Plos One* **10**, (2015).
- Petit, F. et al. The radial expansion of the Diego blood group system polymorphisms in Asia: mark of co-migration with the Mongol conquests. *Eur. J. Hum. Genet.* **27**, 125–132 (2019).
- Tarazona-Santos, E. et al. Population Genetics of GYPB and Association Study between GYPB\*S/s Polymorphism and Susceptibility to P. falciparum Infection in the Brazilian Amazon. *PLoS ONE*. **6**, e16123 (2011).
- McManus, K. F. et al. Population genetic analysis of the DARC locus (Duffy) reveals adaptation from standing variation associated with malaria resistance in humans. *PLoS Genet.* **13**, e1006560 (2017).
- Ma, S. et al. Common PIEZO1 Allele in African Populations Causes RBC Dehydration and Attenuates Plasmodium Infection. *Cell* **173**, 443–455e12 (2018).
- Peng, J. et al. Insights into extensive deletions around the XK locus associated with McLeod phenotype and characterization of two novel cases. *Gene* **392**, 142–150 (2007).
- Gouagna, L. C. et al. Genetic variation in human HBB is associated with Plasmodium falciparum transmission. *Nat. Genet.* **42**, 328–331 (2010).
- Kotepui, K. U., Mahittikorn, A., Masangkay, F. R. & Kotepui, M. Association between ovalocytosis and Plasmodium infection: a systematic review and meta-analysis. *Sci. Rep.* **13**, 7164 (2023).
- Ségurel, L. et al. The ABO blood group is a trans-species polymorphism in primates. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18493–18498 (2012).
- Kitano, T., Blancher, A. & Saitou, N. The functional A allele was resurrected via recombination in the human ABO blood group gene. *Mol. Biol. Evol.* **29**, 1791–1796 (2012).
- Srivastava, K. et al. The DAU cluster: a comparative analysis of 18 RHD alleles, some forming partial D antigens. *Transfus. (Paris)*. **56**, 2520–2531 (2016).
- Carritt, B., Kemp, T. J. & Poulter, M. Evolution of the human RH (rhesus) blood group genes: a 50 year old prediction (partially fulfilled). *Hum. Mol. Genet.* **6**, 843–850 (1997).
- Lalueza-Fox, C. et al. Genetic characterization of the ABO blood group in Neandertals. *BMC Evol. Biol.* **8**, 342 (2008).
- Möller, M., Jöud, M., Storry, J. R. & Olsson, M. L. ErythroGene: a database for in-depth analysis of the extensive variation in 36 blood group systems in the 1000 Genomes Project. *Blood Adv.* **1**, 240–249 (2016).
- Olsson, M. L. & Chester, M. A. Frequent occurrence of a variant O1 gene at the blood group ABO locus. *Vox Sang.* **70**, 26–30 (1996).
- Yazer, M. H. & Olsson, M. L. The O2 allele: questioning the phenotypic definition of an ABO allele. *Immunohematology* **24**, 138–147 (2008).
- Consortium, G. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature* **576**, 106–111 (2019).

35. Harrison, P. W. et al. Ensembl 2024. *Nucleic Acids Res.* **52**, D891–D899 (2024).
36. Consortium, U. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
37. Madeira, F. et al. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279 (2022).
38. Sippert, E., Arnoni, C. P. & Rios, M. Impact of RHCE variability and complexity in transfusion medicine: a narrative review. *Ann. Blood.* **8**, 8–8 (2023).
39. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
40. Karamatic Crew, V., Poole, J., Burton, N. & Daniels, G. Three uncommon KEL alleles in one family with unusual Kell phenotypes explain a 35-year old conundrum. *Vox Sang.* **106**, 242–247 (2014).
41. Storry, J. R., Reid, M. E., Fetcs, S. & Huang, C. H. Mutations in GYPB exon 5 drive the S-s-U+(var) phenotype in persons of African descent: implications for transfusion. *Transfus. (Paris)*. **43**, 1738–1747 (2003).
42. Godefroy, E. et al. Microbiota-induced regulatory T cells associate with FUT2-dependent susceptibility to rotavirus gastroenteritis. *Front. Microbiol.* **14**, 1123803 (2023).
43. Larsson, M. M. et al. Antibody prevalence and titer to norovirus (genogroup II) correlate with secretor (FUT2) but not with ABO phenotype or Lewis (FUT3) genotype. *J. Infect. Dis.* **194**, 1422–1427 (2006).
44. Nordgren, J. & Svensson, L. Genetic Susceptibility to Human Norovirus Infection: An Update. *Viruses* **11**, 226 (2019).
45. Hajdinjak, M. et al. Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry. *Nature* **592**, 253–257 (2021).
46. Liu, Y., Mao, X., Krause, J. & Fu, Q. Insights into human history from the first decade of ancient human genomics. *Science* **373**, 1479–1484 (2021).
47. Moulds, J. M. & Moulds, J. J. Blood group associations with parasites, bacteria, and viruses. *Transfus. Med. Rev.* **14**, 302–311 (2000).
48. Ramachandran, S. et al. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl. Acad. Sci. U A.* **102**, 15942–15947 (2005).
49. Vallini, L. et al. The Persian plateau served as hub for Homo sapiens after the main out of Africa dispersal. *Nat. Commun.* **15**, 1882 (2024).
50. Tamm, E. et al. Beringian standstill and spread of Native American founders. *PLoS One.* **2**, e829 (2007).
51. Vallini, L. et al. Genetics and Material Culture Support Repeated Expansions into Paleolithic Eurasia from a Population Hub Out of Africa. *Genome Biol. Evol.* **14**, evac045 (2022).
52. Soares, P. et al. The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Mol. Biol. Evol.* **29**, 915–927 (2012).
53. Ilumäe, A. M. et al. A Non-trivial Time-Resolved Phylogeography that Cuts across Language Families. *Am. J. Hum. Genet.* **99**, 163–173 (2016). Human Y Chromosome Haplogroup N.
54. Haber, M. et al. A Rare Deep-Rooting D0 African Y-Chromosomal Haplogroup and Its Implications for the Expansion of Modern Humans Out of Africa. *Genetics* **212**, 1421–1428 (2019).
55. Ferrer-Admetlla, A. et al. A natural history of FUT2 polymorphism in humans. *Mol. Biol. Evol.* **26**, 1993–2003 (2009).
56. Yamamoto, F. et al. An integrative evolution theory of histo-blood group ABO and related genes. *Sci. Rep.* **4**, 6601 (2014).
57. Jensen, J. D. et al. The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018: COMMENTARY. *Evolution* **73**, 111–114 (2019).
58. Teske, P. R. et al. Mitochondrial DNA is unsuitable to test for isolation by distance. *Sci. Rep.* **8**, 8448 (2018).
59. Yuan, D. & Huang, S. Genetic equidistance at nucleotide level. *Genomics* **109**, 192–195 (2017).
60. Rees, J. S., Castellano, S. & Andrés, A. M. The Genomics of Human Local Adaptation. *Trends Genet.* **36**, 415–428 (2020).
61. Martin, F. J. et al. Ensembl 2023. *Nucleic Acids Res.* **51**, D933–D941 (2023).
62. Sikora, M. et al. The population history of northeastern Siberia since the Pleistocene. *Nature* **570**, 182–188 (2019).
63. Wang, C. C. et al. Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nat. Commun.* **10**, 590 (2019).
64. Gelabert, P. et al. Genome-scale sequencing and analysis of human, wolf, and bison DNA from 25,000-year-old sediment. *Curr. Biol.* **31**, 3564–3574e9 (2021).
65. Skourtanioti, E. et al. Genomic History of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus. *Cell* **181**, 1158–1175e28 (2020).
66. Flesch, B. K. et al. Molecular Blood Group Screening in Donors from Arabian Countries and Iran Using High-Throughput MALDI-TOF Mass Spectrometry and PCR-SSP. *Transfus. Med. Hemother.* **47**, 396–408 (2020).
67. Reid, M. E., Lomas-Francis, C. & Olsson, M. L. *The Blood Group Antigen Factsbook* (Elsevier Academic, 2012).
68. Wagner, F. F. et al. The DAU allele cluster of the RHD gene. *Blood* **100**, 306–311 (2002).
69. Flegel, W. A. et al. D variants at the RhD vestibule in the weak D type 4 and Eurasian D clusters. *Transfus. (Paris)*. **49**, 1059–1069 (2009).
70. Schoeman, E. M., Roulis, E. V., Perry, M. A., Flower, R. L. & Hyland, C. A. Comprehensive blood group antigen profile predictions for Western Desert Indigenous Australians from whole exome sequence data. *Transfus. (Paris)*. **59**, 768–778 (2019).
71. Albers, P. K. & McVean, G. Dating genomic variants and shared ancestry in population-scale sequencing data. *PLoS Biol.* **18**, e3000586 (2020).
72. Maddison, W. P. Gene Trees in Species Trees. *Syst. Biol.* **46**, 523–536 (1997).
73. Huerta-Sánchez, E. et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014).
74. Lipson, M. et al. Calibrating the Human Mutation Rate via Ancestral Recombination Density in Diploid Genomes. *PLOS Genet.* **11**, e1005550 (2015).
75. Mota, S. D. Imputation of ancient human genomes. *Nat. Commun.* **14**, 3660 (2023).
76. Mallick, S. et al. The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Sci. Data.* **11**, 182 (2024).
77. Chiaroni, J., Peyrard, T. & Pirenne, F. *Les Groupes Sanguins Érythrocytaires – 2e Édition* Paris, (2024).
78. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
79. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
80. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
81. Li, H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* **27**, 718–719 (2011).
82. Yamamoto, F., Clausen, H., White, T., Marken, J. & Hakomori, S. Molecular genetic basis of the histo-blood group ABO system. *Nature* **345**, 229–233 (1990).
83. Estrada-Mena, B. et al. Blood group O alleles in Native Americans: implications in the peopling of the Americas. *Am. J. Phys. Anthr.* **142**, 85–94 (2010).
84. Sayers, E. W. et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **50**, D20–D26 (2022).
85. Hyun, J., Oh, S., Hong, Y. J. & Park, K. U. Prediction of various blood group systems using Korean whole-genome sequencing data. *PLoS One.* **17**, e0269481 (2022).
86. Jarolim, P. et al. Band 3 Memphis: a widespread polymorphism with abnormal electrophoretic mobility of erythrocyte band 3 protein caused by substitution AAG---GAG (Lys---Glu) in codon 56. *Blood* **80**, 1592–1598 (1992).
87. Ségurel, L., Gao, Z. & Przeworski, M. Ancestry runs deeper than blood: the evolutionary history of ABO points to cryptic variation of functional importance. *Bioessays* **35**, 862–867 (2013).

88. Nassar, L. R. et al. The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res.* **51**, D1188–D1195 (2023).
89. Frago, C. A., Heffelfinger, C., Zhao, H. & Dellaporta, S. L. Imputing Genotypes in Biallelic Populations from Low-Coverage Sequence Data. *Genetics* **202**, 487–495 (2016).
90. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
91. Felsenstein, J. P. H. Y. L. I. P. (Phylogeny Inference Package) version 3.7a. Distributed by the author (2009).
92. Nei, M. Genetic Distance between Populations. *Am. Nat.* **106**, 283–292 (1972).

## Acknowledgements

Centre de Calcul Intensif d'Aix-Marseille is acknowledged for granting access to its high-performance computing resources. We are also grateful to Mateja Hajdinjak for transmitting the coordinates of the Neandertal tracts in the admixed *H. sapiens*. Silvana Condemi received support from the project Starch4Sapiens (Projet ANR-20-CE03-0002).

## Author contributions

S.M., S.C., J.C conceived the study, S.M. analyzed and curated the data, S.M., J.C, W.E.N. interpreted the genotype and function, S.C. interpreted the paleoanthropological context, S.M., S.C., J.C drafted, revised and edited the manuscript, all authors read and approved the final version of the manuscript.

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-83023-0>.

**Correspondence** and requests for materials should be addressed to S.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024