ESHG

ARTICLE

Sequence diversity of the Rh blood group system in Basques

André Flores-Bello [1] · David Mas-Ponte[2] · Miruna E. Rosu[3] · Elena Bosch[1] · Francesc Calafell [1] · David Comas[1]

Abstract

Basques show specific cultural, demographic, and genetic characteristics that have placed them as an isolated and unique population within Europe, such as their non-Indo-European language, Euskara. They have historically lived along the Western Pyrenees, between Spain and France, in one of the most important European glacial refugia during the Last Glacial Maximum. The most striking genetic characteristic is their highest frequency of the RhD blood group negative allele, a variant related to the hemolytic disease of the newborn. Both demographic and adaptive processes have been suggested as possible causes of the high frequency of RhD negative in Basques, but neither hypothesis has been clearly demonstrated. While previous studies on the Rh system in Basques have been mostly focused on serological and genotyping diversity, in this work we analyze genotyping and next generation sequencing data in order to provide a general framework of the genetic scenario of the system in Basques. In particular, we genotyped the most relevant variants of the system (D/d, E/e, and C/c), and sequenced three ~6 kb flanking regions surrounding the Rh genes in Basques and also in other populations for comparison. Our results are in agreement with previous studies, with Basques presenting the highest frequency of the RHD deletion (47.2%). Haplotype analyses of D/d, E/e, and C/c variants confirmed an association between the RhC allele, previously suggested to be under positive selection, and the RhD positive variant in non-sub-Saharan populations, including Basques. We also found extreme differentiation for the C/c variant when comparing sub-Saharan to non-sub-Saharan populations.

Introduction

Basques have been the subject of many studies during the last decades due not only to cultural but also to biological characteristics that place them as an isolated and unique population within Europe. Their language, Basque or *Euskara*, is not Indo-European, and it is not related to any other extinct or extant language. They have historically settled the Western Pyrenees, in the Franco-Cantabrian

region, between Spain and France [1]. The large number of archeological sites from different periods, as well as different genetic studies with present-day and ancient samples, showed that this region acted as one of the most densely populated European glacial refugia during the Last Glacial Maximum. Despite controversial studies [2], a large expansion is thought to have originated in the same area around 12,000 years ago (ya), in the Late Glacial when the climatic conditions started to improve in Europe [3, 4]. Many controversies surround the origins and the population history of Basques, which partial or limited population sampling has certainly not helped to settle. Several studies had suggested that Basques are descendants of an Upper-Paleolithic hunter-gatherer population that remained isolated in the glacial refugia and maintained very low contacts with surrounding populations, avoiding gene flow since the Neolithic [5]. However, recent studies point to gene flow between hunter-gatherers and early farmers from Chalcolithic and Bronze Age in the region [6]. In the general framework created by ancient DNA studies, Basque distinctiveness seems to arise from their reduced steppe ancestry as compared to other European populations, indicating limited contacts with Bronze Age migrants [7]. The

✉ David Comas
david.comas@upf.edu

1 Institute of Evolutionary Biology (CSIC – UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain

2 Institute for Research in Biomedicine Barcelona, Barcelona, Spain

3 Department of Viroscience, Postgraduate School Molecular Medicine, Erasmus MC, Rotterdam, The Netherlands

SPRINGER NATURE

genetic uniqueness of Basques within the European continent has been shown from classical markers [8], uniparental lineages [4] and autosomal markers [9]. However, some studies have challenged the genetic differentiation of Basques compared to other European populations [10]. One of the most striking genetic singularity of Basques is related to the Rhesus (Rh) system since they show one of the highest frequencies of the RhD negative allele in human populations [1, 11–13].

The human Rh system is a set of antigens that are expressed on the membrane of red blood cells. These antigens are encoded by two homologous genes, *RHD* and *RHCE*, located at chromosomal position 1p34.1-1p36 with opposite orientations (Figure S1). Both genes show a high sequence identity (98%) since they originated from an ancestral gene duplication within primates [14, 15]. Among the current 61 Rh antigens recognized by the International Society of Blood Transfusion (ISBT) [16], only five are of extensive interest because of their importance in hemolytic reactions from transfusion incompatibilities, the hemolytic disease of the newborn (HDN) and autoimmune diseases. The *RHD* gene encodes the D antigen that defines the RhD positive variant (D), while the RhD negative variant (d) is caused by the direct deletion of the gene in the vast majority of the cases [14, 15, 17]. The RhD negative phenotype is recessive and, thus, only shows when the two copies of the *RHD* gene in the individual are deleted. In the *RHCE* gene, different point mutations define the C/c and E/e antigen pairs [14, 17]. The actual function of the RH genes, beyond their role as antigens, has barely been studied. The available information has been inferred only from their homology with other genes in their family, which are involved in erythrocyte membrane structure and/or transport of ammonia or carbon dioxide [18]. A well-known fact is its association with the HDN. Although other Rh alleles and blood group systems can be related to the HDN, most of the cases are associated with the RhD negative variant. HDN takes place when the immune system of a RhD negative mother reacts producing anti-D antibodies after the exposure to red blood cells from a previous RhD positive pregnancy [19]. The HDN incidence in Europeans was 1 in 20 births to RhD negative mothers with a mortality around 20–40% before 1968, when Rho(D) immune globulin started to be used as preventive treatment for HDN [20, 21]. Given the association of the Rh negative allele with HDN, one would expect this allele to have been selected against (at least until quite recently) and thus to have been maintained at low frequencies or even disappear in populations. Note, though, as discussed below, that only heterozygotes are affected by HDN, and their death removes one copy of each allele, leading to an unstable equilibrium. Thus, the existence of the *RHD* deletion at polymorphic frequencies, and, in particular, its high frequency in Basques, needs to be

explained. Both demographic and adaptive hypotheses have been posited including drift by isolation and heterozygous advantage, respectively, without clear conclusions. As for the advantage for heterozygous, an unknown positive effect may counteract the disadvantage of the association with HDN. However, complete understanding of the adaptive nature of the *RHD* and *RHCE* genes is complicated by the fact that their functions are hardly known.

Since Etcheverry detected for the first time a high frequency of the RhD negative variant in Basque emigrants to Argentina [22], many studies have analyzed this variant, confirming the high frequency of the RhD negative allele using different methods and datasets [1, 11–13]. However, most of these studies were based on classical serologic methods and focused on the description of the antigens without describing the full genetic diversity of the system. In the present study, we aim to analyze the genomic region around the RH genes in Basques and other populations in order to better understand the polymorphism of the system in the Basque population and contrast our results with previous immunological studies.

Materials and methods

Sample collection

DNA was extracted from blood samples collected from 53 Basque, 17 Moroccan, and 12 Catalan unrelated volunteers with their four grandparents born in the geographical region studied. All samples were collected with the appropriate informed consent and the project obtained the ethics approval from the local Institutional Review Board, Comitè Ètic d'Investigació Clínica–Institut Municipal d'Assistència Sanitària (CEIC-IMAS) in Spain (2013/5429/I). In order to compare the Rh data in our samples with other populations, we included previously published HapMap samples from 22 Yorubans (YRI), 21 Han Chinese (CHB) and 20 Central Europeans (CEU) analyzed for the Rh system [23]. The chimpanzee sequence mapped on the human genome GRCh37/hg19 from the UCSC database [24] was used as an outgroup for some of the analyses.

Genotyping and sequencing of the Rh region

RHCE SNPs rs676782 (C/c; polymorphism P103S) and rs609320 (E/e; polymorphism A226P) were genotyped as in Perry et al. [23], performing a co-amplification of the surrounding regions, with FAM and HEX dye-labeled primers, and then a digestion of the products with *Mnl*I restriction enzyme, whereas the *RHD* deletion was determined through a TaqMan quantitative PCR assay performed as in Perry et al. [23]. Since the RhD negative variant is due to the
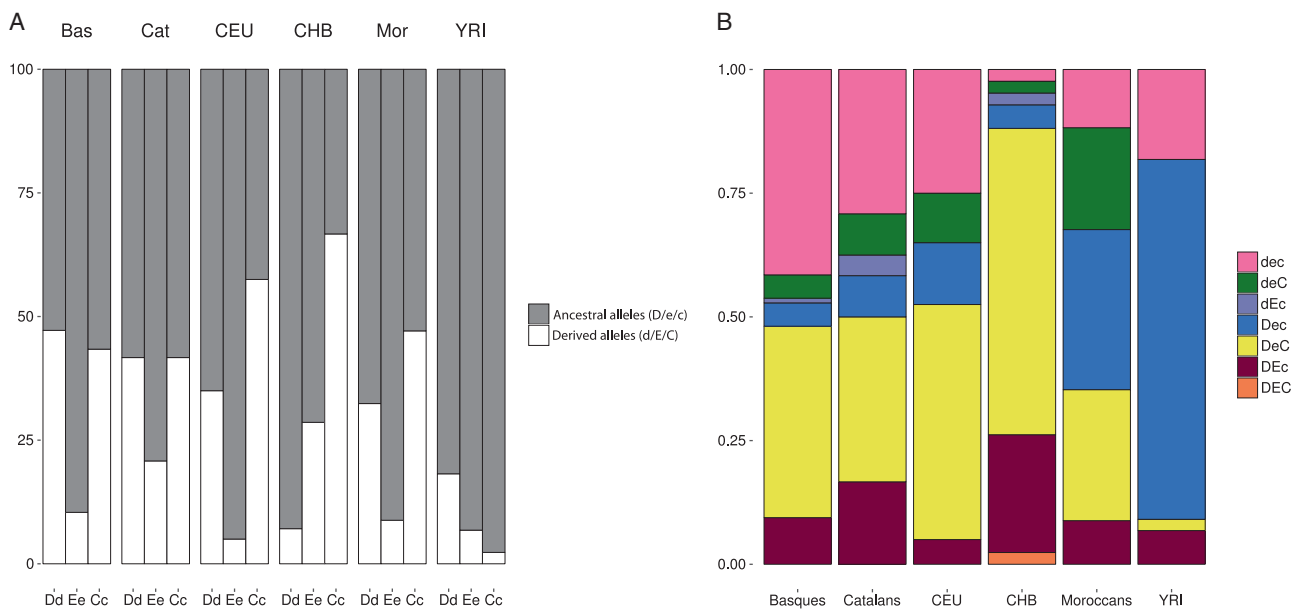
Fig. 1 Frequency plots related to the three analyzed *RHD/RHCE* polymorphisms. **a** Frequencies of the variants within populations. **b** Frequencies of the expected phenotypes within populations. Bas Basque, Cat Catalans, CEU HapMap Europeans, CHB HapMap Chinese, Mor Moroccans, YRI HapMap Yorubans

deletion of the *RHD* gene and both *RHD* and *RHCE* present a high sequence identity, we have focused our analysis on the flaking regions of these genes. We sequenced two regions of 6 kilobases (kb) each, upstream and downstream of the *RHD* gene outside the *Rhesus boxes* (GRCh37/hg19 coordinates chr1:25578836-chr1:25585562 and chr1:25664732-chr1:25671293, respectively) as delimited by Perry et al. [23], and a region of 6 kb downstream the *RHCE* gene (chr1:25682780-chr1:25688684). We amplified these three regions following a Long Touchdown PCR protocol using the primers by Perry et al. [23]. and new designed primers for the downstream region of the *RHCE* gene (Table S1 and Figure S1). After the PCR amplification, the concentration of the products was normalized in order to prepare a library using the *Illumina*® *Nextera XT DNA Library Preparation Kit* before sequencing with *Illumina MiSeq*® with 2 × 250 cycles.

Mapping, variant calling, and phasing

Sequencing reads were mapped with the reference genome GRCh37/hg19 using the *Burrows-Wheeler Aligner* (BWA) software [25]. Mean sequencing coverage for each sample was calculated (Figure S2) in order to ascertain sequence quality. Then, SNP calling was performed with the G*enome Analysis Tool Kit* (GATK) v3.5 software package [26]. To define the Rh phenotype in our samples, we selected a position within the *RHD* gene (chr1:25627957) and created a virtual SNP to represent the lack (T) or presence (C) of the gene copy, referring to the RhD negative or RhD positive phenotype, respectively. Then, we merged our data

with those from Perry et al. [23] and used the SHAPEIT v2.778 software [27, 28] to infer haplotypes. Generated data files are available in: https://figshare.com/s/e0fce29f846e741c601c. Samples and SNP genotyping information is available in Bioproject accession number: PRJNA473473. Sequence data can be found at GenBank accession numbers: MH404266-MH404429 for *RHD* upstream flanking region; MH404430-MH404569 for *RHD* downstream flanking region; and MH404570-MH404669 for *RHCE* downstream flanking region.

Sequence analysis

Summary statistics were calculated using DnaSP v5.10.1 (ref. [29]) and genetic structure measured by F_{ST} values were calculated by region using Arlequin v3.5.2.2. (ref. [30]). Patterns of linkage disequilibrium (LD) within populations were calculated with the Haploview v4.2 (ref. [31]). Each LD block analysis was performed by population, and all three Rh regions were analyzed and plotted together in order to detect LD between different regions. Blocks were estimated by using the *confidence intervals* algorithm defined by Gabriel et al. [32] and Hardy-Weinberg *p*-values and minimum minor allele frequency cut-offs were set at 0.0001 and 0.05, respectively. The virtual RHD SNP was excluded in the LD block analyses. In order to define the relationship of the inferred haplotypes over the regions among populations and Rh genotypes, median-joining networks were obtained for both RHDup and RHDdown regions, together and separately, by using

Table 1 Summary table of the RH flanking regions by population

	Pop.	RHDup				RHDdown				RHCEdown			
		N	SNPs	π	H	N	SNPs	π	H	N	SNPs	π	H
All haplotypes	Bas	106	12	0.0003	0.74	94	9	0.00036	0.82	84	13	0.0005	0.81
	Cat	24	9	0.0003	0.79	22	7	0.00039	0.85	16	7	0.0005	0.87
	Mor	34	12	0.0003	0.77	24	20	0.00064	0.89	—	—	—	—
	CEU	40	10	0.0003	0.72	40	8	0.00043	0.8	—	—	—	—
	CHB	42	6	0.0002	0.6	42	8	0.00036	0.7	—	—	—	—
	YRI	44	24	0.0005	0.9	44	34	0.00099	0.9	—	—	—	—
RhD positive	Bas	56	10	0.0002	0.59	50	8	0.00031	0.72	43	11	0.0004	0.77
	Cat	14	9	0.0003	0.82	13	7	0.00044	0.8	9	6	0.0004	0.86
	Mor	23	8	0.0002	0.67	15	17	0.00056	0.88	—	—	—	—
	CEU	26	7	0.0001	0.41	26	8	0.00031	0.63	—	—	—	—
	CHB	39	5	0.0001	0.53	39	8	0.00035	0.67	—	—	—	—
	YRI	36	23	0.0005	0.91	36	32	0.00092	0.91	—	—	—	—
RhD negative	Bas	50	7	0.0001	0.46	44	6	0.00016	0.61	41	9	0.0002	0.64
	Cat	10	4	0.0002	0.51	9	4	0.00026	0.72	7	5	0.0004	0.76
	Mor	11	6	0.0003	0.82	9	16	0.00073	0.92	—	—	—	—
	CEU	14	4	0.0002	0.73	14	3	0.00018	0.6	—	—	—	—
	CHB	3	0	0	0	3	0	0	0	—	—	—	—
	YRI	8	6	0.0003	0.75	8	18	0.00012	0.89	—	—	—	—

RHDup, RHDdown, and RHCEdown refer to the three Rh flanking sequenced regions: *RHD* upstream, *RHD* downstream, and *RHCE* downstream, respectively. "*N*" is the number of available haplotypes for each region in each population. "SNPs" represents the number of single nucleotide polymorphisms detected after performing the Variant Calling process. "π" indicates the nucleotide diversity. "H" is the haplotype diversity. For *RHCE* only data from Basques and Catalans is available.

the *Fluxus Network* v5 and *Fluxus Network Publisher* v2.1.1.2 programs [33]. F_{ST} values for the D/d, E/e, and C/c variants were calculated as recommended by Bhatia et al. [34] in order to avoid artifacts due to sample size differences, adapting the R script by Di Gaetano et al. [35]. To compare the Rh values to a genomic reference, genome-wide F_{ST} values of 74255 SNPs were estimated in the CEU, CHB and YRI populations from HapMap Phase III [36], Moroccans [37], Basques [38], and Catalans (unpublished data).

Results

RHD and RHCE phenotypes, genotypes, and haplotypes

Genotyping results and genotype frequencies of Basques and the other analyzed populations are available in Table S2. Estimated allele frequencies for the D/d, C/c, and E/e variants and expected phenotypes for the *RHD* and *RHCE* genes are shown in Fig. 1 and Table S3. All three polymorphisms were in Hardy–Weinberg equilibrium after Bonferroni correction in all samples. As previously reported, European populations present the highest estimated worldwide frequencies of the *RHD* deletion (d) [39].

Furthermore, our analyses show that the highest frequency of the *RHD* deletion is found in Basques (47.2%), in agreement with other analyses [12, 22]. However, the frequencies found in our Basque samples are not as extreme as shown in some previous studies, which were mainly based on antigen reactions. In particular, the frequency of the *RHD* deletion in Basques is just slightly higher than that found in Catalans (41.7%, *p*-value = 0.625). Regarding the functional polymorphisms of the *RHCE* gene, the frequency of the derived C allele is very low in sub-Saharan Africans, whereas the C/c alleles showed intermediate frequencies in the remaining populations, as first reported by Mourant in 1954 (ref [13]). Finally, the frequency of the derived E allele is low in most of the analyzed populations (Fig. 1a and Table S3). When we explored the haplotypes defined by the three *RHD/RHCE* polymorphisms (Fig. 1b), the most frequent RhD negative haplotype is dec [14, 39, 40], except in Moroccans where haplotype deC is more common. Among RhD positive haplotypes, the ancestral haplotype Dec [39, 40] is the most frequent in Africans, whereas in non-African populations haplotype DeC is found at higher frequencies [41]. The dEC haplotype was not found in our population dataset, and DEC was only found at low frequencies in the CHB population. These two haplotypes are known to be the most complex derived haplotypes since they have been suggested to be originated by recombination
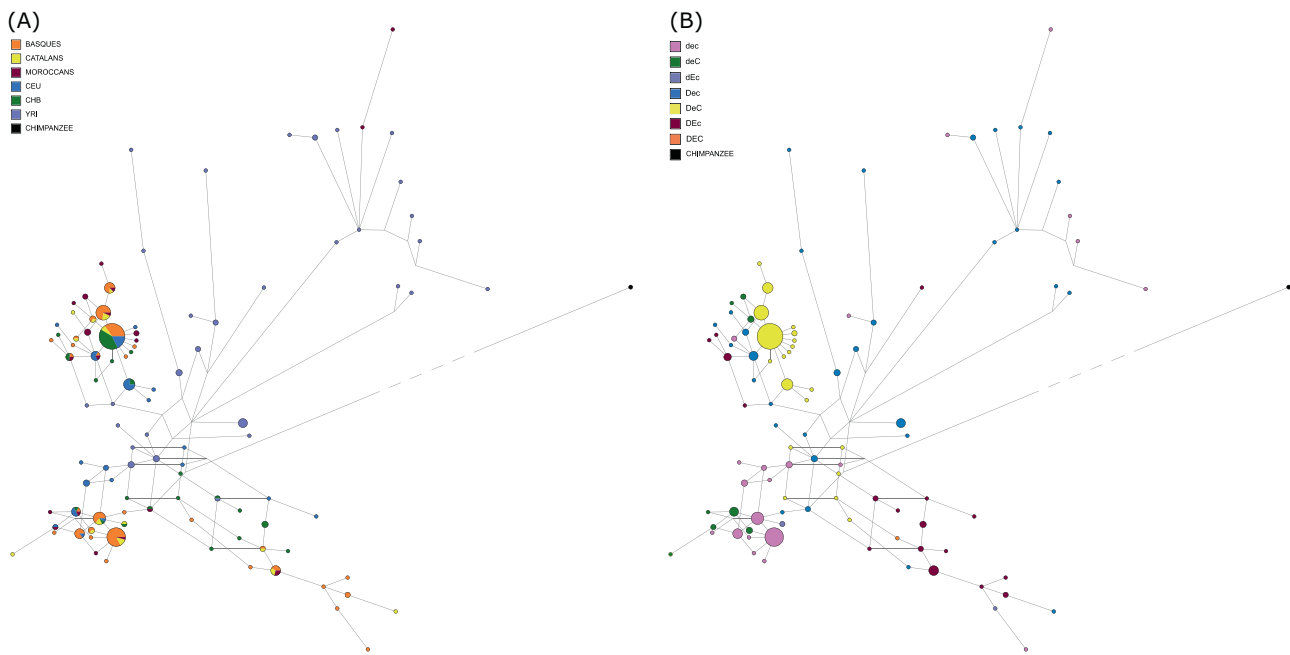
Fig. 2 Network analysis of haplotypes from both RHDup and RHDdown flanking regions. The network was colored by population (**a**) and the three *RHD/RHCE* polymorphisms haplotypes (**b**). Network for the regions separately is shown in Supplementary materials (Figure S4)

of the ancestral forms [14, 39, 40]. Regarding the haplotype frequencies found in Basques, these are concordant, though with small differences, with previous immunological studies. In particular, Basques present the highest frequency of the dec haplotype (41.5%) and the lowest frequencies of the deC (4.7%), dEc (0.9%), and Dec (4.7%) haplotypes among the European populations analyzed here [11, 12].

RHD/RHCE sequence analysis

In the present study we sequenced three ~6 kb regions flanking the *RHD* and *RHCE* genes in Basques and Catalans: the *RHD* upstream (RHDup) region, the *RHD* downstream (RHDdown) region, and the *RHCE* downstream (RHCEdown) region. Additionally, we also sequenced the RHDup and the RHDdown regions in Moroccans and compiled available genotyping and sequencing data for the same two regions in CEU, CHB and YRI from Perry et al. [23]. A total of 47 SNPs were found in the RHDup region, 49 SNPs in the RHDdown and 13 in the RHCEdown (data available only for Basques and Catalans) (SNP information by individual is available in Table S4; inferred haplotypes of each flanking region in the studied populations are shown in Table S5). No indels were detected in any of the regions of our dataset. For the RHDup and the RHDdown regions, sub-Saharan Africans presented a higher number of SNPs as well as higher nucleotide and haplotype diversity compared to the rest of the populations (Table 1) in accordance with the out-of-Africa bottleneck. Both diversity indexes and the corresponding F_{ST} pairwise population comparisons place

sub-Saharan Africans as the most differentiated population (Figure S3, Table S6).

LD analysis sheds light on the forces leading the evolution of a genomic region as well as its geographic subdivision. In order to test if LD patterns could be related to the higher frequency of RhD negative in Basques, we estimated LD blocks by population around the genomic region analyzed. Patterns of LD in the *RHD/RHCE* region are similar in all three European populations analyzed, including Basques (Figure S4). Notably, an LD block in the RHDdown region is found in all populations analyzed. This block includes the *SMP1/TMEM50A* gene, located between both RH genes but apparently functionally unrelated to the Rh system. In Europeans, this block extends upstream and includes the RHDup region. The C/c and E/e alleles were not part of this LD block in all cases.

In order to determine the relationship between the haplotypes in the *RHD/RHCE* region, a network analysis was performed. Similar patterns were found in all obtained networks when analyzing both RHD regions together (Fig. 2) or separately (Figure S5). Mainly, they show more diversity in Africa and clear clusters of sequences can be distinguished according to their Rh phenotypes. Two main differentiated sections corresponding to RhD positive and RhD negative haplotypes are clearly observed, which suggests that the RHD deletion is the result of a major single mutational event [15]. There are other minor cases in the network that could be explained by artifacts in the haplotype reconstruction, mistyping, or recurrent events. The RhD negative section is mainly represented by Basques and
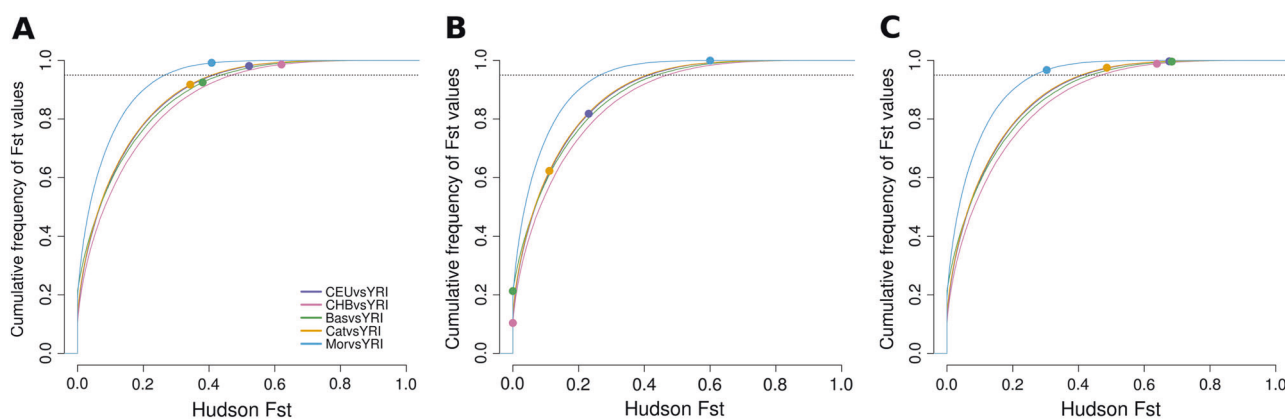
Fig. 3 Unusual patterns of population differentiation for the C/c variant in non-African populations. Lines are the distribution of the genome-wide F_{ST} values calculated between YRI and the remaining populations included in our analysis. Dots represent the F_{ST} values of the C/c variant. Dotted line represents the 95 percentile. The analysis was performed for all our haplotypes together (**a**), and for RhD negative (**b**) and RhD positive (**c**), separately

is less diverse than the RhD positive section (Table 1). This lack of diversity might be expected since the RhD negative phenotypes derive from the ancestral RhD positive form [14, 15, 17]. The ancestral Dec haplotype is mainly distributed around the center of the network.

Unexpectedly, haplotype analysis shows that non-sub-Saharan RhD positive haplotypes are associated with the derived C allele (with high frequencies of the DeC haplotype), while sub-Saharan RHD positive haplotypes are mostly found associated with the ancestral c allele (Fig. 1, Fig. 2, Figure S5 and Table S3). Moreover, variation at the RH flanking regions is quite conserved in the DeC haplotypes, most of them presenting similar sequences that determine four main nodes and several minor haplotypes, conforming a star-like structure in the network. Next, we computed F_{ST} to test whether population-specific positive selection [42, 43] could explain this haplotype pattern. To this end, the F_{ST} values for the C/c variant were compared to the genome-wide distribution in each population taking into account all haplotypes, as well as with the RhD positive and RhD negative considered independently (Fig. 3, Table S7). Our results not only confirmed the extreme F_{ST} values for the C/c variant when comparing YRI to CEU and CHB, respectively, as reported by Perry et al. [23], but also when comparing YRI to Basques, Catalans, and Moroccans. Notably, the F_{ST} values for the C/c variant were extreme, and fell in the 95 percentile of the genomic distribution of F_{ST} values in all population comparisons when considering only RhD positive haplotypes.

Discussion

In our analyses of allele and haplotype frequencies for the *RHD/RHCE* variants in Basques, we obtained similar observations to those shown in previous immunological studies of the Rh system with no striking particular patterns of allele, haplotype or sequence diversity in Basques. Haplotypic composition and allele frequency are very close to other European populations, suggesting that their differentiation, at least based on Rh polymorphisms, is lower than usually expected [12, 22]. It has been shown that the lowest frequencies of RhD negative are found in Asia, in indigenous populations from America and Pacific Area, as well as in Africa. On the other hand, the reported frequencies are higher in Europe, especially in Basques [39]. The RhD negative frequency in our Basque samples (47.2%), despite being the highest among the analyzed populations, is not as extreme as usually suggested, but within the range reported in previous immunological studies around the Franco-Cantabrian region (45–54%) [1, 11–13, 22]. In fact, the frequency of the RhD negative allele is found at a 41.7% in our Catalan sample, and other studies also reported relative similar frequencies in other North Iberian surrounding populations [44]. The small differences we find in the frequency of RhD negative in Basques, as compared to previous analyses, could be due to a sampling bias or the differential accuracy in serological methods and DNA data. Thus, rather than isolated extreme values, high RhD negative frequencies in Basques seem to be the end-point of a cline encompassing also neighboring populations. We confirm in Basques the highest frequency of dec haplotype and lowest frequencies of DEC and dEc. However, we do not observe a lower frequency of DEc in Basques, compared to Western Europeans as reported by previous studies [11, 12]. Instead, this haplotype is showing frequencies similar to those found in other European populations [39]. Even though the origins of the higher frequency of RhD negative variant in Europe, principally in Basques, are still controversial, our results suggest that it is led by a major

deletion event. Thus, a different origin or repeated, independent mutations cannot be invoked to explain the higher frequency of the RhD negative allele in Basques.

Demographic processes have probably been the main factor in the evolution of the Rh polymorphisms frequencies. Bottlenecks and low effective population sizes reduce the effectiveness of potential selection, and, in social contexts where family size is tightly controlled, new pregnancies could compensate the effect of neonatal HDN births [45]. Clear evidence of selection around the system has not been shown yet. Frequencies of RhD negative higher than those of RhD positive are not observed in any population. Moreover, the genotype with the highest frequency is the heterozygote in most of the cases, except in the populations with a very low frequency of RhD negative (Fig. 1, Tables S2 and S3). Hence, studying selection around the Rh system results much more complex than a simplistic analysis of directional selection against or in favor of RhD negative depending on the population. A heterozygote disadvantage scenario has been posited since the potential HDN cases are directly associated with heterozygote children (Dd), from RhD negative mothers (dd) and RhD positive fathers (DD or Dd). This would lead to an unstable equilibrium of the frequencies with a tendency towards the increment of the major allele and its homozygote in a specific population [21]. Otherwise, a balancing selection scenario has been also often suggested to explain the polymorphism and the high frequencies of RhD negative in populations, either by frequency-dependent selection or by some selective advantage linked to the RHD deletion that could overcome its association to the HDN. In the latter case, a possible heterozygote advantage by protection against *Toxoplasma gondii* has been suggested [21, 46, 47]. However, this kind of association has not been clearly demonstrated and analyzing balancing selection is a really complex issue [45–49]. Thus, not only demographic processes but maybe also different selective processes may have been related to the origin of the diverse frequencies of RhD negative/RhD positive in populations. Basques have been characterized by their attractive and unsolved history, probably determined by isolation and drift. Hence, reduced effective population size in a historically isolated population may have decreased the efficiency of the purifying selection associated with HDN, and some events may have generated specific selective pressures and set a different equilibrium state for the *RHD* deletion allele [23, 45, 46, 49].

Finally, our data confirmed an association between the derived RhC allele and the RhD positive variant, as well as extreme values of F_{ST} for the Rh C/c variant when comparing sub-Saharan and non-sub-Saharan populations being consistent with the unexpected result from Perry et al. [23]. In this study, the high population differentiation was suggested to be result from a possible local positive selection process around the *RHCE* region. However, they did not obtain solid evidence of signals of selection. Moreover, as reported by Gardner et al. [50], extreme values of F_{ST} in individual markers do not imply population-specify selection, and the selection signal should extend to the surrounding regions. Furthermore, the scarce information about the function of the *RH* genes makes it difficult to provide hypotheses for the adaptive value of *RH*. It is pivotal to increase the available information with further genetic studies, specially based on the functions of the *RHD/RHCE* system products and the correlation with genotypes, to better understand the evolution and population genetics of the Rh system in humans.

Compliance with ethical standards

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

References

1. Aguirre A, Vicario A, Mazón LI, et al. Are the Basques a single and a unique population? Am J Hum Genet. 1991;49:450–8.

2. Izagirre N, de la Rúa C. An mtDNA analysis in ancient Basque populations: implications for haplogroup V as a marker for a major paleolithic expansion from southwestern europe. Am J Hum Genet. 1999;65:199–207.

3. Torroni A, Bandelt HJ, D'Urbano L, et al. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet. 1998;62:1137–52.

4. Behar DM, Harmant C, Manry J, et al. The Basque paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-neolithic times. Am J Hum Genet. 2012;90:486–93.

5. Cavalli-Sforza L. The Basque population and ancient migrations in Europe. MUNIBE (Antropol Y Arqueol. 1988;6(Suppl): 129–37.

6. Günther T, Valdiosera C, Malmström H, et al. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. Proc Natl Acad Sci. 2015;112:11917–22.

7. Lazaridis I, Patterson N, Mittnik A, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature. 2014;513:409–13.

8. Cavalli-Sforza L, Menozzi P, Piazza A. The history and geography of human genes. Princeton, NJ: Princeton University Press; 1994. P.

9. Rodríguez-Ezpeleta N, Álvarez-Busto J, Imaz L, et al. High-density SNP genotyping detects homogeneity of Spanish and French Basques, and confirms their genomic distinctiveness from other European populations. Hum Genet. 2010;128:113–7.

10. Laayouni H, Calafell F, Bertranpetit J. A genome-wide survey does not show the genetic distinctiveness of Basques. Hum Genet. 2010;127:455–8.

11. Van Der Heide HM, Magnée W, Van Loghem JJ. Blood group distribution in Basques. Am J Hum Genet. 1951;3:356–61.

12. Touinssi M, Chiaroni J, Degioanni A, De Micco P, Dutour O, Bauduer F. Distribution of Rhesus blood group system in the French Basques: a reappraisal using the allele-specific primers PCR method. Hum Hered. 2004;58:69–72.

13. Mourant AE. The distribution of the human blood groups. Oxford, UK: Blackwell Scientific Publications; 1954.

14. Avent ND, Reid ME. The Rh blood group system: a review. Blood. 2000;95:375–87.

15. Wagner FF, Flegel Wa. RHD gene deletion occurred in the Rhesus box. Blood. 2000;95:3662–8.

16. Westhoff CM, Siegel DL. Rh and LW blood group antigens. In: Rossi's principles of transfusion medicine. Chichester, WestSussex: John Wiley & Sons, Ltd.; 2016. p. 176–84.

17. Flegel WA The genetics of the Rhesus blood group system. In: Blood transfusion. 2007. p. 50–57.

18. Westhoff CM. The structure and function of the Rh antigen complex. Semin Hematol. 2007;44:42–50.

19. Ahmad R, de Hass M. Prevention of haemolytic disease of the fetus and newborn with reference to Anti-D. MedCrave Group LLC; 2017.

20. Prevention of haemolytic diseases of the newborn due to anti-D Infective endocarditis. Br Med J (Clinical Res Ed) 1981; 282:676–7.

21. Vogel F, Motulsky AG. Human genetics. Berlin: Springer; 1997. https://doi.org/10.1007/978-3-662-03356-2.

22. Etcheverry MA. El factor rhesus: Su genética y su importancia clínica. Dia Med. 1945;17:1237–59.

23. Perry GH, Xue Y, Smith RS, et al. Evolutionary genetics of the human Rh blood group system. Hum Genet. 2012;131:1205–16.

24. Kent WJ, Sugnet CW, Furey TS, et al. The Human Genome Browser at UCSC. Genome Res. 2002;12:996–1006.

25. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.

26. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.

27. Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. Nat Methods. 2011;9:179–81.

28. O'Connell J, Gurdasani D, Delaneau O, et al. A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. PLoS Genet. 2014;10:e1004234.

29. Librado P, Rozas J. DnaSPv5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–2.

30. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10:564–7.

31. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21:263–5.

32. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science. 2002; 296:2225–9.

33. Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16:37–48.

34. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: The impact of rare variants. Genome Res. 2013; 23:1514–21.

35. Di Gaetano C, Fiorito G, Ortu MF, et al. Sardinians genetic background explained by runs of homozygosity and genomic regions under positive selection. PLoS ONE. 2014;9:e91237 https://doi.org/10.1371/journal.pone.0091237

36. Gibbs RA, Belmont JW, Hardenbol P, et al. The International HapMap Project. Nature. 2003;426:789–96.

37. Arauna LR, Mendoza-Revilla J, Mas-Sandoval A et al. Recent historical migrations have shaped the gene pool of Arabs and Berbers in North Africa. Mol Biol Evol. 2017;34:318–29.

38. Henn BM, Botigué LR, Gravel S, et al. Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. PLoS Genet. 2012;8:e1002397.

39. Daniels G. Human blood groups. 3rd ed. 2013 https://doi.org/10.1002/9781118493595.

40. Carritt B, Kemp TJ, Poulter M. Evolution of the human RH (rhesus) blood group genes: A 50 year old prediction (partially) fulfilled. Hum Mol Genet. 1997;6:843–50.

41. Sheeladevi CS, Suchitha S, Manjunath GV, Murthy S. Hemolytic disease of the newborn due to anti-c isoimmunization: a case report. Indian J Hematol Blood Transfus. 2013;29:155–7.

42. Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. Natural selection has driven population differentiation in modern humans. Nat Genet. 2008;40:340–5.

43. Xue Y, Zhang X, Huang N, et al. Population differentiation as an indicator of recent positive selection in humans: an empirical evaluation. Genetics. 2009;183:1065–77.

44. Calvet R, Pastor JM, Fernández R, Romero JL. Blood groups of the Cantabrian population (Spain). Hum Hered. 1992;42: 120–4.

45. Feldman MW, Nabholz M, Bodmer WF. Evolution of the Rh polymorphism: a model for the interaction of incompatibility, reproductive compensation, and heterozygote advantage. Am J Hum Genet. 1969;21:171–93.

46. Flegr J. Heterozygote advantage probably maintains rhesus factor blood group polymorphism: ecological regression study. PLoS ONE. 2016;11:e0147955.

47. Novotná M, Havlíček J, Smith AP, et al. Toxoplasma and reaction time: role of toxoplasmosis in the origin, preservation and geographical distribution of Rh blood group polymorphism. Parasitology. 2008;135:1253–61. https://doi.org/10.1017/S003118200800485X

48. Fijarczyk A, Babik W. Detecting balancing selection in genomes: limits and prospects. Mol Ecol. 2015;24:3529–45.

49. Flatz G, Brinkmann H-B. A selection model for the rhesus blood-group system. J Hum Evol. 1977;6:135–40.

50. Gardner M, Williamson S, Casals F, et al. Extreme individual marker FST values do not imply population-specific selection in humans: the NRG1 example. Hum Genet. 2007;121:759–62.