Evolution of the *O* alleles of the human ABO blood group gene

Francis Roubinet, Stéphanie Despiau, Francesc Calafell, Fen Jin, Jaume Bertanpetit, Naruya Saitou, and Antoine Blancher

BACKGROUND: To date, at least 40 different alleles *O* have been characterized on the basis of exon 6 and exon 7 sequences but not always for intron 6.

STUDY DESIGN AND METHODS: Among 415 individuals, from four continents (Africa, Europe, South America, and Asia), studied for exon 6 and exon 7 sequences, we selected 46 individuals (of respective phenotypes O [39], AB [3], B [3], or A [1]) for sequencing 1800-bp amplicons spanning exon 6, intron 6, and exon 7. The amplicons were characterized either by direct sequencing or after cloning when required.

RESULTS: We defined 14 new intron 6 *O* allele sequences, including four recombinant alleles. Based on sequence comparison, a phylogenetic network was constructed. It confirmed recombinant allele origins and that most *O* alleles are derived by point mutations from the two worldwide distributed alleles *O01* and *O02*.

CONCLUSION: Allele *O* phylogenetic analysis suggests that the most frequent silencing mutation (deletion of a G in exon 6) appeared once in human evolution in the ancient *O02* allele lineage and that allele *O01* resulted from an interallele exchange between *O02* and *A101*. Assuming constancy of evolutionary rate, diversification of the representative alleles of the three human *ABO* lineages (*A101*, *B101,* and *O02*) was estimated at 4.5 to 6 million years ago.

The ABO system was discovered by Landsteiner in 1901.[1] It is the most important blood group system in transfusion and transplantation practices. It was also the first human genetic system to be applied to anthropologic studies. The *A* and *B* alleles of the *ABO* gene code for glycosyltransferases that either add an *N*-acetylgalactosamine or a galactose to various glycoconjugates. These products result in A or B blood-group-specific antigens. The *O* allele corresponds to a silent (null) allele of the *ABO* gene. The cDNAs of the three major alleles of the human blood group ABO system were first described by Yamamoto and coworkers.[2,3] The human *ABO* locus spanning 18 kb, lies on chromosome 9q34 and encompasses seven exons giving a 1062-bp open reading frame. Exons 1 through 5 encode the amino terminal part, a transmembrane region, and 9 percent of the catalytic domain. Exons 6 and 7 encode 77 percent of the protein and 91 percent of the catalytic active part.[4] Because of

From the Laboratory of Molecular Immunology, Paul Sabatier University, Rangueil Hospital, 31403 Toulouse Cedex 4, France; the Laboratory of Immunohematology, French Establishment of Transfusion of the Pyrenes-Mediterranean, Toulouse Site, Avenue of Great Britain, BP 3210, 31027 Toulouse Cedex, France; Unit of Evolutionary Biology, Faculty of Health Life Sciences, Pompeu Fabra University, Doctor Aiguader, 80, 08003 Barcelona, Catalonia, Spain; the Institute of Genetics and Developmental Biology, Academia Sinica, Beijing, China; and the Division of Population Genetics, National Institute of Genetics, Mishima, 411-8540, Japan.

Address reprint requests to: Antoine Blancher, MD, PhD, Laboratoire d'Immunogénétique Moléculaire, Université Paul Sabatier, Hôpital Rangueil, Faculte de Medicine Bâtiment A2, 133 Route de Norbonne, 31062 Toulouse Cedex 4, France; e-mail: blancher@easynet.fr.

This work was supported by funds from the Agence Française du Sang (Contract 97002551231) from the Etablissement Français du Sang and from the Ministère des Affaires Etrangères, France (Contract EA 3034).

Received for publication October 27, 2003; revision received December 20, 2003, and accepted January 8, 2004.

TRANSFUSION 2004;44:707-715.

that, *ABO* gene polymorphism has been studied, almost exclusively, in exons 6 and 7.

Since this pioneering period, many variants of the classical *A* and *B* alleles have been characterized.[5,6] Moreover, numerous inactive (*O*) alleles have been reported.[5,7,8] The high diversity of inactive alleles in this gene gives a precious opportunity to study convergent loss of function and the role of selection in the maintenance of such diversity in populations.

The most frequent human *O* alleles, *O01* and *O02* (http://www.bioc.aecom.yu.edu/bgmut/abo.htm), differ in exons 6 and 7 by nine nucleotide substitutions[2,9] but share one nucleotide deletion of a G at position 261 in exon 6 of the *A101* sequence (Δ261).[5] This deletion induces a frameshift and creates a premature stop codon (nucleotides 352-354), resulting in a truncated (117 amino acids) protein deprived of any glycosyltransferase activity.[2] Numerous variants of *O01* and *O02* alleles have been described. They differ from *O01* or *O02* by a few point mutations.[5,7,8,10-14] A rare silent allele, initially called "another *O* allele"[15] or "O^2"[16] now referred to as *O03,* was observed only in Europeans and at low frequencies in sub-Saharan Africans.[10-12,17-21] Allele *O03* does not contain the G261 deletion but carries a nonsynonymous inactivating mutation in exon 7.[15,22] Another rare null allele,[23] initially called O^3, and also referred to as *O08*,[5] differs from *O01* and *O02* by the absence of the Δ261 mutation. Allele *O08* is basically an *A2* allele in exons 6 and 7 but displays an inactivating mutation in exon 7, which consists of an insertion of a G in a stretch of G's (between nt 798 and 804).[23] More recently, two other inactivating mutations were identified in very infrequent silent alleles: allele *O4* has a G insertion at nt 87 to 88 in exon 2 of a *A101* allele resulting in a premature stop codon (codon 56), and allele *O5* displays a non-sense mutation in exon 6 (C322T).[7,24]

We previously studied the polymorphism of the *O* allele in five human populations of three continents, Europe, Africa, and South America.[12] Nevertheless, the ABO blood group gene polymorphism was characterized by direct sequencing of exons 6 and 7 only. In this study, we extend the analysis of *O* alleles encountered in the five populations of our previous study by sequencing fragments amplified from the 3′ end of intron 5 to the end of exon 7 (including intron 6), which, as discussed below, is crucial to properly resolve the phylogeny reconstruction of the allelic lineages of this locus. Two Chinese populations of O phenotypes were added in the study. We also analyzed the phylogenetic relationship between human *O* alleles and the two main functional alleles *A101* and *B101* to understand the molecular mechanisms that have generated each *O* allele. This analysis gives a more complete understanding of the diversity and origins of *O* alleles.

MATERIALS AND METHODS

Samples

We used the same human samples as those described in our previous study of exons 6 and 7,[12] that is, one Akan individual of B phenotype and 317 unrelated individuals of O phenotype from five human populations, comprising Akans from Ivory Coast, Moroccan Berbers, Basques from Spain and France, and Cayapas and Aymaras from Bolivia. We also studied 6 Caucasian blood donors from the Toulouse area (southwest France) and one Akan individual to characterize sequences of 14 common alleles (4 *A* alleles, 5 *B* alleles, 2 *O01,* 2 *O02,* and 1 *O03*). Furthermore, individuals of O phenotype from two Han populations in China (47 individuals at Putien in Fujien Province and 43 individuals at Fujiou in Guandong Province, both Han Chinese) were newly examined in this study.

Among these 415 individuals, studied for exon 6 and exon 7 sequences, we selected 46 individuals (of respective phenotypes O [39], AB [3], B [3], or A [1]) for sequencing 1800-bp amplicons spanning exon 6, intron 6, and exon 7. Among individuals of O phenotype for which DNA samples were available, we selected individuals having *O* alleles of each type.

Characterization of *ABO* alleles

Genomic DNA extraction and PCR amplification of a region encompassing the end of intron 5 to the end of exon 7 were performed as indicated in Roubinet et al.[12] Primer locations are indicated in Fig. 1. Primers int.5-dir and 3′UTR-rev were used to amplify the genomic fragment encompassing the end of intron 5, exon 6, intron 6, and exon 7, in 30 cycles of PCR amplification with a 2.5-minute elongation step. PCR products were cloned into the pCR2.1 Topo plasmid vector (Topo TA cloning Kit, Invitrogen, Leek, the Netherlands).[12] Cloned products were sequenced using the fluorescent dye terminator method (Amplitaq FS, PE Applied Biosystems, Foster City, CA) as previously described.[12]

The sequences obtained by cloning amplicons of a given individual were aligned and compared to characterize the two alleles of each individual. As direct sequences were compatible with those deduced from combination of alleles defined by cloning, 17 individuals were studied only by direct sequencing of amplicons from exon 6 to exon 7 (including intron 6). A phylogenetic network was manually constructed as in Saitou and Yamamoto.[25]

Nomenclature of alleles

Alleles were named following the unofficial nomenclature of Yamamoto (http://www.bioc.aecom.yu.edu/bgmut/abo.htm) followed between parentheses, when useful, by the original names given by the authors who described the

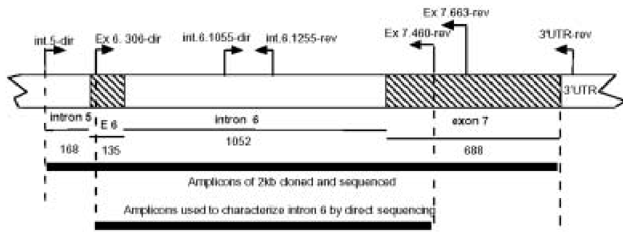

Fig. 1. Schematic representation of the genomic organization of the human *ABO* gene including intron 5, exon 6, intron 6, and exon 7 with primer positions. Primer sequences: int.5-dir, TGATTTGCCCGGTTGGAGTCGC; int.6.1055-dir, GCTCACTGAC CAGAGATAGC; int.6.1255-rev, AGTATGTGTCTGCGGTTGC; Ex6.306-dir, CGACATCCTCAACGAGCAGTTC; Ex7.663-rev, GGAGTCAGGATCTCCACG; Ex7.460-rev, CGGTGAAGACATAG TAGTGGACACG; and 3′UTR-rev, CCAGAGCCCCTGGCAGC CGCT*C*. 168, 135, 1052, and 688 correspond to numbers of bases of the analyzed regions (respectively end of intron 5, exon 6, intron 6, and exon 7).

TABLE 1. Distribution of O genotypes among the two Chinese populations

		Genotype		
Population	Total	*O01/O01*	*O01/O02*	*O02/O02*
Putien, Fujien	47	13	29	5
Fujiou, Guangdong	43	11	22	10

allele for the first time. When alleles are not nomenclatured by Yamamoto we gave only the names given by the original authors.

RESULTS

Nucleotide sequences

Direct sequences obtained from all 90 Chinese individuals studied allocate them all in just three genotypes: *O01/O01*, *O02/O02*, or *O01/O02* (Table 1). Genotype frequencies within O individuals were in Hardy-Weinberg equilibrium. Relative allele frequencies were 0.59 *O01* and 0.41 *O02* in Putien and 0.51 *O01* and 0.49 *O02* in Fujiou. All alleles encountered in the eight populations, but two (*021*($O^{1(C467T)}$ and $O^{1v(A681G\&C1054T)}$) were characterized by direct sequencing and cloning (for at least one allele) from exon 6 to exon 7, including intron 6 (Table 2). Individuals studied by cloning are given without parentheses in Table 2. In total we characterized 24 *ABO* alleles from exon 6 to exon 7. Of these alleles, 15 (14 *O* alleles and 1 variant of *B101* allele) had never been characterized at the level of intron 6. These sequences (comprising the total of exons 6 and 7 and intron 6) are shown aligned in Fig. 2 and allele frequencies in the different population samples are shown in Table 3.

All four sequences of A^1 alleles studied were identical to the human genome sequence AC000397 (GenBank accession number), which is consistent with an *A101* allele.[7] As for *B* alleles, five sequences were identical to allele *B101* from exon 6 to exon 7. The sixth individual had a *B* allele (*B101-var1*) displaying one nucleotide difference in intron 6, when compared to allele *B101* (see Fig. 2).

By reference to our previous study,[12] polymorphism in intron 6 sequences splits three previously determined alleles into two subtypes: *O30* (*Ovar.tlse05*) is split into *Ovar.tlse51* and *Ovar.tlse52*, *O24* (*O1v-B*) into *O1v-B.tlse13* and *O1v-B.tlse14*, and *O13* (O^{v7}) into $O^{v7.1}$ and $O^{v7.2}$. Moreover, the analysis of intron 6 led us to conclude that allele *O06* described in the previous study[12] was in fact allele *O02*. This misidentification of *O06* alleles was due to misinterpretation of direct sequencing results of exon 6 amplicons at the time of our previous study. Indeed, taking into account only e6 and e7, allele *O06* differs from *O02* only at position 297 of e6.

An 8-bp motif is tandemly repeated three times in intron 6, with one additional repeat in three human alleles (*O1v-B.tlse14*, *Ovar.tlse52,* and $O^{v7.2}$; see Fig. 2); this may be considered to be an emergent microsatellite. The mutation had a single origin in allele $O^{v7.2}$ and recombination took it into the other alleles. The same motif has been found repeated three to five times in chimpanzee[26] and three times in gorilla (F. Roubinet, unpublished data).

Recombination events

Some alleles carry sequences that have clearly originated from recombination; these are stretches of identity or near identity with one allele followed by a stretch of identity or near identity with a different allele. These segments usually contain three or more informative polymorphic positions, making mutation as the origin of the allele extremely unlikely. In this study four new recombinant alleles were characterized, and we encountered alleles *O03* and *O05* already characterized as recombinants.[27,28] Five of these alleles (*O1v-B.tlse13, O1v-B.tlse14, Ovar.tlse51*, *O03,* and *O05*) and their two hypothetical parental alleles are depicted in Fig. 3.

Allele *O1v-B.tlse13* can be derived either by two mutations from *B101* or by a single recombination event between *O02* and *B101*. If it originated from two mutations from allele *B101*, one would be the deletion of a G at position 261 in exon 6. As discussed below, it is highly probable that this deletion occurred only once in allele *A101* (giving rise to allele *O01*). The hypothesis of a recombination involving alleles *O02* and *B101* between positions i42 and i88 thus seems more probable (for numbering of nucleotide positions see the legend to Fig. 2). This recombinant seems similar to $O^{1v\text{-}7}$ allele, previously reported in a German family by Bugert et al.[29]

The *O1v-B.tlse14* allele could have derived from point mutations either from *B101* (14 mutations and one inser-

TABLE 2. Determination of genotypes

Individual genotypes				
Identification of allele 1		Identification of allele 2		Individuals numbers and origin*
A101	A	*B101*	B	2, Toulouse
A101	A	*B101-var1*	Bv	1, Toulouse
O01	1	*B101*	B	1, Toulouse
O01	1	*O01*	1	(1, Bolivian and 2 Chinese)
O02	2	*O02*	2	(2, Chinese)
O02	2	*A101*	A	1, Toulouse
O02	2	*B101*	B	1, African
O02	2	*O01*	1	(1, Bolivian)
O03	3	*O01*	1	1, Toulouse cloned (1, Berbers, 2, Basques)
O03	3	*O02*	2	(2, Berbers)
O09	9	*O01*	1	1, African
O09	9	*O02*	2	1, Basque
O09	9	*O31*	31	1, African
O11	11	*O32*	32	1, Bolivian
O11	11	*O11*	11	(1, Bolivian)
O12	12	*O01*	1	3, Basques
O1v-B.tlse13	13	*O02*	2	1, African
O1v-B.tlse13	13	$O^{v7.1}$	7	1, African
O1v-B.tlse13	13	*O28*	28	2, Africans
O1v-B.tlse13	13	*O01*	1	(1, Basque)
O1v-B.tlse14	14	*O01*	1	1, African
Ovar.tlse20	20	$O^{v7.1}$	7	1, African (1, African)
Ovar.tlse20	20	$O^{v7.2}$	8	1, African (1, African)
Ovar.tlse20	20	*O02*	2	(2, Africans)
O26	26	*O02*	2	1, Basque
O27	27	*B101*	1	1, African
O28	28	$O^{v7.1}$	7	1, African
O29	29	*O01*	1	1, African
O33	33	*O02*	2	1, Bolivian
O34	34	*O01*	1	1, Basque
Ovar.tlse51	51	*O01*	1	1, Berber
Ovar.tlse52	52	*O01*	1	1, African

* Outside parentheses are listed the individuals studied by direct sequencing and cloning for E6-i6-E7 region. Between parentheses are listed the individuals studied only by direct sequencing.

tion of an 8-bp repeat in intron 6) or from *Ov7.2* (10 mutations). Nevertheless, a more parsimonious hypothesis is a recombination between $O^{v7.2}$ (derived from *O02*) and *B101* (recombination site between positions i1014 and e525). It should be noted that the *O1v-B.tlse13* and *O1v-B.tlse14* alleles are identical in exon 6 and 7 to the allele previously named O^{1v}-*B*, whose intron 6 sequence had not been sequenced before this study.[12,30]

The third new recombinant allele characterized here is *Ovar.tlse51*. It could have derived from $O^{v7.1}$ by six point mutations but it more probably derived from a single recombination between *O01* (5′) and $O^{v7.1}$ (3′) at a site between positions i236 and i445. Interestingly, *Ovar.tlse51* allele is observed in African populations, where the two parental alleles $O^{v7.1}$ and *O01* are frequent.

Another candidate having originated by recombination is allele *Ovar.tlse20*. If one excludes intron 6, it is identical to allele *O05,* which is a recombinant allele between alleles *O02* and *O01* or *A101*.[11] Nonetheless, allele *Ovar.tlse20* differs clearly from allele *O05* in intron 6 (nine substitutions).[11] Because allele *Ovar.tlse20* carries three mutations not seen in other alleles, as well as three apparent reversions of *O02* mutations and one singleton, it could correspond to a recombinant allele between *O01* and a subtype of *O02* not yet sampled or to an ancient allele that has had a long time to accumulate mutations.

Allele *O03,*[14,27] previously described as a complex recombinant, has also been found. We confirmed that a probable recombination site between *B101* and *A101* lies between e527 and e656.

Phylogenetic network of human alleles

We constructed a phylogenetic network (Fig. 4) for 26 human ABO blood group allele sequences, 23 being *O* alleles, to examine the internal relationship based on variant sequence information shown in Fig. 2. The position of the root of this phylogenetic network was estimated by comparing chimpanzee and gorilla sequences.[26] As analyzed above, six sequences (*Ovar.tlse51*, *O1v-B.tlse14, O1v-B.tlse13*, *Ovar.tlse20, O05,* and *O03*) are compatible with a recombinant origin seen as complex reticulations, a clear footprint of recombination. The two (*Ovar.tlse51* and *O1v-B.tlse14*) have not been incorporated in the network because of the complexity they generate in the graph. Taking into account only sequences from exon 6 to exon 7,

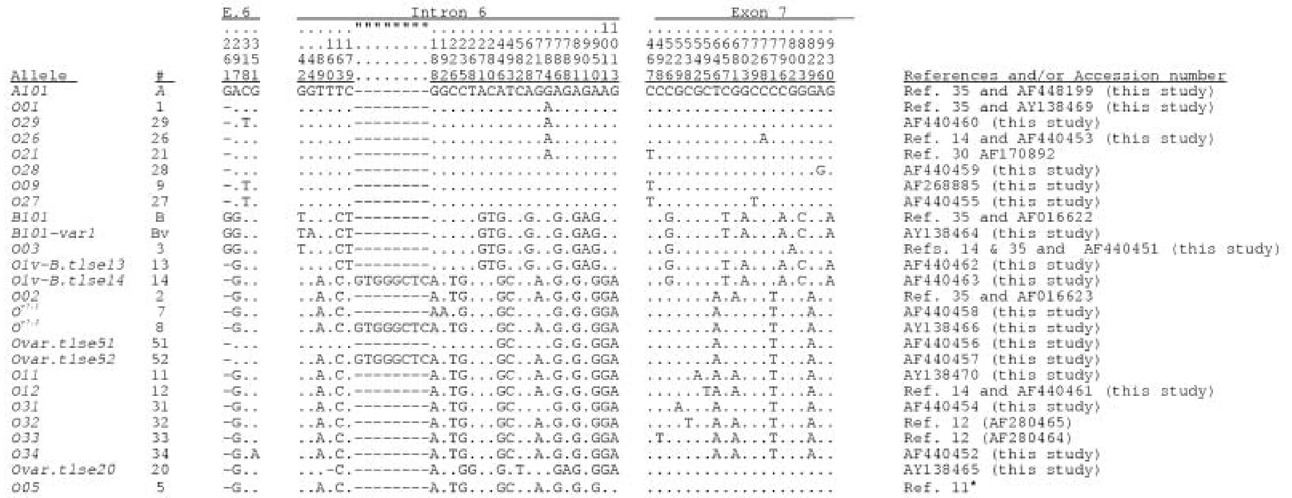

Fig. 2. Multiple alignments of human ABO blood group gene *O* allele sequences found in this study, with *A101* and *B101* major functional alleles. The sequences were aligned through Clustal W and are given by reference to *A101* allele. Sequence AF448499 is identical to the human genome sequence AC000397. Nevertheless, the latter reference does not mention the ABO type of the individual analyzed. Therefore, no *A101* allele sequence from exons 6 to 7 (including the intervening intron 6) was available in the databanks. For that reason, we have deposited our sequence under Accession Number AF448499. Points signify identity with the reference and dashes indicate missing bases. The numbering of nucleotide positions in exons 6 and 7 is given by reference to cDNA numbering (the A of the start codon being 1). As for intron 6 the numbering is given by reference to intron 6 of allele *A101* (position 1 being the first base of the intron). The insertion of 8 bp in certain alleles is not taken into account for the numbering. In the text, nucleotide positions are given by reference to that used in the figure: if the position belongs to an exon, numbers are preceded by the letter "e," and if the position belongs to intron 6, the numbers are preceded the letter "i". *The sequence was deduced from the publication of Ogasawara et al.[11]

TABLE 3. ***O*** **allele frequencies in different human populations***

Population sample	Identification of alleles	Akans (n = 137)	Berbers (n = 78)	Basques (n = 220)	Putien (n = 94)	Fujiou (n = 86)	Cayapas (n = 74)	Aymaras (n = 126)
1	*O01*	26	22	126	55	44	37	34
2	*O02*	26	40	65	39	42	31	75
3	*O03*	0	4	2	0	0	0	0
5	*O05*	3	1	4	0	0	1	0
7	$O^{v7.1}$	16	6	1	0	0	1	0
8	$O^{v7.2}$	10	1	0	0	0	0	0
9	*O09*	26	1	1	0	0	0	0
11	*O11*	0	0	0	0	0	3	15
12	*O12*	0	2	14	0	0	0	0
13	*O1v-B.tlse13*	5	0	1	0	0	0	0
14	*O1v-B.tlse14*	1	0	0	0	0	0	0
20	*Ovar.tlse20*	6	0	0	0	0	0	0
21	*O21*	1†	0	1*	0	0	0	0
26	*O26*	0	0	3	0	0	0	0
27	*O27*	1	0	0	0	0	0	0
28	*O28*	10	0	0	0	0	0	0
29	*O29*	2	0	0	0	0	0	0
31	*O31*	1	0	0	0	0	0	0
32	*O32*	0	0	0	0	0	0	1
33	*O33*	0	0	0	0	0	0	1
34	*O34*	0	0	2	0	0	0	0
51	*Ovar.tlse51*	0	1	0	0	0	0	0
52	*Ovar.tlse52*	2	0	0	0	0	0	0
53	$O^{1v(A681G\&C1054T)}$	1†	0	0	0	0	0	0

* All individuals were studied for exon 6 and exon 7 polymorphism. Individuals bearing representative alleles were studied by cloning and sequencing of both alleles they have. When DNA samples were still available, other individuals were studied by direct sequencing of intron 6.

† No intron 6 sequence was available for these alleles.

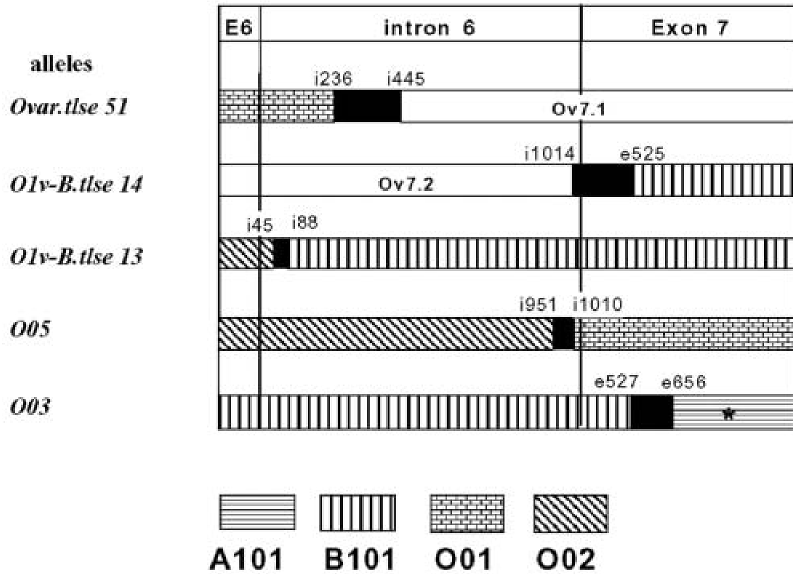

Fig. 3. Schematic representations of five human recombinant alleles. The most probable parental alleles are shown by graphic codes. A recombinant allele described in the text (*Ovar.tlse20*) is not included in the figure because one of the two parental alleles was not sampled in the individuals studied. The possible recombination points flank the black boxes, which correspond to regions where the two parental alleles are identical. *Inactivating mutation.

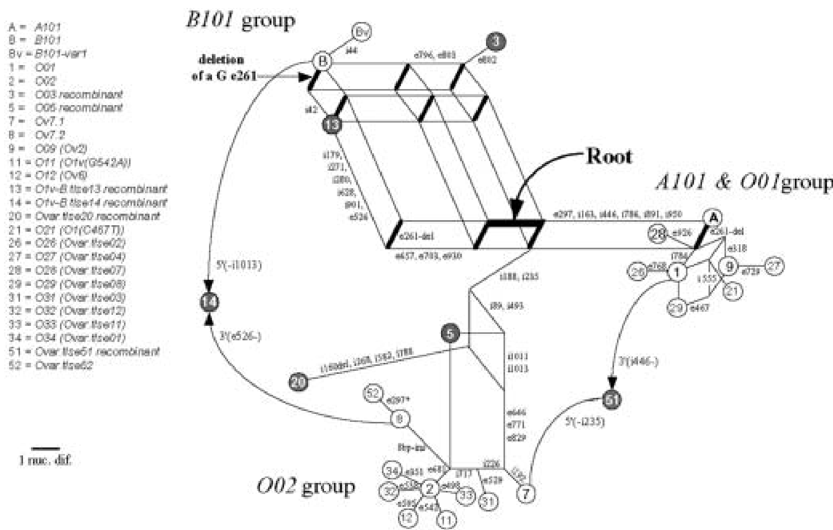

Fig. 4. Phylogenetic network of 26 human *ABO* alleles. 1 nuc. dif. = one nucleotide difference. Numbers in filled circles correspond to recombinant alleles. 3′(e526-) = 3′ end of the allele after position 526 in exon 7; 3′(i446-) = 3′ end of the allele after position 446 in intron 6; 5′(–i1013) = 5′ end of the allele before position 1013 in intron 6; and 5′(–i235) = 5′ end of the allele before position 235 in intron 6. E297* means that mutation at this position for allele *Ovar.tlse51* cannot be represented correctly in the network because the reverse mutation is observed at the same position between the root and *A101* group of alleles.

allele *O03* differs most from a simple recombinant by a single substitution at e802, suggesting a more ancient occurrence of recombination than for other recombinant alleles. In fact, data previously reported from upstream of exon 6 demonstrated that *O03* must be considered as a mosaic of parts of four alleles (*A101*, *B101*, *O01,* and *O02*).[14] Discarding recombinant alleles from the analysis, three major allelic lineages (*A101-O01*, *B101*, and *O02*) are observed in the network. Substantial differences between *O01* and *O02* alleles were already noted from comparison of cDNA sequences, which is confirmed with the information of intron 6.[2]

The proportion of nucleotide differences among these three major allelic lineages goes from 0.009 to 0.012 in the 2-kb region analyzed. The genomic average difference between human and chimpanzee was recently estimated to be 0.0123.[31] If we assume divergence between human and chimpanzee to be 6 million years, the average evolutionary rate becomes 1.0×10^{-9} per site per year for the whole genome. Under the rough assumption of constancy of the evolutionary rate, the diversification of the three major human *ABO* alleles can be estimated at 4.5 to 6.0 million years ago. This unusually large divergence time estimate is consistent with that obtained through cDNA sequence data.[26]

DISCUSSION

Until now, in all populations studied so far, alleles *O01* and *O02* are present although in variable proportions. Apart from these two main *O* alleles, many new alleles have been described, some of them first reported here thanks to the information provided by intron 6. Among the infrequent alleles, even if some have been reported in a single population, they cannot be considered population-specific. For example, allele *O26 (Ovar.tlse02)* was originally found only in Basques but was later reported in a German population.[14] There is a lack of geographic structure owing to their low frequency (drift therefore playing a

major role in shaping the frequencies), the small sample size for most of the studies, and the very deep roots of the O branch in the ABO gene genealogy, much deeper than the genealogy of the present human populations.[32]

It is interesting to note that, as in Iwasaki et al.,[33] among the two Han Chinese populations studied here, we found only two *O* alleles: *O01* and *O02*. Using simple binomial sampling distributions, it can be shown that the maximum frequency of an unobserved allele with a given confidence level α (such as 0.05) in a sample of size N is $p = 1 - \alpha^{1/N}$. Then, the maximum frequency at which alleles other than *O01* or *O02* can exist is 0.062 for the Putien population (n = 47) and 0.067 for Fujiou (n = 43). If pooled, the maximum frequency drops to 0.033. This is highly surprising because even in Amerindians (Cayapas from Ecuador and Aymaras from Bolivia), who have lost *A* and *B* alleles, the number of *O* alleles is at least 5 or 6, and *O* alleles other than *O01* and *O02* add up to 0.11 and 0.135, respectively.[12] These results can be explained by a high genetic homogeneity of some of the Han populations owing to a founding effect, and the difference with the Americas reflects the complex pattern of settlement of this continent that does not justify considering any extant Asian population as being very similar to the Amerindians as if they were the "source" for American settlement. Some alleles found in the Amerindians, like *O11* $(O^{1v(G542A)})$, have not been found in Asians, suggesting a recent and American origin of the allele (after 35,000 before present; see Cavalli-Sforza and Feldman[32]).

In this study, we extended the characterization of *O* alleles by sequencing intron 6 of all alleles encountered in eight human populations. To compare the variability of functional alleles we also cloned and sequenced some individuals of A, B, or AB phenotypes. The description of the intron 6 sequence has allowed some previous alleles to be split into suballeles, but more interestingly, has given a deeper insight into the origin and genetic relationship of alleles. Two types of mechanism are involved in the diversification of *O* alleles: interallelic exchanges and point mutations. From a phylogenetic perspective, if one allele sequence is strictly equivalent to the recombination of the sequences of two alleles found nowadays in human populations, this is compatible with a recent recombination event. Many examples of such recent recombinant alleles have been reported[8,11,26,28-30,34,35] and four new cases are characterized in this study (*O1v-B.tlse14*, *O1v-B.tlse13*, *Ovar.tlse20*, and *Ovar.tlse51*). In contrast, if the potential recombinant allele exhibits point mutations when compared to the theoretical parental alleles, it is compatible with either an ancient recombination event or, less likely, with a recent recombination as yet unsampled alleles, but with recombination involved in any case. This corresponds to *Ovar.tlse20*.

It should be noted that the high divergence among *ABO* alleles facilitates the detection of recombinant alleles (such as recombinants between *A101* or *O01* and *O02* or between or between *B101* and *A101* or *O01*) and, furthermore, allows the region where the putative crossover event took place to be narrowed down to tens of bases. Obviously, recombination between highly similar alleles in the analyzed region (such as *O01* and *A101*) cannot be detected. In the recombinants described here the parental alleles have been detected (Fig. 3), showing that the recombination events do not cluster in particular spots, even though a role in recombination for Ki and Ki-like sequences in intron 6 had been suggested.[8,24] Recombination as a generator of variability in the ABO should not be exceptionally high, because the local recombination rate is slightly under twice the genomewide average according to the recent genetic map[36] where recombination rates were estimated along the genome by comparing genetic to physical distances for a large number of microsatellite markers. *ABO* lies 200 kb upstream from one such marker, namely D9S754, around which the estimated recombination rate is 1.07 cM per million bases; for the 1875-bp stretch we sequenced, that would translate to a recombination rate of 2.01×10^{-5} per meiosis. It should also be noted that the recombinant alleles could result either from nonconservative interallelic exchange (crossing over) or from conservative interallelic exchanges through gene conversion.

If one excludes *O03*, all human nonfunctional *O* alleles observed in the eight populations we studied share the single-nucleotide deletion at position 261 of exon 6. It is highly improbable that such a deletion occurred at exactly the same position more than once during human evolution especially because it is not located in a high mutation rate context, like a stretch of poly(G). As a matter of fact, point deletions are rarely observed among human alleles: among all human *ABO* alleles reported so far, only one, *Ovar.tlse20*, has a point deletion in intron 6, seven alleles (*A201*,[37] *A206* (A^{v1}),[10] *A302*,[38] *Aw01*, *Aw03*, *Aw05*,[6] and *Aw07*[14]) share one point deletion at nucleotide 1060 in exon 7, and allele *Ael03*[13] has a point deletion at nucleotide 804 in exon 7. If there is a common origin for deletion Δ261, then it remains to be explained how this mutation is observed in such divergent alleles as *O01* and *O02* (see Fig. 4). Allele *O02* is clearly the more ancient human allele harboring the single nucleotide deletion, because it is very different from the extant functional alleles *A101* and *B101* and it must have been an ancient and independent lineage. Since then, allele *O02* would have accumulated many mutations without functional impact because it is a null allele. In that context, all other alleles sharing this deletion with *O02* must be more recent and could have been transferred from the *O02* to the *O01* lineage either by recombination between *O02* and *A101* or by conversion between the two previous alleles. Nevertheless, upstream of exon 6, the sequences of *O01* and *O02* are quite divergent,[14] suggesting that gene conversion is the most likely

mechanism for the presence of Δ261 in *O01*. The transfer of the deletion to the *A101* allele created an intermediate allele that has not been found in any population studied so far. Then, a point mutation in intron 6 (i784) appeared in this intermediate recombinant allele to create *O01*.

It remains to be explained, however, how the new recombinant *O01* allele frequency increased, leading this allele to coexist with allele *O02* in a wide range of human populations.[10-12,29] Because *O01* and *O02* are nonfunctional and therefore have identical selective values, only chance can explain the persistence of the two alleles in all human populations and both had to be present in the main human dispersal events, including the one out of Africa. In spite of the fact that balancing selection is often invoked to explain the maintenance of human ABO polymorphism, the persistence of the two main *O* silent alleles in humans must be explained by genetic drift. Because the number of differences between *O01* and *O02* is comparable to the number of differences between *A101* and *B101*, we may speculate that the persistence of A/B polymorphism is also due to chance, because the number of substitutions in the three main lineages (*A*, *B*, and *O*) is similar. In this context it should be recalled that chimpanzees lack the *B* allele, whereas gorillas lack both *A* and *O* alleles.[39] Moreover, the two *O* alleles characterized in chimpanzees[40] are totally different from those observed in humans, indicating that *O* silent alleles appeared independently during the evolution of humans and chimpanzees.

It is thus clear that, even if lineage *O01* is ancient, it cannot be considered as a founding lineage in humans, contrary to the description of Seltsam et al.[14] This is also the case of allele *O03*, which is not among the deepest branches of the ABO gene genealogy. As first suggested by Irshaid and coworkers[27] and Seltsam and colleagues[14] it is a complex recombinant, making *O03* necessarily more recent than the other alleles. Moreover, it has accumulated less variation, has generated fewer alleles, and has a more restricted geographic distribution than *A101*, *B101*, *O01*, or *O02*. In fact, the position of *O03* in a neighbor-joining tree or in other representations (see Fig. 4) should not be read as the result of a long process of mutation accumulation, but rather as an artifact caused by recombination.

The evolution of the ABO system, beyond the dynamics involving mutation and recombination from the three main branches, shows extremely long roots in the human lineage (5-6 millions years), around the same age as time of separation with chimpanzee lineage.

ACKNOWLEDGMENTS

The authors are very grateful to I. Desmoulins (EFS Aquitaine Limousin, Site de Biarritz, France), A. Chaventré (University Bordeaux II, Bordeaux, France), A. Pitte (Abidjan, Ivory Coast), R. Calderon (Universidad del Pais Vasco, Bilbao, Spain), G.F. De Stefano (University Tor Vergata, Rome, Italy), A. Baali and M. Cherkaoui (University Cadi Ayyad, Marrakech, Morocco), and M. Villena (Instituto Boliviano de Biologia de Altitudes, La Paz, Bolivia) for providing human blood or DNA samples. The authors are also indebted to W.W. Socha and G. Dubreuil for providing primate samples. We thank Marianne Dutaur and Pierre Tisseyre for their technical assistance.

REFERENCES

1. Landsteiner K. Uber agglutinationserscheinungen normalen menschlichen blutes. Wien Klin Wochenschr 1901;14:1132-4.
2. Yamamoto F, Clausen H, White T, Hakomori S. Molecular genetic basis of the histo-blood group ABO system. Nature 1990;345:229-33.
3. Yamamoto F, Marken J, Tsuji T, et al. Cloning and characterization of DNA complementary to human UDP-GalNAc:Fucα1-2Galα1,3-GalNAc transferase (histo-blood group A transferase) mRNA. J Biol Chem 1990;265:1146-51.
4. Yamamoto F, McNeill PD, Hakomori S. Genomic organization of human histo-blood group *ABO* genes. Glycobiology 1995;5:51-8.
5. Yamamoto F. Molecular genetics of ABO. Vox Sang 2000;78(Suppl 2):91-103.
6. Olsson ML, Irshaid NM, Hosseini-Maaf B, et al. Genomic analysis of clinical samples with serologic ABO blood grouping discrepancies: identification of 15 novel A and B subgroup alleles. Blood 2001;98:1585-93.
7. Chester MA, Olsson ML. The ABO blood group gene: a locus of considerable gene diversity. Transfus Med Rev 2001;15: 177-200.
8. Yip SP. Sequence variation at the human ABO locus. Ann Hum Genet 2002;66:1-27.
9. Olsson ML, Chester MA. Frequent occurrence of a variant *O1* gene at the blood group *ABO* locus. Vox Sang 1996;70:26-30.
10. Yip SP. Single-tube multiplex PCR-SSCP analysis distinguishes 7 common ABO alleles and readily identifies new alleles. Blood 2000;95:1487-92.
11. Ogasawara K, Yabe R, Uchikawa M, et al. Recombination and gene conversion-like events may contribute to ABO gene diversity causing various phenotypes. Immunogenetics 2001;53:190-9.
12. Roubinet F, Kermarrec N, Despiau S, et al. Molecular polymorphism of O alleles in five populations of different ethnic origins. Immunogenetics 2001;53:95-104.
13. Seltsam A, Hallensleben M, Kollmann A, et al. Systematic analysis of the ABO gene diversity within exons 6 and 7 by PCR screening reveals new ABO alleles. Transfusion 2003;43:428-39.
14. Seltsam A, Hallensleben M, Kollmann A, Blasczyk R. The nature of diversity and diversification at the *ABO* locus. Blood 2003;102:3035-42.
15. Yamamoto F, McNeill PD, Yamamoto M, et al. Molecular

genetic analysis of the ABO blood group system. 4. Another type of O allele. Vox Sang 1993;64:175-8.
16. Grunnet N, Steffensen R, Bennett EP, Clausen H. Evaluation of histo-blood group ABO genotyping in a Danish population: frequency of a novel *O* allele defined as *O2*. Vox Sang 1994;67:210-5.
17. Franco RF, Simoes BP, Zago MA. Relative frequencies of the two O alleles at the histo-blood ABH system in different racial groups. Vox Sang 1995;69:50-2.
18. Ogasawara K, Bannai M, Saitou N, et al. Extensive polymorphism of ABO blood group system: Three major lineages of the alleles for the common ABO phenotypes. Hum Genet 1996;97:777-83.
19. Kang SH, Fukumori Y, Ohnoki S, et al. Distribution of ABO genotypes and allele frequencies in a Korean population. Jpn J Hum Genet 1997;42:331-5.
20. Fukumori Y, Ohnoki S, Shibata H, Nishimukai H. Suballeles of the ABO blood group system in a Japanese population. Hum Hered 1996;46:85-91.
21. Olsson ML, Santos SE, Guerreiro JF, et al. Heterogeneity of the *O* alleles at the blood group *ABO* locus in Amerindians. Vox Sang 1998;74:46-50.
22. Yamamoto F, McNeill PD. Amino acid residue at codon 268 determines both activity and nucleotide-sugar donor substrate specificity of human histo-blood group A and B transferase. J Biol Chem 1996;271:10515-20.
23. Olsson ML, Chester MA. Evidence of a new type of *O* allele at the ABO locus, due to a combination of the *A2* nucleotide deletion and the *Ael* nucleotide insertion. Vox Sang 1996;71:113-7.
24. Olsson ML, Chester MA. Polymorphism and recombination events at the ABO locus: a major challenge for genomic ABO blood grouping strategies. Transfus Med 2001;11:295-313.
25. Saitou N, Yamamoto F. Evolution of primate ABO blood group genes and their homologous genes. Mol Biol Evol 1997;14:399-411.
26. Kitano T, Noda R, Sumiyama K, et al. Gene diversity of chimpanzee ABO blood group genes elucidated from intron 6 sequences. J Hered 2000;91:211-4.
27. Irshaid NM, Chester MA, Olsson ML. Allele related variation in minisatellite repeats involved in the transcription of the blood group ABO gene. Transfus Med 1999;9:219-26.
28. Ogasawara K, Yabe R, Uchikawa M, et al. Molecular genetic analysis of variant phenotypes of the ABO blood group system. Blood 1996;88:2732-7.
29. Bugert P, Rütten L, Goerg S, Klüter H. Characterization of a novel O^1 variant allele at the ABO blood group locus. Tissue Antigens 2001;58:422-4.
30. Olsson ML, Guerreiro JF, Zago MA, Chester MA. Molecular analysis of the *O* alleles at the blood group *ABO* locus in populations of different ethnic origins reveals novel crossing-over events and point mutations. Biochem Biophys Res Commun 1997;234:779-82.
31. Fujiyama A, Watanabe H, Toyoda A, et al. Construction and analysis of a human-chimpanzee comparative clone map. Science 2002;295:131-4.
32. Cavalli-Sforza LL, Feldman MW. The application of molecular genetic approaches to the study of human evolution. Nat Genet 2003;33:266-75.
33. Iwasaki M, Kobayashi K, Suzuki H, et al. Polymorphism of the ABO blood group genes in Han, Kazak and Uygur populations in the silk route of northwestern China. Tissue Antigens 2000;56:136-42.
34. Suzuki K, Iwata M, Tsuji H, et al. A de novo recombination in the ABO blood group gene and evidence for the occurrence of recombination products. Hum Genet 1997;99:454-61.
35. Olsson ML, Chester MA. Heterogeneity of the blood group *Ax* allele: genetic recombination of common alleles can result in the A_x phenotype. Transfus Med 1998;8:231-8.
36. Kong A, Gudbjartsson DF, Sainz J, et al. A high-resolution recombination map of the human genome. Nat Genet 2002;31:241-7.
37. Yamamoto F, McNeill PD, Hakomori S. Human histo-blood group A2 transferase coded by A2 allele, one of the A subtypes, is characterized by a single base deletion in the coding sequence, which results in an additional domain at the carboxyl terminal. Biochem Biophys Res Commun 1992;187:366-74.
38. Barjas-Castro ML, Carvalho MH, Locatelli MF, et al. Molecular heterogeneity of the A_3 subgroup. Clin Lab Haematol 2000;22:73-8.
39. Blancher A, Socha WW. The ABO, Hh and Lewis blood group in humans and nonhuman primates. In: Blancher A, Klein J, Socha WW, editors. Molecular biology of evolution of blood group and MHC in primates. Berlin/Heidelberg/New York: Springer-Verlag; 1997
40. Kermarrec N, Roubinet F, Apoil PA, Blancher A. Comparison of alleles *O* sequences of the human and non human primate ABO system. Immunogenetics 1999;49:517-26.