

# The nature of diversity and diversification at the *ABO* locus

Axel Seltsam, Michael Hallensleben, Anke Kollmann, and Rainer Blasczyk

In this study we analyzed the complete genomic sequences, except intron 1, and 2 regulatory regions of 6 common (*ABO*\*A101, *ABO*\*A201, *ABO*\*B101, *ABO*\*O01, *ABO*\*O02, and *ABO*\*O03) and 18 rare *ABO* alleles, 3 of which were new. This was done by phylogenetic analysis and correlating sequence data with the *ABO* phenotypes. The study revealed multiple polymorphisms in noncoding regions. The intron-based phylogenetic analysis revealed 5 main lineages: *ABO*\*A, *ABO*\*B, *ABO*\*O01, *ABO*\*O02, and

*ABO*\*O03. The genomic sequences of most rare *ABO* alleles differed slightly from those of the common alleles. Singular mutations or hybrid alleles were most common, but a few exhibited mosaic sequence pattern containing multiple exon and/or intron motifs from other *ABO* lineages. Thus, both an accumulation of mutations as well as an assortment of the mutations by recombination seems to be responsible for the *ABO* gene diversity. The prevalence of replacement mutations indicates positive selection for allelic di-

versity. Phenotype-genotype correlation showed that sequence variations within the complete coding sequence can affect A- and B-antigen expression. All variant *ABO*\*A/B alleles and one new *ABO*\*O03-like allele were associated with weak *ABO* phenotypes. These findings are suggestive of the requirement of a comprehensive coding sequence database for sequence-based phenotype prediction. (Blood. 2003;102:3035-3042)

© 2003 by The American Society of Hematology

## Introduction

The *ABO* blood group is the most important blood group system in transfusion and transplantation medicine. Its antigenic determinants are oligosaccharides located on glycoproteins and glycolipids expressed on erythrocytes and tissue cells and occur in various body fluids and secretions. Depending on an individual's *ABO* blood type, immunoglobulin M (IgM) antibodies directed against the missing A and/or B antigens are regularly present in serum; they constitute an immunologic barrier against incompatible blood transfusion and organ transplantation. The *ABO* gene codes for the glycosyltransferases that transfer specific sugar residues to H substance, resulting in the formation of blood group A and B antigens. This gene maps to chromosome 9, position 9q34.1-q34.2. It consists of 7 exons, ranging in size from 28 to 688 base pairs (bp), and 6 introns with 554 to 12 982 bp (Figure 1).<sup>1-3</sup> The last 2 exons (6 and 7), which comprise 823 of 1062 bp of the transcribed mRNA, encode for the catalytic domain of *ABO* glycosyltransferases.

The 6 common *ABO* alleles in white individuals are *ABO*\*A101 (*A*<sup>1</sup>), *ABO*\*A201 (*A*<sup>2</sup>), *ABO*\*B101 (*B*<sup>1</sup>), *ABO*\*O01 (*O*<sup>1</sup>), *ABO*\*O02 (*O*<sup>1v</sup>), and *ABO*\*O03 (*O*<sup>2</sup>). In exons 6 and 7 they differ by only a few base positions. *ABO*\*A201, which is responsible for blood group A<sub>2</sub>, is identical to *ABO*\*A101 apart from a nonsynonymous substitution at nucleotide (nt) position 467 and a single deletion (1060delC) in exon 7. This deletion results in disruption of the stop codon and an A-transferase product with an extra 21 amino acid (AA) residue at the C-terminus. *ABO*\*B101 is distinguishable from *ABO*\*A101 at 7 nt positions: 3 synonymous mutations at positions 297, 657, and 930; and 4 nonsynonymous mutations at positions 526, 703, 796, and 803. The nt sequence of *ABO*\*O01 differs from that of *ABO*\*A101 by a single base deletion at position 261 in exon

6; this deletion shifts the reading frame, thus generating a premature stop codon. *ABO*\*O01 is thought to be either silent or translated into a truncated and catalytically inactive peptide. In contrast, the *ABO*\*O03 allele lacks the 216delG polymorphism but possesses nonsynonymous mutations that may abolish the protein's enzyme activity by altering the nt sugar binding site.

Eighty-three *ABO* alleles discriminated at 52 polymorphic sites within the coding region of the *ABO* gene have been reported in the literature so far.<sup>4-10</sup> In most cases the investigators analyzed only exons 6 and 7. The number of described *ABO* alleles increases to 88 when nt differences within intron 6 are also considered. It has been shown that studies of the nt sequence of intron 6 are crucial for elucidation of the origin of some novel haplotypes.<sup>11-13</sup> To our knowledge, there is no information available on sequence variation of the noncoding regions upstream from exon 6 and little data on mutations within the first 5 exons of the *ABO* gene and their relevance for the *ABO* phenotypes.<sup>14</sup> In the present study, we therefore examined the complete exon/intron sequences (except for the huge intron 1 comprising 12 982 bp) and 2 regulatory regions of common and rare *ABO* alleles to evaluate the genetic diversity and diversification at the *ABO* locus. The genomic sequence data were first correlated with the associated *ABO* phenotypes then used for lineage definition.

## Patients, materials, and methods

### Collection of peripheral blood samples

A total of 55 peripheral blood samples were analyzed. Twenty-five blood samples were obtained from unrelated healthy blood donors with

From the Department of Transfusion Medicine, Hannover Medical School, Hannover, Germany.

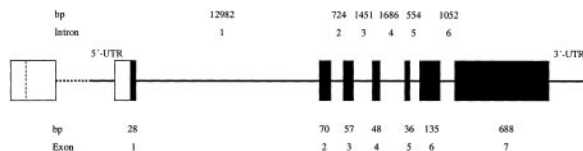
Submitted March 27, 2003; accepted June 18, 2003. Prepublished online as *Blood* First Edition Paper, June 26, 2003; DOI 10.1182/blood-2003-03-0955.

**Reprints:** Rainer Blasczyk, Department of Transfusion Medicine, Hannover

Medical School, Carl-Neuberg-Str 1, D-30625 Hannover, Germany; e-mail: blasczyk.rainer@mh-hannover.de.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

© 2003 by The American Society of Hematology



**Figure 1. Schematic representation of the genomic organization of the *ABO* gene.** The exons (■) and regulatory regions (□) are drawn to scale, as are the intervening introns, although the scale of the latter is 10 times smaller. The calculated numbers of nucleotides (nts) in the exons and introns are shown. The upstream regulatory region, which includes a CBF/NF- $\kappa$ B binding motif, is located around nt  $-3800$  and, depending on the *ABO* haplotype, 215 or 344 base pairs (bp) in size; the regulatory region in the 5' untranslated region (UTR) is located from nt  $-118$  to  $-1$ .

common ABO phenotypes, and 30 were obtained from unrelated patients ( $n = 19$ ) or healthy blood donors ( $n = 11$ ) with variant ABO blood groups. Blood was obtained with informed consent as approved by the local ethics committee of the Hannover Medical School. Samples from 18 of the latter individuals were already included in previous *ABO* genotyping studies.<sup>6,10</sup> The ethnic origin of the patients/donors was predominantly German ( $n = 49$ ). The remaining individuals originated from Africa ( $n = 1$ ), Bosnia ( $n = 1$ ), India ( $n = 1$ ), Italy ( $n = 2$ ), and Turkey ( $n = 1$ ). Data on the patients' diagnoses were not collected.

### Blood group serology

*ABO* typing of red blood cells (RBCs) was carried out as described previously, and the subgroups were classified according to current recommendations.<sup>6,10,15</sup> The following commercial antisera and lectins were used: anti-A (immuClone, monoclonal; Immucor, Roedermark, Germany; Seraclone; Biotest, Dreieich, Germany; BioClone, monoclonal IgM, murine; Ortho-Clinical Diagnostics, Raritan, NJ), anti-B (immuClone, monoclonal; Immucor; BioClone, monoclonal IgM, murine; Ortho-Clinical Diagnostics), anti-AB (Seraclone Anti-AB; Biotest),  $A_{\text{hel}}$  from *Helix pomatia* (Immucor), anti-A1 from *Dolichus biflorus* (Lectin; Mast Diagnostics, Reinfeld, Germany), and monoclonal anti-H (Mast Diagnostics). All reagents were used according to the manufacturers' instructions. In cases where the subgroup status could not be determined due to the lack of saliva samples, the variant ABO phenotypes were assigned the more general terms  $A_{\text{weak}}$  and  $B_{\text{weak}}$ .

### *ABO* genotyping of exons 6 and 7

The *ABO* genotype of all blood samples was first determined by screening exons 6 and 7 using a polymerase chain reaction method with sequence-specific primers (PCR-SSP).<sup>10</sup> Twenty-five healthy blood donors with common ABO blood groups and genotypes *ABO*\*A101/*ABO*\*A101 ( $n = 3$ ), *ABO*\*A201/*ABO*\*A201 ( $n = 2$ ), *ABO*\*A101/*ABO*\*A201 ( $n = 2$ ), *ABO*\*B101/*ABO*\*B101 ( $n = 2$ ), *ABO*\*A101/*ABO*\*B101 ( $n = 2$ ), *ABO*\*O01/*ABO*\*O01 ( $n = 4$ ), *ABO*\*O02/*ABO*\*O02 ( $n = 3$ ), *ABO*\*O01/*ABO*\*O02 ( $n = 4$ ), and *ABO*\*O02/*ABO*\*O03 ( $n = 3$ ) were thereby selected.

### PCR amplification and sequencing

**Strategy.** Our aim was to analyze all genomic sequences (except intron 1) and 2 regulatory regions (promoter, from nt  $-118$  to  $-1$ ; enhancer CBF/NF- $\kappa$ B, a 215- or 344-bp region around nt  $-3800$ ) of the *ABO* gene of each individual. First, blood samples homozygous for *ABO*\*A101, *ABO*\*B101, *ABO*\*O01, or *ABO*\*O02 (samples homozygous for the less frequent *ABO*\*O03 allele were not available) were investigated using a generic sequence strategy (templates for sequence analysis were generated by amplifying both *ABO* alleles simultaneously) to establish an intermediate database. Then, samples heterozygous for common or rare *ABO* alleles were sequenced, first by generic sequencing then by haplotype-specific sequencing (templates for sequence analysis were prepared using allele-specific primers) of at least 1 of the 2 alleles to define the *cis/trans* linkage of the polymorphic sites. The design of the allele-specific amplification primers was based on the growing intermediate sequence database. Thus, each allele was sequenced from at least 2 independent PCR products. Most

sequences were obtained by forward and reverse sequencing. All amplification and sequencing primers as well as primer pairs used for amplification are listed in Tables 1-3.

**PCR amplification.** The PCR reaction mixture (final volume: 20  $\mu$ L) contained 200 ng genomic DNA, PCR buffer ( $5 \times$  buffer: 300 mM Tris-HCl, 75 mM ammonium sulfate, 12.5 mM MgCl<sub>2</sub>, pH 9.0), 200  $\mu$ M of each dNTP (deoxynucleoside triphosphate), 5.0 pmol of each amplification primer, and 0.3 units Taq DNA polymerase (Platinum Taq; GIBCO BRL, Karlsruhe, Germany). PCR amplification was carried out in a GenAmp PCR system 9600 (Perkin Elmer, Norwalk, CT). After an initial denaturation step at 94°C for 2 minutes, the samples were subjected to 30 cycles of amplification, consisting of ten 2-temperature cycles (94°C for 10 s and 65°C for 60 s) followed by twenty 3-temperature cycles (94°C for 10 s, 61°C for 50 s, and 72°C for 30 s). The amplification products were separated electrophoretically and visualized on a 2.5% agarose gel prestained with ethidium bromide (0.2  $\mu$ g/mL). If 2 fragments occurred in the enhancer CBF/NF- $\kappa$ B region, the amplification products were purified from agarose gel for direct sequencing using the QIAquick gel extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

**Sequencing of PCR products.** Sequencing was carried out by dye terminator cycle sequencing chemistry (DYEnamic ET terminators, Amersham Pharmacia Biotech, Freiburg, Germany) as described previously.<sup>5,9</sup> Briefly, the sequencing reactions were performed in a 20- $\mu$ L sample volume using 4  $\mu$ L dye terminator mix containing fluorescence-labeled dideoxynucleoside triphosphate (ddNTPs), dNTPs, Thermo Sequenase II DNA polymerase, and 3 pmol biotinylated sequencing primer. This mixture was subjected to twenty-five 3-temperature cycles (96°C for 10 s, 50°C for 5 s, and 60°C for 4 min). For purification, the biotinylated products were attached to streptavidin-coated paramagnetic beads (Dyna, Oslo, Norway) according to the manufacturer's instructions. Samples were washed once with 60  $\mu$ L of 70% ethanol and air dried. The single-stranded sequencing fragments attached to the beads were then resuspended in 4  $\mu$ L loading dye (Amersham Pharmacia Biotech). The mixture was heated at 90°C for 2 minutes, and 1.5  $\mu$ L was loaded onto a 0.2-mm-thick layer of 5% polyacrylamide-7M urea gel. Electrophoresis was run at constant 2700 V for 4 hours on an ABI 377 sequencer (PE Biosystem, Foster City, CA).

### Phylogenetic analysis

The intron sequences of the 110 *ABO* alleles examined in this study were phylogenetically analyzed. The sequences were aligned using CLUSTAL W software (European Molecular Biology Laboratory, Heidelberg, Germany) run on the graphic interface program BioEdit version 5.0.9 (North Carolina State University, NC).<sup>16,17</sup> Phylogenetic trees showing the phylogenetic relationship between the alleles were constructed according to the neighbor-joining method using Treecon version 1.3b (Ghent University, Ghent, Belgium).<sup>18,19</sup> The trees were unrooted to get a topology representing the relative genetic distances between the *ABO* alleles.

### Classification and nomenclature of *ABO* alleles

The *ABO* alleles were named according to the unofficial nomenclature used in the Blood Group Antigen Gene Mutation Database (<http://www.bioc.aecom.yu.edu/bgmut/abo.htm>).

## Results

A total of 24 different *ABO* alleles were analyzed in this study (Tables 4-5). The number of alleles for each *ABO* allele specificity detected ranged from 1 to 21. Six common alleles (*ABO*\*A101, *ABO*\*A201, *ABO*\*B101, *ABO*\*O01, *ABO*\*O02, *ABO*\*O03) and 18 rare *ABO* alleles, including 3 new alleles (*ABO*\*Aw07, *ABO*\*Aw08, *ABO*\*Bw08), were studied. Most of the *ABO* alleles analyzed originated from Germans. The rare alleles *ABO*\*Aw06, *ABO*\*Ae103, *ABO*\*Bw07, and *ABO*\*O06 and the common alleles *ABO*\*A201,

**Table 1. Amplification primers**

Name	Sequence (5'-3')	Length, bp	Location
ABO003s	gatctggactgggtttggag	20	Intron 1, nt 12797 to 12816
ABO007as	atgcagcagaagtgatgctg	20	Intron 3, nt 1059 to 1078
ABO012as	ggtcaaaactgaagctccag	20	Intron 4, nt 194 to 213
ABO017as	tgggcctctgaattcagatg	20	Intron 5, nt 129 to 148
ABO020s*	gcacgcctctctccatgtg	19	Intron 5, nt 532 to 551
ABO021s*	GGAAGGATGTCCTCGTGGTA	20	Exon 6, nt 242 to 261
ABO023s*	GGAAGGATGTCCTCGAGGTG	20	Exon 6, nt 242 to 261
ABO024as	ACAATGGGAGCCAGCCAAG	19	Exon 6, nt 266 to 284
ABO026s	gctgaatcagagactctgag	20	Intron 4, nt 1473 to 1492
ABO028s*	CATTGTCTGGGAGGCACA	19	Exon 6, nt 279 to 297
ABO080as*	AGAAATCGCCCTCGTCTTA	20	Exon 7, nt 771 to 790
ABO090as*	CCGACCCCCGAAGAACC	18	Exon 7, nt 803 to 820
ABO118as	cctaggcttcagttactcac	20	3'-UTR, nt 103 to 122
ABO127as	GACTTCTGGGGCTTAGGAC	19	Exon 3, nt 106 to 124
ABO128as	AGACTTCTGGGGCTTAGGAA	20	Exon 3, nt 106 to 125
ABO129s	gagctttgctgagaacttgc	20	Intron 2, nt 564 to 583
ABO130s	AACCTGACCATCTGCAGCG	19	Exon 4, nt 170 to 188
ABO131s	GAACCTGACCATCTGGAGCA	20	Exon 4, nt 169 to 188
ABO137s	gcacttgtgccctaaatcct	20	Intron 3, nt 1390 to 1409
ABO164s	tcctttctgcagTTACGGGT	20	Intron 2, nt 713 to exon 3 nt 106
ABO184s	cttagtcctcttgccagttta	22	Intron 4, nt 1446 to 1467
ABO205s	accctctgcccaactcca	19	Intron 2, nt 521 to 539
ABO212s	cagggtcctaaggacagttaa	21	Intron 2, nt 107 to 127
ABO214as	ggtagaccatcctgctagcg	20	Intron 4, nt 1678 to 1697
ABO217s*	ggcgccgtcccttctctag	18	5'-UTR, nt -188 to -171
ABO269s	tcacgtggcgcgctttgc	19	Intron 4, nt 1660 to 1678
ABO270as	GGGAGCCAGCCAAGGGGTA	19	Exon 6, nt 261 to 279
ABO274as*	cctgcggttagcgctccct	19	Intron 1, nt 33 to 51
ABO296s*	ggaacaacaatcctaccctac	21	5'-UTR, nt -3988 to -3968
ABO297as*	gtgctgcctgtgcctgttac	20	5'-UTR, nt -3664 to -3645

Primer sequences corresponding to noncoding sequences are shown in lowercase. Bold letters indicate artificial mismatches to increase PCR specificity. Nucleotide (nt) positions refer to the mRNA sequence for primers located in the coding region. Primers in the noncoding regions are numbered starting from the beginning of each intron or untranslated region (UTR). The locations of the ABO296s and ABO297as primers used to amplify the Enhancer (CBF/NF-Y) are indicated relative to the transcription start site. s indicates sense primer; as, antisense primer; and nt, nucleotide.

\*Previously published primers.<sup>6,10,14</sup>

ABO\*B101, ABO\*O01, and ABO\*O02 were identified in individuals from Germany and in persons from Bosnia, India, Italy, and Turkey, respectively. All rare ABO\*A and ABO\*B alleles, including the new alleles, were associated with weak A or B phenotypes. The

number used to denote the nt positions of the exons refers to the mRNA sequence starting from the beginning of the coding region. The number used to denote the nt positions of the introns refers to each intron separately.

**Table 2. Primer combinations used for amplification**

Sense primer	Antisense primer	Amplified region	Amplified ABO alleles (ABO*)
ABO296s	ABO297as	Enhancer (CBF/NF-Y)	All
ABO217s	ABO274as	Promoter, exon 1	All
ABO003s	ABO007as	Exon 2-exon 3	All
ABO003s	ABO127as	Exon 2-intron 2	All except for O02, O06, O12, O40
ABO003s	ABO128as	Exon 2-intron 2	O02, O06, O12, O40
ABO212s	ABO012as	Intron 2-exon 4	O03, Aw08
ABO129s	ABO012as	Exon 3-exon 4	All
ABO205s	ABO012as	Exon 3-exon 4	O01, O011, O26, O40, O03, Aw08
ABO164s	ABO012as	Intron 3-exon 4	O02, O06, O12, O40
ABO137s	ABO017as	Exon 4-exon 5	All
ABO130s	ABO214as	Intron 4	O01, O011, O26
ABO131s	ABO017as	Intron 4-exon 5	O02, O06, O12, O40
ABO184s	ABO017as	Exon 5	O02, O06, O12, O39, O03, Aw08
ABO026s	ABO024as	Exon 5-intron 5	All
ABO026s	ABO270as	Exon 5-intron 5	O01, O011, O26, O21, O02, O06, O12, O39, O40
ABO269s	ABO270as	Exon 5-intron 5	O01, O011, O26
ABO020s	ABO118as	Exon 6-exon 7	All
ABO023s	ABO118as	Exon 6-exon 7	A101, A201, Aw04, Aw06, Aw07, Ael01, Ael03, Ax08, B101, Bw07, Bw08, B(A)03, O03, Aw08, O41
ABO021s	ABO118as	Exon 6-exon 7	O01, O011, O26, O21, O02, O06, O12, O39, O40
ABO028s	ABO118as	Exon 6-exon 7	A101, A201, Aw04, Aw06, Aw07, Ael01, Ael03, Ax08, O41, O01, O011, O26, O21, O06

**Table 3. Sequencing primers\***

Name	Sequence (5'→3')	Length, bp	Location
ABO004s	gatgacctgaattacaaggcg	20	Intron 1, nt 12894 to 12913
ABO005as	atgctccacctgctcttcc	19	Intron 3, nt 11 to 29
ABO008s	acgtgtctggatgaatgtgtg	20	Intron 3, nt 1184 to 1203
ABO013s	cagGATGGTCTACCCCCAG	19	Intron 4, nt 1684 to exon 5 nt 16
ABO024as	ACAATGGGAGCCAGCCAAG	19	Exon 6, nt 266 to 284
ABO026s	gctgaatcagagactctgag	20	Intron 4, nt 1473 to 1492
ABO032s	CAGTTCAGGCTCCAGAACAC	20	Exon 6, nt 322 to 341
ABO035as*	acctcaatgtccacagctcac	20	Intron 6, nt 46 to 65
ABO037s	tcaactgaccagagatagcag	20	Intron 6, nt 332 to 351
ABO039s*	cgtcgcctgccttgacag	18	Intron 6, nt 1035 to 1052
ABO042as	GGTGGCCACCATGAAGTG	19	Exon 7, nt 418 to 436
ABO105as*	tcccagagcccctggcag	18	3'-UTR, nt 5 to 22
ABO138s	CGCACTTCGACCTATGATCC	20	Exon 2, nt 45 to 64
ABO139as	CAGGCTTCCTGGCATTAGAC	20	Exon 3, nt 122 to 141
ABO140as	ctgtgttctcattctgcagc	20	Intron 3, nt 1433 to exon 4, nt 156
ABO141s	AAGCCCAGAAAGTCTAATGC	20	Exon 3, nt 111 to 130
ABO142as	gggaggcactgacattatac	20	Intron 4, nt 1 to 20
ABO143s	ggacgggcctcctgcag	17	Intron 6, nt 983 to 999
ABO145s	ctacactgggttcccaaatc	20	Intron 3, nt 306 to 325
ABO147s	aagtgattctcctgcctcag	20	Intron 4, nt 330 to 349
ABO149as	ccctcccatcaagagatg	19	Intron 3, nt 1110 to 1138
ABO165as	aaccgaccttaataccttc	20	Intron 5, nt 52 to 71
ABO186as	cagctactcaggaggetgc	19	Intron 4, nt 324 to 342
ABO214as	GGTAGACCATCctgcaagcg	20	Intron 4, nt 1678 to exon 5, nt 11
ABO217s*	ggcgcctgccttctcag	18	5'-UTR, nt 1866 to 1883
ABO219s	ccccagctcaccattccatg	20	Intron 4, nt 851 to 870
ABO274as*	cctgcggtagcggctcctc	19	Intron 1, nt 33 to 51
ABO296s*	ggaacaatcctacccttac	21	5'-UTR, nt -3988 to -3968
ABO297as*	gtgctgctgtgctgttac	20	5'-UTR, nt -3664 to -3645

Primer sequences corresponding to noncoding sequences are shown in lowercase. Nucleotide (nt) positions refer to the mRNA sequence for primers located in the coding region. Primers in the noncoding regions are numbered starting from the beginning of each intron or untranslated region (UTR). The locations of the ABO296s and ABO297as primers used to amplify the Enhancer (CBF/NF-Y) are indicated relative to the transcription start site.

s indicates sense primer; as, antisense primer; nt, nucleotide.

\*Previously published primers.<sup>6,10,14</sup>

### Sequence diversity of the coding region of the ABO gene

A total of 30 nt mutations were found in the coding region of the ABO gene. The mutations consisted of 15 transitions, 10 transversions, 3 single base deletions, and 2 single base insertions, most of which were located within exon 7. Twenty-one of the nt mutations

led to alteration of the AA residue of ABO glycosyltransferase (data not shown).

Three new mutations within the nt sequence of exon 7 were identified in this study (Table 4). A new phenotypically relevant nt mutation was detected in a new ABO\*A allele named ABO\*Aw07.

**Table 4. Exon mutations of the new ABO alleles detected in this study**

Allele	Nucleotide positions of exons														
	5	6	7												1061
ABO*	220	297	467	488	526	592	657	703	796	802	803	863	930	1037	1061
A101	C	A	C	C	C	C	C	G	C	G	G	T	G	A	C
Aw07	.	.	T	.	.	T	.	.	.	.	.	.	.	.	*
Aw08	T	G	.	T	G	.	.	A	A	A	.	.	.	.	.
Bw08	.	G	.	.	G	.	T	A	A	.	C	.	A	T	.
Amino acid (AA) position and change	Pro74Ser		Pro156Leu	Thr163Met	Arg176Gly	Arg198Trp		Gly235Ser	Leu266Met	Gly288Arg	Gly288Ala	Met288Arg		Lys346Met	Elongated by 21 AA

The ABO\*A101 sequence was used as the reference sequence. The numbers on top refer to the nucleotide (nt) positions of the exon polymorphisms of the ABO glycosyltransferase gene. Amino acid (AA) changes resulting from nonsynonymous exon nt mutations are listed on the bottom line. Nt position 1061 of exon 7 represents the string of 3C's at nts 1059-1061. The numbers assigned to the nt positions of the exons refer to the mRNA sequence, starting from the beginning of the coding region. Numbering of the AA positions starts with methionine encoded by the initiation codon. The new exon mutations detected in this study and the resulting AA changes are indicated in bold-face type. The nt sequence data of the 3 new ABO alleles have been assigned the accession numbers AJ536130 (ABO\*Aw07), AJ536153 (ABO\*Aw08), and AJ536138 (ABO\*Bw08).

Dot (.) indicates identity to the reference sequence; and —, synonymous nucleotide mutations.

\*Deletions compared with the consensus nts.

This allele was found in a blood sample categorized as A<sub>weak</sub> and had the same nt sequence as *ABO*\*A201 except for an additional mutation at position 592, where C was replaced by T, causing Arg198Trp substitution. Another new nt mutation, 1037A>T, was found in an *ABO*\*B101-like allele named *ABO*\*Bw08, causing Lys346Met substitution. The patient who carried the new *ABO*\*B allele showed very weak B-antigen expression and had weak reacting anti-B antibodies in his serum. A very interesting new mutation, 488C>T, was found in a blood donor initially genotyped as homozygous for *ABO*\*O03. The predicted AA substitution, Thr163Met, was associated with weak A-antigen expression that was detectable only by adsorption/elution. All 3 new *ABO* alleles were found in individuals originating from Germany.

All nt sequence variations detected in the coding region upstream from exon 7, including 4 single point mutations (3 nonsynonymous, 1 synonymous) and a very rare insertion in exon 2, were already known. The 4 point mutations were present in exon 3 (106G>T), exon 4 (188G>A, 189C>T), and exon 5 (220C>T) of *ABO*\*O02-like alleles. Only the 220C>T base substitution could be detected in *ABO*\*O03-like alleles. The insertion, 87-89insG, was identified in exon 2 of an *ABO*\*O allele that was tentatively called *ABO*\*O41. This insertion causes a frame shift that produces a stop codon after nt 165 and premature termination of translation after AA 55. This abnormal *ABO*\*O allele has been described only once before.<sup>4</sup> We detected it in a German blood group O donor.

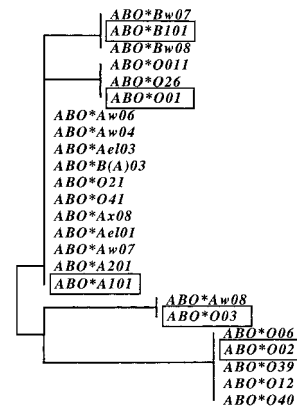
#### Sequence diversity of noncoding regions of the *ABO* gene

The noncoding regions of the *ABO* gene were found to contain 61 polymorphic sites, most of which were located in introns 4 and 6. These polymorphisms consisted of simple point mutations (40 transitions, 15 transversions), one 2-bp deletion, and 5 insertions of up to 13 bp (Table 5). The nt sequences in the promoter region of all common and rare *ABO* alleles were identical. Allele-specific motifs were found in the enhancer CBF/NF-Y region, as was described previously.<sup>5</sup> All non-*ABO*\*A101- and non-*ABO*\*O03-like alleles exhibited a 344-bp fragment containing four 43-bp repetitive motifs. The *ABO*\*O02-like alleles and 51% of the *ABO*\*O01-like alleles also displayed a 41G>C mutation in the first (5') 43-bp sequence. The *ABO*\*A101- and *ABO*\*O03-like alleles, on the other hand, had a 215-bp enhancer CFB-NFY fragment with only a single copy of the 43-bp sequence motif bearing a 41G>A substitution.

As can be seen in Table 5, we identified 5 different allelic groups by a number of recurrent mutations, each group consisting of one major *ABO* allele and a few rare alleles. One *ABO*\*A-, *ABO*\*B-, *ABO*\*O01-, *ABO*\*O02-, and *ABO*\*O03-like allele group, respectively, was distinguished, corresponding to 5 main lineages identified in the phylogenetic sequence analysis: *ABO*\*A, *ABO*\*B, *ABO*\*O01, *ABO*\*O02, and *ABO*\*O03. These lineages were best discriminated by analyzing introns 5 and 6, both of which contained highly conserved lineage-specific sequence motifs (Figure 2). In introns 2 to 4, the *ABO*\*B alleles clustered together with the *ABO*\*A alleles, whereas in intron 6 they clustered with the *ABO*\*O03 group.

#### Inter- and intra-lineage allelic diversity at the *ABO* locus

All sequence variations found in this study are summarized in Figure 3. We discriminated 5 different allelic groups, each of which displayed characteristic sequence motifs. The *ABO*\*B allelic group differed from the *ABO*\*A allelic group in that it had a series of additional mutations downstream from exon 5. The only exception was the rare *ABO*\*B(A)03 allele, which lacked the characteristic *ABO*\*B mutation in intron 5 but had a unique point mutation in intron 3. The *ABO*\*O01-like alleles were characterized by 261delG



**Figure 2. Phylogenetic analysis of intron 5 of the *ABO* gene.** The phylogenetic analysis has been done using Treecon version 1.3b.<sup>18,19</sup> The unrooted tree has been used in order to describe the relative genetic distance between the sequences. The length of the branches represents the extent of nucleotide differences between the lineages. The topology of the tree suggests 5 main lineages for nucleotide sequence diversity in intron 5: *ABO*\*A, *ABO*\*B, *ABO*\*O01, *ABO*\*O02, and *ABO*\*O03. The major *ABO* alleles are indicated by rectangles.

polymorphism in exon 6 and by 9 additional point mutations, predominantly in intron 2. The *ABO*\*O02-like alleles had multiple mutations extending from intron 2 to exon 7, including the A-glycosyltransferase activity abolishing deletion at nt position 261. The 2 *ABO*\*O03 group alleles exhibited a patchwork sequence pattern containing *ABO*\*A-, *ABO*\*B-, *ABO*\*O01-, and *ABO*\*O02-specific exon/intron sequence motifs and additional characteristic mutations (eg, deletion of 2 bp at nt positions 1095 and 1096 in intron 3).

A number of rare alleles differed only slightly from the major *ABO* alleles by single exon and/or intron mutations (eg, *ABO*\*A201, *ABO*\*B(A)03, *ABO*\*O011, and *ABO*\*O06). By contrast, the 2 rare *ABO* alleles *ABO*\*Ax08 and *ABO*\*O12 seemed to be hybrid alleles that occur due to single recombination events; they were classified as *ABO*\*101-*ABO*\*O02 (*ABO*\*Ax08) and as *ABO*\*A101-*ABO*\*O02 or *ABO*\*B101-*ABO*\*O02 (*ABO*\*O12), respectively. Their crossover positions were located between nts 261 and 297 of exon 6 (*ABO*\*Ax08) and between nt 437 of intron 2 and nt 106 of exon 3 (*ABO*\*O12), respectively. The sequence pattern of the rare *ABO*\*O alleles *ABO*\*O39 and *ABO*\*O40 were more complex. Each of these alleles exhibited 2 exon and/or intron nt sequence motifs upstream from intron 5 that were specific for *ABO*\*A and *ABO*\*O01 group alleles, respectively. In 2 very unusual *ABO*\*O alleles, *ABO*\*O21 and *ABO*\*O41, both belonging to the *ABO*\*A group, only singular polymorphisms were detected, including A-transferase activity abolishing mutations in exon 6 (261delG) and exon 2 (88insG), respectively.

## Discussion

In the present study, we investigated the entire genomic sequence (except for intron 1) and 2 regulatory regions of 110 common and variant *ABO* haplotypes to identify variations in the noncoding regions and polymorphisms in the coding region possibly responsible for common and variant ABO phenotypes. Only known nt polymorphisms were detected in the coding region outside exon 7. Twenty-one known and 3 new *ABO* alleles were thereby identified. Each new allele was characterized by the addition of a new single base substitution in exon 7. A 592C>T substitution found on an *ABO*\*A201 background was associated with very weak A-antigen

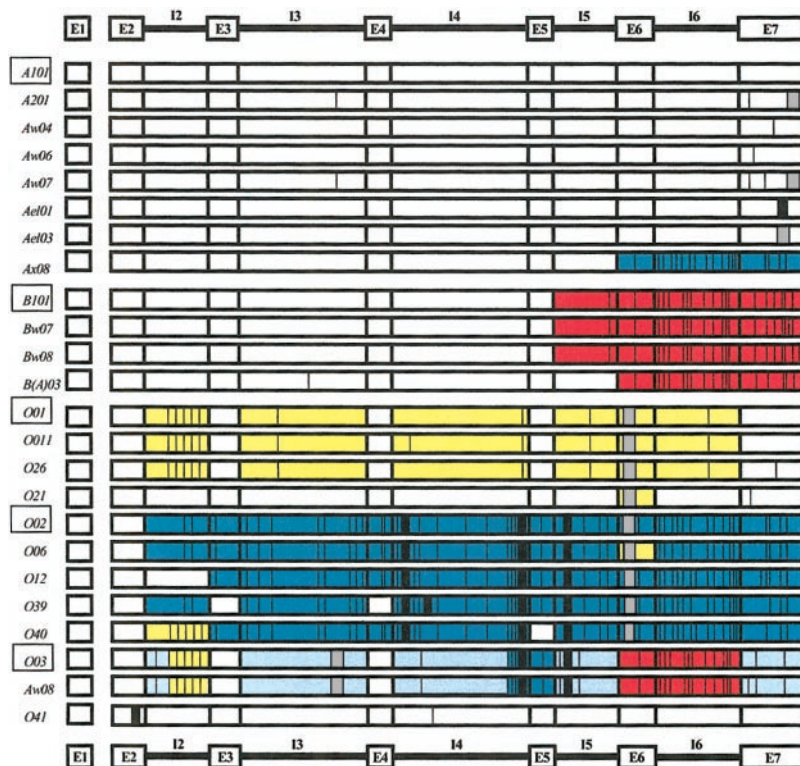
**Table 5. Polymorphic nucleotide positions within introns 2 to 6 of the ABO gene**

Allele	Nucleotide positions in introns																																				
	2					3							4																								
ABO*	127	362	369	396	437	539	205	479	525	575	858	969	1063	1095	1096	1170	1244	1278	28	73	102	114	163	252	346	375	560	738	1176	1467	1496	1511	1623	1678			
A101	.	C	C	T	C	C	C	C	A	C	C	T	A	C	A	C	C	T	G	—	C	C	G	—	T	C	G	T	C	G	T	T	—	T			
A201	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Aw04	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Aw06	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Aw07	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Ae101	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Ae103	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
O41	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	
Ax08	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B101	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bw07	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bw08	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B(A)03	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
O01	.	T	G	C	T	A	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
O011	.	T	G	C	T	A	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
O26	.	T	G	C	T	A	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
O21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
O02	.	.	.	C	T	.	T	T	T	.	.	C	G	.	.	T	C	C	+	.	T	A	.	G	.	.	G	T	A	C	G	C	.	.	.		
O06	.	.	.	C	T	.	T	T	T	.	.	C	G	.	.	T	C	C	+	.	T	A	.	G	.	.	G	T	A	C	G	C	.	.	.		
O12	.	.	.	.	.	.	T	T	T	.	.	C	G	.	.	T	C	C	+	.	T	A	.	G	.	.	G	T	A	C	G	C	.	.	.		
O39	.	.	.	C	T	.	T	T	T	.	.	C	G	.	.	T	C	C	+	.	T	A	TG	G	.	.	G	T	A	C	G	C	.	.	.		
O40	.	T	G	C	T	A	T	T	T	.	.	C	G	.	.	T	C	C	+	.	T	A	.	G	.	.	G	T	A	C	G	C	.	.	.		
O03	A	T	G	C	T	A	.	.	.	.	.	.	.	*	*	.	.	.	.	.	.	.	.	.	.	T	.	.	.	A	C	G	C	.	.		
Aw08	A	T	G	C	T	A	.	.	.	.	.	.	.	*	*	.	.	.	.	.	.	.	.	.	.	T	.	.	.	A	C	G	C	.	.		

The ABO\*A101 sequence was used as the reference sequence. The numbers on top refer to the nucleotide (nt) positions of the intron polymorphisms of the ABO glycosyltransferase gene. The nt positions of an intron are numbered starting from the beginning of each intron. The nt sequence data of the listed ABO alleles have been assigned the accession numbers AJ536122-AJ536153. + indicates an insertion of 13 bp: GTGTGGACAGAAG; dot (.), identity to the reference sequence; and —, insertions compared with the reference sequence.

\*Deletions compared with the consensus nts.

†n = 1 unless otherwise indicated.



**Figure 3. Overview of ABO gene sequence variations observed in this study.** ABO\*A101 was used as the reference sequence. Point mutations, deletions, and insertions are indicated using different bars. The major ABO alleles are represented by rectangles. Parts of the ABO gene showing identical or closely related sequence motifs are highlighted using the following colors: red, ABO\*B; yellow, ABO\*O01; teal, ABO\*O02; and light blue, ABO\*O03. This figure is schematic and the boundaries of the colored areas do not indicate the exact locations of a possible recombination event. ■ indicates deletion; ▣, insertion; |, point mutation.

**Table 5. Polymorphic nucleotide positions within introns 2 to 6 of the ABO gene (continued)**

Geographic origin of the individuals (n) found to carry the respective allele†	Nucleotide positions in introns																Allelic group																		
	5								6																										
	91	103	266	306	336	450	527	530	42	89	163	179	188	226	235	271		280	446	493	628	717	784	786	891	901	950	1011	1013						
Germany (12)	T	-	G	C	G	C	G	A	G	T	T	C	G	C	C	A	C	A	T	A	G	G	A	A	G	A	A	G	A	A	G	A	A		
Germany (9), India, Italy	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Germany (2)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Bosnia, Germany, Turkey	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Germany	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Germany (2)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Africa	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Germany	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Germany	.	.	.	.	.	.	.	.	.	A	C	.	A	T	G	.	.	G	C	.	A	.	G	G	.	G	G	A	.	.	.	.	.		
Germany (11), Bosnia	.	.	.	.	.	.	.	G	T	.	C	T	.	.	.	G	T	G	.	G	.	.	G	G	A	G	.	.	.	.	.	.	.	B	
India	.	.	.	.	.	.	.	G	T	.	C	T	.	.	.	G	T	G	.	G	.	.	G	G	A	G	.	.	.	.	.	.	.		
Germany	.	.	.	.	.	.	.	G	T	.	C	T	.	.	.	G	T	G	.	G	.	.	G	G	A	G	.	.	.	.	.	.	.		
Germany (6)	.	.	.	.	.	.	.	T	.	C	T	.	.	.	G	T	G	.	G	.	.	G	G	A	G	.	.	.	.	.	.	.	.		
Africa, Germany (18), Italy (2)	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	O01
Germany (3)	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	
Germany	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	
Germany	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
Germany (19), Turkey	.	CCC	.	T	.	A	A	.	.	A	C	.	A	T	G	.	.	G	C	.	A	.	G	G	.	G	G	A	.	.	.	.	.	O02	
Italy	.	CCC	.	T	.	A	A	.	.	A	C	.	A	T	G	.	.	G	C	.	A	.	G	G	.	G	G	A	.	.	.	.	.		
Germany	.	CCC	.	T	.	A	A	.	.	A	C	.	A	T	G	.	.	G	C	.	A	.	G	G	.	G	G	A	.	.	.	.	.		
Germany	.	CCC	.	T	.	A	A	.	.	A	C	.	A	T	G	.	.	G	C	.	A	.	G	G	.	G	G	A	.	.	.	.	.		
Germany (2)	.	CCC	.	T	.	A	A	.	.	A	C	.	A	T	G	.	.	G	C	.	A	.	G	G	.	G	G	A	.	.	.	.	.		
Germany (4)	C	CCC	A	.	.	.	.	T	.	C	T	.	.	.	G	T	G	.	G	.	.	G	G	A	G	.	.	.	.	.	.	.	.	O03	
Germany	C	CCC	A	.	.	.	.	T	.	C	T	.	.	.	G	T	G	.	G	.	.	G	G	A	G	.	.	.	.	.	.	.	.		

expression (*ABO\*Aw07*). This is the first time that a mutation relevant for ABO subgroup expression was detected between nts 548 and 641 of the coding sequence. This finding suggests that changes in this region can also affect the sugar binding domain of ABO glycosyltransferase. However, the question of whether the new mutation acts alone or in combination with typical *ABO\*A201* polymorphisms to yield altered A-antigen expression is still unanswered. The second nonsynonymous new mutation (1037A>T) was located near the 3' end of the gene. The occurrence of this mutation in an *ABO\*B101*-like allele was associated with a diminution of B-antigen expression (*ABO\*Bw08*). This was not surprising since an alteration of B-antigen expression has already been described for a mutation at position 1036 in a variant *ABO\*B* allele.<sup>14</sup> The third phenotypically relevant mutation was found in a genotype with *ABO\*O03*-defining mutations (297A>G, 526C>G, 802G>A) in homozygosity. While the mutation at position 802 is known to abolish the enzymatic activity of the ABO glycosyltransferase, the simultaneous presence of a new mutation at position 488 near the 5' end of exon 7 seemed to reactivate the A-transferase activity. The resulting new *ABO\*A* allele (*ABO\*Aw08*) is the first reported *ABO\*O*-like allele with an additional mutation that seems to be responsible for "reactivation" of A-antigen expression. Although mutations in the 5' region of exon 7 were thought to be incapable of drastically altering enzymatic activity or specificity, our recent finding of a single base substitution at nt position 502 in 3 individuals with identical weak A phenotypes strongly suggests that changes in the AA structure in this region can have a significant effect on A-transferase activity.<sup>6</sup>

Some sequence variations in noncoding regions of the *ABO* gene have already been reported. These include variations in minisatellite repeats in the 5' untranslated region (UTR) and single-nt polymorphisms in intron 6 and the 3' UTR.<sup>11-13,20</sup> The pattern of these variations was concordant with the major *ABO*

alleles. Moreover, the discovery of *ABO* allele-specific haplotypes of intron 6 allowed for identification of break points of recombination events and localization of recombination hot spots around intron 6 and exon 7.<sup>11,13</sup> However, allelic variations in introns 1 to 5 have not yet been examined. In this study, haplotype-specific sequence analysis of introns 2 to 6 of 24 different common and rare *ABO* alleles revealed extensive sequence variations in all investigated introns. While only a few of the observed intron mutations were specific for single rare alleles, highly conserved sequence motifs closely resembling the major *ABO* alleles were found in each intron. Accordingly, our phylogenetic analysis revealed 5 main *ABO* allele lineages: *ABO\*A*, *ABO\*B*, *ABO\*O01*, *ABO\*O02*, and *ABO\*O03*. This finding is largely consistent with those of Saitou and Yamamoto,<sup>21</sup> who phylogenetically analyzed the coding sequences of human and primate *ABO* alleles. However, our intron data may carry more weight than the exon data because substitutions in introns are presumably less susceptible to selection pressure, so they may reflect the phylogeny of the alleles more reliably.<sup>22</sup>

The frequency of replacement mutations (73%) indicates the occurrence of positive selection for allelic diversity. Two mechanisms are considered to be responsible for genetic diversity and diversification at the *ABO* locus: point mutations and genetic recombination (crossovers and gene conversion).<sup>9</sup> By extending the sequence analysis to the entire coding region and to 5 of the 6 introns of the *ABO* gene, we were able to confirm the important role of both mechanisms in generating the extensive sequence variations at the *ABO* locus. The origin of a number of rare alleles (eg, *ABO\*Ael03*, *ABO\*Bw07*, *ABO\*O26*, *ABO\*Aw08*) can simply be explained by single mutations that occurred on the background of common *ABO* alleles, whereas other alleles (eg, *ABO\*Ax08*, *ABO\*O40*) arise mainly or solely due to one or more recombination events. Most interestingly, the *ABO\*O03*-like alleles exhibited

a mosaic structure containing large exon and/or intron sequence motifs from each of the other 4 allelic groups. Since only a few additional *ABO\*O03*-specific polymorphisms were found, we hypothesize that *ABO\*O03* is the most recent allelic group and that it arose by recombination of *ABO\*A101*, *ABO\*B101*, *ABO\*O01*, and *ABO\*O02* lineages. The fact that *ABO\*O03*-like alleles have only been found at low frequencies in various white populations and in blacks but not yet in Chinese and American Indians supports this theory.<sup>4</sup> Due to the apparently exceptional role of recombination in generating new alleles at the *ABO* locus, the lack of group-specific mutations in some rare alleles (eg, 703G>A in exon 7 of *ABO\*B(A)03*) or the sharing of sequence motifs between lineages (eg, 261delG in exon 6 of *ABO\*O21*) is most probably due to an inter-lineage nt sequence transfer rather than backward mutation or convergent evolution, as was previously suggested for the *RH* and *HLA* loci.<sup>23-25</sup> Given the rarity of variant *ABO* alleles, the theory that such alleles are intermediate products of common *ABO* allele formation may be less plausible. Chi sequences and chi-like motifs in the 3' end of intron 6 have been considered to be an initiator of crossover events.<sup>26</sup> However, our findings in rare *ABO* alleles indicate that intragenic recombination events outside intron 6 may also account for genetic diversity at the *ABO* locus.

In conclusion, the comprehensive database of *ABO* sequences provided by this study indicates that there is a high level of recombinant activity in almost all parts of the *ABO* gene. The

presence of a large number of recurrent mutations is characteristic for the considerable diversity of the *ABO* gene. Phenotype-genotype correlation revealed that an extensive heterogeneity underlies the molecular basis of various alleles that generate serologic ABO subgroups. *ABO* sequence variations also include phenotypically relevant replacement mutations in exons 2 to 5. Thus, *ABO* genotyping strategies would have to consider all variations distributed across the entire coding region to achieve safe phenotype prediction. Therefore, *ABO* genotyping remains mainly reserved as a complement to serology for determination of inherited ABO subgroups and exclusion of *ABO\*B* allele markers in the acquired B phenotype. The data on highly conserved and lineage-specific intron sequence motifs provide a powerful base for elucidating the origin of variant *ABO* alleles and may prove valuable for anthropologic studies on the origins and movements of populations.

## Acknowledgments

We are grateful to Dr Volker Lenhard (Biotest AG, Dreieich, Germany) for supplying multiple blood samples with rare ABO phenotypes. Part of this work will appear in Anke Kollmann's doctoral thesis.

## References

1. Yamamoto F, Clausen H, White T, Marken J, Hakamori S. Molecular genetic basis of the histo-blood group system. *Nature*. 1990;345:229-233.
2. Yamamoto F, McNeill PD, Hakamori S. Genomic organization of human histo-blood group ABO genes. *Glycobiology*. 1995;5:51-58.
3. Bennett EP, Steffensen R, Clausen H, Weghuis DO, van Kessel AG. Genomic cloning of the human histo-blood group ABO locus. *Biochem Biophys Res Commun*. 1995;206:318-325.
4. Chester MA, Olsson ML. The ABO blood group gene: a locus of considerable genetic diversity. *Transfus Med Rev*. 2001;15:177-200.
5. 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.
6. Seltam A, Hallensleben M, Eiz-Vesper B, Lenhard V, Heymann G, Blasczyk R. A weak blood group A phenotype caused by a new mutation at the ABO locus. *Transfusion*. 2002;42:294-301.
7. Roubinet F, Janvier D, Blancher A. A novel *cis AB* allele derived from a B allele through a single point mutation. *Transfusion*. 2002;42:239-246.
8. Yu L-C, Twu Y-C, Chou M-L, Chang C-Y, Wu C-Y, Lin M. Molecular genetic analysis for the B<sup>3</sup> allele. *Blood*. 2002;100:1490-1492.
9. Yip SP. Sequence variation at the human ABO locus. *Ann Hum Genet*. 2002;66:1-27.
10. Seltam A, Hallensleben M, Kollmann A, Burkhart J, Blasczyk R. Systematic analysis of the ABO gene diversity within exons 6 and 7 by PCR-screening revealed new ABO alleles. *Transfusion*. 2003;43:428-439.
11. 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;99:454-461.
12. Olsson ML, Santos SE, Guerreiro JF, Zago MA, Chester MA. Heterogeneity of the O alleles at the blood group ABO locus in Amerindians. *Vox Sang*. 1998;74:46-50.
13. 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;53:190-199.
14. 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;98:1585-1593.
15. Daniels G. *Human Blood Groups*. Oxford, United Kingdom: Blackwell Scientific; 2002.
16. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acid Res*. 1994;22:4673-4680.
17. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 1999;41:95-98.
18. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987;4:406-425.
19. Van de Peer Y, De Wachter R. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput Appl Biosci*. 1994;10:569-570.
20. 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;9:219-226.
21. Saitou N, Yamamoto F. Evolution of primate ABO blood group genes and their homologous genes. *Mol Biol Evol*. 1997;14:399-411.
22. Kimura M. *The Neutral Theory of Molecular Evolution*. Cambridge, United Kingdom: Cambridge University Press; 1983.
23. Omi T, Takahashi J, Tsudo N, et al. The genomic organization of the partial D category D<sup>15a</sup>: the presence of a new partial D associated with the D<sup>15a</sup> phenotype. *Biochem Biophys Res Commun*. 1999;254:786-794.
24. Elsner HA, Blasczyk R. Identification of the novel allele HLA-B\*1545: assessment of alloreactivity in case of mismatch with other B\*15 alleles. *Tissue Antigens*. 2000;55:80-82.
25. 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;6:843-850.
26. Olsson ML, Chester MA. Heterogeneity of the blood group A<sub>x</sub> allele: genetic recombination of common alleles can result in the A<sub>x</sub> phenotype. *Transfus Med*. 1998;8:231-238.