Review Article

Molecular genetics and genomics of the ABO blood group system

Fumiichiro Yamamoto

Immunohematology and Glycobiology, Josep Carreras Leukaemia Research Institute (IJC), Badalona, Spain

Correspondence to: Fumiichiro Yamamoto. Immunohematology and Glycobiology, Josep Carreras Leukaemia Research Institute (IJC), Badalona, Spain. Email: fyamamoto@carrerasresearch.org.

Abstract: The A and B oligosaccharide antigens of the ABO blood group system are produced from the common precursor, H substance, by enzymatic reactions catalyzed by A and B glycosyltransferases (AT and BT) encoded by functional *A* and *B* alleles at the *ABO* genetic locus, respectively. In 1990, my research team cloned human *A*, *B*, and *O* allelic cDNAs. We then demonstrated this central dogma of ABO and opened a new era of molecular genetics. We identified four amino acid substitutions between AT and BT and inactivating mutations in the *O* alleles, clarifying the allelic basis of ABO. We became the first to achieve successful ABO genotyping, discriminating between *AA* and *AO* genotypes and between *BB* and *BO*, which was impossible using immunohematological/serological methods. We also identified mutations in several subgroup alleles and also in the *cis-AB* and *B(A)* alleles that specify the expression of the A and B antigens by single alleles. Later, other scientists interested in the ABO system characterized many additional *ABO* alleles. However, the situation has changed drastically in the last decade, due to rapid advances in next-generation sequencing (NGS) technology, which has allowed the sequencing of several thousand genes and even the entire genome in individual experiments. Genome sequencing has revealed not only the exome but also transcription/translation regulatory elements. RNA sequencing determines which genes and spliced transcripts are expressed. Because more than 500,000 human genomes have been sequenced and deposited in sequence databases, bioinformaticians can retrieve and analyze this data without generating it. Now, in this era of genomics, we can harness the vast sequence information to unravel the molecular mechanisms responsible for important biological phenomena associated with the ABO polymorphism. Two examples are presented in this review: the delineation of the *ABO* gene evolution in a variety of species and the association of single nucleotide variant (SNV) sites in the *ABO* gene with diseases and biological parameters through genome-wide association studies (GWAS).

Keywords: Blood group ABO system; *ABO* genes; A and B glycosyltransferases; A and B oligosaccharide antigens

Received: 10 November 2020; Accepted: 22 February 2021; Published: 25 September 2021.
doi: 10.21037/aob-20-71
View this article at: http://dx.doi.org/10.21037/aob-20-71

The dawn of the molecular genetics of the ABO blood group system

Immunohematology and serology studying erythrocyte antigens and their antibodies, respectively, were the initial fields of ABO research. However, it subsequently expanded into a wide variety of scientific disciplines. Antigens A and B can also be expressed on epithelial cells and endothelial cells, in addition to red blood cells (RBCs), depending on the blood types of the individuals. Consequently, ABO compatibility is also essential in cell/tissue/organ transplantation, as well as blood transfusion. The antigens A and B are oligosaccharides, GalNAcα1-3(Fucα1-2)Gal and Galα1-3(Fucα1-2)Gal, respectively. Therefore, they are subjects of glycobiology. In the late 1950s, a hypothesis was put forward on the biosynthetic pathways of these antigens: *A* and *B* alleles at the *ABO* genetic locus encode A and B transferases (AT and BT), which catalyze the biosynthesis of A and B antigens by transferring an N-acetyl-D-galactosamine (GalNAc) and a galactose (Gal), respectively, to the same substrates, H substances (Fucα1-2Gal), which are present on the RBCs of individuals of

Exon	6		7														
Nucleotide Position	261	297	467	526	646	657	681	703	771	796	802	803	829	871	930	1054	1060
A alleles																	
A101	G	A	C	C	T	C	G	G	C	C	G	G	G	G	G	C	C
A102	*	*	T	*	*	*	*	*	*	*	*	*	*	*	*	*	*
A201	*	*	T	*	*	*	*	*	*	*	*	*	*	*	*	*	Δ
A301	*	*	*	*	*	*	*	*	*	*	*	*	*	A	*	*	*
Ax01	*	*	*	*	A	*	*	*	*	*	*	*	*	*	*	*	*
B alleles																	
B101	*	G	*	G	*	T	*	A	*	A	*	C	*	*	A	*	*
B301	*	G	*	G	*	T	*	A	*	A	*	C	*	*	A	T	*
AB alleles																	
cis-AB01	*	*	T	*	*	*	*	*	*	*	*	C	*	*	*	*	*
B(A)01	*	G	*	G	*	*	*	*	*	A	*	C	*	*	A	*	*
O alleles																	
O01	Δ	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
O02	Δ	G	*	*	A	*	A	*	T	*	*	*	A	*	*	*	*
O03	*	G	*	G	*	*	*	*	*	*	A	*	*	*	*	*	*
Possible Amino Acid Change	Frameshift	No change	P156L	R176G	F216I	No change	No change	G235S	No change	L266M	G268R	G268A	V277M	D291N	No change	R352W	Frameshift

Figure 1 Twelve *ABO* alleles characterized molecularly in our group. Nucleotide and deduced amino acid sequences were compared, and only differences from the *A101* allele are shown. This table was modified from previous publications (8,9).

group O. Therefore, ABO has become subjects of genetics, biochemistry and enzymology. A and/or B antigens can also be found in secretion, such as saliva and seminal fluid, as well as hair and skin. Biological samples containing A/B antigens can provide crucial evidence in crime scene investigation. In fact, the ABO polymorphism was the most widely used evidence to exclude innocent people from possible suspects before DNA typing became popular in the last decade. Furthermore, the ABO polymorphism is not limited to humans. Consequently, ABO is the subject of research on genetic evolution.

In 1990, 90 years after the discovery of ABO blood groups, Clausen, Hakomori, and I cloned the *A*, *B*, and *O* allelic cDNAs from stomach and colon cancer cell line cells expressing differential ABO phenotypes, and determined their nucleotide sequences and deduced amino acid sequences (1,2). By correlating the sequence differences with the expression of A/B antigens, we were able to demonstrate, what is now called, the Central Dogma of ABO. AT and BT are the same in size with 354 amino acids (353 in alternatively spliced transcripts), but 4 amino acid substitutions differentiate these two enzymes with different sugar specificities. They are arginine (R), glycine (G), leucine (L), and glycine at codons 176, 235, 266, and 268 in AT encoded by *A* allele (*A101*), while they are glycine, serine (S), methionine (M), and alanine (A) in BT encoded by *B* allele (*B101*). We also identified a single nucleotide inactivating deletion of guanine at nucleotide 261, 261delG, in most *O* alleles (*O01*), while some other *O* alleles (*O02*) have additional nucleotide substitutions, in addition to 261delG. Using allele-specific differences in nucleotide sequences and restriction fragment length polymorphism (RFLP), we successfully achieved the first ABO genotyping. *AA* and *AO* genotypes, as well as *BB* and *BO* genotypes, were also discriminated. Later, we also found another type of *O* allele (*O03*) lacking 261delG, but containing an inactivating glycine to arginine substitution at codon 268, G268R, and a functionally insignificant arginine to glycine substitution at codon 176, R176G (3). We also identified a single nucleotide deletion (1060delC) at the C-terminus and a substitution of proline to leucine at codon 156 (P156L) in *A2* alleles (*A201*) (4), an aspartic acid to asparagine at codon 291 (D291N) in an *A3* allele (*A301*) (5), a phenylalanine to isoleucine substitution at codon 216, F216I, in an *Ax* allele (*Ax01*) (6), and an arginine to tryptophan substitution at codon 352, R352W, in a *B3* allele (*B301*) (5). We also demonstrated that a *cis-AB* allele (*cis-AB01*) and a *B(A)* allele (*B(A)01*) that specify the expression of the A and B antigens by single alleles encode AT-BT chimeras (6,7). The alleles that we characterized are summarized in *Figure 1*.

Gold hunting for additional *ABO* alleles

When analyzing the sequences of the cDNA clones, we found that several clones contained unspliced intron sequences. Taking advantage of the sequence information, we showed that PCR and DNA sequencing of the amplified fragments could be used for the characterization of the majority of the coding sequence of the *ABO* gene. In fact, we have shown that the coding sequences of the last two coding exons (exons 6 and 7), which correspond to approximately 80% of the complete coding sequence of the soluble form of AT, could be amplified by PCR. The DNA fragment spanning exon 6 was amplified, using a pair of oligo primers (a sense strand primer in intron 5 and an antisense strand primer in intron 6), while the exon 7 sequence was amplified in two overlapping DNA fragments, using a sense primer in intron 6 and an antisense primer in exon 7, and also using a sense primer in exon 7 and an antisense primer in the 3'-untranslated region. We used these primers to characterize mutations in the alleles of A and B subgroups and mutations in the *cis-AB* and *B(A)* alleles described above (3-7). Our characterization of twelve *ABO* alleles was the beginning of the search for new *ABO* alleles. Other researchers followed suit, employing the same strategy. Some used the same primers for PCR and characterized additional *ABO* alleles, while others used primers of their own design, all based on our published sequences. We also found at least one primer that had the same nucleotide sequence as ours, but with a different primer name. Anyway, thanks to the availability of the combined strategy of PCR and sequencing, mutations in dozens of *ABO* alleles were characterized even before ABO genomic cloning clarified the entire genetic organization of the *ABO* gene in 1995 (10).

Many researchers contributed to the characterization of these additional *ABO* alleles. Among them are Olsson and Chester (Sweden), Ogasawara, Hosoi, Suzuki and Fukumori (Japan), Yu and Lin (Taiwan), Yip (China), and Seltsam and Blasczyk (Germany) [see the list of references in review articles (11-14)]. In addition to the ABO system, other blood group systems were also molecularly characterized. Originally, we used an allele nomenclature combining the phenotype with the number showing the order of discovery in parentheses, *A1*(1) for example to indicate the first *A1* allele identified (8). As the number of alleles identified increased, it became apparent that discoveries were often made without the knowledge of others. Along with the use of different nomenclatures by different research teams, this caused a confusion. To alleviate chaos, in 1999 Blumenfeld established the Blood Group Antigen Gene Mutation Database (BGMUT) through an initiative of the Human Genome Variation Society (HGVS) (15), and I was invited to be a curator for the ABO section of the database. Assuming that 99 alleles for individual ABO phenotypes would be sufficient for some time, we employed a nomenclature of alleles per phenotype followed by a 2-digit number showing the order of discovery. The nucleotide and deduced amino acid sequences of the first characterized *A1* allele, *A101*, were used as standards (9,11). In 2006, BGMUT became a part of NCBI's dbRBC (database Red Blood Cells) resource at NIH. More than 250 *ABO* alleles were deposited before the unfortunate closure of the database in 2016 due to the passing of Dr. Blumenfeld. Meanwhile, the International Society of Blood Transfusion (ISBT) established a Working Party on Red Cell Immunogenetics and Blood Group Terminology chaired by Storry, and prepared catalogs of blood group antigens and alleles. Olsson used *ABO* allele terminology, which was similar to ours based on phenotype followed by a number (http://www.isbtweb.org/working-parties/red-cell-immunogenetics-and-blood-group-terminology/).

Functional characterization of *ABO* mutations

Thanks to the efforts of many, a variety of *ABO* alleles were molecularly characterized. These alleles include *A1*, *A2*, *A3*, *Ael*, *Aint*, *Am*, *Aw*, *Ax*, *cis-AB*, *B(A)*, *B*, *B3*, *Bel*, *Bw*, *Bx*, and *O*. Different types of mutations were found, including synonymous mutations, missense mutations, nonsense mutations, frameshift mutations, and splicing mutations. Later studies identified mutations in the regulatory elements of transcription. Furthermore, several alleles were found to contain different types of mutations. See review articles (11-14) for individual references to the original articles.

Identifying potential mutations is one thing, but proving their functional importance is another. Therefore, it was necessary to examine the effects of the identified mutations on the appearance of A and/or B antigens. For this evaluation, we developed a functional assay system that employs DNA transfection of AT/BT expression constructs, as well as those containing mutations, and subsequent immunological detection of A/B antigens. We used HeLa cells derived from uterine cancer. We first performed ABO genotyping of these cells and found that Ms. Henrietta Lacks had a group O phenotype based on the presence of

Construct	Possible Gene	Expression	
		A antigen	B antigen
A construct	A1	++++	-
A(P156L)	A1	++++	-
A(1060delC)	A2	++	-
A(P156L, 1060delC)	A2	++	-
A(D291N)	A3	++	-
A(F216I)	Ax	ND	ND
A(261delG)	O	-	-
A(R176G, G268R)	O	-	-
B construct	B	-	++++
B(R352W)	B3	ND	ND

Figure 2 Functional analysis of mutations in *O*, *A2*, and *A3* alleles. The appearance of A and/or B antigens was examined in cells transfected with DNA from the AT, BT and their derivative constructs containing mutations specific to *O*, *A2*, and *A3* alleles. The word "ND" means "not determined". This table was modified from our previous publication (8).

261delG in both alleles of the *ABO* gene. We also showed that HeLa cells express H substance on the cell surface (16). It was assumed that if a eukaryotic expression construct encodes a protein with AT and/or BT activities, a GalNAc and/or a Gal can be transferred to H substances, and A and/or B antigens can be produced, respectively. And these antigens can be detected immunologically, using anti-A and/or anti-B antibodies.

To examine the feasibility of the approach, we prepared the cDNA expression constructs of *A101* (AT) and *B101* (BT) in a eukaryotic expression plasmid vector pSG5 (16). We then transfected the DNA of these constructs into HeLa cells and examined the appearance of the A and B antigens. As we anticipated, we observed the expression of A and B antigens, respectively (*Figure 2*). Once the AT and BT activities of the original AT and BT constructs were confirmed, AT constructs containing 261delG specific for the *O01/O02* alleles or R176G and G268R substitutions specific for the *O03* allele were prepared, and their AT activity was examined. We did not detect any appearance of A antigens, which showed that these are inactivating mutations (3). Although it was not determined whether both substitutions of R176G and G268R were necessary for inactivation in that experiment, a single G268R substitution was found to be sufficient to abolish AT activity in a later experiment (17). Similarly, we introduced the *A201* allele-specific P156L substitution and 1060delC deletion into the *A101* construct and observed a decrease in AT activity. Because P156L substitution was also found in some *A1* alleles and the introduction of P156L alone did not decrease AT activity, we concluded that 1060delC was responsible for the A2 phenotype (4). We also introduced D291N substitution specific for an *A301* allele in AT and observed a decrease in AT activity (8).

In addition to subgroup-specific mutations, we also used this system to analyze differences in sugar specificity. First, we determined which of the four amino acid substitutions are responsible for the differential sugar specificity between AT and BT. Are all the four amino acid substitutions required? Or just a few are enough? We answered these questions. We constructed 14 AT-BT chimeras that are different at those four positions having the amino acid of AT or BT, and examined the sugar specificities of GalNAc/Gal (16). The following results were obtained as shown in *Figure 3*. When positions 3 and 4 have amino acids of AT (AA), only AT activity was observed. When they were from BT (BB), only BT activity was observed. When they were AB in this order, the amino acid in the 2nd position determined the specificity. Only AT activity was detected when it was from AT (AAB), while weak BT activity was also present, in addition to AT activity, when it was from BT (BAB). And finally, when positions 3 and 4 were BA in this order, strong AT and BT activities were observed. We concluded that amino acids at codons 176 and 235 have no and slight effects, respectively, while the amino acids at codons 266 and 268 are decisive in determining the sugar specificity. Despite our success in evaluating the functional significance of mutations in *ABO* alleles that we characterized, this approach was not widely used, possibly due to the inability to conduct molecular and cellular biological investigations in immunohematology/serology laboratories. Consequently, the functionality of many of the mutations found in the *ABO* gene has yet to be determined.

ABO polymorphism in the era of genomics

In the current ISBT Blood Group Allele Table (ISBT 001 v1.1 171023), 207 *ABO* blood group alleles are listed, including 84 *A*, 49 *B*, 12 *cis-AB/B(A)*, and 62 *O*. However, recent advances in next-generation sequencing (NGS) have generated sequence data for more than half a million human genomes. If the sequences of two alleles in a genome were counted separately, data could have been generated for more than a million *ABO* genes. *Figure 4* shows the nucleotide and deduced amino acid sequences of the coding region of the reported human *ABO* gene cDNA (as of August 20, 2020). To prepare this figure, the original (transcript: ABO-203 ENST00000611156.4 in the Ensembl

Plasmid	Experiment		Experiment		Experiment		Activity
DNA	1		2		3		
	A	B	A	B	A	B	
pAAAA	41,5	0,0	3,8	0,0	14,0	0,1	A
pAAAB	17,5	0,3	1,2	0,0	7,4	0,1	A
pAABA	NT	NT	1,0	0,6	2,4	1,1	AB
pAABB	0,2	26,3	0,1	1,4	0,1	5,6	B
pABAA	27,5	0,2	5,7	0,1	11,6	0,1	A
pABAB	21,3	3,0	1,8	0,1	5,8	0,2	A(B)
pABBA	17,0	22,1	0,8	1,4	2,1	2,9	AB
pABBB	0,1	31,1	0,1	1,6	0,1	5,5	B
pBAAA	29,1	0,1	2,9	0,1	10,3	0,0	A
pBAAB	10,0	0,1	0,5	0,0	4,8	0,1	A
pBABA	NT	NT	0,5	0,4	3,1	1,3	AB
pBABB	0,1	20,7	0,0	1,0	0,0	5,4	B
pBBAA	12,7	0,1	4,8	0,0	12,3	0,0	A
pBBAB	29,5	2,9	1,4	0,0	8,0	0,4	A(B)
pBBBA	NT	NT	1,0	0,6	3,2	2,0	AB
pBBBB	0,1	30,6	0,0	2,5	0,1	3,3	B
no DNA	0,0	0,1	0,0	0,0	0,0	0,0	

Figure 3 AT/BT specificity/activity of the AT-BT chimeras. The appearance of A and/or B antigens was examined in cells transfected with DNA from the AT, BT, and 14 AT-BT chimeric expression constructs. This table was modified from our previous publication (16).

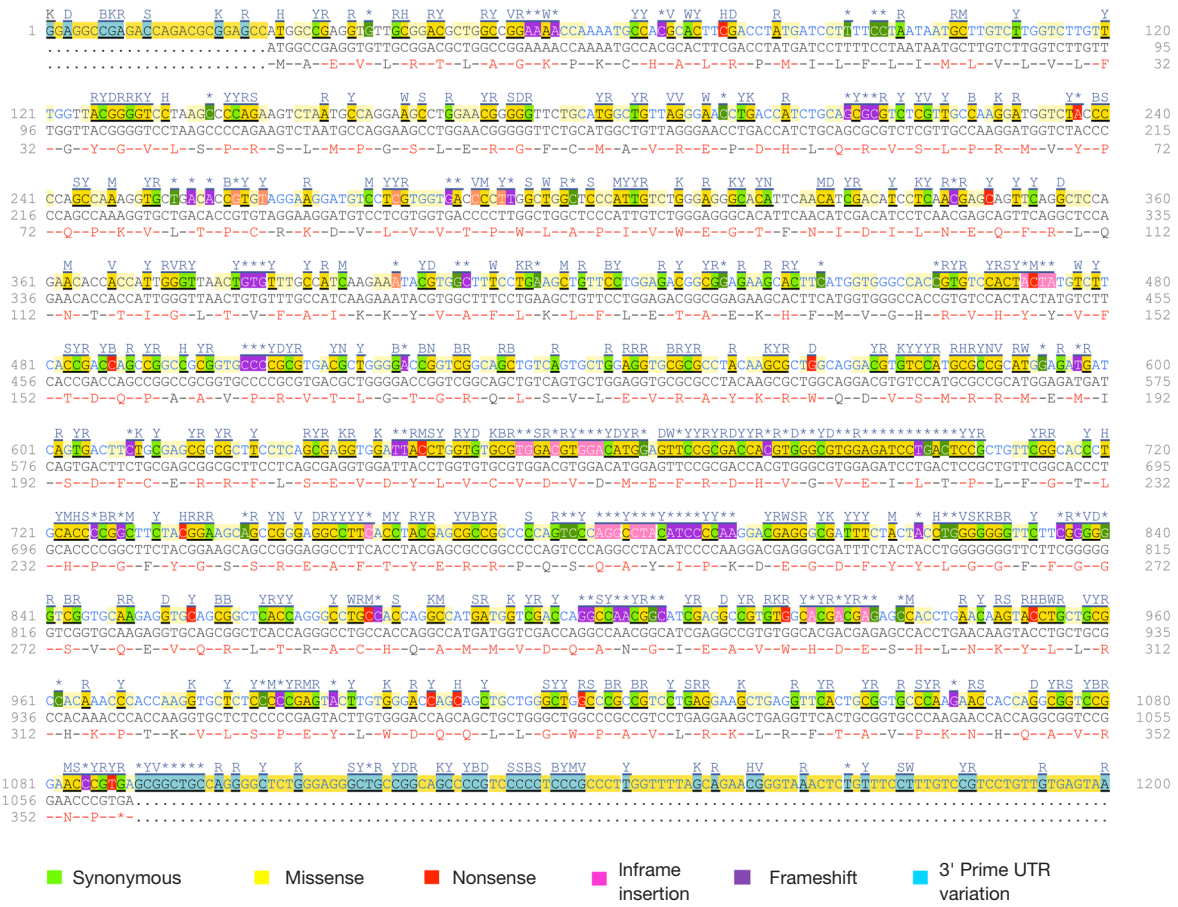


Figure 4 SNVs in the *ABO* gene cDNA sequence. SNVs are shown in the *O* allele transcript (ABO-203 ENST00000611156.4) from the Ensembl database, which was modified to encode AT. Different types of SNVs are highlighted in different colors. SNV, single nucleotide variant.

database) was added with three nucleotides and one amino acid surrounding the 261delG. This manipulation was necessary because the reference sequence in the database was from an *O* allele with the 261delG deletion, and because two nucleotides were mistakenly removed as an intron to fit in the reading frame after the deletion. Different types of single nucleotide variants (SNVs) are shown in different colors.

Many SNVs have been identified within the coding region of the *ABO* gene. Non-synonymous SNVs that produced a change in the amino acid sequence of the encoded protein were observed at 549 nucleotides out of 1,065 nucleotides in the coding sequence including the termination codon. The number of SNVs is higher if synonymous differences are added. And the number will increase even more if the SNVs found in introns and regulatory regions are added, because the human *ABO* gene spans over 24,818 base pairs (chromosome 9: 133,250,401–133,275,219) where the majority is non-coding sequence. Assuming that alleles are defined as genes that occupy the same genetic locus but have different sequences, the number of alleles can be in the millions, because they can have different combinations of SNVs. It was clear that the allele nomenclature in use, which is based on numbering, would soon disappear due to a rapid increase in new alleles identified from NGS. In predicting this situation, in 2012 we proposed the use of allele nomenclature that lists SNVs that are different from the nucleotide and amino acid sequences of the standard *A101* allele (14).

Predominance of ABO phenotyping over ABO genotyping

The ABO polymorphism is specified by a single genetic locus *ABO*. This is correct in a general sense. However, it can also be wrong. For example, individuals of the Bombay and para-Bombay phenotypes may be wrongly genotyped to exhibit an apparent A/B/AB phenotype if functional *A* and/or *B* alleles are present. This is due to the fact that these individuals lack functional alleles in the *FUT1* (*H/h*) and/or *FUT2* (*Se/se*) genes that encode α1,2-fucosyltransferases to synthesize H substances, the precursor substrates for A and B transferases. Without H substances, even functional AT and BT are unable to synthesize A and B antigens, respectively. Although the incidence of Bombay or para-Bombay individuals is extremely low, this illustrates a potential difficulty in ABO genotyping. In other words, the ABO phenotype can be deduced from the *ABO* gene sequence determined by ABO genotyping, but it may not be accurate. Even if there are no mutations in the coding region of the *ABO* gene, mutations may be present in regulatory elements. Even in the presence of functional *FUT1/FUT2* and *A/B* genes, donor nucleotide-sugars (UDP-GalNAc/UDP-Gal) are necessary for the synthesis of A and/or B antigens to occur in the *trans*-Golgi apparatus.

In addition, the ABO phenotypes of most of the samples, whose genomic sequences were determined, have not been characterized. Furthermore, the functional significance of most of the SNVs that were identified in the *ABO* gene has not been determined. In this circumstance, the ABO phenotypes deduced from the *ABO* gene sequences may not be completely correct, although the A, B, or O phenotype can be determined with greater than 99% precision, using SNV data characterized by function. Nonetheless, useful information can be obtained on the possible molecular mechanisms that explain differential specificity/activity. It is also true that DNA typing has been used successfully to treat hemolytic disease of the fetus and newborn (HDFN) and to identify optimally matched donations for patients with pre-existing antibodies or predisposed to alloimmunization. However, a cross-match test is required prior to transfusion to determine whether the donor blood is compatible with the blood of the intended recipient because incorrect determination of ABO phenotypes by genotyping can result in very detrimental clinical outcomes. In conclusion, ABO phenotyping appears to continue to prevail over ABO genotyping.

ABO genes in species other than *Homo sapiens*

The ABO polymorphism was initially identified in humans. It was later found in other animals. Moor-Jankowski performed an extensive analysis of the ABO polymorphism of primates (18,19). It was found that chimpanzees analyzed were either A or O phenotype, while gorillas were B phenotype. These anthropoid ape species express A/B antigens on RBCs like humans do, but lower primates do not. Still, they can express these antigens in respiratory/digestive epithelium and/or in secretions such as saliva. We determined partial nucleotide sequences and deduced amino acid sequences of the *ABO* genes from several primate species (20). We observed that the amino acids corresponding to codons 266 and 268 of human AT/BT were conserved in primates, depending on the A/B status. They were leucine (L) and glycine (G) in the *A* alleles of chimpanzee, orangutan, macaque, and baboon, while they were methionine (M) and alanine (A) in the *B*

(I). G at codon 268				(II). A at codon 268				(III). Additional			
Codons (266-268)	A Activity	B Activity	A/B Specificity	Codons (266-268)	A Activity	B Activity	A/B Specificity	Codons (266-268)	A Activity	B Activity	A/B Specificity
AGG	+++++	-	A	**AGA**	+++++	++	AB	**AAA**	+++++	-	A
CGG	+++++	-	A	CGA	++++	+++	AB	**AAN**	-	-	NEG
DGG	++	++	AB	DGA	-	+++	B	**AAS**	++++	+++	AB
EGG	++++	-	A	EGA	-	++++	B	**MAA**	-	+++++	B
FGG	-	++++	B	FGA	-	++++	B	**MGP**	-	+++	B
GGG	++++	-	A	**GGA**	++++	+++	AB	**MGS**	-	+++++	B
HGG	-	++++	B	HGA	-	++++	B	**QGC**	-	+++++	B
IGG	++++	++++	AB	**IGA**	-	++++	B	SSE	-	-	NEG
KGG	-	-	NEG	KGA	-	+++	B	TAS	-	-	NEG
LGG	+++++	-	A	**LGA**	+++++	+	AB	**TEA**	-	-	NEG
MGG	++++	++++	AB	**MGA**	-	++++	B	**TGC**	++++	-	A
NGG	+++++	+	AB	NGA	++++	++	AB	**TGF**	-	-	NEG
PGG	+++++	-	A	PGA	++++	-	A	TSE	-	-	NEG
QGG	++++	+++	AB	QGA	-	+++++	B				
RGG	-	-	NEG	RGA	-	-	NEG	(263-268)			
SGG	++++	-	A	**SGA**	++++	+++	AB	**AYVYGS**	-	-	NEG
TGG	+++++	-	A	**TGA**	++++	+++	AB	FYFTSE	-	-	NEG
VGG	++++	-	A	VGA	++++	+++	AB	HYYMGG	++++	++++	AB
WGG	++	+	AB	WGA	-	++++	B	YYYAGG	+++++	-	A
YGG	-	++++	B	YGA	-	++	B	YYYMGG	+++++	+++	AB
								YYYTGS	+++++	-	A
								YYYTSE	-	-	NEG
								YYYTSG	+++++	-	A

Figure 5 The amino acid motif—A/B specificity code table. Sugar specificity was determined experimentally for human AT with a variety of amino acid substitutions around codons 266–268 or 263–268 to associate amino acid motifs with A/B specificity. Amino acids are shown in one-letter symbols. This table was taken from our previous publication (24).

alleles of gorilla and baboon. The results confirmed our previous finding from functional assays that demonstrated the importance of these amino acids for differential sugar specificities of AT/BT. Antigens A and/or B are also expressed in animals other than primates (21). Due to its usefulness for animal experimentation, we characterized the mouse *ABO* gene (22). We showed that the murine gene encodes a *cis*-AB transferase capable of transferring both GalNAc and Gal *in vitro*, although A/B expression was very weak *in vivo*. We also studied the swine AO system (23). Pigs exhibit A or O phenotype. We cloned porcine AT cDNA and demonstrated AT activity. We also showed that the porcine *O* allele lacks most of the structural gene. It was in contrast to human *O* alleles that have minimal alterations, such as a single nucleotide deletion (261delG) (2) or an inactivating single amino acid substitution (G268R) (3).

In the era of genomics, genomes of many species of organisms have been sequenced. And this sequence data is available in public databases, such as Ensembl and Genbank, along with annotations and other useful information. We retrieved the nucleotide and deduced amino acid sequences of *ABO* genes from a variety of species and generated phylogenetic trees to study the evolution of *ABO* genes (24). In addition, we used our expertise in biochemistry, immunology and glycobiology. We previously demonstrated that codons 266 and 268 of human AT and BT are crucial in determining the sugar specificity of GalNAc and Gal (16,17). We also showed that the corresponding codons are also crucial in determining the sugar specificity of the enzymes encoded by the *ABO* genes in primates, mice, and pigs (20,22,23,25). Therefore, we constructed several dozen human AT expression constructs that possess the amino acid substitutions at codons 266 to 268, and also 263 to 268 in some, corresponding to codons in the *ABO* genes of various species. We then examined the specificity and activity. The results are shown in a code table that associates the amino acid sequences with the potential AT/BT specificity/activity (*Figure 5*) (24). Since amino acids at codon 235 were known to affect specificity/activity, there was no guarantee that the specificity/activity assignment using the table was 100% accurate. However, combined with other information, we hoped to gain new insight. In fact, we made several unexpected discoveries. In humans and primates, excluding primordial primates, *A*, *B*, and/or *O* genes are located at the single genetic locus *ABO* as alleles. Therefore, the ABO polymorphism of primates is the result of allelism. However, it was shown that there are species that contain multiple *ABO* genes and/or gene fragments linked in tandem. There are also some species with *ABO* genes in different chromosomal regions and/or different chromosomes (*Figure 6A*). There are even some that have non-allelic *A* and *B* genes. For example, rats can

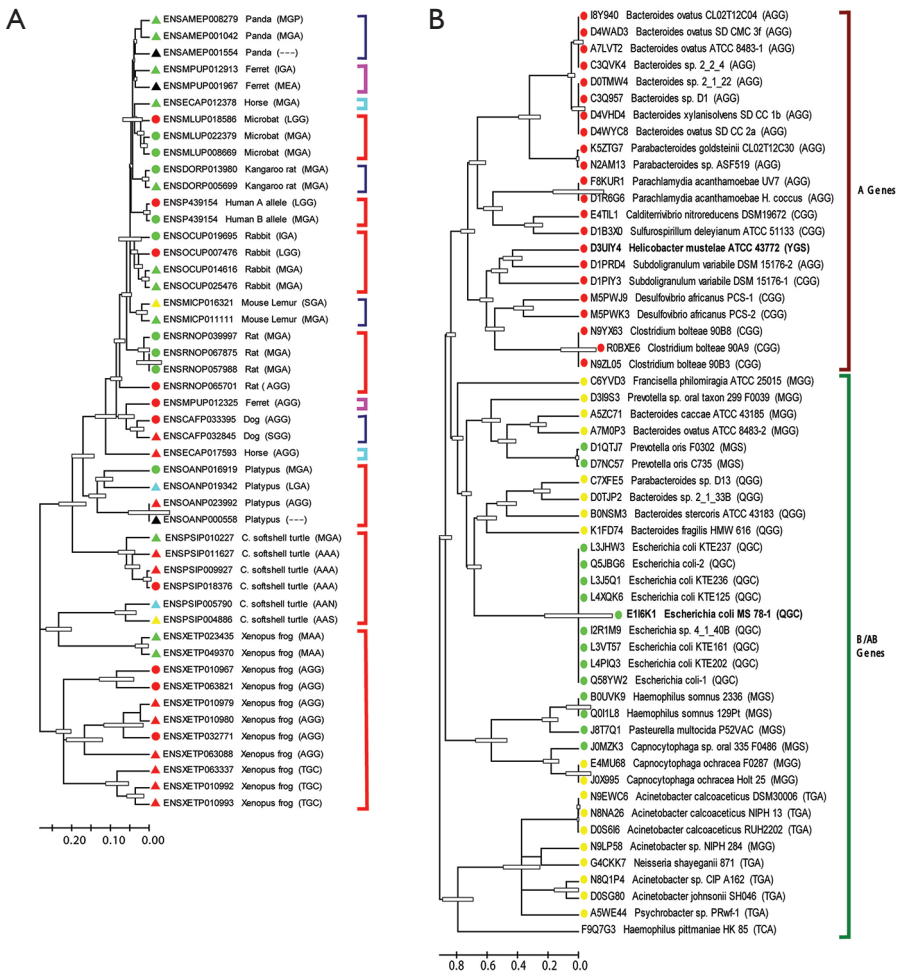


Figure 6 Phylogenetic trees of mammalian and bacterial *ABO* genes. (A) Phylogenetic tree of mammalian *ABO* genes possessing multiple gene copies. Round and triangle symbols indicate seemingly complete and partial gene sequences. The colors red, green, yellow, and blue indicate the deduced A, B, AB, and O specificities, respectively, while the color black indicates "undetermined". Tripeptide sequences corresponding to codons 266–268 of human AT/BT are shown in one-letter symbols in parentheses after the species names. (B) Phylogenetic tree of bacterial *ABO* genes. The A/B specificity is color coded as shown above in (A). These figures were duplicated from our previous publication (24).

exhibit AO polymorphism, having a functional *A* allele or a non-functional *O* allele at the same genetic locus. However, rats also possess additional *B* gene sequences at different chromosomal locations.

We also studied the evolution of bacterial *ABO* genes. Some strains of bacteria were known to express A, B, AB, or O phenotype based on studies conducted by Springer in the 1960s (26). He is famous for his elucidation of the sensitization by A/B antigens expressed in bacteria in the intestinal flora as a possible acquisition of what are called "natural antibodies" (27). In that experiment he demonstrated that chickens grown under sterile conditions did not develop anti-B antibodies, while chickens fed diets contaminated with the O_{86} strain of *Escherichia coli* bacteria that express the B antigen developed anti-B antibodies. Bacterial *B* and *A* genes were cloned from *Escherichia coli* O_{86} strain and *Helicobacter mustelae*, respectively, in 2005 and 2008 (28,29). We used the deduced amino acid sequences of the bacterial *ABO* genes from a few dozen bacterial genome sequences deposited in databases, and generated the phylogenetic trees. We then assigned the specificity of AT/BT based on the motif of the amino acid sequence, using

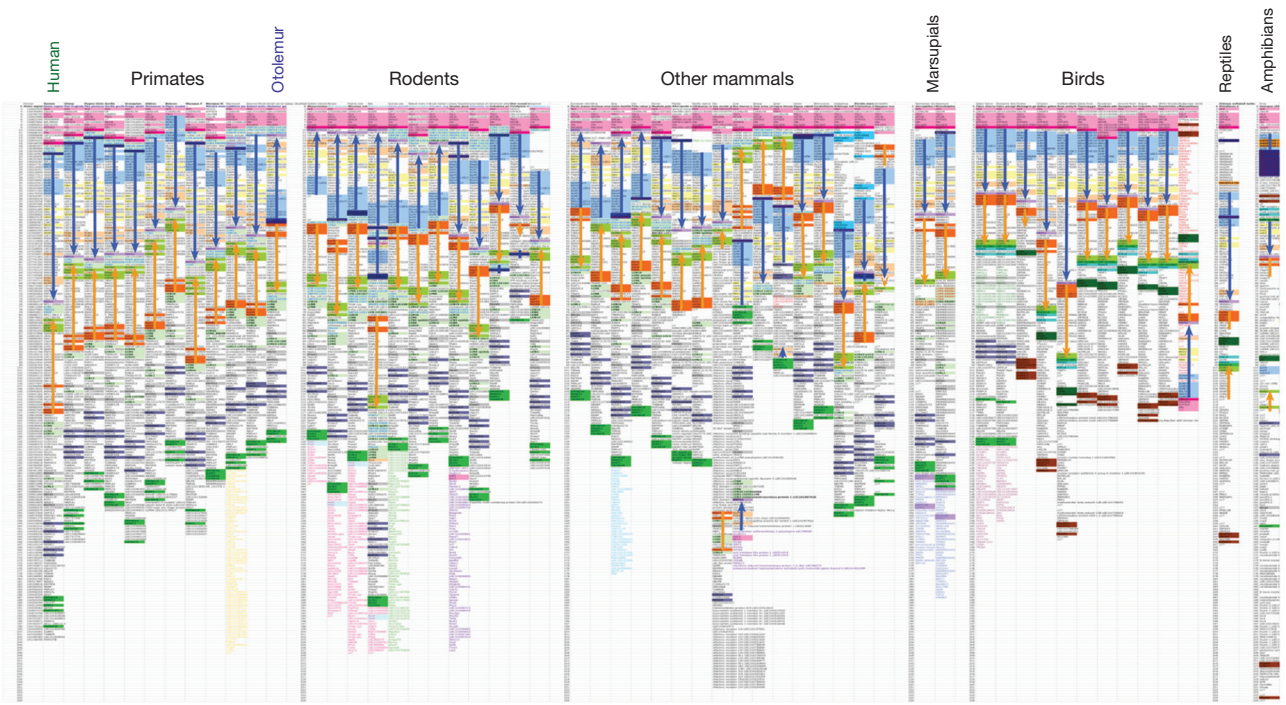


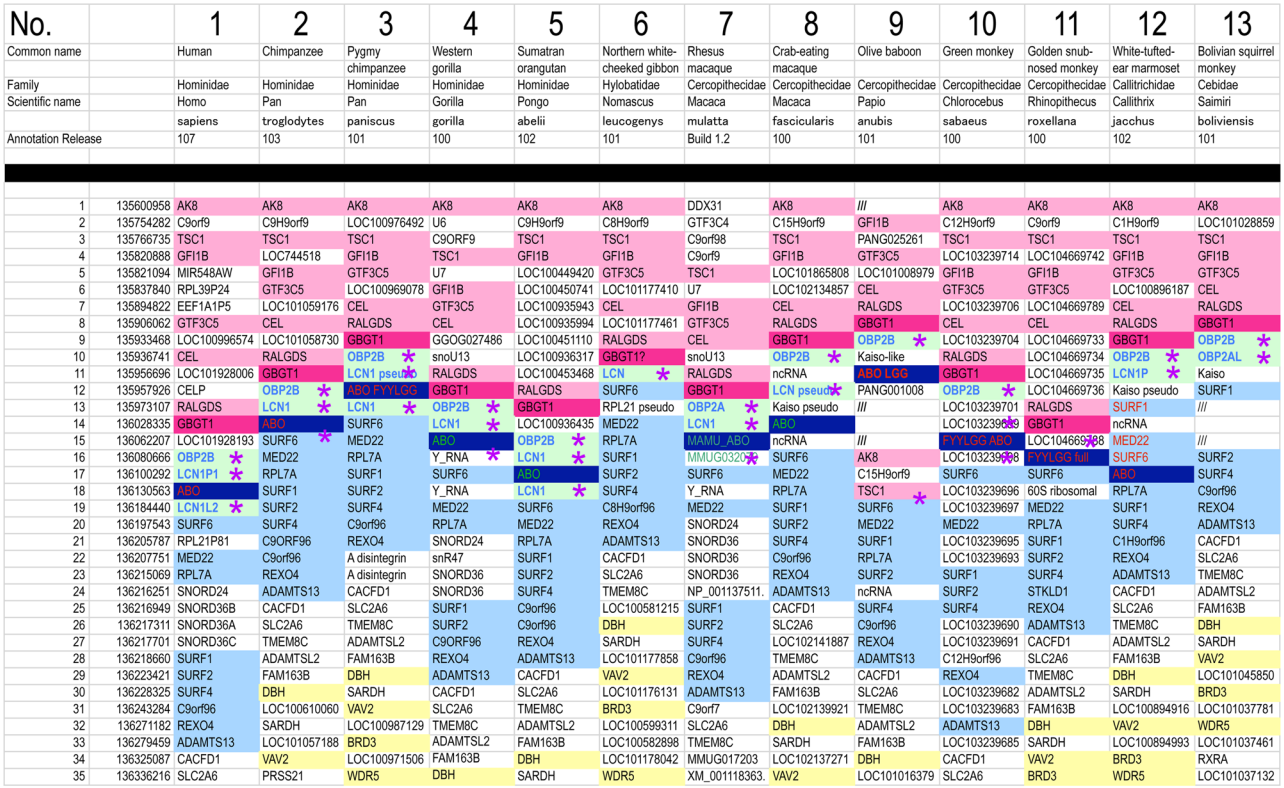
Figure 7 Orientations of two chromosomal fragments. The blue and orange arrows indicate the directions of two chromosomal fragments. Various combinations were observed, possibly due to translocations/inversions that occurred during the evolution of the species. The *ABO* and *GBGT1* genes are found at the ends of the fragments. The figure was duplicated from our previous publication (35).

our code table. Bacterial *ABO* genes formed two separate clusters, one with *A* genes and the other with *B/AB* genes, showing vertical transmission (24). However, *Bacteroides* genes were found in both clusters, implying that horizontal transmission may have occurred in the evolution of the bacterial *ABO* gene. The results are shown in *Figure 6B*.

Generation of the species-dependent ABO polymorphism

The repertoire of the ABO polymorphism varies, depending on the species, as mentioned above. Primates that exclude primordials have single gene ABO polymorphism, while several other species, such as rats, have multiple gene ABO polymorphism. We looked for the molecular mechanisms that generated changes in the number of genes and that converted the multigenic polymorphism into unigenic polymorphism. In humans, the *ABO* gene is mapped together with an evolutionarily related *GBGT1* gene on chromosomal region 9q34 (human GRCh38/hg38: 133,250,401–133,275,219 for *ABO* and 133,152,948–133,163,945 for *GBGT1*). In some species, functional *GBGT1* genes encode Forssman glycolipid synthases (FSs) that catalyze the biosynthesis of Forssman glycolipid antigen (FORS1) of the FORS blood group system (30). However, the human *GBGT1* gene is not functional (31), having two inactivating amino acid substitutions G235S and Q296R (32), excluding rare individuals showing the A_{pae} phenotype, who have the R296Q reversion (33,34).

We analyzed the genes in the chromosomal regions neighboring the *ABO* and *GBGT1* genes in other species (35). Two chromosomal fragments were found flanking these genes in different combinations of orientations, indicating that chromosomal translocations/inversions occurred during the evolution of the species (*Figure* 7). Importantly, the *ABO* and *GBGT1* genes were found at the ends of the chromosomal fragments. Because genetic alterations, such as duplications and deletions, are frequent at the junctions of chromosomal rearrangement, we hypothesized that this may be the molecular force that has driven the species-dependent divergence of the *ABO* gene. Furthermore, we observed that the orientations of



No.		1	2	3	4	5	6	7	8	9	10	11	12	13
Common name		Human	Chimpanzee	Pygmy chimpanzee	Western gorilla	Sumatran orangutan	Northern white-cheeked gibbon	Rhesus macaque	Crab-eating macaque	Olive baboon	Green monkey	Golden snub-nosed monkey	White-tufted-ear marmoset	Bolivian squirrel monkey
Family		Hominidae	Hominidae	Hominidae	Hominidae	Hominidae	Hylobatidae	Cercopithecidae	Cercopithecidae	Cercopithecidae	Cercopithecidae	Cercopithecidae	Callitrichidae	Cebidae
Scientific name		Homo sapiens	Pan troglodytes	Pan paniscus	Gorilla gorilla	Pongo abelii	Nomascus leucogenys	Macaca mulatta	Macaca fascicularis	Papio anubis	Chlorocebus sabaeus	Rhinopithecus roxellana	Callithrix jacchus	Saimiri boliviensis
Annotation Release		107	103	101	100	102	101	Build 1.2	100	101	100	100	102	101
1	135600958	AK8	AK8	AK8	AK8	AK8	AK8	DDX31	AK8	///	AK8	AK8	AK8	AK8
2	135754282	C9orf9	C9H9orf9	LOC100976492	U6	C9H9orf9	C8H9orf9	GTF3C4	C15H9orf9	GFI1B	C12H9orf9	C9orf9	C1H9orf9	LOC101028859
3	135766735	TSC1	TSC1	TSC1	C9ORF9	TSC1	TSC1	C9orf98	TSC1	PANG025261	TSC1	TSC1	TSC1	TSC1
4	135820888	GFI1B	LOC744518	GFI1B	TSC1	GFI1B	GFI1B	C9orf9	GFI1B	GTF3C5	LOC103239714	LOC104669742	GFI1B	GFI1B
5	135821094	MIR548AW	GFI1B	GTF3C5	U7	LOC100449420	GTF3C5	TSC1	LOC101865808	LOC101008979	GFI1B	GFI1B	GTF3C5	GTF3C5
6	135837840	RPL39P24	GTF3C5	LOC100969078	GFI1B	LOC100450741	LOC101177410	U7	LOC102134857	CEL	GTF3C5	GTF3C5	LOC100896187	CEL
7	135894822	EEF1A1P5	LOC101059176	CEL	GTF3C5	LOC100935943	CEL	GFI1B	CEL	RALGDS	LOC103239706	LOC104669789	CEL	RALGDS
8	135906062	GTF3C5	CEL	RALGDS	CEL	LOC100935994	LOC101177461	GTF3C5	RALGDS	GBGT1	CEL	CEL	RALGDS	GBGT1
9	135933468	LOC100996574	LOC101058730	GBGT1	GGOG027486	LOC100451110	RALGDS	CEL	GBGT1	OBP2B *	LOC103239704	LOC104669733	GBGT1	OBP2B *
10	135936741	CEL	RALGDS	OBP2B *	snoU13	LOC100936317	GBGT1?	snoU13	OBP2B *	Kaiso-like	RALGDS	LOC104669734	OBP2B *	OBP2AL *
11	135956696	LOC101928006	GBGT1	LCN1 pseudo *	RALGDS	LOC100453468	LCN *	RALGDS	ncRNA	ABO LGG	GBGT1	LOC104669735	LCN1P *	Kaiso
12	135957926	CELP	OBP2B *	ABO FYLGG	GBGT1	RALGDS	SURF6	GBGT1	LCN pseudo *	PANG001008	OBP2B *	LOC104669736	Kaiso pseudo	SURF1
13	135973107	RALGDS	LCN1 *	LCN1 *	OBP2B *	GBGT1	RPL21 pseudo	OBP2A *	Kaiso pseudo	///	LOC103239701	RALGDS	SURF1	///
14	136028335	GBGT1	ABO	SURF6	LCN1 *	LOC100936435	MED22	LCN1 *	ABO		LOC103239[illegible]	GBGT1	ncRNA	
15	136062207	LOC101928193	SURF6 *	MED22	ABO	OBP2B *	RPL7A	MAMU_ABO	ncRNA	///	FYLGG ABO	LOC104669[illegible]	MED22	///
16	136080666	OBP2B *	MED22	RPL7A	Y_RNA *	LCN1 *	SURF1	MMUG0320[illegible] *	SURF6	AK8	LOC103239[illegible]	FYLGG full	SURF6	SURF2
17	136100292	LCN1P1 *	RPL7A	SURF1	SURF6	ABO	SURF2	SURF6	MED22	C15H9orf9	SURF6	SURF6	ABO	SURF4
18	136130563	ABO	SURF1	SURF2	Y_RNA	LCN1 *	SURF4	Y_RNA	RPL7A	TSC1 *	LOC103239696	60S ribosomal	RPL7A	C9orf96
19	136184440	LCN1L2 *	SURF2	SURF4	MED22	SURF6	C8H9orf96	MED22	SURF1	SURF6	LOC103239697	MED22	SURF1	REXO4
20	136197543	SURF6	SURF4	C9orf96	RPL7A	MED22	REXO4	SNORD24	SURF2	MED22	MED22	RPL7A	SURF4	ADAMTS13
21	136205787	RPL21P81	C9ORF96	REXO4	SNORD24	RPL7A	ADAMTS13	SNORD36	SURF4	SURF1	LOC103239695	SURF1	C1H9orf96	CACFD1
22	136207751	MED22	C9orf96	A disintegrin	snR47	SURF1	CACFD1	SNORD36	C9orf96	RPL7A	LOC103239693	SURF2	REXO4	SLC2A6
23	136215069	RPL7A	REXO4	A disintegrin	SNORD36	SURF2	SLC2A6	SNORD36	REXO4	SURF2	SURF1	SURF4	ADAMTS13	TMEM8C
24	136216251	SNORD24	ADAMTS13	CACFD1	SNORD36	SURF4	TMEM8C	NP_001137511.	ADAMTS13	ncRNA	SURF2	STKLD1	CACFD1	ADAMTSL2
25	136216949	SNORD36B	CACFD1	SLC2A6	SURF1	C9orf96	LOC100581215	SURF1	CACFD1	SURF4	SURF4	REXO4	SLC2A6	FAM163B
26	136217311	SNORD36A	SLC2A6	TMEM8C	SURF2	C9orf96	DBH	SURF2	SLC2A6	C9orf96	LOC103239690	ADAMTS13	TMEM8C	DBH
27	136217701	SNORD36C	TMEM8C	ADAMTSL2	C9ORF96	REXO4	SARDH	SURF4	LOC102141887	REXO4	LOC103239691	CACFD1	ADAMTSL2	SARDH
28	136218660	SURF1	ADAMTSL2	FAM163B	REXO4	ADAMTS13	LOC101177858	C9orf96	TMEM8C	ADAMTS13	C12H9orf96	SLC2A6	FAM163B	VAV2
29	136223421	SURF2	FAM163B	DBH	ADAMTS13	CACFD1	VAV2	REXO4	ADAMTSL2	CACFD1	REXO4	TMEM8C	DBH	LOC101045850
30	136228325	SURF4	DBH	SARDH	CACFD1	SLC2A6	LOC101176131	ADAMTS13	FAM163B	SLC2A6	LOC103239682	ADAMTSL2	SARDH	BRD3
31	136243284	C9orf96	LOC100610060	VAV2	SLC2A6	TMEM8C	BRD3	C9orf7	LOC102139921	TMEM8C	LOC103239683	FAM163B	LOC100894916	LOC101037781
32	136271182	REXO4	SARDH	LOC100987129	TMEM8C	ADAMTSL2	LOC100599311	SLC2A6	DBH	ADAMTSL2	ADAMTS13	DBH	VAV2	WDR5
33	136279459	ADAMTS13	LOC101057188	BRD3	ADAMTSL2	FAM163B	LOC100582898	TMEM8C	SARDH	FAM163B	LOC103239685	SARDH	LOC100894993	LOC101037461
34	136325087	CACFD1	VAV2	LOC100971506	FAM163B	DBH	LOC101178042	MMUG017203	LOC102137271	DBH	CACFD1	VAV2	BRD3	RXRA
35	136336216	SLC2A6	PRSS21	WDR5	DBH	SARDH	WDR5	XM_001118363.	VAV2	LOC101016379	SLC2A6	BRD3	WDR5	LOC101037132

Figure 8 Lipocalin genes (*LCN1/3/4*) neighboring the *ABO* genes in many primates. The chromosomal regions surrounding the *ABO* and *GBGT1* genes are shown. The *LCN1/3/4* gene sequences are indicated with purple asterisks. The figure was duplicated from our previous publication (35).

these two chromosome fragments are the same in primates, excluding *Otolemur garnettii*. Closer examination showed that lipocalin genes (*LCN1/3/4*) are located on both sides of the *ABO* gene in most primate species (*Figure 8*). We therefore reasoned that the recombination events flanking these genes may have played an important role in the establishment of *ABO* alleles in primates from non-allelic *ABO* genes, and that the ABO polymorphism has been inherited in a *trans*-species manner since then.

ABO association with diseases observed by genome-wide association studies (GWAS)

ABO-incompatible blood transfusion can lead to hemagglutination and lysis of RBCs, kidney failure, and occasional death of the recipients. ABO-incompatible cell/tissue/organ transplantation can lead to acute rejection. However, these are reactions against artificial medical practices. Although ABO-incompatible pregnancy can cause HDFN, symptoms are mild and do not generally require any treatment. Since the discovery of ABO, dozens of diseases have been associated with the ABO polymorphism (36). In most of these studies, statistics were used to assess the significance of the association, comparing the ABO distributions between the diseased population and the corresponding healthy population. However, the selection of healthy population may have the potential to introduce bias. After the Human Genome Project determined the nucleotide sequence of the first human genome in 2003, many human genomes have been sequenced. This resulted in the identification of hundreds to thousands of SNVs in the *ABO* gene as mentioned above, as well as millions of SNVs throughout the entire genome. Recent advances in genomics have made it possible to conduct GWAS that examine the association of numerous SNVs distributed in more than 25,000 genes throughout the human genome with specific diseases and/or biological parameters. In contrast to the *ABO* gene-driven approach whereby the

Diseases (OR and P value)					
Venous thromboembolism	1.55	2E-52	Cerebral malaria	1.48	4E-21
Coronary artery disease	1.10	4E-14	Duodenal ulcer	1.32	1E-10
Graves' disease	1.14	2E-10	Pancreatic cancer	1.20	5E-8
Type 2 diabetes	1.05	4E-12	COVID-19	1.32	5E-8
Biological parameters (beta-coefficient and P value)					
CEA	0.21	7E-105	ICAM-1	17.30	3E-91
E-Selectin	9.71	2E-82	ALP	0.31	4E-59
vWF	0.56	5E-57	Galectin-3	0.07	4E-47
P-Selectin	14.00	2E-41	TNFA	NR	7E-40
FVIII	0.46	2E-25	IL-6	NR	9E-25
Cholesterol (LDL)	2.05	8E-22	Cholesterol (total)	2.30	9E-21
EA, RBCC	0.02	9E-18	Campesterol	8.00	9E-13
TSH	0.06	4E-10	Hb	0.09	1E-11
Ht	0.08	6E-10	Angiotensin-C enzyme	4.9	3E-8

Figure 9 Diseases and biological parameters associated with SNVs in the *ABO* gene by GWAS. Representative diseases and biological parameters found associated by GWAS with SNVs in the *ABO* gene are shown, along with the odds ratio (OR), β-coefficient and the P values. CEA, carcinoembryonic antigen; vWF, von Willebrand factor; FVIII, factor VIII; EA, European ancestry; RBCC, red blood cell count; TSH, thyroid-stimulating hormone; Ht, hematocrit; ICAM-1, intercellular adhesion molecule 1; ALP, alkaline phosphatase; TNFA, tumor necrosis factor alpha; IL-6, interleukin 6; Hb, hemoglobin; SNVs, single nucleotide variants; GWAS, genome-wide association studies.

association between ABO polymorphism and a given disease is investigated, GWAS identifies anonymous SNVs with high association and is therefore more objective with less chance of introducing bias.

Figure 9 shows some representative diseases and biological parameters that were found associated with SNVs in the *ABO* gene. Diseases include pancreatic cancer, venous thromboembolism (VTE), coronary artery disease (CAD), severe cerebral malaria, Graves' disease, stomach/duodenal ulcers, type 2 diabetes, and coronavirus disease 2019 (COVID-19). Serum levels of more than a dozen soluble glycoproteins have been associated with SNVs in the *ABO* gene, including von Willebrand factor (vWF) and coagulation factor VIII (FVIII) (37), intercellular adhesion molecule 1 (ICAM-1) (38), tumor necrosis factor alpha (TNF-alpha) (although assay specific) (39), alkaline phosphatase (40,41), E-selectin (42,43), P-selectin (44) and angiotensin-converting enzyme (45).

The association of ABO polymorphism to disease can be direct or indirect. The latter may be mediated by differential concentrations of factors. In this review, only cardiovascular diseases including VTE, malaria, ulcer and COVID-19 are detailed below, restricted to diseases whose molecular mechanisms of the ABO association have been determined in some way by functional analysis.

A differential plasma level of coagulation FVIII causing differential susceptibility to VTE

The *ABO* locus was found associated with VTE and CAD by GWAS (46-48). In the VTE study, the authors analyzed 317,000 SNVs in 453 VTE cases and 1,327 controls, and identified three SNVs with significant association (P value below 1.7×10^{-7}). One SNV was located at the coagulation factor V (*FV*) locus ($P=8.1\times10^{-10}$) and the other two (rs505922 and rs657152) were located at the *ABO* locus ($P=1.5\times10^{-14}$ and 2.2×10^{-13}, respectively) (46). ABO genotyping of additional 1,700 VTE cases and 1,400 controls was performed, analyzing three additional SNVs that specify *A2* (rs8176750), *B* (rs8176746), and *O* (rs8176719) alleles, which showed that individuals in groups O and A2 have a lower risk of VTE. The *ABO* gene has also been associated with myocardial infarction in the presence of coronary atherosclerosis (47). Eleven SNVs with the highest associations were mapped at the *ABO* locus. The odds ratio (OR) for A/B/AB *vs.* O was 1.44. A meta-analysis of 14 GWAS studies of CAD comprising 22,233 cases and 64,762 controls of European descent, followed by genotyping of top association signals in 56,682 additional individuals, identified SNV rs579459 at the *ABO* locus as having the 5th highest association with an OR of 1.1 (48).

The different rates of synthesis, secretion and

clearance of vWF lead to differential plasma levels of vWF. Because FVIII is stable in blood only when it binds to vWF, the plasma concentration of vWF dictates the plasma concentration of FVIII. And the major determinant of the plasma levels of these glycoproteins is the ABO polymorphism, and individuals in group O have levels approximately 25% lower than those in group A (49). Consequently, the molecular cause of the ABO association with cardiovascular disease can be explained, at least partially, by ABO-dependent differential plasma concentrations of vWF and FVIII. A high level of plasma FVIII may result in an increased risk of ischemic heart disease and VTE in the non-O groups, while a low level may cause excessive bleeding in the O group (50,51). In addition to vWF and FVIII, some other factors showing ABO dependence may also be involved in susceptibility to cardiovascular disease.

Stronger adhesion to capillaries of group A erythrocytes infected with *Plasmodium falciparum* parasites than group O erythrocytes infected with parasites

In 1967, Athreya and Coriell reported that group B confers a selective advantage over malaria infection (52). Since then, many articles have been published linking ABO phenotypes with susceptibility to malaria. It was even proposed that malaria had played an important role in shaping the current distribution of ABO polymorphism in the world (53). However, the association was not evident in some studies. Critically analyzing the literature that reported the association/non-association between ABO and *Plasmodium falciparum* malaria, Cserti and Dzik showed that individuals in group O tend to exhibit a favorable outcome compared to individuals in group A (54). In 2008, using SNVs rs8176719 (A/B *vs.* O) and rs8176746 (A/O *vs.* B), the association between severe malaria and SNVs at the *ABO* locus was examined (55). The *A/B* alleles that produce functional transferases were associated with an increased risk of severe malaria compared to the *O* allele with 261delG (OR =1.18, $P=2\times10^{-7}$). The following year, a GWAS study on severe malaria was published (56). In that study, 19 genetic loci exhibiting a significant association with the threshold of $P<10^{-4}$ were identified, including β-hemoglobin gene (*HBB*). The signal peak in *HBB* coincided with the position of the causative variant of S-hemoglobin (HbS). In that study, none of the SNVs in the *ABO* gene were associated with malaria. However, later studies associated severe malaria with the *ABO* gene (57-60).

RBCs infected with *P. falciparum* roll on and adhere to the microvascular endothelium and then disappear from the circulation (sequestration) to form aggregates (rosettes) with uninfected RBCs and/or platelets. By obstructing blood microcirculation and reducing the supply of oxygen and substrates, rosette formation results in cerebral malaria (61). There is experimental evidence showing that these processes depend on ABO polymorphism. A higher binding affinity of non-O group erythrocytes than O group is reflected in larger rosettes (62). The terminal mono- and tri-saccharides of A and B antigens, H disaccharide, and fucose specifically inhibited rosettes formation of group A/B RBCs infected with the parasites. Selective enzymatic digestion of the A antigen from the surface of uninfected RBCs totally eliminated the parasite's preference to form rosettes with these RBCs (63,64). The *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) encoded by the parasite gene mediates this adhesion to interact with a repertoire of human host proteins. The semi-conserved head structure of the parasite protein was shown to bind A antigen (65).

Greater susceptibility of group O individuals to ulcers by more abundant Lewis b expressed on the duodenal epithelium

GWAS identified two susceptibility loci: one in the *PSCA* gene encoding the prostate stem cell antigen in chromosomal region 8q24 and the other in the *ABO* gene at 9q34, analyzing a total of 7,035 individuals with duodenal ulcer and 25,323 controls in Japan (66). SNV rs505922 in *ABO* was associated with duodenal ulcer in a recessive model (OR =1.32; $P=1.15\times10^{-10}$). In fact, the ABO and ulcer association was one of the first associations identified using a targeted approach, in which individuals in group O were shown to have a higher susceptibility (67). The relative incidences reported were 0.73 (A *vs.* O) and 0.80 (B *vs.* O) for duodenal ulcers and 0.87 for gastric ulcers (both A *vs.* O and B *vs.* O). In 1984, Marshall demonstrated that infection with the bacterium *Helicobacter pylori* causes gastritis and ulceration of the stomach/duodenum and that patients can be cured of peptic ulcer by eradicating the bacteria with antibiotics and acid secretion inhibitors (68). Subsequently, the relative risk of non-secretors/secretors was calculated to be 1.9 (69). In 1993, Boren demonstrated that fucosylated H type 1 and Lewis b (Le^b) antigens mediate the attachment of *H. pylori* to the human gastric mucosa and that the

soluble glycoproteins that present Le^b or antibodies against the Le^b antigen inhibit bacterial binding (70). Furthermore, the conversion of Le^b to ALe^b by the addition of a terminal GalNAc by AT decreased bacterial binding. Based on the observation, the lower availability of Le^b receptors in non-O groups individuals was assumed to be responsible for the reduced infectivity. However, subsequent studies showed that different *H. pylori* strains exhibit different sugar specificity, complicating understanding of the interactions between host blood group antigens and bacterial blood group antigen-binding adhesion (BabA). In fact, it was later discovered that more than 95% of the strains bound to fucosylated blood group antigens without showing a preference for group O. Instead, these generalists bind to A and B antigens, as well as H substance. Only 5% of the strains specifically bind to H substances. However, these specialists occupy a higher percentage (60%) among the Amerindian strains of South America (71). The specialization of *H. pylori* appears to have coincided with the unique predominance of group O in Amerindians. Selection cycles to increase or decrease bacterial adherence were proposed to contribute to BabA diversity and to replace generalists by specialists gradually. In other words, positive selection for pathogens due to the abundance of a specific host population group was suggested, as opposed to negative selection for a specific host population group caused by infectious agents.

ABO blood groups and SARS-CoV-2 infectivity/ COVID-19 progression

Unfortunately, we are in the midst of a COVID-19 pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Although most infected people have a few or no symptoms, some progress to severe pneumonia, multiple organ failure, and even death (72). Older people are at high risk, and eighty percent of deaths have occurred in people with at least one underlying comorbidity, particularly cardiovascular disease/ hypertension, overweight/obesity, and diabetes (73).

A GWAS study of COVID-19 has recently been published (74). More than 8 million SNVs of 835 patients with respiratory failure due to severe COVID-19 and 1,255 control participants from Italy and 775 patients and 950 control participants from Spain were analyzed. Two chromosomal regions with significant associations were identified at 3p21.31 and 9q34.2. Furthermore, the frequency of risk alleles for lead variants was higher in the more severely ill and mechanically ventilated patients compared to the less severely ill patients who received only supplemental oxygen. The association in 9q34.2 was mapped to the *ABO* gene. A higher and lower risk was calculated for individuals in group A (OR =1.45, $P=1.48\times10^{-4}$) and group O (OR =0.65, $P=1.06\times10^{-5}$) by group-specific analysis. A more recent GWAS study of critical illness in COVID-19 has identified several additional genetic loci with a significant association (75).

Prior to the GWAS work, the ABO association to COVID-19 was reported in a couple of manuscripts posted on medRxiv, the preprint server for health sciences (76-78). In the article recently published in the *British Journal of Hematology*, Li compared ABO blood group distribution among 265 SARS-CoV-2 infected patients and 3,694 healthy controls. A significantly higher proportion of group A individuals was found in patients than in healthy controls (39.3% *vs.* 32.3%, P=0.017), while the proportion of group O individuals was significantly lower in patients (25.7% *vs.* 33.8%, P<0.01) (79). This trend was also observed in other studies (75-78,80,81). Furthermore, genetic testing company 23andMe posted online preliminary unpublished data from its ongoing COVID-19 study (82). It analyzed the genetic and survey data of more than 750,000 participants, and calculated the percentages of individuals with different ABO groups who reported COVID-19: 1.3%, 1.4%, 1.5%, and 1.5% for groups O, A, B, and AB, respectively, among all participants, and 3.2%, 3.9%, 4.0% and 4.1% among health professionals. A statistically significant protective effect of group O was observed (OR =0.86, P<0.0001 against acquisition, OR =0.81, P=0.05 hospitalization) in the entire population and also among health professionals (OR =0.81, P<0.0001 against acquisition). This new approach based on mass genome analysis and survey questions has proven very powerful in identifying significant genetic associations.

Inhibition of SARS-CoV-2 infection by natural antibodies

SARS-CoV-2 exhibits a broad organ tropism, infecting and proliferating in the epithelial cells of the respiratory and digestive tracts where the A and/or B antigens are expressed depending on the ABO phenotype of the individual. SARS-CoV-2 is membrane encapsulated and its infection in human cells is mediated by the binding of viral Spike (S) glycoproteins embedded in the membrane with angiotensin-converting enzyme 2 (ACE2) receptors present on the cell surface. S proteins can carry A and/or B glycan antigens,

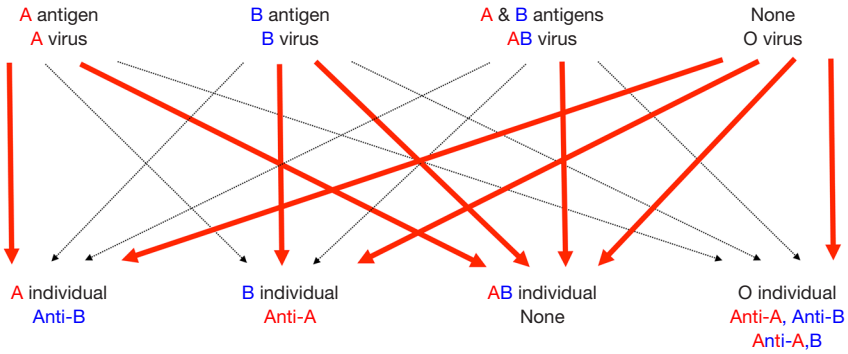


Figure 10 ABO-dependent differential inhibition of SARS-CoV-2 infection. With and without inhibition of the SARS-CoV-2 infection by natural antibodies is shown schematically with dotted black and solid red arrows, respectively. This figure was modified from our previous publication (87).

reflecting the ABO phenotype of the cells where viruses are produced. In an experimental cellular model of SARS (and not SARS-CoV-2), mouse monoclonal or human polyclonal anti-A antibodies were shown to partially inhibit the physical interaction between viral S proteins carrying A antigens and cellular ACE2 proteins (83). Taking into account that the ABO blood group polymorphism was shown to influence susceptibility to SARS (highest and lowest risk for individuals in groups A and O, respectively) (84), it was suggested that this molecular mechanism is responsible for the ABO-dependent differential susceptibility to SARS. In fact, similar experimental observations have been reported of HIV and measles viruses expressing A or B antigens (85,86).

The infectivity of SARS-CoV-2 among individuals with various ABO phenotypes is shown schematically in *Figure 10*. On the one hand, viruses produced in individuals of groups A, B, AB, and O can express A, B, A and B antigens, and none, respectively. On the other hand, individuals in groups A, B, AB, and O have anti-B, anti-A, none, and anti-A/anti-B/anti-A,B antibodies, respectively, following Landsteiner's law. Thus, like "matched" and "mismatched" blood transfusion, these antibodies are expected to react with corresponding antigens on viral particles and inhibit, at least partially, interpersonal infection in certain combinations (88). For example, SARS-CoV-2 viruses produced in cells of group A individuals (group A viruses) can express A antigens and infect group A or AB individuals without such antigen-antibody reactions, as shown by red arrows. However, infection of group B or O individuals who possess anti-A antibodies may be somewhat inhibited (as shown by black dotted arrows). The important thing here is that the inhibition is directional and that it may or may not be 100% efficient. Therefore, infection can occur even if the corresponding antibodies are present. More importantly, once infection is established, newly produced SARS-CoV-2 viruses exhibit the same ABO phenotype as the infected individual, and these antibodies no longer inactivate them. In other words, natural antibodies are only relevant for the prevention of initial attacks, but irrelevant for the subsequent productive infection. Ironically, more protected group O individuals can produce group O SARS-CoV-2 viruses that are capable of effectively infecting individuals with any ABO phenotypes, including group O, without the protection of natural antibodies. In other words, having the group O phenotype does not protect against group O viruses. This protection works better in the ABO-heterogeneous population than in ABO-homogeneous populations. As a result, countries with the highest frequency of O phenotype, such as Ecuador (75%) and Peru (70%), also suffer from the COVID-19 pandemic. Natural antibodies of the IgA class may be primarily responsible for mucosal immunity, although natural antibodies of other classes, especially anti-A,B IgG, may also function. The latter class of antibodies may explain a greater protective effect reported in group O than in group B ($P<0.001$) (89).

In addition to natural antibodies that affect the infectivity of SARS-CoV-2, the ABO polymorphism can indirectly affect the progression of COVID-19. As mentioned above, individuals in group O have 25% lower plasma levels of vWF and FVIII, and have a lower risk of thrombosis, pulmonary embolism, and VTE (46). Most of these glycoproteins are produced in vascular endothelial cells where A and/or B antigens are expressed, depending

on the ABO phenotype of the individual. Because severe COVID-19 can dysregulate vascular tone and permeability and induce cytokine storms and redox stress, the ABO polymorphism may influence COVID-19 differently. However, the progression of the disease also depends on other factors, several of which are more relevant than the ABO polymorphism, with much higher ORs. For example, the ORs for people over the age of 85 are 13 and 630 for hospitalization and death, respectively, compared to the young adults aged between 18 and 29 years (73). In this circumstance, the effects of ABO on disease severity and mortality can be easily masked. Recently, a couple of articles have been published that reported a question about the association between ABO and COVID-19 (90,91). For example, Boudin compared ABO blood group distributions in 1,279 crew members confirmed/suspected of being infected with SARS-CoV-2 and 409 crew members who were exposed but not infected at the same time and in the same location on a French aircraft carrier (92,93). Univariate analysis did not show a statistically significant association between ABO polymorphism and SARS-CoV-2 infection, although a slightly higher distribution was observed for group A (40.7% *vs.* 37.4%) and lower for group O (43.2% *vs.* 46.2%) in the infected population. It is one thing that the ABO blood group polymorphism inhibits SARS-CoV-2 infection and affects the progression of COVID-19. Another is whether or not these effects are actually observed in a population.

Conclusions

Recent advances in molecular genetic/genomic analysis of the ABO system have answered many important biological and medical questions. These studies have also contributed to the development of ABO genotyping and the unbiased identification of diseases associated with SNVs in the *ABO* gene by GWAS, which are beneficial for clinical work. There are still a variety of problems to solve, as well as technical and economical limitations to overcome. However, future studies are expected to open a novel venue in our better understanding of the ABO blood group system and useful clinical translations.

Acknowledgments

The author acknowledges the efforts of many who have contributed to the advanced understanding of the molecular genetic/genomic basis of the ABO blood group system. He apologizes for not citing many relevant studies due to space limitations.
Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the Guest Editor (Yann Fichou) for the series "Molecular Genetics and Genomics of Blood Group Systems" published in *Annals of Blood*. The article has undergone external peer review.

Conflicts of Interest: The author has completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/aob-20-71). The series "Molecular Genetics and Genomics of Blood Group Systems" was commissioned by the editorial office without any funding or sponsorship. The author has no other conflicts of interest to declare.

Ethical Statement: The author is accountable for all aspects of the manuscript and ensure that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Yamamoto F, Marken J, Tsuji T, et al. Cloning and characterization of DNA complementary to human UDP-GalNAc: Fuc alpha 1->2Gal alpha 1->3GalNAc transferase (histo-blood group A transferase) mRNA. J Biol Chem 1990;265:1146-51.
2. Yamamoto F, Clausen H, White T, et al. Molecular genetic basis of the histo-blood group ABO system. Nature 1990;345:229-33.
3. 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.
4. Yamamoto F, McNeill PD, Hakomori S. Human

histo-blood group 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.
5. Yamamoto F, McNeill PD, Yamamoto M, et al. Molecular genetic analysis of the ABO blood group system: 1. Weak subgroups: A3 and B3 alleles. Vox Sang 1993;64:116-9.
6. Yamamoto F, McNeill PD, Yamamoto M, et al. Molecular genetic analysis of the ABO blood group system: 3. A(X) and B(A) alleles. Vox Sang 1993;64:171-4.
7. Yamamoto F, McNeill PD, Kominato Y, et al. Molecular genetic analysis of the ABO blood group system: 2. *cis*-AB alleles. Vox Sang 1993;64:120-3.
8. Yamamoto F. Molecular genetics of the ABO histo-blood group system. Vox Sang 1995;69:1-7.
9. Yamamoto F. Molecular genetics of ABO. Vox Sang 2000;78 Suppl 2:91-103.
10. Yamamoto F, McNeill PD, Hakomori S. Genomic organization of human histo-blood group ABO genes. Glycobiology 1995;5:51-8.
11. Yamamoto F. Cloning and regulation of the ABO genes. Transfus Med 2001;11:281-94.
12. Yip SP. Sequence variation at the human ABO locus. Ann Hum Genet 2002;66:1-27.
13. Yamamoto F. Review: ABO blood group system--ABH oligosaccharide antigens, anti-A and anti-B, A and B glycosyltransferases, and ABO genes. Immunohematology 2004;20:3-22.
14. Yamamoto F, Cid E, Yamamoto M, et al. ABO research in the modern era of genomics. Transfus Med Rev 2012;26:103-18.
15. Blumenfeld OO. Mutation databases and other online sites as a resource for transfusion medicine: history and attributes. Transfus Med Rev 2002;16:103-14.
16. Yamamoto F, Hakomori S. Sugar-nucleotide donor specificity of histo-blood group A and B transferases is based on amino acid substitutions. J Biol Chem 1990;265:19257-62.
17. 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 transferases: *In vitro* mutagenesis study. J Biol Chem 1996;271:10515-20.
18. Moor-Jankowski J, Wiener AS, Rogers CM. Human blood group factors in non-human primates. Nature 1964;202:663-5.
19. Moor-Jankowski J, Wiener AS. Blood group antigens in primate animals and their relation to human blood groups. Primates Med 1969;3:64-77.
20. Kominato Y, McNeill PD, Yamamoto M, et al. Animal histo-blood group ABO genes. Biochem Biophys Res Commun 1992;189:154-64.
21. Oriol R, Candelier JJ, Taniguchi S, et al. Major carbohydrate epitopes in tissues of domestic and African wild animals of potential interest for xenotransplantation research. Xenotransplantation 1999;6:79-89.
22. Yamamoto M, Lin XH, Kominato Y, et al. Murine equivalent of the human histo-blood group ABO gene is a *cis*-AB gene and encodes a glycosyltransferase with both A and B transferase activity. J Biol Chem 2001;276:13701-8.
23. Yamamoto F, Yamamoto M. Molecular genetic basis of porcine histo-blood group AO system. Blood 2001;97:3308-10.
24. Yamamoto F, Cid E, Yamamoto M, et al. An integrative evolution theory of histo-blood group ABO and related genes. Sci Rep 2014;4:6601.
25. Saitou N, Yamamoto F. Evolution of primate ABO blood group genes and their homologous genes. Mol Biol Evol 1997;14:399-411.
26. Springer GF, Williamson P, Brandes WC. Blood group activity of Gram-negative bacteria. J Exp Med 1961;113:1077-93.
27. Springer GF, Horton RE. Blood group isoantibody stimulation in man by feeding blood group-active bacteria. J Clin Invest 1969;48:1280-91.
28. Yi W, Shao J, Zhu L, et al. *Escherichia coli* O86 O-antigen biosynthetic gene cluster and stepwise enzymatic synthesis of human blood group B antigen tetrasaccharide. J Am Chem Soc 2005;127:2040-1.
29. Yi W, Shen J, Zhou G, et al. Bacterial homologue of human blood group A transferase. J Am Chem Soc 2008;130:14420-1.
30. Haslam DB, Baenziger JU. Expression cloning of Forssman glycolipid synthetase: a novel member of the histo-blood group ABO gene family. Proc Natl Acad Sci U S A 1996;93:10697-702.
31. Xu H, Storch T, Yu M, et al. Characterization of the human Forssman synthetase gene. An evolving association between glycolipid synthesis and host-microbial interactions. J Biol Chem 1999;274:29390-8.
32. Yamamoto M, Cid E, Yamamoto F. Molecular genetic basis of the human Forssman glycolipid antigen negativity. Sci Rep 2012;2:975.
33. Svensson L, Hult AK, Stamps R, et al. Forssman expression on human erythrocytes: biochemical and

genetic evidence of a new histo-blood group system. Blood 2013;121:1459-68.
34. Hult AK, McSherry E, Moller M, et al. GBGT1 is allelically diverse but dispensable in humans and naturally occurring anti-FORS1 shows an ABO-restricted pattern. Transfusion 2018;58:2036-45.
35. Yamamoto F. Evolutionary divergence of the ABO and GBGT1 genes specifying the ABO and FORS blood group systems through chromosomal rearrangements. Sci Rep 2017;7:9375.
36. Mourant AE, Kope*c AC, Domaniewska-Sobczak K. Blood groups and diseases: a study of associations of diseases with blood groups and other polymorphisms. Oxford monographs on medical genetics. Oxford; New York: Oxford University Press, 1978.
37. Smith NL, Chen MH, Dehghan A, et al. Novel Associations of Multiple Genetic Loci With Plasma Levels of Factor VII, Factor VIII, and von Willebrand Factor. Circulation 2010;121:1382-92.
38. Paré G, Chasman DI, Kellogg M, et al. Novel association of ABO histo-blood group antigen with soluble ICAM-1: results of a genome-wide association study of 6,578 women. PLoS Genet 2008;4:e1000118.
39. Melzer D, Perry JR, Hernandez D, et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). PLoS Genet 2008;4:e1000072.
40. Yuan X, Waterworth D, Perry JR, et al. Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. Am J Hum Genet 2008;83:520-8.
41. Kamatani Y, Matsuda K, Okada Y, et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat Genet 2010;42:210-5.
42. Paterson AD, Lopes-Virella MF, Waggott D, et al. Genome-wide association identifies the ABO blood group as a major locus associated with serum levels of soluble E-selectin. Arterioscler Thromb Vasc Biol 2009;29:1958-67.
43. Qi L, Cornelis MC, Kraft P, et al. Genetic variants in ABO blood group region, plasma soluble E-selectin levels and risk of type 2 diabetes. Hum Mol Genet 2010;19:1856-62.
44. Barbalic M, Dupuis J, Dehghan A, et al. Large-scale genomic studies reveal central role of ABO in sP-selectin and sICAM-1 levels. Hum Mol Genet 2010;19:1863-72.
45. Chung CM, Wang RY, Chen JW, et al. A genome-wide association study identifies new loci for ACE activity: potential implications for response to ACE inhibitor. Pharmacogenomics J 2010;10:537-44.
46. Trégouët DA, Heath S, Saut N, et al. Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. Blood 2009;113:5298-303.
47. Reilly MP, Li M, He J, et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. Lancet 2011;377:383-92.
48. Schunkert H, Konig IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet 2011;43:333-8.
49. Gill JC, Endres-Brooks J, Bauer PJ, et al. The effect of ABO blood group on the diagnosis of von Willebrand disease. Blood 1987;69:1691-5.
50. O'Donnell J, Laffan MA. The relationship between ABO histo-blood group, factor VIII and von Willebrand factor. Transfus Med 2001;11:343-51.
51. Jenkins PV, O'Donnell JS. ABO blood group determines plasma von Willebrand factor levels: a biologic function after all? Transfusion 2006;46:1836-44.
52. Athreya BH, Coriell LL. Relation of blood groups to infection. I. A survey and review of data suggesting possible relationship between malaria and blood groups. Am J Epidemiol 1967;86:292-304.
53. Weatherall DJ. Genetic variation and susceptibility to infection: the red cell and malaria. Br J Haematol 2008;141:276-86.
54. Cserti CM, Dzik WH. The ABO blood group system and *Plasmodium falciparum* malaria. Blood 2007;110:2250-8.
55. Fry AE, Griffiths MJ, Auburn S, et al. Common variation in the ABO glycosyltransferase is associated with susceptibility to severe *Plasmodium falciparum* malaria. Hum Mol Genet 2008;17:567-76.
56. Jallow M, Teo YY, Small KS, et al. Genome-wide and fine-resolution association analysis of malaria in West Africa. Nat Genet 2009;41:657-65.
57. Timmann C, Thye T, Vens M, et al. Genome-wide association study indicates two novel resistance loci for severe malaria. Nature 2012;489:443-6.
58. Band G, Le QS, Jostins L, et al. Imputation-based meta-analysis of severe malaria in three African populations. PLoS Genet 2013;9:e1003509. Erratum in: PLoS Genet. 2013 Jun;9(6). doi:10.1371/annotation/adc2beaf-4bee-4e22-925b-6788d62fe029. Malaria Genomic Epidemiological Network [corrected to Malaria Genomic Epidemiology Network].
59. Malaria Genomic Epidemiology Network, Band G,

Rockett KA, et al. A novel locus of resistance to severe malaria in a region of ancient balancing selection. Nature 2015;526:253-7.
60. Malaria Genomic Epidemiology Network. Insights into malaria susceptibility using genome-wide data on 17,000 individuals from Africa, Asia and Oceania. Nat Commun 2019;10:5732.
61. White NJ, Ho M. The pathophysiology of malaria. Adv Parasitol 1992;31:83-173.
62. Carlson J, Wahlgren M. *Plasmodium falciparum* erythrocyte rosetting is mediated by promiscuous lectin-like interactions. J Exp Med 1992;176:1311-7.
63. Barragan A, Kremsner PG, Wahlgren M, et al. Blood group A antigen is a coreceptor in *Plasmodium falciparum* rosetting. Infect Immun 2000;68:2971-5.
64. Rowe JA, Handel IG, Thera MA, et al. Blood group O protects against severe *Plasmodium falciparum* malaria through the mechanism of reduced rosetting. Proc Natl Acad Sci U S A 2007;104:17471-6.
65. Chen Q, Heddini A, Barragan A, et al. The semiconserved head structure of *Plasmodium falciparum* erythrocyte membrane protein 1 mediates binding to multiple independent host receptors. J Exp Med 2000;192:1-10.
66. Tanikawa C, Urabe Y, Matsuo K, et al. A genome-wide association study identifies two susceptibility loci for duodenal ulcer in the Japanese population. Nat Genet 2012;44:430-4, S1-2.
67. Aird I, Bentall HH, Mehigan JA, et al. The blood groups in relation to peptic ulceration and carcinoma of colon, rectum, breast, and bronchus; an association between the ABO groups and peptic ulceration. Br Med J 1954;2:315-21.
68. Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet 1984;1:1311-5.
69. Dickey W, Collins JS, Watson RG, et al. Secretor status and *Helicobacter pylori* infection are independent risk factors for gastroduodenal disease. Gut 1993;34:351-3.
70. Borén T, Falk P, Roth KA, et al. Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. Science 1993;262:1892-5.
71. Aspholm-Hurtig M, Dailide G, Lahmann M, et al. Functional adaptation of BabA, the H. pylori ABO blood group antigen binding adhesin. Science 2004;305:519-22.
72. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395:497-506.
73. Prevention CfDCa. Coronavirus Disease 2019 (COVID-19), People at Increased Risk for Severe Illness. Available online: https://wwwcdcgov/coronavirus/2019-ncov/need-extra-precautions/indexhtml (accessed on 20 August 2020) 2020.
74. Severe Covid-19 GWAS Group, Ellinghaus D, Degenhardt F, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N Engl J Med 2020;383:1522-34.
75. Pairo-Castineira E, Clohisey S, Klaric L, et al. Genetic mechanisms of critical illness in COVID-19. Nature 2021;591:92-8.
76. Zhao J, Yang Y, Huang H, et al. Relationship between the ABO blood group and the COVID-19 susceptibility. Clin Infect Dis 2021;73:328-31.
77. Zietz M, Tatonetti NP. Testing the association between blood type and COVID-19 infection, intubation, and death. medRxiv Preprint 2020. doi: 10.1101/2020.04.08.20058073.
78. Zeng X, Fan H, Lu D, et al. Association between ABO blood groups and clinical outcome of coronavirus disease 2019: Evidence from two cohorts. medRxiv Preprint 2020. doi: 10.1101/2020.04.15.20063107.
79. Li J, Wang X, Chen J, et al. Association between ABO blood groups and risk of SARS-CoV-2 pneumonia. Br J Haematol 2020;190:24-7.
80. Hoiland RL, Fergusson NA, Mitra AR, et al. The association of ABO blood group with indices of disease severity and multiorgan dysfunction in COVID-19. Blood Adv 2020;4:4981-9.
81. Barnkob MB, Pottegard A, Støvring H, et al. Reduced prevalence of SARS-CoV-2 infection in ABO blood group O. Blood Adv 2020;4:4990-3.
82. 23andMe. 23andMe finds evidence that blood type plays a role in COVID-19. 2020. Available online: https://blog.23andme.com/andme-research/andme-finds-evidence-that-blood-type-plays-a-role-in-covid-19/ (accessed on 20 August 2020).
83. Guillon P, Ruvoen-Clouet N, Le Moullac-Vaidye B, et al. Association between expression of the H histo-blood group antigen, alpha1,2fucosyltransferases polymorphism of wild rabbits, and sensitivity to rabbit hemorrhagic disease virus. Glycobiology 2009;19:21-8.
84. Cheng Y, Cheng G, Chui CH, et al. ABO blood group and susceptibility to severe acute respiratory syndrome. JAMA 2005;293:1450-1. Erratum in: JAMA. 2005 Aug 17;294(7):794. Cheng, Yufeng [corrected to Cheng, Yunfeng].
85. Arendrup M, Hansen JE, Clausen H, et al. Antibody to histo-blood group A antigen neutralizes HIV produced

by lymphocytes from blood group A donors but not from blood group B or O donors. AIDS 1991;5:441-4.
86. Preece AF, Strahan KM, Devitt J, et al. Expression of ABO or related antigenic carbohydrates on viral envelopes leads to neutralization in the presence of serum containing specific natural antibodies and complement. Blood 2002;99:2477-82.
87. Yamamoto F, Yamamoto M, Muñiz-Diaz E. Blood group ABO polymorphism inhibits SARS-CoV-2 infection and affects COVID-19 progression. Vox Sang 2021;116:15-7.
88. Yamamoto F. ABO blood groups and SARS-CoV-2 infection. ResearchGate 2020. doi: 10.13140/RG.2.2.24833.20329. Available online: https://www.researchgate.net/publication/340209045_ABO_Blood_Groups_and_SARS-CoV-2_Infection
89. Gérard C, Maggipinto G, Minon JM. COVID-19 and ABO blood group: another viewpoint. Br J Haematol 2020;190:e93-4.
90. Latz CA, DeCarlo C, Boitano L, et al. Blood type and outcomes in patients with COVID-19. Ann Hematol 2020;99:2113-8.
91. Dzik S, Eliason K, Morris EB, et al. COVID-19 and ABO blood groups. Transfusion 2020;60:1883-4.
92. Boudin L, Janvier F, Bylicki O, et al. ABO blood groups are not associated with risk of acquiring the SARS-CoV-2 infection in young adults. Haematologica 2020;105:2841-3.
93. Flegel WA. COVID-19: risk of infection is high, independently of ABO blood group. Haematologica 2020;105:2706-8.

doi: 10.21037/aob-20-71
Cite this article as: Yamamoto F. Molecular genetics and genomics of the ABO blood group system. Ann Blood 2021;6:25.