

Primate *ABO* Gene is under Weak Positive Selection

Eliane B. Santos EVANOVICH*, Maria Lúcia HARADA

Universidade Federal do Pará, Laboratório de Biologia Molecular, Departamento de Genética, Centro de Ciências Biológicas, Guamá, Belém, PA, CEP 66075-110, Brazil; evanovic@ufjf.br (*corresponding author)

Abstract

ABO locus presents three main alleles: *A*, *B* and *O*. *A* and *B* encode glycosyltransferases that catalyze the addition of an N-GalNac and D-galactose to a precursor substance (H substance), producing A and B antigens, while the *O* allele does not produce a functional protein. The presence of A and B antigens have been associated to resistance against infectious agents which could use them as attachment factors increasing the virulence of some parasitic agents. As these antigens are not restrict to humans, analyses them in others species, for instance non-human primates, may be crucial to understand the relationship between pathogens and ABO phenotypes. Despite of the relevance of this issue, in the last decade few studies have addressed, mainly in New World Monkeys (NWM), natural reservoir of tropical diseases in Amazon Region. In order to understand the evolution of the ABO system in the primates, it has been obtained the partial sequence of the most important exon of *ABO* gene (exon 7), in platyrrhini families: Atelidae, Pitheciidae and Cebidae. Then, it has been compared the sequences obtained those present in the literature, and measured the selective pressure. The present results shown that residues 266 and 268 are also crucial to distinguish A and B phenotypes in the platyrrhines, such as in catarrhines, and the 266 codon is under positive selection, although the most site codons are under action of purifying selection.

Keywords: *ABO*, adaptive evolution, platyrrhini, positive selection and purifying selection

Introduction

The ABO histo-blood system is the most well known and clinically important blood group and was one of the first genetic traits applied to human population studies, but its biological and functional role has remained an enigma. The co-dominant alleles *A* and *B*, encode A and B glycosyltransferases that add N-acetyl-D-galactosamine (GalNac) or (and) D-galactose (galactose) to a precursor substance (H substance), producing the A or (and) B antigens, respectively. *O* allele is a recessive and null allele produces a non-functional enzyme due one single nucleotide deletion that changes the frame of codons (Bennett *et al.*, 1995; Yamamoto *et al.*, 1990, 1995; Yazer, 2005). After the identification of the molecular structure of the *ABO* gene, many molecular polymorphisms were found for this locus in human and in other primates. A current theory on this genetic variability it is that may be driven by infectious agents which use specific sugar of the erythrocyte surface such as the A and B antigens as receptors for invasion (Baum *et al.*, 2002; Gagneux and Varki 1999), diversity would then be a response to constant pressure in the host to combat the pathogen (Haldane, 1949; Pfenig, 2001).

Indeed there is a growing number of publications that indicate the relationship between ABO group (and others blood systems) to infectious agents as *Escherichia coli*, *Plasmodium falciparum*, *P. vivax*, *Candida spp.*, *Helicobacter pylori*, *Vibrio cholerae*, *Human immunodeficiency virus* (HIV), parvovirus and influenza virus (Aho

et al., 1980; Anstee, 2010; Germain *et al.*, 2011; Neil *et al.*, 2005; Rowe *et al.*, 2007; Tregouet *et al.*, 2009). A well established example of selective pressure exerted by pathogens on ABO blood group is the high frequency of the O phenotype in the areas endemic for *P. falciparum* and the prevalence of groups A and B in the areas with *V. cholerae*, due to vulnerability related to ABO specific phenotypes (Harris *et al.*, 2005; Williams, 2006).

No doubt, study the blood groups in non-human primates may clarify important points on the tolerance and selective pressure of the pathogens on human population, mainly because some of parasitological diseases are common to most primates. Despite of the relevance of this issue, in the last decade few studies have addressed, specialty in platyrrhini, known also as New World Monkeys (NWM), an important natural reservoir of tropical diseases in Amazon Region.

Because of the pertinence and importance of this approach, it has been clarified the main region of the *ABO* gene (exon 7) in 14 species of the platyrrhini families: Atelidae, Cebidae and Pitheciidae. Some of these samples have ABO blood type unknown, but based on molecular characteristics of the gene region studied, it was possible estimate the possible phenotypes of them. These samples were compared to sequences of catarrhine primates (human, apes and Old World monkey) present in the literature in order of identify differences among primates, and to measure the selective pressure under each lineage and codons in this *ABO* gene region.

Materials and methods

Extraction, PCR amplification and sequencing of DNA

The genomic DNA from 22 samples from NWM (acquired in Blood Sample Bank of Laboratory of Molecular Biology-UFPA) was obtained by standard phenol-chloroform extraction as described by Sambrook *et al.* (1989).

Only the exon 7 was amplified because it encodes 344 amino acids corresponding to 65% of catalytic domain of the ABO protein, and presents the critical residues located at codon 266 and 268 that distinguish A and B transferase in the catarrhine primates. The exon 6 shorter with only 45 amino acids, it was not used in this analysis, but was sequenced. The primers were constructed based on human sequence described by Yamamoto *et al.* (1990). They are as follows: ABO6.3: 5'-CACCATTGGGTAACTG-3', ABO1403: 5'-AGTGCTTCTCCGCCGTCTCC-3', ABO 7a1: 5'-CCACTACTATGTCTTCACCG-3' and ABO 7c2: 5'-CAAACCCACCAAGGTGCTC-3'. The polymerase chain reaction (PCR) was performed using 1 X reaction buffer, 100 ng DNA, 0.4 mM of each primer, 0.03 U/ μ l Taq DNA polymerase, 1.4 mM MgCl₂ and 10 mM of each dNTP. Amplification was performed with the following temperature: 5 min at 95°C followed by 35 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min. The PCR product was purified using Kit Wizard PCR Preps (PROMEGA) and sequenced utilizing the dideoxynucleotide chain termination method (Sanger *et al.*, 1977) the ABI 377 (Applied Biosystem) automatic sequencer.

Sequence alignment

The sequences obtained are shown in the Tab. 1. The following GenBank sequences were also used in the analysis: *Gorilla gorilla* (AB048862), *Homo sapiens* (AF134412, AH007588, AF134436 and EU707567), *Pan paniscus* (AB031246), *Pan troglodytes* (AB021844), one sequence of *Pongo pygmaeus* described by Martinko *et al.* (1993), *Macaca fascicularis* (AH008342, AH008343 and AH008220), *Macaca mulatta* (AF071830) and *Papio hymadryas* (AH006696, AH006697 and AH006698).

These sequences were aligned using the MUSCLE of the MEGA 5.05, an algorithm that allows align the nucleotide based in the amino acid sequences (Edgar, 2004; Tamura *et al.*, 2011). *Mus musculus* (AK035261) and *Rattus norvegicus* (AB081649) sequences were included to data set for compare the amino acid residues.

Analysis of molecular evolution

This analysis was performed using dN/dS ratio (non-synonymous/synonymous substitution rates, known as ω parameter) obtained with Maximum likelihood framework. According to the value of ω is possible to estimate the natural selection on a protein, then $\omega < 1$ means negative purifying selection; $\omega = 1$, neutral evolution, and $\omega > 1$, mean positive selection. In order of detect positive selection on the primate lineages and under the codon sites, it

Tab. 1. Samples, theirs genbank access numbers and ABO phenotype

Primate Group/ Species	Accession number	Phenotype
PRIMATE		
CATARRHINI		
CERCOPITHECIDAE		
<i>Macaca fascicularis</i>	AH008342	A
<i>Macaca fascicularis</i>	AH008343	B
<i>Macaca fascicularis</i>	AH008220	O
<i>Macaca mulatta</i>	AF071830	O
<i>Papio hymadryas</i>	AH006696	A
<i>Papio hymadryas</i>	AH006697	B
<i>Papio hymadryas</i>	AH006698	O
HOMINIDAE		
<i>Homo sapiens</i>	AF134412	A1
<i>Homo sapiens</i>	AH007588	A2
<i>Homo sapiens</i>	AF134436	O
<i>Gorilla gorilla</i>	AB048862	B
<i>Pan paniscus</i>	AB031246	Unknown
<i>Pan troglodytes</i>	AY138474	O
PLATYRRHINI		
ATELIDAE		
<i>Alouatta belzebul</i>	FJ377681	B
<i>Brachyteles arachnoides</i>	FJ377682	Unknown
PITHECIIDAE		
<i>Callicebus brunneus</i> (1)	FJ377686	A
<i>C. brunneus</i> (2)	FJ377687	A
<i>C. moloch</i>	FJ377689	Unknown
<i>C. personatus</i>	FJ377692	Unknown
<i>Pithecia irrorata</i> (1)	FJ377693	A
<i>P. irrorata</i> (2)	FJ377694	A
CEBIDAE		
<i>Cebus apella apella</i>	FJ377683	Unknown
<i>C. a. apella</i>	FJ377684	O
<i>C. a. paraguayanus</i>	FJ377685	A
<i>C. nigrovittatus</i>	FJ377690	Unknown
<i>C. olivaceus</i>	FJ377691	Unknown
<i>Callitrix humeralifer</i>	FJ377688	Unknown
<i>Saimiri boliviensis boliviensis</i>	FJ377695	A
<i>S. sciureus</i>	FJ377697	Unknown
<i>S. sciureus</i>	FJ377698	A
<i>S. sciureus macrodon</i> (1)	FJ377699	A
<i>S. sciureus macrodon</i> (2)	FJ377700	A
<i>S. sciureus macrodon</i>	FJ377701	AB
<i>S. ustus</i>	FJ377702	Unknown
<i>Saguinus mystax</i>	FJ377696	Unknown
RODENTIA		
MURIDAE		
<i>Mus musculus</i>	AK035261	A
<i>Rattus norvegicus</i>	AF296762	A

has been used the methods Branch Models and Site Models, respectively. For these analyses, it has been utilized primate data sets, excluding Rodentia sequences.

Branch models

This model also known as “Free-ratio” assumes an independent ω ratio for each branch of a phylogeny, detecting positive selection on particular lineages (Yang *et al.*, 1998). To make this analysis, it has been assumed a specie-tree topology described by Chatterjee *et al.* (2009), and the relationship phylogenetic among main human allele sequences (A^1 , A^2 , B and O) was obtained in Ogasawara *et al.* (1996). The samples of same specie and phenotype were deleted of the data set, so were used only 35 samples in the whole. The phylogenetic tree was manually constructed in newick format and was inputted in the Codeml, a program of the package PAML Version 4.5 (Yang, 2007).

Site models

The analysis to identify the sites under positive selection were performed using Codeml (Yang, 2007), which estimates the different parameters for M1a, M2a, M7, M8 and M8a, site models that are compared to pairs to test of neutrality and selection hypotheses. M1a, M7 and M8a are models for neutral evolution or null hypothesis, as M2a and M8 are models for positive selection, or alternative hypotheses. In each model, the data set and the phylogenetic tree were input in the Codeml and a log-likelihood value, and parameters were estimated. The fit of the pairs M1a-M2a, M7-M8 and M8-M8a were compared using the χ^2 distribution based on log-likelihood scores obtain in each model (likelihood ratio test-LRT) (Nielsen and Yang, 1998; Yang and Bielawski, 2000). It has been used 2 degree of freedom (df) for M1a-M2a and M7-M8 pairs, and 1 df for M8-M8a pair, a suggested alternative and conservative to use of the distribution with a 50:50 mixture of point mass 0 and χ^2_1 (Self and Liang, 1987). A site was admitted as under positive selection when the posterior probability on a codon belonging to the $\omega > 1$ class was greater than 95% using the Bayes Empirical Bayes approach-BEB (Yang *et al.*, 2005).

Result and discussion

Exon 7 of ABO gene of the New World Monkeys

In this work, it has been examined 184 codons (126 to 309 codon) that include all site codons considered as most important for distinguish A and B transferases, i.e. 176, 235, 266 and 268 codons. The comparison among amino acid sequences from primates showed quite variations, mainly among platyrrhine sequences, that presented 25% of divergence. In the catarrhine sequences this variation is lower, of 9.78%. The greatest diversity in exon 7 of platyrrhini could be because of its ancient age, about of 35 million years (MYA), 10 MYA older than catarrhine (Glazko and Nei, 2003; Schrago, 2007).

In the spite of this, the majority of amino acids are conserved and some substitutions are shared among lineages, for instance the amino acids alanine, lysine and serine at position 166, 215, and 226 are common in NWM and rodents, seeming to have ancestor origin, because these residues are encoded by same codon or a similar codon i.e.,

for NWM and OWM, the amino acid alanine at position 166 is code by GCN (T or C or G or A) codons, lysine by AAG (at 215) and serine at codon 226 by TCC for NWM and TCA for rodents.

The specificity of the A and B transferases in NWM is determinate by codons 266 and 268, such as in the ABO glycosyltransferases of catarrhine. Among typed samples, *Callicebus brunneus* is an exception presenting methionine and serine in the codon position 266 and 268 and possesses A phenotype detected by saliva inhibition test (Rocha *et al.*, 1992). The same molecular pattern was observed in others samples of *Callicebus* (*C. personatus* and *C. moloch*), but theirs phenotypes are unknown. In opposition, an experiment with A and B transferase chimeras performed by Yamamoto and Hakomori (1990) indicated that methionine and serine at positions 266 and 268 results in both A and B transferase activities. Thus, the complex pattern of A transferase of *Callicebus* could happen due: (1) the weak activity of B determinant that could be insufficient for be detected by inhibition technique or, (2) the 266 position is not crucial to determinate the enzymatic specificity in this genera.

A or O “molecular patterns” with leucine and glycine at codons 266 and 268 respectively, it is predominant among the NWM analysed, including those with unknown phenotypes, therefore it is more likely that these NWM have A or O phenotypes. This observation is coherent with earlier works that found more specie platyrrhines with A or O phenotype (Corvelo *et al.*, 1985; Hamel *et al.*, 1988; Rocha *et al.*, 1992; Schneider *et al.*, 1993). However, *Cebus olivaceus* seem has AB phenotype because has methionine and glycine at 266 and 268.

Our results on the reconstruction ancestral sequences (performed by Codeml) corroborate the hypothesis suggested by Saitou and Yamamoto (1997) that B alleles arose independently in the primates, and A allele is ancestral form.

Analysis of adaptive molecular evolution

Free-ratio model

The results of the “Free ratio” M1 model revealed, in general very low dN/dS ratios, well below 1 to the most of the branches, indicating purifying selection. But these values vary considerably within some lineage primates, i.e., some branches (of *Brachyteles* and B allele human) indicated evidence of evolution under relaxed purifying selection (when $\omega > 0.5$), and others, shown ω value infinity (when dS = 0), under adaptive selection. So, it is likely that ABO gene had evolved under different kinds of selective pressure in different primate lineages, because they are adapt to particular environmental conditions with vary demographic history and biological, beyond they may have differential tissue expression of A and B antigenic epitope, trait that may influence the selective pressure under each lineage (Kosiol *et al.*, 2008). The phylogenetic tree and ω value obtained to each branch by “Free-ratio” model is illustrated in Fig. 1.

Tab. 2. Tests of positive selection and positively selected codons in the exon 7 from ABO gene of primates according to pair M1a-M2a and M7-M8 and M8-M8a

Model	Likelihood ^a	LRT ^d	Df ^e	Parameters ^{b,c}	Positively selection codons ^f	dN/dS and SE ^g	Probability ^h
M1a	-2283.96	1.50	2	$\kappa=6.49, p_0=0.93, p_1=0.06, \omega_0=0.04, \omega_1=1.00$	Not Allowed		
M2a	-2283.20			$\kappa=6.49, p_0=0.93, p_1=0.05, p_2=0.01, \omega_0=0.04, \omega_1=1.00, \omega_2=3.09$	238 G 266 L	2.02 ± 1.44 2.39 ± 1.38	0.66 0.92
M7	-2283.64	8.52	2	$\kappa=6.25, p=0.19, q=1.65$	Not Allowed		
					210 V	1.16 ± 0.46	0.55
M8	-2279.38		1	$\kappa=6.46, p_0=0.99, (p_1=0.01), p=0.23, q=2.34, \omega=2.77$	238 G 266 L 268 G	1.67 ± 1.16 2.39 ± 1.38 1.15 ± 0.50	0.75 0.99 0.53
M8a	-2276.48	5.80		$\kappa=6.28, p_0=0.95, (p_1=0.04), p=0.411, q=6.84, \omega=1.00$	Not Allowed		

^aThe likelihood indicates the fit of the data to the different models; ^bPi denotes the proportion of sites falling in sites class ω_i ; ^cParameters p and q are the shape parameters of the beta distribution which underlies M7, M8 and M8a models; ^dResult of the Likelihood ratio test among the pair models M1a-M2a, M7-M8 and M8-M8a; ^eDegree of freedom for χ^2 test; ^fCodons under positive selection determinate by Bayes empirical Bayes (BEB); ^gPosterior probabilities for site classes by BEB, values > 95% do is not significantly, but is shown

erant to *H. pylori* infection than ones with others ABO groups. An interesting fact is that B phenotype is prevalent in the regions even in human populations (Malekasgar, 2005).

A support of the action of balancing selection on *ABO* gene of primates is the maintenance of different amino acids in varies site codons. Ferrer-Admetlla *et al.* (2008) described that the balancing selection is fundamental for the evolution of innate immunity genes. So, it is likely that it also act on others genes which take part in mechanism of protection against infectious agents, as *ABO* gene.

Conclusions

In the present study it has been detected that the majority of sites from ABO glycotransferases are under purifying selection, and one only site is under positive selection, thereby the present results agree with Anisimova *et al.* (2001) which observed that the adaptive selection occurs at only few sites and not in all the sites of a functional gene. The results obtained lead to three main conclusions: (1) codons 266 and 268 are important for distinguish A and B transferases also in platyrrhini; (2) a allele is ancestral allele in primates, (3) and adaptive and purifying selections act on *ABO* gene of primates, the first creates diversity and the second keeps the function and structure of the ABO glycosyltransferases.

Acknowledgements

This paper is part of the master of Eliane Evanovich at Centro de Ciências Biológicas-Universida de Federal do Pará. Eliane Evanovich was sponsored by CAPES from the Education Ministry, Brazil.

References

- Aho K, Pyhala R, Visakorpi R (1980). ABO associated genetic determinant in H1N1 influenza. *Tissue Antigens* 16(4):310-313.
- Anisimova M, Bielawski JP, Yang Z (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol* 18(8):1585-1592.
- Anstee DJ (2010). The relationship between blood groups and disease. *Blood* 115(23):4635-4642.
- Baum J, Ward RH, Conway DJ (2002). Natural selection on the erythrocyte surface. *Mol Biol Evol* 19(3):223-229.
- Bennett EP, Steffensen R, Clausen H, Weghuis DO, Vankessel AG (1995). Genomic cloning of the human histo-blood group ABO locus. *Biochem Bioph Res Co* 206(1):318-325.
- Chatterjee HJ, Simon YW, Barnes I, Groves C (2009). Estimating the phylogeny and divergence times of primates using a supermatrix approach. *BMC Evol Biol* 9(259):1-9.
- Corvelo TC, Hamel MLH, Schneider H (1985). Grupos sanguíneos ABO em *Alouatta belzebul* da região de Tucuruí-Pará. *Cienc Cult* 37(3):705.
- Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res* 32(5):1792-1797.
- Ferrer-Admetlla A, Bosch E, Sikora M, Marquès-Bonet T, Ramírez-Soriano A, Muntasell A, Navarro A, Lazarus R, Calafell F, Bertranpetit J, Casals F (2008). Balancing selection is the main force shaping the evolution of innate immunity genes. *J Immunol* 181(2):1315-1322.
- Gagneux P, Varki A (1999). Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology* 9(8):747-755.
- Germain M, Saut N, Greliche N, Dina C, Lambert JC, Perret C, Cohen W, Oudot-Mellakh T, Antoni G, Alessi MC, Zelenika D, Cambien F, Tiret L, Bertrand M, Dupuy AM, Letenneur L, Lathrop M, Emmerich J, Amouy P, Trégouët DA, Morange PE (2011). Genetics of venous thrombosis: Insights from a new genome wide association study. *PLoS ONE* 6(9):1-11.
- Glazko GV, Nei M (2003). Estimation of divergence times for major lineages of primate species. *Mol Biol Evol* 20(3):424-434.

- Hahn MW, Rockman MV, Soranzo N, Goldstein DB, Wray AG (2004). Population genetic and phylogenetic evidence for positive selection on regulatory mutations at the Factor VII locus in humans. *Genetics* 167(2):867-877.
- Haldane JBS (1949). Disease and Evolution. *La Ricerca Scient* (Suppl) 19:68-76.
- Hamel MLH, Schneider H, Matayoshi T, Nagle C, Nasazzi N, Howling H (1988). ABH antigens in population of capuchin monkeys (*Cebus apella paraguayanus*) from Paraguay. *Brazil J Genet* 11(2):317-327.
- Harris JB, Khan AI, LaRocque RC, Dorer DJ, Chowdhury F, Faruque ASG, Sack DA, Ryan ETR, Qadri F, Calderwood SB (2005). Blood group, immunity, and risk of infection with *Vibrio cholerae* in an area of endemicity. *Infect Immun* 73(11):7422-7427.
- Kosiol C, Vinař T, Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A (2008). Patterns of positive selection in six mammalian genomes. *Plos Genet* 4(8):1-17.
- Lindén S, Mahdavi J, Semino-Mora C, Olsen C, Carlstedt I, Borén T, Andre Dubois A (2008). Role of ABO secretor status in mucosal innate immunity and *H. pylori* infection. *Plos Pathog* 4(1):6-13.
- Malekasgar AM (2005). ABO blood group prevalence in spontaneously repeated abortion. *Turk J Hematol* 21(4):181-187.
- Martinko JM, Vincek V, Klein D, Klein J (1993). Primate ABO glycosyltransferase: Evidence for trans-species evolution. *Immunogenetics* 37(4):274-278.
- Neil S, McKnight A, Gustafsson K, Weiss RA (2005). HIV-1 incorporates ABO histo-blood group antigens that sensitize virions to complement-mediated inactivation. *Blood* 105(12):4693-4699.
- Nielsen R, Yang Z (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148(3):929-936.
- Ogasawara K, Bannai M, Saitou N, Yabe R, Nakata K, Takenaka M, Fujisawa K, Uchikawa M, Ishikawa Y, Juji T, Tokunaga K (1996). Extensive polymorphism of ABO blood group gene: three major lineages of the alleles for the common ABO phenotypes. *Hum. Genet* 97(6):777-783.
- Pfennig KS (2001). Evolution of pathogen virulence: the role of variation in host phenotype. *Proc R Soc Lond B* 268(1468):755-760.
- Rocha RM, Harada ML, Schneider H (1992). ABO blood group in Cebidae (Platyrrhini, Primates) from Jamari region. *Primates* 33(3):391-397.
- Rowe AJ, Handel IG, Mahamadou AT, Deans AM, Lyke KE, Koné A, Diallo DA, Raza A, Kai O, Marsh K, Plowe CV, Doumbo OK, Moulds JM (2007). Blood group O protects against severe *Plasmodium falciparum* malaria through the mechanism of reduced rosetting. *Proc Natl Acad Sci* 104(44):17471-17476.
- Saitou N, Yamamoto F (1997). Evolution of primate ABO blood group genes and their homologous genes. *Mol Biol Evol* 14(4):399-411.
- Sambrook J, Fritsch EF, Maniatis T (1989). *Molecular cloning: A laboratory manual*, 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1659 p.
- Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain-termination inhibitors. *Proc Natl Acad Sci* 74(12):5463-5468.
- Schneider H, Harada-Hamel MLH, Encarnación F, Montoya E (1993). Comparison of ABO blood groups from two species of peruvian squirrel monkeys (*Saimiri sciureus macrodon*, *Saimiri boliviensis peruviansis*) with a natural population of *S. sciureus macrodon*-*S. boliviensis peruviansis* hybrids. *Braz J Genet* 16(3):661-669.
- Schrago CG (2007). On the time scale of New World primate diversification. *Am J Phys Anthropol* 132(3):344-354.
- Self SG, Liang K (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc* 82(398):605-610.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28(10):2731-2739.
- Tregouet DA, Heath S, Saut N, Biron-Andreani C, Schved JF, Pernod G, Galan P, Drouet L, Zelenika D, Juhan-Vague I, Alessi MC, Tiret L, Lathrop M, Emmerich J, Morange PM (2009). Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. *Blood* 113(21):5298-5303.
- Williams TN (2006). Human red cell polymorphisms and malaria. *Curr Opin Microbiol* 9(4):388-394.
- Yamamoto F, Clausen H, White T, Marken J, Hakomori S (1990). Molecular genetic basis of the histo-blood group ABO system. *Nature* 345(6272):229-233.
- Yamamoto F, Hakomori S (1990). Sugar-nucleotide donor specificity of histo-blood group A and B transferases is based on amino acid substitutions. *J Biol Chem* 265(31):19257-19262.
- Yamamoto F, McNeil PD, Hakomori S (1995). Genomic organization of human histo-blood group ABO genes. *Glycobiology* 5(1):51-58.
- Yang Z (2007). PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586-1591.
- Yang Z, Bielawski JP (2000). Statistical methods for detecting molecular adaptation. *Trends Ecol Evolut* 15(12):496-503.
- Yang Z, Nielsen R, Hasegawa M (1998). Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol Biol Evol* 15(12):1600-1611.
- Yang ZH, Wong WSW, Nielsen R (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22(4):1107-1118.
- Yazer MH (2005). What a difference 2 nucleotides make: a short review of ABO genetics. *Transfus Med Rev* 19(3):200-209.