The relationship between blood groups and disease

David. J. Anstee

Bristol Institute for Transfusion Sciences,

National Health Service (NHS) Blood and Transplant, Bristol, United Kingdom.

Correspondence: david.anstee@nhsbt.nhs.uk

Abstract
The relative contribution of founder effects and natural selection to the observed distribution of human blood groups has been debated since blood group frequencies were shown to differ between populations almost a century ago. Recent advances in our understanding of the migration patterns of early man from Africa to populate the rest of the world obtained using Y chromosome and mitochondrial DNA markers do much to inform this debate. There are clear examples of protection against infectious diseases from inheritance of polymorphisms in genes encoding and regulating the expression of ABH and Lewis antigens in bodily secretions particularly in respect of Helicobacter pylori, Norovirus and Cholera infections. However, available evidence suggests surviving malaria is the most significant selective force affecting the expression of blood groups. Red cells lacking or having altered forms of blood group-active molecules are commonly found in regions of the world where malaria is endemic, notably the Fy(a-b-) phenotype and the S-s- phenotype in Africa and the Ge negative and SAO phenotypes in South East Asia. Founder effects provide a more convincing explanation for the distribution of the D negative phenotype and the occurrence of Hemolytic Disease of the Fetus and Newborn in Europe and Central Asia.

Introduction
Hirszfeld and Hirszfeld (1) showed the frequencies of blood groups A and B differ between populations. Their observations raised fundamental questions regarding the causes of these differences which were eloquently summarized by Mourant et al. (2).

"were the differences the result of random genetic drift and founder effects, in small populations which later multiplied and stabilized the original, fortuitous, frequencies, or were they the result of natural selection, arising from differences in fitness between the various blood groups, fitnesses which themselves depended upon locally determined features of the external environment?"

Mourant et al concluded "most workers now agree that both processes are operative, but their relative importance remains in question."

We now have detailed information concerning almost all the genes giving rise to blood group polymorphisms, the structure of the gene products and the antigens themselves and in many cases functional information sufficient to delineate mechanisms of interaction with external agents. (3,4,5).

In addition, studies on the tracking of Y chromosome and mtDNA haplotypes in human populations provide us with unprecedented information concerning the
significance of genetic drift and founder effects in determining the genetic background of different world populations (6).

Given this new information it seems an appropriate time to revisit these questions and ask if we are any nearer understanding the relative importance of natural selection and founder effects in determining the distribution of human blood groups.

**Infectious Diseases and Selection for ABO Blood Group Antigens**

The molecular basis of the ABO blood group system was elucidated in 1990 (7). The gene encodes a glycosyltransferase which transfers N-acetyl D-galactosamine (Group A) or D-Galactose (Group B) to the non-reducing ends of glycans on glycoproteins and glycolipids. The group O phenotype results from inactivation of the A1 glycosyltransferase gene and the non-reducing ends of the corresponding glycans in group O individuals express the blood group H antigen (figure 1a). The ABH antigens are not confined to red cells, but widely expressed in body fluids and tissues. The biological significance of the A/B transferase has not been clearly demonstrated but it would be expected that loss of this functional protein in group O individuals would have some deleterious consequences for individuals of this blood type.

One of the most significant disease associations described for non-O (individuals of group A, B, or AB) versus O individuals is susceptibility to arterial and venous thromboembolism (8,9). Non-group O individuals have a higher risk of venous thromboembolism (VTE) than individuals of group O and have higher levels of von Willebrand factor (vWWF) and factor VIII (FVIII) (8,10). The risk of VTE is probably related to the level of vWWF and FVIII because individuals of group A2 have lower levels of these proteins than A1, B and AB and have a lower risk of VTE (9). A,B and H blood group antigens are expressed on N-glycans of vWWF and influence the half life of the protein (10hr for group O and 25hr for non-O subjects) providing an explanation for the higher levels on non-O individuals, (11). These observations raise the possibility that a greater propensity for blood clot formation in non-O individuals conferred a survival advantage to early man. Such an argument has been made for the occurrence of the prothrombotic mutations Factor V Leiden and Prothrombin 20210G>A commonly found in whites dated as occurring 20,000-24,000 years ago towards the end of the last ice age (12). It is proposed that mutations like FV Leiden lower the risk of hemorrhage and/or severe infections and thereby the risk of death during pregnancy (13). A similar hypothesis could explain the function A, B antigens on vWWF.

What then was the stimulus causing the inactivation of this gene and creating the group O phenotype which is so prevalent throughout the world? Evidence supporting the view that blood group O provides a selective advantage against severe malaria has been recently reviewed (14 -16). The argument is persuasive. Group O is presumed to have arisen in Africa prior to the migration of early humans. Severe malaria results in the death of millions each year before they reach child bearing age and so selects survival genes (17). Experimental support for the hypothesis is provided by Fry et al (18) and by Rowe et al (19). Rowe et al report reduced rosetting of *P falciparum* isolates from group O Malian children compared with non-O blood groups. Parasitized red cells form rosettes with uninfected red cells and adhere to vascular endothelium causing vasocclusion and effecting severe disease.

There are other examples of infectious diseases where severity of infection can be directly linked to ABO phenotype. Numerous studies have shown that once an
individual is infected with cholera (**Vibrio cholerae** strains O1; El Tor and O139) the phenotype group O confers a greater likelihood of severe infections than a non-O blood group phenotypes (20). Glass et al (21) suggest that the low prevalence of group O and high prevalence of group B in the Ganges Delta in Bangladesh is directly related to selective pressure from cholera. Almost all recent cholera pandemics have emanated from this region of the world (22). Individuals of group O were more susceptible in an outbreak of gastrointestinal infections due to **Escherichia coli** O157 in Scotland in 1996. 87.5% of patients who died were group O (23). However, suggestions that smallpox selects against A thereby explaining the high frequency of group A in Europe and that the low frequency of O in ancient plague centres in Mongolia and the Middle East is also a reflection of selection are not supported by adequate data (Vogel et al 1960 cited in 2, p18; 24). More recent studies have linked the high frequency of the HIV-1 resistance mutation CCR5 delta 32 in Europe with protection against smallpox and Black Death (25). This proposal has also been questioned (26). The mutations A to O and CCRdelta32 occurred much earlier in human evolution than the plague and smallpox epidemics of medieval times. As discussed above, the A to O mutation was likely driven by malaria in Africa prior to migration of early human to Europe and CCR delta32 has been described in bronze age skeletons (25). A combination of selection against infectious diseases like plague and smallpox and genetic drift and founder effects in small populations resulting from migration patterns of early man may ultimately explain the allele frequencies observed today.

The expression of ABH antigens in tissues and body fluids other than blood cells is regulated by the Secretor gene (**FUT2**) which encodes an alpha 1,2-fucosyltransferase capable of transferring L-fucose to carbon 2 of Galactose (beta, 1-3) N-Acetyl D-Glucosamine containing glycans. In the absence of an active FUT2 gene (non-secretor) the structure created is the Le^a_ antigen (27). The product of the Le gene is an alpha 1,3/4 fucosyltransferase (**FUT3**) which transfers L-fucose to carbon 4 of the penultimate N-acetyl-D-glucosamine residue of the same glycans (28). The structure created in tissues by the combined action of FUT2 and FUT3 is the Le^b_ antigen A and B antigens can only be formed in the tissues of individuals with an active FUT2 by the action of alpha-glycosyltransferases capable of transferring N-acetyl D-galactosamine or D-Galactose to carbon 3 of the same glycans (figure 1b). The secretions and tissues of an individual with an active FUT2 (a secretor) can express A, B, H and Le^b_ antigens in those secretions according to the glycosyltransferase genes inherited. In European and African non-secretors homozygous inheritance of a nonsense mutation (G428A) inactivating FUT2 denoted se^428_ is frequently found (29. 20% of Europeans). In the Far East and Pacific regions the commonest mutation in FUT2 (A385T, se^385_) causes a single amino acid change (Ile129Phe) in the stem region of the fucosyltransferase resulting in a five-fold reduction in active enzyme and a weak Le(a+b+) phenotype. (30). Sequencing FUT2 in 732 individuals from 39 populations confirmed the widespread occurrence of the se^428_ allele in Europe, Central Asia and Africa and the se^385_ allele the Far East and Pacific and mapped two further se alleles with a more restricted distribution (se^302_ and se^571_) to Central and South Asia and Cambodia respectively (31). Possession of homozygosity for a non-secretor phenotype has a demonstrable survival advantage for some infectious diseases.

One of the first proven associations of a blood group polymorphism with disease was that between group O and peptic ulceration (32,33). The gastric pathogen **Helicobacter pylori** is now known to be a causative agent leading to peptic ulceration
and gastric cancer. *H pylori* is said to “colonise the stomachs of about half the worlds population” (34). Early studies demonstrated that a South American strain of *H pylori* P466 bound to blood group O Le<sup>b</sup> but not ALe<sup>b</sup> structures on the gastric epithelium thereby providing a clear explanation for the greater susceptibility of Group O secretors (35). More recent studies on strains of *H pylori* from different parts of the world have shown that not all strains are so specific for O Le<sup>b</sup> with many strains from outside South America having binding capabilities for ALe<sup>b</sup> in addition to OLe<sup>b</sup>. Nevertheless these strains have a higher binding affinity for OLe<sup>b</sup> compared with ALe<sup>b</sup> (ca. fivefold (median) higher (36). Sequence analysis of the bacterial surface molecule responsible for binding to gastric epithelium BabA (blood group antigen binding adhesin) from different strains of *H pylori* showed that Peruvian strains were closely related to Spanish but not to Asian strains raising the intriguing possibility that the OLe<sup>b</sup>-specific strains found in South American may have arisen following European colonization of South America in the 16<sup>th</sup> century and represent adaptation to a population which is almost entirely of blood group O phenotype (36).

Susceptibility to Norovirus infection is also closely linked to the expression of ABH and Le antigens in the gastrointestinal tract. Noroviruses are the commonest cause of acute gastroenteritis in humans estimated to account for 60-85% of all gastroenteritis outbreaks in developing countries (37). They are transmitted by consumption of contaminated food particularly oysters which appear able to concentrate the virus, or exposure to contaminated water (37). The pivotal role of secretor status in determining susceptibility to Norovirus has been clearly demonstrated by Thorven et al (38) who compared susceptibility to gastroenteritis in patients and medical staff involved in hospital outbreaks in Sweden. The results demonstrated that only those individuals homozygous for non-secretor were protected from infection. Further studies demonstrated significantly lower antibody titers to norovirus GGII in non-secretors compared with secretors (39). There are many different strains of norovirus however and some strains bind to non-secretor Le<sup>a</sup> structures and cause symptomatic infection (40,41). The variable specificity of different strains for ABH and Le<sup>b</sup> structures reported reflects a similar diversity to that of *H pylori* above. Evidence for greater susceptibility of Secretors to influenza viruses, rhinoviruses, respiratory syncytial virus and echoviruses has also been presented (42). Reduced risk of HIV type 1 infection was found in Senegalese commercial sex workers with the non-secretor type (43). Slow disease progression of HIV-1 in non-secretors was also reported by Kindberg et al (44).

Non-secretors appear more susceptible to infections by *Haemophilus influenzae* (45), *Nesseria meningitidis* and *Streptococcus pneumoniae* (46) and urinary tract infection caused by *Escherichia coli* (47).

A mutation (delta F508) in the cystic fibrosis transmembrane regulator gene (*CFTR*) is common in Europeans and was present in Europe during the Paleolithic period more than 10,000 years ago (48). The possibility that differences in A, B and H antigen expression in the airway mucus might lead to differences in microbial binding and predispose to more severe lung disease was investigated in 808 patients homozygous for deltaF508. No association with ABO, Se or Le genotype was observed (49).

**Rh Blood Groups and the origin of Hemolytic Disease of the Fetus and Newborn (HDFN)**
The major clinical disease associated with the Rh blood group system is HDFN. HDFN usually arises when a mother who is blood group D negative carries a fetus who is blood group D positive and fetal red cells released into the maternal circulation immunize the mother to make antibody to D which traverses the placenta and damages the fetus. Prior to the introduction of a successful prophylactic treatment in 1968 the frequency of the disease in England and North America was about 1 per 170 births (3,p524). Recognition of the disease as a single entity was slow to emerge. In severe cases anti-D crosses the placenta and causes death of the fetus in utero a condition known as hydrops fetalis. More commonly, disease occurs in the neonatal period where severe and acute anemia and severe jaundice is fatal, a condition known as icterus gravis neonatorum. Roberts (50,p13) cites an account of Louise Bourgeois, a midwife of Marie de Medici, who published in 1609 what is probably the earliest account of hydrops fetalis in one twin and neonatal jaundice in the other and credits Auden (1905) with a number of key observations relating to neonatal jaundice, in particular, its appearance in successive children of the same parents. The recognition that hydrops fetalis and neonatal jaundice were manifestations of the same disease gradually emerged during the 1920’s and anti-D was shown to be the causative agent in 1939 (51).

There is now a formidable body of evidence to support the hypothesis that man originated in Africa and inform the timescale of various migrations out of Africa which have led to the world populations we have today (6). Simply by overlaying the known distribution of blood group frequencies on the world map of human migrations the potential significance of genetic drift and founder effects is apparent. Wells (6,p141) argues for example, that it is possible to account for all of the mtDNA and Y-chromosome types in Native Americans with a founding population of 10-20 individuals. Little wonder then that Native Americans are almost exclusively blood group O (52) or that the Di² blood group polymorphism tracks the migration of man from East Asia to the Americas (52). The occurrence of the Di² antigen in South East Poland also provides a measure of the extent to which the Mongol invasions penetrated Europe in more recent times (53,54).

In Europe a similar founder effect can be invoked to explain the high frequency of the D negative phenotype. The emergence of Paleolithic ancestors surviving the last ice age from refuges in the Basque region of Northern Spain and Southern France and the Ukraine 10,000-15,000 years ago and subsequent interbreeding of these survivors with Neolithic migrants from the Middle East provides an explanation for the occurrence of HDFN. In order to explain the high frequency of the D negative allele in Europe Mourant (55) proposed a mixing of two populations one essentially D negative and the other D positive. He noted that D negative frequency was very high in the Basques and postulated mixing of Paleolithic peoples from the Basque region with Neolithic migrants as the cause. This hypothesis has been largely ignored in the succeeding years but recent observations using mtDNA and Y chromosome markers have led to wide acceptance of the population mixing hypothesis (56,57; figure 2 ). Tracking of haplotypes emerging from the Basque and Ukrainian refuges has shown that these peoples migrated throughout Europe and Central Asia and into India and Pakistan (58). HDFN is found in all these regions. Mourant (55, p80) also suggested a link between the Basques and the Berbers of North Africa because of the high frequency of D negative among Berbers. This hypothesis is now supported by evidence from maternal DNA markers showing ancestral Berbers occupied the Basque refuge area and migrated back into North Africa (59). In Western Europe the
D negative phenotype results from a complete deletion of *RHD* (60). The molecular basis of D negative phenotype has not been formally determined for Ukrainian D negative individuals. Loss of RhD protein does not appear to be of significant detriment to red cell function. The best available structural models for RhD protein and its homologue RhCE protein indicate they do not function as transport proteins but rather serve to facilitate the assembly of the band 3 protein gas transport complex in the red cell membrane. These observations suggest there is considerable functional redundancy with D and CE proteins effectively substituting for each other (figure 3, 61).

A counter argument to the population mixing hypothesis, would be provided by a clear demonstration of selection for D negative by environmental factors. A thorough review of early studies seeking to identify associations between the D polymorphism and diseases revealed no convincing associations (2). More recently, two studies have reported an association of the D polymorphism with disease.

Busquets et al (62) in a study from Barcelona report an increased incidence of biliary complications in transplant recipients of livers mismatched for D. The presence of biliary complications in D-non-identical graft-host cases (23/76, 30%) was higher than in D-identical grafts (47/269, 17%). Rh polypeptides are not expressed in liver (63) and so the mechanism of such an association is not clear. The fact that the study took place in the Basque region of Spain where D negative is very common and may result from ancestral paleolithic settlements may be very relevant to the interpretation of these findings since it raises the possibility that other genes more relevant to transplantation and also occurring more commonly in Basques than in other populations may be influencing the results. In this context it is interesting to note evidence that donor HLA-C genotype has a profound impact on the outcome of liver transplants (64). HLA-C is the major inhibitory ligand for Killer cell immunoglobulin-like receptors (KIRs). KIR genes are highly polymorphic and are expressed on Natural Killer (NK) cells and a subset of T lymphocytes (65,66). Several KIR genes (KIR2DS5, KIR3DS1, KIR2DL2) are significantly different in frequency in Basques and three novel haplotypes were identified by Santin et al (67).

Flegr et al (68) in a study from the Czech Republic report an association of the polymorphism with *Toxoplasma* infection whereby individuals (military conscripts) with the D negative phenotype who were *Toxoplasma* infected (11 of 181; 6.08%) had slower reaction times and consequently were involved in more road traffic accidents than D positive *Toxoplasma* infected (17 of 709; 2.4%) individuals. Rh D protein is not reported to be expressed in brain so the likely mechanism of such an association is obscure and given the small numbers of *Toxoplasma* infected individuals involved in the study a much larger cohort study will be required to prove the validity of this association.

Considering the evidence thus far it seems most likely that the frequency of D positive and negative phenotypes in Europe and Central Asia is a reflection of genetic drift and migration rather than natural selection with the early colonists of Europe emerging from Africa with a deletion of *RHD* (figure 2). There remains the possibility the original stimulus driving this deletion occurred in Africa as a result of selection. The concomitant occurrence of D negative phenotype in Africans resulting from a different molecular mechanism (69,70) may be suggestive of some ancient selective pressure.
**Malaria – Evidence for selection of blood group phenotypes which are rare outside areas where malaria is endemic.**

It seems likely that the most devastating effects of malaria on human populations coincided with a change of life style from hunter gatherer to more sedentary agricultural practices ca.10,000 years ago (15). Clearing of trees from forest areas creating potential for pools of stagnant water and breeding grounds for the mosquitoes carrying parasites.

The clearest examples of selection in the face of malaria are reflected in the widespread distribution of inherited anemias, particularly sickle cell anemia and alpha thalassemia and the occurrence of hemoglobin C in regions of the world where malaria is endemic (71,72). The mutation giving rise to sickle cell disease (HbS) may have arisen at three different sites in Africa (Atlantic West Africa, Central West Africa, Bantu-speaking central and Southern Africa) with expansion of the mutation occurring 2,000-2,500 years ago (73). In this case, individuals who inherit an HbS gene from both parents have SCD while those who are heterozygous inheriting the HbS gene from one parent and the normal HbA gene from the other parent have substantial protection against malaria. A similar protective effect for the heterozygote seems likely in South East Asia where HbE is very common and red cells from individuals of genotype HbAE are markedly less susceptible to malaria parasite invasion *in vitro* (74).

Further illustrations of the diversity of mutations that have arisen in response to malaria are deficiency of glucose-6 phosphate dehydrogenase, widespread in Mediterranean and India (71) and a polymorphism in the promoter of inducible nitric oxide synthase(75).

**Plasmodium vivax and the blood group Fy(a-b-) phenotype**

Complete absence from red cells of the molecule carrying the Duffy blood group antigens (aka DARC) is found in almost 100% of West Africans and this absence is clearly and unambiguously demonstrated to provide protection from *Plasmodium vivax* (76). The molecular basis of of this Duffy deficiency is a point mutation in the binding site for the transcription factor GATA-1 (77). GATA-1 is a DNA-binding protein essential for erythropoiesis and its failure to bind to the Duffy gene promoter means that the Duffy protein is absent from the red cells of affected individuals. In Africans the mutation occurs on a Duffy allele that would otherwise generate a Fy(b+) phenotype. The same GATA-1 mutation appears to have occurred on a second occasion in South East Asia where it occurs on a Duffy allele that would otherwise generate a Fy(a+) phenotype (78). Another mutation creating weak expression of Duffy (Fyx) may also be relevant to malaria but relevant population studies have not been reported (79). Recently, evidence for the emergence of *P vivax* strains capable of invading Fy (a-b-) red cells has emerged in South American and East Africa (80,81).

The protective effect of the Fy(a-b-) phenotype against *P vivax* is clear and unambiguously established. Not so clear are any deleterious consequences of this mutation for the individuals expressing the phenotype. Duffy protein is expressed on endothelial cells in these individuals but not on red cells (82) so any attempt to understand the consequences of red cell Duffy deficiency must take account of the functional role of endothelial Duffy. The Duffy protein is a member of the seven membrane spanning chemokine receptor family (figure 3) but unlike most chemokine receptors does not effect intracellular signalling through G-proteins. It binds several
pro-inflammatory chemokines of both the CXC and CC subfamilies but does not bind homeostatic chemokines (83). Recent evidence suggests Duffy protein on endothelial cells binds chemokines and facilitates leukocyte extravasation contributing to disease pathogenesis through inflammation (84). Evidence for upregulation of Duffy expression in the vascular endothelium during infection and transplant rejection supports this view (85,86).

The lack of Duffy on red cells in Fy(a-b-) individuals alters the balance of proinflammatory chemokines in the body because the very large capacity of red cell binding is absent but the consequences of this change are presently unclear. Lee et al. (87) provide evidence that red cell and endothelial Duffy regulate the kinetics of chemokine bioavailability between the circulation and extravascular sites during inflammation. Clearly this regulation would be altered in individuals of Fy(a-b-). In a mouse model, inflammation induced by polycytidylic acid significantly enhanced alloimmunization to red cells (88). In this context it is interesting to note that patients with SCD are predominantly of the Fy(a-b-) phenotype and the production of multiple red cell alloantibodies upon transfusion (usually with blood from white donors) is a frequent and significant problem encountered by Blood Banks seeking to provide compatible blood for the patients (reviewed in 89). SCD patients in sickle cell crisis and mouse models of human sickle cell disease have many indicators of an inflammatory response (90). These data suggest that the enhanced propensity for alloimmunization in SCD patients is related to inflammation and also pose the question as to the significance of Fy(a-b-) in this process. Are Fy(a-b-) SCD patients more likely to make alloantibodies in response to transfusion than SCD patients of normal Fy phenotype? Is there a link between regulation of proinflammatory chemokine availability by red cell Fy and the adaptive immune response? The data of Afenyi-Annan et al (91) provide evidence that SCD patients with the Fy(a-b-) phenotype are more susceptible to chronic organ damage and proteinuria than SCD patients of normal Fy phenotype and are consistent with such an hypothesis. Interpretation is likely also influenced by genetic differences of immune response and cytokine genes in Africans compared with other world populations (92,93) but the genetic backgrounds of SCD patients with normal and Fy(a-b-) phenotype may be sufficiently comparable to allow conclusions regarding alloimmunisation and the role of Fy to be drawn. Should Fy(a-b-) individuals be more susceptible to alloimmunization then the potential use anti-inflammatory therapies in the treatment of vasocclusion (94,95) may have the added bonus of reducing rates of red cell alloimmunization and provide a much needed alternative approach to a major transfusion problem.

A further consequence of selection for the Fy(a-b-) phenotype in Africa may be to alter the kinetics of HIV-1 infection in those with this phenotype. Several HIV-1 strains bind to Duffy on normal red cells facilitating the transfer of HIV-1 to its target cells (CD4+/CCR5+ T lymphocytes) with 5-12-fold greater efficiency than Fy(a-b-) red cells (96). He et al. (96) calculate individuals with the Fy(a-b-) phenotype have a 40% greater likelihood of acquiring HIV than those lacking the phenotype however the disease once acquired had a slower progression than in infected individuals of normal Fy type. They conclude that these differences are related to loss of competition for binding HIV-1 between plasma chemokine CCR5 and Fy on red cells in Fy(a-b-) individuals and consequent changes in the inflammatory state those infected. The findings of this study have been contested by Walley et al. (97) who used different methodology to analyse a different cohort of HIV positive and HIV
negative African-Americans and found no association between Fy genotype and progression to AIDS or risk for HIV acquisition. He et al. point out that the number of HIV negatives used by Walley et al. was much smaller (227 versus 814) and suggest this may be a major factor affecting the analysis (98).

Different strains of *P. falciparum* use different blood group proteins as receptors.

The dual availability of *in vitro* culture systems to study the invasion of human red cells by *P. falciparum* and well characterized rare blood group phenotypes made it possible to identify red cell receptors utilized by different parasite strains. Early studies on cells lacking Glycophorin A (Ena-cells,99) and Glycophorin B (S-s-cells,100) provided evidence that these sialic-acid rich red cell surface glycoproteins were parasite receptors and these observations have been confirmed (101-104). Glycophorins C and D (Gc – red cells) are also receptors for some strains of *P. falciparum* (105-107). Glycophorins are major proteins at the red cell surface (figure 3). GPA and the major anion transport protein (AE1,Band3) with ca. $10^6$ copies/red cell are the most abundant red cell surface proteins with Glycophorins B,C and D together accounting for a further 450,000 copies/red cell (108,109). Perhaps surprisingly, there is little experimental evidence to suggest selection against the expression of Glycophorin A has occurred in response to *P. falciparum* infection. Red cells from individuals having the hybrid GPA-GPB protein Dantu which is common in certain parts of Africa are reported to resist invasion (110) and it has been suggested that elevated expression of band 3 occurring in individuals with the GPB-GPA-GPB MiIII protein common in South East Asia may be relevant to malaria survival (111). The importance of sialic acid on GPA in forming a receptor for *P. falciparum* (101) suggests that red cells expressing glycosylation variants of GPA found commonly in Africans in which N-Acetyl D-glucosamine is present in some of the sialic acid-rich O-glycans at the N-terminus (individuals with the M1 antigen (112) may be relevant to malaria susceptibility. In contrast to the situation with GPA, individuals lacking Glycophorin B are found in high frequency in central Africa (104). Individuals with red cells lacking GPC and GPD (Ge negative,Leach phenotype) are very rare but individuals with Ge negative red cells having an altered GPC resulting from deletion of exon 3 of *GYPC* (113,114) are common in Melanesians, most notably in Papua New Guinea and the resulting phenotype provides protection against *P. falciparum* (107,figure 4). Clearly, different strains of *P. falciparum* target glycophorins associated with one or other of the membrane complexes providing key cytoskeletal linkages maintaining the stability of the red cell membrane (figure 3) and selection resulting in loss or alteration of glycophorins in either of these sites confers a survival advantage.

Melanesians also exhibit another example of selection affording protection against cerebral malaria, a phenotype known as South East Asian Ovalocytosis (SAO). SAO cells, as the name implies have an abnormal shape. They are also characterized by weakened expression of a large number of blood group antigens including antigens found on band 3,GPA and the Rh blood group proteins (115). In this case selection favours the heterozygote. Heterozygotes inherit a normal band 3 gene together with a mutant inactive band 3 gene resulting from a deletion causing a loss of nine amino acids at the point at which the cytoplasmic N-terminal domain enters the cytosolic face of the lipid bilayer. (reviewed in 116). Homozygous inheritance of this mutation would result in total band 3 deficiency. Since band 3 is essential for respiration (
Cl/HCO₃ exchange) and for maintaining the integrity of the red cell membrane it must be assumed in evolutionary terms that such an inheritance is incompatible with survival. Rare individuals with total band 3 deficiency states have been described but survival depends on extensive medical support particularly in the neonatal period (117).

Complement receptor 1 (CR1, figure 3) carries antigens of the Knops blood group system. CR1 expression is very variable between individuals and red cells expressing less than 100 copies CR1/ cell show reduced rosetting with *P falciparum* strain R29R as do red cells expressing the Sla⁻ blood group phenotype. The Sla⁻ phenotype which results from a single nucleotide polymorphism (R1601G) in long homologous repeat D occurs in only 1% whites but reaches 70% in Malians (118-120).

**Conclusions**

The significance of human blood groups can now be seen more clearly in the context of population movement, and the constant battle between man and infectious disease. Evidence for selection by infectious diseases at the level of the ABO and Secretor genes is persuasive but for other blood group antigens founder effects seem more likely to account for the distribution of blood group polymorphisms except that is, in parts of the world where malaria is endemic. Available data suggest survival from malaria has been the most significant selective force acting on the blood groups.

Rare blood group phenotypes revealed through compatibility testing in Transfusion Centres and Blood Banks throughout the world have provided powerful tools with which to investigate the mechanisms whereby malarial parasites invade human red cells. However, a comprehensive study of the distribution of known blood group polymorphisms in areas where malaria is endemic has not been undertaken. Now that almost all the blood group genes have been cloned and the molecular bases of most antigens determined it is feasible to conduct such a study using high throughput DNA-based methods (120-124). Furthermore, the availability of rapid DNA sequencing methodologies presages an era in which mass screening of genes encoding red cell membrane proteins could be used to identify new polymorphisms of relevance to malarial invasion. Studies of this type, focussed in tropical Africa, South East Asia and Latin America would provide a valuable database of new information about blood group diversity in populations inhabiting these regions not only for malarial epidemiologists but also for those investigating man's susceptibility to new emerging infectious diseases given that zoonoses from wildlife in these regions have been identified as the most significant growing threat to global health of all emerging infectious diseases (125).
Acknowledgements

I thank Lesley Bruce for preparing figure 3, David Briggs for helpful discussions regarding HLA-C and KIR receptors and Geoff Daniels for critical reading the manuscript.

David Anstee wrote this manuscript. The author declares no competing financial interests.

References


52. Mourant AE, Kopec AC, Domaniewska-Sobczak K. The distribution of the human blood groups and other polymorphisms. 2nd edn.OUP. 1976


82. Hadley TJ, Lu Z, Wasniowska K et al. Postcapillary venule endothelial cells in kidney express a multispecific chemokine receptor that is structurally and functionally identical to the erythroid isoform, which is the Duffy blood group antigen. J Clin Invest 1994;94 (3):985-91.


Legends to Figures.

**Figure 1.** (a). **Structure of ABO and H antigens on human red cells.** H antigen formed by the action of **FUT1** on oligosaccharide precursor chains in which the terminal D-galactose residue is linked to carbon four of the penultimate N-Acetyl D-glucosamine residue (Type II chain).

(b). **Structure of Le blood group antigens in bodily secretions.** Secretor gene (**FUT2**) regulates the production of H antigen which can be converted to A or B antigen if the corresponding active ABO glycosyltransferase is present. The ABH, Le^b^-active structures are formed on oligosaccharide precursor chains in which the terminal D-galactose residue is linked to carbon 3 of the penultimate N-Acetyl D-glucosamine residue (Type I chain) if **FUT2** is deficient the Le^a^ active structure predominates.

**Figure 2.** Paleolithic settlers from the last glacial maximum may be the source of the high frequency of D negative allele in Europeans (upper panel). European location of Paleolithic refuges at the time of the last glacial maximum. Note migration of population containing marker M173 (from 126, Gibbons A. Europeans trace ancestry to paleolithic people. *Science* 2000;290 (5494):1080-1081. Reprinted with permission from AAAS). (lower panel). Distribution of the D negative allele in Europe (from 52, Figure 22. Rhesus blood group system. Distribution of the d gene in the indigenous populations of Europe, Western Asia, and Africa, from “Distribution of the Human Blood Groups and other Polymorphisms” by Mourant AE, Kopec AC, Domaniewska-Sobczak K (1976). Reprinted by permission of Oxford University Press).

**Figure 3.** Structure of the human red cell membrane showing the major surface proteins and minor proteins Fy and CR1. Two major membrane complexes linked to the underlying red cell skeleton are depicted. The band 3 complex containing glycophorins A (GPA) and B (GPB) and Rh proteins, Rh associated protein (RhAG), CD47, LW glycoprotein (ICAM-4) and the junctional complex comprising glycophorins C and D (GPC,GPD), Kell glycoprotein, XK glycoprotein and Duffy (Fy) glycoprotein. Aquaporin 1 (AQP1), the glucose transporter (GLUT1), Decay accelerating factor (DAF,CD55) and complement receptor 1 (CR1) are also shown. ABH active oligosaccharides known to be present on all major surface proteins except Rh proteins are not depicted.

**Figure 4.** Distribution of rare blood group phenotypes selected by malaria in Africa and South East Asia. The location of rare blood group phenotypes lacking glycophorin B (S-s-), having altered glycophorin C (Ge-,Gerbich negative), Fy (Duffy) blood group null allele (Fy), Sl(a-) allele of complement receptor 1 (CR1) and the band 3 mutation causing South East Asian Ovalocytosis (SAO) in comparison with the distribution of HbS and HbE alleles (71).
Figure 2

Map showing the spread of various Y markers in Europe:
- "Aurignacian" Y marker M173 30,000 to 35,000 years ago
- "Gravettian" Y marker M170 25,000 years ago
- Neolithic Y markers 5,000 to 9,000 years ago

Color coding for frequency of D negative allele:
- 20-30%
- 30-35%
- 35-40%
- 40-45%
- 45-50%
- Over 50%
Figure 4
The relationship between blood groups and disease

David J. Anstee