Common variation in the ABO glycosyltransferase is associated with susceptibility to severe *Plasmodium falciparum* malaria

Andrew E. Fry1,* Michael J. Griffiths1,2, Sarah Auburn1, Mahamadou Diakite1, Julian T. Forton1, Angela Green1, Anna Richardson1, Jonathan Wilson1, Muminatou Jallow3, Fatou Sisay-Joof3, Margaret Pinder3, Norbert Peshu2, Thomas N. Williams2,4, Kevin Marsh2,4, Malcolm E. Molyneux5,7, Terrie E. Taylor6,8, Kirk A. Rockett1 and Dominic P. Kwiatkowski1,9

1The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK, 2Kenya Medical Research, Institute Centre for Geographical Medicine Research (Coast), PO Box 230, Kilifi, Kenya, 3Medical Research Council, PO Box 273, Banjul, The Gambia, 4Nuffield Department of Medicine, John Radcliffe Hospital, Oxford OX3 9DS, UK, 5The Malawi–Liverpool–Wellcome Trust Programme of Clinical Tropical Research, 6Blantyre Malaria Project, College of Medicine, PO Box 30096, Blantyre, Malawi, 7Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK, 8Department of Internal Medicine, College of Osteopathic Medicine, Michigan State University, East Lansing, Michigan 48824, USA and 9Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK

Received October 9, 2007; Revised November 7, 2007; Accepted November 9, 2007

There is growing epidemiological and molecular evidence that ABO blood group affects host susceptibility to severe *Plasmodium falciparum* infection. The high frequency of common ABO alleles means that even modest differences in susceptibility could have a significant impact on the health of people living in malaria endemic regions. We performed an association study, the first to utilize key molecular genetic variation underlying the ABO system, genotyping >9000 individuals across three African populations. Using population- and family-based tests, we demonstrated that alleles producing functional ABO enzymes are associated with greater risk of severe malaria phenotypes (particularly malarial anemia) in comparison with the frameshift deletion underlying blood group O: case–control allelic odds ratio (OR), 1.2; 95% confidence interval (CI), 1.09–1.32; *P* = 0.0003; family-studies allelic OR, 1.19; 95% CI, 1.08–1.32; *P* = 0.001; pooled across all studies allelic OR, 1.18; 95% CI, 1.11–1.26; *P* = 2 × 10⁻⁷. We found suggestive evidence of a parent-of-origin effect at the ABO locus by analyzing the family trios. Non-O haplotypes inherited from mothers, but not fathers, are significantly associated with severe malaria (likelihood ratio test of Weinberg, *P* = 0.046). Finally, we used HapMap data to demonstrate a region of low *F*<sub>ST</sub> (−0.001) between the three main HapMap population groups across the ABO locus, an outlier in the empirical distribution of *F*<sub>ST</sub> across chromosome 9 (∼99.5–99.9th centile). This low *F*<sub>ST</sub> region may be a signal of long-standing balancing selection at the ABO locus, caused by multiple infectious pathogens including *P. falciparum*.

INTRODUCTION

A link between the ABO blood group system and malaria susceptibility has long been suspected. Significant associations between blood group and *Plasmodium falciparum* malaria have been reported from cross-sectional and case–control studies in Brazil, Gabon, India, Sri Lanka and Zimbabwe (1–7) [see also recent reviews and references therein (8,9)]. However, other studies in Colombia, India, Sudan and Nigeria could not find an association between malaria and blood group (10–15). The positive association studies have...
consistently suggested that blood group O individuals are relatively protected from severe malaria. Human erythrocytes infected with mature forms of the *P. falciparum* parasite adhere to uninfected red blood cells, endothelia and other components of the vascular space (16). This adhesive behavior is mediated by *P. falciparum* erythrocyte membrane protein 1 (PfEMP1), which is encoded by a family of highly variant parasite genes and subject to switching during the course of an infection (17). A range of human host molecules binding to PfEMP1 have been identified including CD36 (18), intercellular adhesion molecule (ICAM)-1 (19) and complement receptor-1 (CR1) (20). The ABO A and B antigens have been implicated in the formation of ‘rosettes’, the process by which infected red blood cells (iRBGs) surround themselves with uninfected erythrocytes (21,22). PfEMP1 has been identified as the rosetting ligand of the parasite (20) and the blood group A antigen has been shown to bind to the semiconserved head structure of PfEMP1 (23). Erythrocyte rosetting is linked to the pathogenesis of severe malaria phenotypes such as cerebral malaria (CM) and severe malarial anemia (SA) (24,25). Indeed, fresh isolates of *P. falciparum* from Kenyan children with severe malaria bind A antigen more frequently than strains from children with mild disease (26). Recent work on Kenyan and Malian isolates suggests a significant reduction in rosetting in blood group O individuals compared with non-O blood groups. Furthermore, parasites required rosetting activity for blood group O to offer protection from severe disease (27). Wild and laboratory parasite strains demonstrate blood group preferences, which appear to vary geographically. In general, rosette formation has been shown to occur preferentially with blood group A, B or AB erythrocytes, and particularly, groups A and AB in studies of African strains (21,22,28,29).

Previous epidemiological studies have employed the phenotype of serology to determine the host ABO genotype. Although convenient, serological studies of the ABO system are unable to discriminate all genotypes, e.g. AO heterozygotes from AA homozygotes, and cannot assess subtler coding or non-coding variation. We took the novel approach of investigating the molecular genetic variation underlying this system. The ABO glycosyltransferase performs the terminal step in the biosynthesis of the ABO macromolecule, adding sugar residues to the precursor H antigen. The enzyme adds either N-acetylgalactosamine, to form the A antigen, galactose to form the B antigen, or is functionless leaving the H (or O) antigen unmodified. A variety of polymorphisms have been reported in the ABO glycosyltransferase gene (30). The common key functional variants are: (i) a one nucleotide deletion in exon 6 (codon 87) leading to a reading frameshift and premature termination of the polypeptide before the N-terminal catalytic domain, producing the functionless O allele (31). Alternative ‘non-deletional’ O alleles exist but are rare in African populations (32). (ii) Four non-synonymous single nucleotide polymorphisms (SNPs) (altering residues 176, 235, 266 and 268), which switch enzyme function from A transferase to B transferase activity. The third and fourth SNPs (codons 266 and 268) have the greatest effect on determining the nucleotide–sugar donor used by the transferase, the second amino acid (235) also affects nucleotide–sugar specificity but to a lesser degree, finally the most 5' SNP (codon 176) has very little influence on donor specificity (33).

We designed genotyping assays for the O allele frameshift mutation (rs8176719) and the three key A/B nonsynonymous SNPs in codons 235 (rs8176743), 266 (rs8176746) and 268 (rs8176747) in the catalytic domain. We conducted a multi-centred study design employing samples from three African regions: The Gambia (West Africa), Malawi (South Central Africa) and Kenya (East Africa). We used two study types: population-based association studies in each of the regions employing children with severe malarial disease and local cord blood controls and family-based association studies at each site looking for transmission distortion of ABO alleles between parents and affected offspring, an approach that has the advantage of being robust to population stratification. Our aims were: (i) given the presence of substantial linkage disequilibrium (LD) among the functional SNPs in the ABO locus to define the most efficient marker set for large-scale genotyping; (ii) to perform SNP- and haplotype-based tests for association with severe malarial phenotypes; (iii) to employ our family-based genetic data to check for parent-of-origin effects, an approach not possible with previous case–control and cross-sectional study designs and (iv) to consider the population genetic implications of an association between ABO variation and severe falciparum malaria.

**RESULTS**

**Substantial LD between functional SNPs in the ABO glycosyltransferase gene**

The four key functional polymorphisms that we tested in the ABO locus are reported to be in substantial LD (30). To confirm this, we genotyped 1320 Gambian parent–offspring trios, and 30 Yoruba parent–offspring trios from Ibadan in Nigeria [cell line DNA used by International Haplotype Map (HapMap) Project] (34). This allowed us to look at the fine structure of the ABO locus, combining our functional SNPs with the high-density HapMap marker dataset (Fig. 1).

Genotyping in both sample sets demonstrated near perfect LD ($r^2$, 0.9–1.0) among the three nonsynonymous SNPs that distinguish A and B haplotypes. The frameshift deletion underlying O alleles is in moderate LD with the three nonsynonymous SNPs ($D'$, ~0.9; $r^2$, ~0.3). The majority of O alleles occur on an A haplotypic background, although the minority of recombinants with B haplotypes are still expected to produce truncated products. The implication of this haplotypic structure is that two markers, rs8176719 and one of the three nonsynonymous SNPs, can generally distinguish the three ABO alleles.

**The common frameshift mutation underlying blood group O is associated with protection from severe malaria phenotypes, particularly SA**

Two markers, rs8176719 and rs8176746, were genotyped in 3906 cases of severe malaria plus population and family
controls from The Gambia, Kenya and Malawi. Case–control analyses were performed on 2127 cases of severe malaria and 1931 population controls, and family-based association tests were performed on a different set of 1779 cases of severe malaria and their parents. Further details of severe malaria phenotypes and sample sizes from each population are given in Materials and Methods.

Our analysis took the statistical convention of using the commonest category (i.e. most frequent allele, haplotype or blood group) as reference, and comparing other alleles, haplotypes and inferred blood groups against this category. Single SNP analysis revealed that the minor allele of rs8176719, an insertion relative to the reference sequence (although almost certainly the ancestral allele), is associated with increased risk of severe malaria phenotypes. Consistent trends were found in both family and case–control studies, and across study sites (Table 1). However, some individual studies did not reach statistical significance; probably because of lack of power (see Materials and Methods for power calculations). Overall, the case–control allelic odds ratio (OR) was 1.2 and 95% confidence interval (CI) was 1.09–1.32, \( P = 0.0003 \); family studies allelic OR 1.19, 95% CI 1.08–1.32, \( P = 0.001 \); pooling data across all our studies, both family- and population-based, suggested an allelic OR of 1.18, 95% CI of 1.11–1.26, \( P = 2 \times 10^{-7} \) for severe disease (Fig. 2). The results suggest that the full-length allele of rs8176719 may be associated with a particular risk of anemia during severe malaria infections. For the phenotype of SA, the allelic OR and 95% CI were 1.30 and 1.12–1.5 (\( P = 0.0004 \)) and 1.34, 1.1–1.64 (\( P = 0.004 \)) for the case–control and the family-based studies, respectively.

The situation with rs8176746 is more complex. The major allele, although defining the putatively ‘high risk’ A haplotype, occurs with the ‘low-risk’ frameshift deletion upstream in about two-thirds of chromosomes. Thus, no simple genetic model is significantly associated with disease when this SNP is considered in isolation.

**Figure 1.** Linkage disequilibrium (LD) around the ABO glycosyltransferase gene. Strong LD exists between the four key functional SNPs. (A) Yoruba HapMap parent–offspring trios from Ibadan in Nigeria. (B) 1320 Gambian parent–offspring trios. Both of these West-African populations have near perfect LD between the three nonsynonymous SNPs that differentiate A and B ABO alleles, and moderate but lower LD with the common deletion that generates the O allele. (C) \( r^2 \) values for the Yoruba population across the ABO gene. Phased genotypes (HapMap, July 2006) with additional genotyping for four functional polymorphisms in ABO. The ABO gene (total region shown chromosome 9, \( \sim 24 \) kb, 133156822–133180999, NCBI Build 35) is illustrated 3’ to 5’ with exons 6 and 7 on the left. The SNPs differentiating A and B blood groups are in a high LD block, and indicated by three arrows under exon 7. The frameshift mutation responsible for the O allele (arrow under exon 6) is on the 3’ edge of another high LD block.

**ABO haplotypes producing A and B antigens are associated with susceptibility to severe malaria**

We inferred two SNP haplotypes using rs8176719 and rs8176746. In comparison with O haplotypes, both A (case–control OR, 1.27; 95% CI, 1.12–1.44, \( P = 0.0003 \); family OR, 1.16; CI, 1.03–1.31, \( P = 0.018 \)) and B (case–control OR, 1.13; 95% CI, 0.99–1.28, \( P = 0.06 \); family OR, 1.20; CI, 1.05–1.37, \( P = 0.009 \)) haplotypes are significantly associated with severe malaria (Table 1).

Using an individual’s haplotypes we inferred their ABO blood group (A and B haplotypes codominant, O haplotype recessive). Blood group A, B and AB individuals appear to be at significantly greater risk of severe malaria in comparison with blood group O (e.g. blood group A individuals: case–control OR, 1.33; 95% CI, 1.13–1.56, \( P = 0.00065 \); family OR, 1.29; 95% CI, 1.09–1.53, \( P = 0.003 \) (Table 2). Trends in the data suggest that blood group B individuals may be at subtly lower risk than blood group A, whereas blood group AB individuals are probably at the greatest risk of severe disease (case–control OR, 1.59; 95% CI, 1.15–2.21 \( P = 0.006 \); family OR, 1.46; 95% CI, 1.05–2.04, \( P = 0.025 \)).

Non-O blood groups demonstrate particularly high risk of SA (e.g. blood group A: case–control OR, 1.54; 95% CI, 1.22–1.96, \( P = 0.00039 \); family OR, 1.51, 95% CI, 1.09–2.09, \( P = 0.014 \)). The higher risk of SA experienced by individuals with non-O blood groups may reflect a pathophysiological effect, e.g. accelerated clearance of erythrocytes bound to iRBC. Frequencies of genotypes, haplotypes and blood groups for all cases, controls and parents are documented in Supplementary Material, Table S1.

**ABO glycosyltransferase alleles and a possible parent-of-origin effect in disease susceptibility**

Previous association studies of the ABO system in severe malaria have examined serological data from unrelated
individuals. In this study, having analyzed genetic data from pedigrees, we were in the unique position of being able to look for parent-of-origin effects. We were surprised to find a substantial difference in transmission of high-risk alleles relat-
ging A and B haplotypes to identify the three ABO alleles, and repeating
for maternal and paternal A haplotypes is significant ($P = 0.023$).

There was no significant difference in risk estimate for severe malaria between AO and AA genotypes in the case–control study (AO: OR, 1.35; 95% CI, 1.13–1.60; AA: OR, 1.23; 95% CI, 0.84–1.79; Wald test $P = 0.64$) or family study (AO: OR, 1.31; 95% CI, 1.10–1.55; AA: OR, 1.13; 95% CI, 0.76–1.67; Wald test $P = 0.45$). The same was true for BO versus BB genotypes, and grouping both non-O haplotypes. The phase of AO genotypes was distinguishable in the family study and, as noted above, suggested different risks of severe disease ($A_{\text{mat/Opat}}$: OR, 1.64; 95% CI, 1.27–2.12; $O_{\text{mat/Apat}}$: OR, 1.04; 95% CI, 0.81–1.33; Wald test $P = 0.01$). Differences between these genotypes and the AA genotype were not significant ($A_{\text{mat/Opat}}$ versus AA; Wald test $P = 0.07$; $O_{\text{mat/Apat}}$ versus AA; Wald test $P = 0.92$).

Low levels of population differentiation at the ABO locus represent a signal of long-standing balancing selection

Given that one ABO allele is protective against a major selection pressure such as $P. falciparum$ malaria, it is important to consider the reasons why the ABO system remains polymorphic in Africa. The answer may relate to the balance of protection offered by the specific ABO alleles to other infectious disease. ABO antigens have been implicated not only in malaria, but also to the pathogenesis of Baschieria coli (36), Helicobacter pylori (37), Norwalk virus (38), Hepatitis C (39) and respiratory infections (40). There has also been speculation about whether historical agents such as smallpox and plague may have shaped the global distribution of ABO alleles (41, 42).
Although more attention is generally given to high FST values, FST (using a sample-size weighted metric of population differing sequences 20–30 kb upstream, a region of 85 SNPs, has an locus in HapMap populations (Fig. 4A). The ABO locus includ-
rather than coding sequence itself. Interestingly, we observed score from the ABO locus (45) may reflect positive selection
zygosity signal, which is consistent with a long-standing
SNPs appear to be associated with an extended haplotype homo-
aced selection in all three populations, and would be consist-
whether ABO gene expression is affected by parent of origin.
expression patterns immediately upstream of the gene. Although
empirical distribution of FST in windows across chromo-
some 9 and found the ABO region to be around the 99.5–

cering selection in all three populations, and would be consist-
ance of species-specific mutations underlying the non-human
ancestral enzyme had A transferase activity and that the pre-
species. Sequence analysis has suggested that the common
moter hypermethylation is common in certain forms of malig-
tering selection is also shaping the cis-acting regu-
resembling selection in Africa from P. vivax malaria (47), here the low FST might reflect simultaneous bal-
sequence (while stopping just 3 kb downstream) may indicate that the balancing selection is also shaping the cis-acting regu-
parents, we measured the empirical distribution of FST in windows across chromo-
are of the blood group locus under selection in Africa from P. vivax malaria (47)]. The size of the population differen-
the Duffy blood group locus under selection in Africa from P. vivax malaria (47)].

tations among African, Asian and European populations) of

### DISCUSSION

Our analysis strongly supports the hypothesis that blood group O individuals are relatively protected from severe malaria in comparison with other blood groups, particularly blood group A and AB. In addition to population-based studies, we conducted the first family-based association studies of ABO variation, with severe malaria. The analysis of parent–
offspring trios is robust to artifacts of population stratification and allowed us to examine the possibility of parent-of-origin effects. We found that full-length alleles at rs8176719, particu-
larly A haplotypes, inherited from a mother lead to greater risk for offspring than similar alleles inherited from a father. The analysis suggests a phenomenon such as genomic imprinting rather than an effect of maternal genotype. The tissue- and
cell-type specific expression of ABO is controlled by epige-
nic signals, particularly at a 5’ CpG island in the promoter of the gene (48). The loss of ABO expression mediated by pro-
mechanism is common in certain forms of malignancy (49). There has also been a small case series documenting preferential loss of the maternally derived ABO allele in adult leukemia (50). Our results would be con-
sistent with altered expression of the paternal ABO allele because of imprinting, leaving less target ligand for iRBCs to bind to, or shifting the pattern of sequestration to a ‘safer’ distribution. Paternal allele expression is not abolished, as demonstrated by the existence of blood group AB. However, it is interesting to speculate that during gestation it could be beneficial to modulate or suppress the paternal ABO allele, so as to minimize maternal–fetal ABO incompatibility and the risk of hemolytic disease of the newborn. To our knowl-
edge, this is the first report suggestive of genomic imprinting in the genetics of malaria susceptibility. However, despite our sample size, we have limited power to detect the subtle differ-
ence in risk between maternal and paternally derived alleles, it remains possible that this suggestive result is in fact type I error. Further experiments are required to confirm this finding, and could include family-based studies of ABO vari-
ation in malaria and other complex diseases, or functional experiments with family-derived samples to determine whether ABO gene expression is affected by parent of origin.

The sizes of our individual study datasets give us limited power to analyze the differences between regions. The differences in OR between a family study and a case–control study in a single population are of similar magnitude to the
differences seen between regions, suggesting much of the variation observed is because of statistical fluctuation or methodological issues. Factors possibly contributing to the diversity of risk estimates reported between our populations, and in comparison with other reports include variation in phenotype definition and geographic differences in parasite strain, e.g. differing affinities for ABO antigens. The use of cord blood controls instead of age-matched controls could affect

### Table 2. Association analysis between severe malaria phenotypes and inferred blood group

<table>
<thead>
<tr>
<th>Study/region</th>
<th>Phenotype</th>
<th>Blood group</th>
<th>A OR</th>
<th>95% CI</th>
<th>P-value</th>
<th>B OR</th>
<th>95% CI</th>
<th>P-value</th>
<th>AB OR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case–control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gambia</td>
<td>CM</td>
<td></td>
<td>1.13</td>
<td>0.78–1.61</td>
<td>0.52</td>
<td>0.96</td>
<td>0.67–1.37</td>
<td>0.83</td>
<td>2.14</td>
<td>1.12–4.07</td>
<td>0.021</td>
</tr>
<tr>
<td>Kenya</td>
<td></td>
<td></td>
<td>1.42</td>
<td>0.98–2.07</td>
<td>0.07</td>
<td>1.17</td>
<td>0.8–1.72</td>
<td>0.42</td>
<td>0.93</td>
<td>0.34–2.54</td>
<td>0.89</td>
</tr>
<tr>
<td>Malawi</td>
<td></td>
<td></td>
<td>1.53</td>
<td>1.11–2.12</td>
<td>0.01</td>
<td>1.09</td>
<td>0.8–1.5</td>
<td>0.58</td>
<td>1.32</td>
<td>0.75–2.34</td>
<td>0.34</td>
</tr>
<tr>
<td>All</td>
<td></td>
<td></td>
<td>1.36</td>
<td>1.11–1.66</td>
<td>0.003</td>
<td>1.07</td>
<td>0.87–1.31</td>
<td>0.53</td>
<td>1.51</td>
<td>1.02–2.22</td>
<td>0.04</td>
</tr>
<tr>
<td>Gambia</td>
<td>SA</td>
<td></td>
<td>1.52</td>
<td>1.02–2.26</td>
<td>0.041</td>
<td>1.22</td>
<td>0.82–1.83</td>
<td>0.33</td>
<td>1.76</td>
<td>0.79–3.91</td>
<td>0.17</td>
</tr>
<tr>
<td>Kenya</td>
<td></td>
<td></td>
<td>1.51</td>
<td>1.06–2.16</td>
<td>0.024</td>
<td>1.55</td>
<td>1.1–2.19</td>
<td>0.013</td>
<td>2.05</td>
<td>0.98–4.29</td>
<td>0.06</td>
</tr>
<tr>
<td>Malawi</td>
<td></td>
<td></td>
<td>1.66</td>
<td>0.96–2.86</td>
<td>0.07</td>
<td>1.15</td>
<td>0.65–2.02</td>
<td>0.63</td>
<td>1.33</td>
<td>0.5–3.5</td>
<td>0.57</td>
</tr>
<tr>
<td>All</td>
<td></td>
<td></td>
<td>1.54</td>
<td>1.22–1.96</td>
<td>0.00039</td>
<td>1.36</td>
<td>1.07–1.72</td>
<td>0.012</td>
<td>1.76</td>
<td>1.1–2.81</td>
<td>0.019</td>
</tr>
<tr>
<td><strong>Family-based study</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gambia</td>
<td>CM</td>
<td></td>
<td>1.27</td>
<td>0.95–1.7</td>
<td>0.10</td>
<td>1.07</td>
<td>0.8–1.42</td>
<td>0.65</td>
<td>2.04</td>
<td>1.16–3.58</td>
<td>0.013</td>
</tr>
<tr>
<td>Kenya</td>
<td></td>
<td></td>
<td>1.28</td>
<td>0.99–1.65</td>
<td>0.06</td>
<td>1.22</td>
<td>0.95–1.57</td>
<td>0.12</td>
<td>1.43</td>
<td>0.79–2.6</td>
<td>0.24</td>
</tr>
<tr>
<td>Malawi</td>
<td></td>
<td></td>
<td>1.49</td>
<td>1.08–2.04</td>
<td>0.014</td>
<td>1.06</td>
<td>0.78–1.45</td>
<td>0.70</td>
<td>1.31</td>
<td>0.75–2.3</td>
<td>0.34</td>
</tr>
<tr>
<td>All</td>
<td></td>
<td></td>
<td>1.33</td>
<td>1.13–1.56</td>
<td>0.00065</td>
<td>1.13</td>
<td>0.96–1.33</td>
<td>0.15</td>
<td>1.59</td>
<td>1.15–2.21</td>
<td>0.006</td>
</tr>
</tbody>
</table>

*P*-values are from logistic regression with covariates of ethnic group, gender and HbS genotype (case-control) or case–pseudo-control approach and conditional logistic regression based on parental genotypes (family). Commonest category (O blood group) taken as reference.

Phenotypes: CM, cerebral malaria; SA, severe anemia; Severe, all severe cases.

### Table 3. Evidence of a parent-of-origin effect at the ABO locus

<table>
<thead>
<tr>
<th>Variation/region</th>
<th>Origin</th>
<th>rs8176719</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CLR OR</td>
<td>95% CI</td>
<td>P-value</td>
<td>Wald</td>
<td>TDT</td>
<td>Observed</td>
<td>Expected</td>
<td>TDT ( \chi^2 )</td>
<td>P-value</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gambia</td>
<td>Maternal</td>
<td>1.35</td>
<td>1.1–1.65</td>
<td>0.004</td>
<td>–</td>
<td>–</td>
<td>233</td>
<td>202</td>
<td>9.51</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paternal</td>
<td>1.05</td>
<td>0.86–1.28</td>
<td>0.65</td>
<td>0.13</td>
<td>–</td>
<td>225</td>
<td>212</td>
<td>2.22</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>1.17</td>
<td>1.04–1.32</td>
<td>0.008</td>
<td>1.43</td>
<td>0.94–2.17</td>
<td>0.10</td>
<td>1.46</td>
<td>0.69–3.09</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>Malawi</td>
<td>Maternal</td>
<td>1.51</td>
<td>0.94–2.4</td>
<td>0.09</td>
<td>–</td>
<td>49</td>
<td>40</td>
<td>4.05</td>
<td>0.044</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paternal</td>
<td>1.42</td>
<td>0.74–1.93</td>
<td>0.46</td>
<td>0.55</td>
<td>43</td>
<td>37.5</td>
<td>1.61</td>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>1.35</td>
<td>1.01–1.78</td>
<td>0.04</td>
<td>0.71</td>
<td>113</td>
<td>98.5</td>
<td>4.27</td>
<td>0.039</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>Maternal</td>
<td>1.48</td>
<td>0.92–2.35</td>
<td>0.1</td>
<td>–</td>
<td>46</td>
<td>39</td>
<td>2.51</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paternal</td>
<td>0.89</td>
<td>0.53–1.5</td>
<td>0.67</td>
<td>0.21</td>
<td>31</td>
<td>31</td>
<td>0.00</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>1.17</td>
<td>0.87–1.56</td>
<td>0.3</td>
<td>0.71</td>
<td>98</td>
<td>91</td>
<td>1.08</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>Maternal</td>
<td>1.38</td>
<td>1.16–1.65</td>
<td>0.0002</td>
<td>–</td>
<td>328</td>
<td>281</td>
<td>15.72</td>
<td>0.00007</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paternal</td>
<td>1.05</td>
<td>0.88–1.24</td>
<td>0.6</td>
<td>0.047</td>
<td>299</td>
<td>280.5</td>
<td>2.44</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>1.19</td>
<td>1.08–1.32</td>
<td>0.001</td>
<td>809</td>
<td>743.5</td>
<td>11.54</td>
<td>0.0007</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABO haplotype</td>
<td>Maternal</td>
<td>1.45</td>
<td>1.16–1.8</td>
<td>0.001</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paternal</td>
<td>0.99</td>
<td>0.8–1.22</td>
<td>0.894</td>
<td>0.023</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maternal</td>
<td>1.30</td>
<td>1.03–1.63</td>
<td>0.024</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paternal</td>
<td>1.17</td>
<td>0.92–1.49</td>
<td>0.203</td>
<td>0.58</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CLR, conditional logistic regression; Wald, \( P \)-value for the Wald test comparing estimated parameters for maternal and paternal transmission; TDT, transmission disequilibrium test (66). Observed and expected rates of transmission of the minor allele along with classical TDT statistics are presented for comparison.
is longstanding. The lack of a pronounced EHH SNP in the N-terminal catalytic domain of ABO, one of the functional variants included in model. 

**Table 4. Parent-of-origin likelihood ratio test for rs8176719**

<table>
<thead>
<tr>
<th>Test</th>
<th>Model</th>
<th>−2LL</th>
<th>χ²</th>
<th>df</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>−</td>
<td>−</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent-of-origin effect</td>
<td>Child and maternal genotypes, and PO</td>
<td>1365.582</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Maternal genotype effect</td>
<td>Child genotype and PO only</td>
<td>1368.714</td>
<td>3.133</td>
<td>2</td>
<td>0.21</td>
</tr>
</tbody>
</table>

LL, log likelihood; χ², log ratio test χ² value; df, degrees of freedom, P-value, P-value for the likelihood ratio test; PO, parent-of-origin effect included in model.

**Figure 3.** Extended haplotype homozygosity plots around functional SNPs in the ABO locus. Alleles that have risen rapidly in frequency because of recent positive or balancing selection (e.g. a partial selective sweep) can be surrounded by a region of similar haplotypes that can extend for hundreds of kilobases (45,46). This occurs because recombination has had insufficient time to swap variation between the selected haplotype and others in the population. The decay of homozygosity (EHH) on phased haplotypes partitioned by the alleles of the functional ABO SNPs. (A) rs8176719, the frameshift mutation in exon 6 of the ABO gene, marking O haplotypes, which is associated with protection from severe malaria. (B) rs8176746, a nonsynonymous coding SNP in the N-terminal catalytic domain of ABO, one of the functional variants determining A/B glycosyltransferase activity. The lack of a pronounced EHH signal suggests that the balanced selection affecting variation at the ABO gene is longstanding.

association results. Cord blood samples represent the distribution of genotypes in the populations at birth; since a minority will go on to develop severe disease, this will tend to underestimate the true OR of any genetic susceptibility or resistance allele. In contrast, if blood group O individuals are depleted from the population between birth and mid-childhood (the age of the cases) because of non-malarial pathogens, this would lead to an over-estimation of protection from blood group O. However, given the consistent results between the population- and family-based studies, the similar allele frequencies between cord blood samples and untransmitted parental alleles and the limited effect size found, such biases are probably limited.

Our study employed samples from children with severe malaria phenotypes and it has been noted that the ABO effect is harder to detect in samples incorporating adults with malaria (8). The gradual accumulation of antibodies against a range of iRBC surface antigens is thought to explain a component of the immunity to severe life-threatening malaria that develops in late childhood in endemic regions (51). It is possible that a limited repertoire of parasite epitopes binding A and B antigens could lead to a high prevalence of effective antibodies against these variants in adults, reducing the significance of host ABO blood group in older age groups. It is interesting to speculate whether vaccination, leading to effective immunity to A and B antigen binding PIEMP1 variants, could reduce morbidity among blood group A, B and AB children.

The modest effect size found is typical of validated associations with complex disease in humans. The effect size is likely, in part, to explain why some previous studies have been unable to demonstrate significant results, despite trending toward protection from blood group O, particularly when sample size was low. However, the global health impact of the frameshift deletion, underlying blood group O, needs to be put in context by considering its high frequency. Roughly half the peoples of Sub-Saharan Africa, and many other human populations, at risk of life-threatening disease caused by *P. falciparum*, are homozygotes for this null mutation and are protected by being blood group O (52).

**MATERIALS AND METHODS**

**Subjects**

Patient samples were collected as part of ongoing epidemiological studies of severe malaria at the Royal Victoria Hospital, Banjul, The Gambia; the Queen Elizabeth Central Hospital, Blantyre, Malawi and Kilifi District Hospital, Kilifi, Kenya. Populations in these study sites are exposed to endemic malaria, with the burden of life-threatening disease being experienced by children (<5 years old). Nuclear family trios comprised two parents and one affected child. All DNA samples were collected and genotyped following approval from the relevant research ethics committees and informed consent from participants. Controls were cord blood samples obtained from birth clinics in the same locality as the cases. See Supplementary Material, Table S2 for further demographics of patients and controls.

**Phenotype definition**

All cases were children admitted to hospital with evidence of *P. falciparum* on blood film and clinical features of severe malaria (53,54). In our analyses of sub-phenotypes, we use Blantyre coma score of ≤2 as a criterion of CM, and hemoglobin <5 g/dl or packed cell volume <15% as a criterion of SA. Some individuals had both CM and SA. Of the severe malaria cases that were not CM or SA by these criteria, most had lesser degrees of coma (Blantyre coma score 3) or...
anemia (Hb 5–6 g/dl), or other complications such as respiratory distress. Our samples comprised:

1. 701 Gambian cases (324 with CM and 217 with SA) and 624 controls.
2. 718 Malawian cases (640 CM, 101 SA) and 405 controls.
3. 708 Kenyan cases (216 CM, 270 SA) and 902 controls.
4. 1320 Gambian trios (512 probands with CM and 343 with SA).
5. 225 Malawian trios (216 CM, 39 SA).
6. 234 Kenyan trios (114 CM, 85 SA).

**Power calculations**

Power calculations were performed using the Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc) (55). A single regional case–control study had ~61% power, based on a sample size of 700 cases, 700 unselected controls, allelic OR of 1.2 (our risk estimate for non-O alleles compared with O alleles), high-risk allele frequency of 0.3 and a type I error rate of 0.05. Across all case–control studies, we would expect 97% power. The smaller family trio studies (n = 230 trios) had ~25% power, whereas the larger Gambian study (n = 1320 trios) had ~87% power.

**Sample preparation and genotyping**

Genomic DNA samples underwent whole-genome amplification through either primer extension pre-amplification (PEP) (56) or multiple displacement amplification (MDA) (57), before genotyping on a Sequenom MassArray genotyping platform (58). All assays achieved high rates of genotyping success and no deviations from Hardy–Weinberg equilibrium were encountered (Supplementary Material, Table S2).

**Statistical analysis**

Analysis was performed using STATA (v9.2 for windows) and the genassoc package (http://www-gene.cimr.cam.ac.uk/clayton/software/) written for STATA by David Clayton. In general, the results for the multiplicative model are presented. Although full-length alleles are dominant with regard to blood group phenotype, we found a trend toward increased susceptibility for blood group AB individuals (homozygotes for the full-length allele at rs8176719), which makes this model valid. Case–control association analysis was undertaken by logistic regression and included the covariates of ethnic group, gender and Sickle status. DNA Sequenom genotyping for the Hemoglobin S (HbS) variant was performed for all samples. The HbS results for a proportion of the Gambian samples have been published previously (59). Gender was included in the regression analysis to control for sex-linked traits (e.g. G6PD). With regards to other variation thought to affect malaria susceptibility, Hemoglobin C was absent from our study populations and the range of deletions underlying the Thalassemias are not easily amenable to the high-throughput genotyping technology used.

The family-based association analysis was performed using a case–pseudo-control approach and conditional logistic regression (60). Trios were drawn from a larger pool of samples checked for relationship misspecification. All family samples were genotyped for 48 SNP markers and 15% of trios (indicated by their Mendelian errors rates) excluded from further analysis.

In the population-based studies, two SNP haplotypes (and therefore the ABO alleles) were reconstructed using the snp2hap function, whereas in family studies, phase can often be tracked from parent and child in the creation of the case–pseudo-control dataset. Data from the case–control studies were pooled in a single logistic regression analysis including covariates of ethnic group, gender and Sickle status, whereas family studies were combined in a single case–pseudo-control analysis. Pooling across all case–control and family studies is less straightforward. Here, we used the UNPHASED application version 3.0 (http://www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/) (61,62), which employs a retrospective likelihood framework for performing genetic association analysis, and can be used to combine data from nuclear families and unrelated subjects. Ethnic origin was found to be a significant confounder and was retained as a covariate in the UNPHASED analysis.

**Integrating phased genotypes from the HapMap project**

Phased ABO locus haplotypes for the 30 Yoruba parent–offspring trios from Ibadan in Nigeria were downloaded from the HapMap website (http://www.hapmap.org) (34). Cell lines derived from the Yoruba individuals (obtained from the

---

**Figure 4.** A region of low FST region around the ABO gene. (A) 400 Single SNP FST values for three HapMap populations (CEU, YRI and combined Asian) surrounding the ABO gene. FST drops to a level of ~0.001 in an 85 SNP window across the gene and the noncoding sequence ~20–30 kb upstream. (B) Histogram representing an empirical distribution of FST determined by screening similarly sized windows across chromosome 9. Three different window sizes were used based on either marker numbers (85 SNPs), genetic distance (0.054 cM) or physical distance (50 kb), only windows containing more than two markers were included. By all three distributions, the region around the ABO locus is a relative outlier ~97.5–99.9th centile for low FST.
Coriell Cell Repositories, http://ccr.coriell.org/) were genotyped for the four key SNPs. The additional SNP data was combined with the known phased genotypes using PHASE version 2.1 (63,64). LD patterns between our genotyped SNPs and between these genotypes and HapMap markers were visualized using HAPLOVIEW, version 3.32 (http://www.broad.mit.edu/mpg/haploview/) (65). Extended haplotype homozygosity values (46) were calculated using SWEEP, version 1.0 (http://www.broad.mit.edu/mpg/sweep/index.html).

F_{ST} calculation and window-based screen of chromosome 9

Single nucleotide polymorphism genotyping data for chromosome 9 was downloaded from the HapMap website (Release 21a/Phase 2, Jan 2007). The dataset was derived from three populations: the 30 Yoruba trios; 30 US parent–offspring trios of northern and western European origin, collected by the Centre d’Etude du Polymorphisme Humain (CEPH); and a combined dataset of 45 unrelated individuals from Tokyo, Japan and 45 unrelated individuals from Beijing, in China (34). A total of 94,678 SNPs across chromosome 9 were polymorphic in all three populations. The F_{ST} among the three populations was estimated, in windows across chromosome 9, using the HapMap Consortial publication (34).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

Conflict of Interest statement. None declared.

FUNDING

A.E.F. and M.J.G. were funded by Wellcome Trust Clinical Research Training Fellowships and T.W. by a Wellcome Trust Senior Fellowship. The sample collections and resources used in this study received funding from the UK Medical Research Council, the Wellcome Trust, the Bill and Melinda Gates Foundation, the National Institutes of Health and the European Union Network 6 BioMalPar Consortium. This manuscript is published with the permission of the director of the Kenya Medical Research Institute.

REFERENCES


