REVIEW

Population genetics of malaria resistance in humans

PW Hedrick

School of Life Sciences, Arizona State University, Tempe, AZ, USA

The high mortality and widespread impact of malaria have resulted in this disease being the strongest evolutionary selective force in recent human history, and genes that confer resistance to malaria provide some of the best-known case studies of strong positive selection in modern humans. I begin by reviewing JBS Haldane's initial contribution to the potential of malaria genetic resistance in humans. Further, I discuss the population genetics aspects of many of the variants, including globin, G6PD deficiency, Duffy, ovalocytosis, ABO and human leukocyte antigen variants. Many of the variants conferring resistance to malaria are 'loss-of-function' mutants and appear to be recent polymorphisms from the last 5000–10 000 years or less. I discuss estimation of selection coefficients from case–control data and make predictions about the change for S, C and G6PD-deficiency variants. In addition, I consider the predicted joint changes when the two β-globin alleles S and C are both variable in the same population and when there is a variation for α-thalassemia and S, two unlinked, but epistatic variants. As more becomes known about genes conferring genetic resistance to malaria in humans, population genetics approaches can contribute both to investigating past selection and predicting the consequences in future generations for these variants.
Heredity (2011) **107**, 283–304; doi:10.1038/hdy.2011.16; published online 23 March 2011

Introduction

Today, the knowledge that specific genetic variants can contribute to disease resistance in humans is well known (Dronamraju, 2004; Kaslow *et al.*, 2008). Haldane (1949a) is widely credited with first proposing that genetic resistance to a disease was potentially an important evolutionary force in humans. Before discussing malaria resistance further, it is worthwhile to consider the initial historical contribution and insights of JBS Haldane, one of the founders of population genetics, to the population genetics of malaria resistance in humans.

Acknowledgment to the general review entitled 'Disease and Evolution' by Haldane (1949a) is confusing because there is nothing discussed about human resistance to disease in the body of this manuscript from a meeting in 1948 in Milan. Only in the subsequent discussion from G Montalenti at the end of the article (in Italian), and then from Haldane's response to Montalenti's comment, is this hypothesis put forward. As a result, it is useful to present below a translation of these comments from the Italian and Haldane's brief response.

> MONTALENTI. Stresses the importance of the views expressed by Professor Haldane. Recalls the case of microcitemia or thalassemia studied by Silvestroni, Bianco and Montalenti. Recalls a gene, lethal in the

homozygous state (Cooley's disease) occurs, in the heterozygous state, with such frequency in some populations (more than 10%) that one must admit that it represents in this condition an advantage for individuals that carry it. Because of some research, far from complete, it appears that the gene is more frequent in malarial areas. Professor Haldane has suggested in verbal communication that individuals with microcitemico, who among other characteristics have increased globular resistance, may be more resistant to malaria. However, this is an interesting case of heterosis, which relates to what was described by Professor Haldane.

> HALDANE. 1: I agree with Dr Montalenti's project. Another possibility is that (by analogy with the advantage possessed by vermillion *Drosophila* on media deficient in tryptophan) microcythemic heterozygotes may be at an advantage on diets deficient in iron or other substances, thus leading to anaemia.

In addition, at the VIII International Congress of Genetics in Stockholm in 1948, actually previous to the Milan meeting, Haldane (1949b) outlined his hypothesis in more detail. This was in response to the idea put forth by Neel and Valentine (1947) that different human populations may have different mutation rates.

> Finally, Neel and Valentine (1947) have calculated a mutation rate for the gene which when homozygous gives rise to the lethal condition of thalassemia major (Cooley's anemia), while the heterozygote has a mild microcytic anemia with decreased fragility of the

Correspondence: *Dr PW Hedrick, School of Life Sciences, Arizona State University, Tempe, AZ 85287-4501, USA.*
E-mail: *philip.hedrick@asu.edu*

erythrocytes. The gene is found in about 4.1% of the people of Italian origin in Rochester, NY. If the heterozygote had normal viability, equilibrium would be secured by a mutation rate of 4×10^{-4}. On the other hand, if the heterozygotes had an increased fitness of only 2.1% there would be equilibrium without any mutation at all.

Neel and Valentine believe that the heterozygote is less fit than normal, and think that the mutation rate is above 4×10^{-4} rather than below it. I believe that the possibility that the heterozygote is fitter than normal must be seriously considered. Such increased fitness is found in the case of several lethal and sublethal genes in *Drosophila* and *Zea*. A possible mechanism is as follows. The corpuscles of the anemic heterozygotes are smaller than normal and more resistant to hypotonic solutions. It is at least conceivable that they are also more resistant to attacks by the sporozoa that cause malaria, a disease prevalent in Italy, Sicily and Greece, where the gene is frequent. Similarly, the gene, which causes an anemia similar to that of iron deficiency, might be harmless or even useful to persons on an iron-deficient diet, although causing a relative anemia when the diet is more generous. Numerous other similar hypotheses could be framed. On many of them, the gene would lower fitness in America, but might, at least in the heterozygous condition, have the opposite effect in Greece or Sicily. Until more is known of the physiology of this gene in various environments, I doubt if we can accept the hypothesis that it arises very frequently by mutation in a small section of the human species.

From examining these passages, it seems appropriate to cite Haldane (1949b) as the first source of the idea that genetic variants in humans may confer resistance to malaria (see also the discussion in Crow, 2004; Weatherall, 2004).

At present, it is difficult to imagine that Haldane in 1948 was the first to articulate the potential importance of genetic variants in humans for resistance to malaria (actually, Beet, 1946 found that there were lower rates of malaria in sickle-cell heterozygotes than in normal individuals). The first generally recognized evidence for genetic resistance to malaria was published only a few years later by Allison (1954, 2004) showing that the densities of the malarial parasite, *Plasmodium falciparum*, were significantly lower in sickle-cell hemoglobin heterozygotes *AS* than in normal *AA* individuals. Allison (1955) then discussed the evolutionary implications of his findings in the influential 1955 Cold Springs Harbor Symposium on Quantitative Biology, probably the initiation of the large effort, both medically and evolutionarily, to understand genetic resistance to malaria in humans (see also Allison, 1964). Overall, the original 'malaria hypothesis' of Haldane that diseases such as thalassemia are polymorphisms with an advantage to heterozygotes because of the balance of the benefits of resistance to malaria and detrimental effects due to the disease has been proven correct; however, the actual mechanisms of protection for different genetic variants have often proven to be complicated to understand.

My purpose here is to discuss some of the recent evolutionary genetic developments in human genetic resistance to malaria. Fortunately, in recent years there have been excellent reviews of the mechanisms of malaria resistance, as well as surveys of the many genes that may putatively confer malaria resistance (Kwiatkowski, 2005; Verra *et al.*, 2009; López *et al.*, 2010). Therefore, I will not discuss in detail the molecular basis of the mechanisms of resistance or enumerate all the potential malaria resistance genes, but will focus on the genetic variants in which there is strong evidence for resistance, and data useful for population genetic approaches and analysis. Understanding the geographic patterns of these variants is fundamentally influenced by resistance selection, but would also include consideration of gene flow, mutation, genetic drift and the interaction of all these factors. I will mention these other evolutionary factors as appropriate, but will focus here on the effect of selection on these advantageous variants. I will also discuss the population genetics of two examples in which there are two malarial resistance variants in the same population, the two alleles *S* and *C* at the β-globin locus and then alleles at two different genes, the *S* allele at the β-globin locus and the α^+ thalassemia variant at the α-globin locus.

General background

Surprising to many today, malaria was endemic in southern parts of Europe and North America less than a century ago and was not eliminated from many locations around the Mediterranean Sea until after World War II in the late 1940s. Today malaria is still found across most of sub-Saharan Africa and in extensive regions in Asia and South America. To illustrate the human impact of malaria, in 2008 there were 243 million clinical cases and an estimated 863 000 deaths from malaria (WHO, 2009), many of whom were children, making it one of the leading causes of death worldwide. Most of this mortality is from 'severe malaria,' which is due to *P. falciparum*, begins about 6–14 days after infection from mosquitoes, has the biggest impact on children and pregnant women, progresses rapidly and can cause death within hours or days (here, I will generally be considering the impact of severe malaria). The past mortality rate from malaria was even higher, particularly for populations without previous exposure to malaria. For example, on the west coasts of Africa in the early 1800s, mortality rates in Europeans often exceeded 50% per year (Curtin, 1989). After quinine was introduced in the later 1800s, the mortality rate reduced to about 25% of that before its introduction, indicating that the great majority of the mortality was due to malaria.

It is generally assumed that death from malaria increased between 10 000 and 5000 years ago with the development of agriculture and settlements. These changes are thought to have increased both the density of the malaria mosquito vector and human population density, thereby greatly facilitating the transmission of malaria (Carter and Mendis, 2002). Support for this hypothesis comes from some molecular analysis of *P. falciparum* strains that suggests that the African malaria population expanded around 10 000 years ago and spread to other areas (Joy *et al.*, 2003). In addition, chromosomal data from the mosquito vector Anopheles gambiae support contemporaneous speciation (Coluzzi

et al., 2002 see also; Lawniczak *et al.*, 2010 and Neafsey *et al.*, 2010). Recent analysis suggests that *P. falciparum* in humans is the result of cross-species transmission from gorillas (Liu *et al.*, 2010). Although it is not possible to date this transmission precisely at this point, it appears recent and may be coincidental with the advent of strong selection for malaria resistance in humans.

Although I will generally not discuss the mosquito vector of malaria here, Carter and Mendis (2002) summarized some relevant information about vector–host preference. In most areas where endemic malaria is prevalent, *Anopheles* mosquitoes are generally highly zoophilic, rather than anthropophilic, that is, they greatly prefer animals to humans for blood meals. However, in sub-Saharan Africa, these mosquito vectors are highly anthropophilic. Livingstone (1958) suggested that this was an adaptation to preference for humans in the new agrarian environments that had both high densities of humans and provided numerous small water pools for mosquito development. On the other hand, in Asia and the Middle East, there was an abundance of domesticated animal species during the rise of agriculture, which served as a blood source for mosquitoes; there were generally no such domesticated animal sources in Africa.

P. falciparum is the malarial species that is the most important cause of mortality in humans but, in addition, nearly 3 billion people, mainly in Asia and South America, are at risk from another malaria species, *P. vivax*, which causes chronic disease and deadly complications (Guerra *et al.*, 2010). Three more species, *P. ovale, P. malariae* and *P. knowlesi* (Cox-Singh *et al.*, 2008) can also cause malaria in humans. Infections with malaria may reach very high levels, for example, Mueller *et al.* (2009) found that 73% of a Papua New Guinea population were infected. When the species of malaria infecting individuals were identified, 54.9, 35.7, 13.4 and 4.8% of the individuals were infected by *P. falciparum, P. vivax, P.malariae* and *P. ovale*, respectively, and a significant excess of mixed infections over expectations was detected.

As a result of the high mortality and widespread impact of malaria, it is thought to be the strongest evolutionary selective force in recent human history (Kwiatkowski, 2005). In fact, genes that confer resistance to malaria provide some of the best-known case studies of strong positive selection in modern humans. For example, maintenance of the sickle-cell hemoglobin variant is the classic example of heterozygote advantage, G6PD deficiency is an example of very strong selection at an X-linked locus, variants at β-globin (*S, C* and *E*) and those causing G6PD deficiency (*A-, Med* and *Mahidol*) illustrate selective advantage from a single-nucleotide change, *HLA-B53* provides an example of a selective advantage of a gene conversion product and α-thalassemia provides an example of selection on duplicated loci. Before I examine resistance from some specific genetic variants, let me summarize several important general findings related to genetic resistance to malaria in humans.

(1) **There is extensive overlap of the historical geographical distribution of malaria and human genetic variants that confer resistance to malaria.** For example, *P. falciparum* is found across Africa and Asia, as are the resistance variants of hemoglobin and G6PD. Further, in regions of high endemic malaria, such as sub-

Saharan tropical Africa and lowland Melanesia, there are generally more than one resistant variant present. Resistant variants are generally not present in regions where there is no history of malaria, except in recent immigrants (O'Shaughnessy *et al.*, 1990). For example, there does not appear to be any of the common malaria-resistant variants in native people of the new world (Livingstone, 1985), apparently because their Asian ancestors were not exposed to malaria and malaria was only brought to the new world by early Spanish explorers in the 1600s and 1700s. In addition, there are convincing examples of microgeographic variation in β-thalassemia with higher frequencies at lower altitudes in Sardinia where malaria was historically endemic compared with higher altitudes (Siniscalco *et al.*, 1961) and variation of α-thalassemia (Flint *et al.*, 1986) and β-thalassemia (Hill *et al.*, 1988) in Melanesia correlated with altitude, which is highly correlated with malaria endemicity.

(2) **The presence of strong selective pressure for resistance to malaria has resulted in the high frequency of some detrimental genetic diseases, such as sickle-cell anemia, thalassemia, G6PD deficiency and ovalocytosis.** In fact, Weatherall (2010) has suggested that there is now a global health burden because more than 300 000 children are born each year with a severe inherited hemoglobin disorder that is in high frequency because it confers, or has conferred, malarial resistance. In fact, this strong selective pressure explains why disorders of the red blood cell are by far the most common genetic diseases in humans (Weatherall, 2008). Further, it is possible that some of our other immunological, inflammatory and other chronic diseases found today may be due to the pleiotropic effects of the malaria resistance variants in high frequency because of the evolutionary pressure that malaria has exerted on humans (Kwiatkowski, 2005).

(3) **Many of the variants conferring resistance to malaria are 'loss-of-function' mutants and result in changed, generally reduced, expression or an altered product.** In some cases, these mutants have a large pleiotropic cost associated with them and result in disease, as discussed above. In other cases, the change in expression or in gene product does not appear to have a major cost, as in the Duffy 'null' variants, the O in the ABO system and allele *C* for the β-globin gene.

(4) **Variants at some loci conferring resistance to malaria are nucleotide- or codon-specific whereas others are quite general.** For example, the two known independent Duffy 'null' variants in African and Papua New Guinea are at the exact same nucleotide position, 33 nucleotides upstream from the start codon apparently on different genetic backgrounds. Moreover, two of the three structural polymorphic β-globin variants that confer resistance to malaria, *S* and *C*, are at the same codon. The only other structural variant that appears to provide resistance to malaria, *E*, may provide protection because it alters β-globin expression and functions similar to a thalassemia variant. On the other hand, there are multiple variants that alter expression levels and provide protection within each of the categories known as α-thalassemia, β-thalassemia and G6PD deficiency.

(5) **It appears that many of the genetic variants that provide resistance to severe malaria that have been**

dated are recent, not ancient, polymorphisms and that they have arisen in the last 5000–10 000 years or less (see Table 2). That is, these variants appear to date to a recent time when severe malaria became a strong selective force in humans. This means that these variants did not provide resistance to malaria in other species, that is, they are not trans-specific polymorphisms, and that they have been generated by new mutations in our recent human ancestors. Either there does not appear to be homologous resistance variants in other primates (Verrelli *et al.*, 2006; MacFie *et al.*, 2009) or the malaria resistance variants found in primates are not identical to those in humans (Tung *et al.*, 2009).

(6) **There are many different unlinked genes that appear to confer resistance to malaria in humans.** Because the selection pressure from malaria through high mortality and widespread exposure is great, genes that provide resistance to malaria would be selected even though there may be a cost to resistance from genetic disease. In addition, there appears to be many different types of genes that can confer resistance. For example, Kwiatkowski and Luoni (2008) list potential resistance genes in the categories of hemoglobin genes, erythrocyte surface molecules, other erythrocyte-related proteins, adhesion molecules, mediators of innate immunity and acquired immunity genes. Recent genome-wide studies show the potential to identify many other genes that have a smaller effect on malarial resistance. Further, although much of the variation in malaria resistance appears because of genetic effects, only a small proportion of this variation appears because of well-known genetic factors, such as sickle-cell and α-thalassemia (Mackinnon *et al.*, 2005).

(7) **Some of the genes that confer resistance to malaria are among the most variable genes in the human genome.** The human leukocyte antigen (HLA), G6PD, globin and ABO genes are all among the most variable genes in the human genome and all appear to have variants that confer resistance to malaria. This high variation may be because genomic surveys have shown that disease resistance genes in general have high standing variation and that some of this variability provides resistance to malaria. It may also be because these regions have a high enough mutation rate by which they can generate enough variants quickly enough to provide variation for positive selection for malaria resistance.

(8) **Different populations appear to have different levels of resistance to malaria even when they both live in the same, or similar, area where malaria is prevalent.** These differences could result from non-genetic factors such as culture, nutritional status or patterns of disease transmission or incidence, but in some cases they appear to be from genetic differences such as different frequencies of resistance variants (for example, high frequency of Duffy 'null' in Africa confers resistance to *P. vivax*, see below; high frequency of α-thalassemia in Nepalese populations, Modiano *et al.*, 1991), epistasis with other host resistance variants or differences in unknown genetic resistance factors (Dolo *et al.*, 2005).

Resistance mutants

Many different genes are involved with genetic resistance to malaria (Kwiatkowski, 2005; Verra *et al.*, 2009) and these genetic factors may vary over populations because of the presence or absence of specific resistant genetic variants, different species or strains of malaria and different effects of specific variants in given host genotypes. Other factors such as population structure and chance, as well as the experimental design or the phenotype(s) of malaria diagnosed in a given study, may impact the conclusions. Efforts to use genomics approaches and genome-wide association analysis have been initiated (Ayodo *et al.*, 2007; Timmann *et al.*, 2007; The Malaria Genomic Epidemiology Network, 2008; Jallow *et al.*, 2009; Eid *et al.*, 2010) and some new regions of potential malaria resistance have been identified.

In an estimate of the extent of malaria resistance explained by genetic factors in Kenya using pedigree analysis, 25% was attributed to additively functioning host genes and 29% to unidentified household effects (Mackinnon *et al.*, 2005). Only 2.1% of the variation was attributed to sickle-cell variation and, using an unrelated approach, only about 2% was attributed to α-thalassemia. In other words, a large part of the variation in malaria resistance was attributed to additive genetic effects, but only a small proportion could be attributed to the most well-known genetic factors in the population.

Below I will concentrate on genes for which there is strong evidence for resistance and data that are useful for population genetic approaches and analysis. In general, I will discuss resistance variants that lower mortality from severe malaria, although resistance variants may reduce mortality or morbidity when malaria is not severe or have other impacts, such as reduction in the rate of malaria infection. Some other important genes not discussed here are those that are concerned with the innate immune response (as distinct from the innate resistance) and cytoadherence (Kwiatkowski and Luoni, 2008). Some recent references to genes that are not discussed here are for the tumor necrosis factor (*TNF*) gene (Clark *et al.*, 2009), the toll-like receptor (*TLR*) genes (Ferwerda *et al.*, 2007) and the genes that cause *CD36* deficiency (Fry *et al.*, 2009).

Hemoglobin mutants

Hemoglobinopathies, inherited disorders of hemoglobin structure and function, consist of structural hemoglobin variants and the thalassemias, inherited defects that reduce production in the synthesis of the α- and β-globins of the human adult hemoglobin (Weatherall and Clegg, 2001; Weatherall, 2008). Two α and two β subunits compose the tetrameric protein backbone of adult hemoglobin, and the α and β subunits are made up of 141 and 146 amino acids, respectively, each with three exons. The genes that code for α- and β-globins, *HBA* (there are actually two genes, *HBA1* and *HBA2*, with identical amino-acid sequences) and *HBB*, are unlinked and are located on chromosomes 11 and 16, respectively. Hundreds of structural variants of adult hemoglobin have been documented (Giardine *et al.*, 2007), but only three, all at the *HBB* locus, *S*, *C* and *E*, have reached substantial frequencies in any populations. Two of these three, *S* and *C*, are different mutants (with different phenotypes) at the same amino-acid position 6 and, although *E* is a structural variant at amino-acid position 26, it also affects regulation in a manner similar to the thalassemias. Surprisingly then, the apparently advantageous structural variants are only at two codons in gene

HBB, suggesting that mutants at other sites at the *HBB* locus, and all sites at the *HBA* loci, are either neutral or detrimental in malarial environments (or have not undergone mutation in malarial environments). Finally, every region in the world where malaria was prevalent has hemoglobinopathies, but no two areas have the same hemoglobinopathy or combination of hemoglobino-pathies (Flint *et al.*, 1998). I will discuss some of these patterns below, but here I suggest that they appear to be caused primarily by differential mutation originating in various regions where malaria is prevalent and poten-tially by subsequent selective interaction between the variants.

Sickle cell, *S* or β^s or HbS

The sickle-cell allele (for simplicity I will symbolize it by *S*) was one of the first human genetic variants that was associated with a specific molecular-caused disease (Pauling *et al.*, 1949; Ingram, 2004). The sickle-cell allele is widely known as a variant that causes red blood cells to be deformed into a sickle shape when deoxygenated in *AS* heterozygotes, in which *A* indicates the non-mutant form of the β-globin gene, and also provides resistance to malaria in *AS* heterozygotes. In *SS* homozygotes, *S* causes the severe disease sickle-cell anemia. It is generally assumed that individuals with genotype *SS* had very low or zero fitness when, or if, there was no modern medical care. The greatest documented impact of *S* is that it protects against death or severe malarial disease in heterozygotes, and it appears that *S* has less effect on the rate of infection (Weatherall, 2008). Multiple mechanisms of malaria protection in *AS* individuals have been proposed and supported, including that the growth of malarial parasites is suppressed in sickle cells.

The *S* allele differs from the normal *A* allele at one nucleotide at codon position 6 in the β-globin molecule, which translates into a change from the amino acid glutamic acid to valine (β6 Glu→Val and GAG→GTG; Table 1). The frequency of allele *S* is up to 0.2 in some parts of sub-Saharan Africa, Greece and India. There appears to be five independent origins, or mutations, of the *S* allele from the *A* allele, based on their existence in nearly non-overlapping geographical regions, four in Africa, Bantu (Central Africa Republic, CAR), Benin, Cameroon and Senegal, and another one in both India and Saudi Arabia, and the extensive differences in linked variants resulting in association with different classical haplotypes. Mediterranean *S* alleles appear to descend from the African *S* alleles, many of them from the Benin haplotype (Flint *et al.*, 1998). Recently, Ralph and Coop (2010) have shown that such parallel adaptation of multiple *S* alleles is not inconsistent with their theoretical spatial model and general estimates of mutation rate, gene flow, selection and population size, a finding consistent with that of Karasov *et al.* (2010) who suggest that mutation does not limit adaptation in many situations. On the other hand, it has been proposed that a plausible explanation for much of the haplotype diversity around *S* alleles is that it reflects past recombi-nation and gene conversion (Flint *et al.*, 1998).

A general approach to determine the risk of indivi-duals with a given genotype getting a disease, relative to that in the rest of the population, is to calculate the odds ratio (*OR*) as

$$OR = \frac{f_d(1 - f_c)}{f_c(1 - f_d)} \qquad (1a)$$

where f_c and f_d are the frequencies of the genotype in control and diseased groups, respectively. Another measure of the protectiveness of a genotype from a disease is the relative risk (*RR*), which is

$$RR = \frac{f_d}{f_c} \qquad (1b)$$

OR and *RR* are closest to each other when both f_c and f_d are in low frequency (for an introduction to *OR* and *RR*, see Agresti, 2007).

In a large study of HLA variation and malaria in Gambia, Hill *et al.* (1991) also examined the effect of sickle-cell variation on the presence of severe malaria as a standard to measure the impact of HLA variation. In 619 children with severe malaria only seven (0.012) were *AS*, carriers of the *S* allele. On the other hand, in 510 other children, who were outpatients and termed 'mild controls', many more, that is, 66 (0.129) were *AS*. Further in this study, *OR* = 0.082 and *RR* = 0.093, indicating very great protection (often calculated as 1—*OR* or 1—*RR*) from severe malaria for *AS* genotypes. Ackerman *et al.* (2005) compared this 'case–control' method with a family-based method in more recent samples from Gambia and found similar *OR* values of 0.10 and 0.11.

To understand the dynamics of genetic change from adaptation for resistance to malaria, it is useful to estimate the selection coefficients for the different

Table 1 The variants at the β- and α-globin loci that confer resistance to malaria and information about them, including their chromosomal location, molecular basis of the mutations, estimated number of independent mutations, extent of anemia and their general highest frequency

Gene (chromosome)	Mutation		Anemia in homozygotes	High frequency
	Change	Number		
HBB or β-globin (11p15.5)				
S	Glu6Val GAG/GTG	2–5	Severe	0.2
C	Glu6Lys GAG/AAG	2	None	0.5
E	Glu26Lys GAG to AAG	1–3	Mild	0.7
β-thalassemia	Many	Many	Severe	0.1
HBA or α-globin (16p13.3)				
α-thalassemia	α$^+$ or α one gene deleted	Many	Very mild	0.8
	α^0 or −−both genes deleted	Many	Severe or fatal	0.03

genotypes, values that can be estimated from *OR* data; but first, let me show how *OR* can be used in a population genetics context. Let me define the probability that malaria is present, given the *AS* heterozygote, relative to the probability of malaria when another genotype is present, as *w*. Using the traditional population genetics equation for viability selection (without reproduction; Hedrick 2011a, p. 129), the frequency of the genotype *AS* in the disease group is expected to be

$$f_d = \frac{f_c w}{f_c w + (1 - f_c)} \qquad (2a)$$

This expression can be solved for *w* and is identical to *OR*

$$w = \frac{f_d(1 - f_c)}{f_c(1 - f_d)} = OR \qquad (2b)$$

The selection coefficient (*s*) for genotype *AS* can be estimated as

$$s = m(1 - OR) \qquad (3)$$

where *m* is the rate of non-genotype-specific mortality due to malaria (generally the mortality given that an individual has severe malaria) and 1—*OR* is the protection from malaria conferred by the genotype (after Hill, 1991; Hedrick, 2004). (Note that both Hill (1991) and Hedrick (2004) used the symbol *RR* in this equation, but in fact used the definition of *OR* given in equation (1a), reflecting the similarity of these measures for rare genotypes and the less specific use of these terms in earlier studies). For example, for the *AS* genotype if *OR* = 0.082 as above and *m* = 0.1 (see below), then *s* = 0.092.

Hill (1991) suggested a value of 0.07 for *m* whereas from studies by both WHO (1998) and Carter and Mendis (2002), a value of at least *m* = 0.1 may be appropriate. However, recognizing that the appropriate value of *m* here should reflect the mortality in the evolutionary period over the many centuries when malaria was endemic and that it should account for the cumulative mortality of individuals with malaria before reproduction (and selection should include the impact on reproductive and mating success), an even higher value may be appropriate. For example, Rowe *et al*. (2006) found that in rural West Africans in whom malaria is hyperendemic and there is inadequate medical care, ~20% of the children die of malaria by the age of 5. As a result, I will use *m* = 0.1 here as a general value for illustration, with the caveat that *m*, and the consequent selection coefficient, may actually have been significantly higher in the past.

Because this selection gives an advantage to the *AS* genotype, the relative fitness of the *AS* genotype is the highest (1) and the relative fitness of genotype *AA* is 1—*s* or 0.908; therefore, the relative fitness levels of genotypes *AA*, *AS* and *SS* become 0.908, 1 and 0. From these values, the expected equilibrium frequency of allele *S* (Hedrick, 2011a, p. 136) is 0.092/(0.092 + 1.0) = 0.084, not unlike that observed in many sub-Saharan populations.

The age of two of the putative *S* alleles has been estimated by examining the extent of linkage disequilibrium surrounding the *HBB* locus. Currat *et al*. (2002) examined the *S* Senegal haplotype in a sample from eastern Senegal, in which there has not been recent admixture, and found that all haplotypes were identical. Using a selective advantage of 0.152 for *AS* genotypes over *AA*, they estimated that the age of *S* Senegal alleles was between 45 and 70 generations (1125 and 1750 years using 25 years per generation; see Table 2). Currat *et al*. (2002) suggested that the *S* mutation may have been at low frequency in the population and that this age estimate reflects the time since selection for malaria resistance increased this Senegal haplotype. Modiano *et al*. (2008) estimated the age of the Benin *S* haplotype by estimating the number of generations taken to decay to the observed linkage disequilibrium from the maximum expected when the *S* mutation occurred. They estimated that the age of the mutation was very recent and between 10 and 28 generations (250–700 years).

What is the expected future frequency of *S* alleles, because malaria has been eradicated in many areas and because modern medical care has increased the survival of individuals with sickle-cell anemia? First, heterozygotes *AS* do not appear to have symptoms of anemia or other effects (costs) of carrying the *S* allele in non-malarial environments (Sheng *et al*., 2010), although some negative effects for *AS* individuals during strenuous exercise have been suggested (Connes *et al*., 2008; Tsaras *et al*., 2009). As a result, the National Collegiate Athletic Association in the USA has instituted mandatory testing for sickle-cell carrier status for student athletes in its Division 1 sports (Bonham *et al*., 2010). Second, with modern medical care, individuals with *SS* can live up to adulthood and reproduce with a much less reduction in fitness than without modern medical care. For example, Platt *et al*. (1994) found that the median age at death for individuals with sickle-cell anemia in the United States was 42 and 48 years for males and females, respectively (see similar results in Jamaica in Wierenga *et al*., 2001). As a result, in the absence of malaria and with modern medical care, the relative fitness levels of genotypes *AA*, *AS* and *SS* can generally be indicated by 1, 1 and 1—*s*. With these fitness levels, we can calculate how long it would take for selection (against *SS* homozygotes) to reduce frequency of the *S* allele from q_0 in generation 0 to q_t in generation *t* using the following equation

$$t = \frac{1}{s}\left(\frac{q_0 - q_t}{q_0 q_t} + \ln\frac{q_0(1 - q_t)}{q_t(1 - q_0)}\right) \qquad (4)$$

(Hedrick, 2011a, p. 126).

In African-Americans, the frequency of *S* is about 0.05 and the proportion of individuals with sickle-cell anemia is about 1 in 400. How long would it take for the frequency of the *S* allele to be halved to 0.025 (this decreases the expected proportion of individuals with sickle-cell anemia to 1 in 1600)? With $q_0 = 0.05$ and $q_t = 0.025$, the term in brackets in equation (4) is 20.7. Therefore, the number of generations to halve the frequency of allele *S* is about 20.7/*s*. If we assume that *s* is still quite large, say 0.5, then it will take 41.4 generations, or 1035 years, to halve the frequency, quite a long time (of course if *s* is smaller, the decline would be even slower). The basis for this very slow rate of decrease is that most *S* alleles are in heterozygotes. Assuming that the frequency of alleles *A* and *S* are *p* and *q*, the

Table 2 Estimated age of alleles in generations and years (assuming a 25-year generation length) that confer resistance to malaria

Gene allele	Age		Selection coefficient (s)	Reference
	Generations	Years		
β-globin				
S	45–70	1125–1750	0.152*	Currat et al. (2002)
	10–28	250–700	—	Modiano et al. (2008)
C	75–150	1875–3750	0.04–0.09	Wood et al. (2005)
	38–120	950–3000	—	Modiano et al. (2008)
E	100.3 (62–222)	2508 (1550–5550)	0.079 (0.035–0.099)	Ohashi et al., 2004
G6PD				
A-	254.3	6357 (3840–11 760)	0.044	Tishkoff et al. (2001)
A-	93	2325 (1200–3862)	—	Sabeti et al. (2002)
A-	1800	45 000 (25 000–65 000)	—	Verrelli et al. (2002)
A-	40	1000	0.25	Slatkin (2008)
Med	133.2	3330 (1600–6640)	0.034	Tishkoff et al. (2001)
Med	400	10 000 (0–35 000)	—	Verrelli et al. (2002)
Mahidol	63	1575	0.23	Louicharoen et al. (2009)
Duffy				
ES (null)	1323	33 075 (6500–97 200)	—	Hamblin and Di Rienzo, 2000
ES (null)	310	7750 (3625–13 125	—	Seixas et al. (2002)
ES (null)	490	12 250 (4250–26 500)	—	Seixas et al. (2002)
ABO				
O01	—	1.15 million	—	Calafell et al. (2008)
O02	—	2.5 million	—	Calafell et al. (2008)
HLA-B				
B53	100	2500	0.028*	Hill et al., 1991
B53	86	2150	0.041*	This paper

Here, the age of the allele generally corresponds to the time as strong selection for malaria resistance due to high mortality began the increase in frequency of a given variant. Also given is the estimated selection coefficient where — indicates no selection (neutrality) is assumed in the estimation and the selection coefficients with an * are assumed, not estimated, values.

proportion of S alleles in heterozygotes is $pq/(pq + q^2) = p$. In other words, 95% of the S alleles in this example are in heterozygotes in which selection cannot act against them. Consistent with this prediction, Hanchard et al. (2007) found no significant difference between haplotype frequencies from Jamaica, where there has not been malaria since 1963 (about two generations), and the ancestral African haplotypes. This is in contrast to when selection favored S because of protection from malaria and selection acted favorably on the proportion p of the S alleles that were in heterozygotes and consequently resulted in a fast increase in allele frequency.

Gouagna et al. (2010) showed that there is higher transmission from the human host to the Anopheles mosquito vector of P. falciparum for S- and C-carrying individuals. This is an example of how host genetics can influence the efficiency of malaria transmission, but here the same host genotypes that confer protection against malaria appear to increase its transmission. However, individuals with these protective genotypes also use anti-malarial drugs less frequently because they may have more chronic, asymptomatic infections and therefore may become higher transmitters of infection (Gouagna et al., 2010). Overall, it is not clear as to how these effects influence disease incidence and mortality in genotypically different individuals.

C or βᶜ or HbC
Allison (1956) examined the effects and frequency of the hemoglobin C variant as well as sickle cell in his early studies in Africa. C is much more localized in central West Africa and reaches an allele frequency up to 50% in the Ivory Coast. C provides protection against P. falciparum malaria in both heterozygotes AC and homozygotes CC, although the estimated level of protection in heterozygotes appears to vary in different studies (Modiano et al., 2008 see discussion below). The protective effect of C may result from increased immune clearance of infected erythrocytes. From examination of haplotypes and from its geographic distribution, there appears to be a single origin of C in Africa (Flint et al., 1998). The C variant has also been found in Thailand on a different, non-African haplotype, indicating a second, independent origin (Sanchaisuriya et al., 2001). Because in routine hemoglobin typing, C is often mistaken for variant E, which is common in parts of Thailand, the frequency of C in Thailand is not known precisely.

The change in the β-globin molecule for C is at the same amino-acid position as for S, but results in a change from glutamic acid to lysine rather than to valine (β6 Glu→Lys or GAG→AAG). There does not appear to be any cost from anemia or other traits to either the heterozygote AC or the homozygote CC, although CC homozygotes have symptoms comparable with a mild form of thalassemia. Individuals with the genotype SC have anemia, but it is milder than sickle-cell anemia (the much lower fitness of SC than AS indicates that C is different from the wild-type A allele). As a result, it appears that C could eventually go to fixation in malarial environments and could even replace S (Modiano et al.,

2001; Hedrick, 2004). However, Modiano *et al.* (2008) suggested that *C* is not as widespread as *S* because it is recessive in its advantage for malaria resistance (see below for a discussion).

Wood *et al.* (2005) used a coalescent-based approach to jointly estimate the time since the *C* allele began to increase in frequency because of strong selection and the selection coefficient utilizing data on the present allele frequency (0.15), effective population size and recombination rate. They estimated that the age of the allele was between 75 and 150 generations (1875 and 3750 years) and that the selection coefficient *s* was between 0.04 and 0.09 (assuming that the relative fitness levels of genotypes *AA*, *AC* and *CC* were 1, $1+s$, and $1+2s$, respectively). Modiano *et al.* (2008) calculated the age of the *C* haplotype in Burkina Faso by estimating the number of generations taken for linkage disequilibrium to decay to the observed level from the maximum expected when the *C* mutation occurred. They estimated that the age of the mutation was between 38 and 120 generations (950–3000 years), at the lower range of the estimate of Wood *et al.* (2005).

E or β^E or HbE

Hemoglobin *E* is the most common structural hemoglobin variant in Southeast Asia and has reached an allele frequency of up to 70% in some areas of northern Thailand and Cambodia. In general, neither hemoglobin *S* nor *C* variants are present in Southeast Asia and hemoglobin *E* is generally absent from populations in which *S* and *C* are present. This appears to be the result of historical differences resulting from mutational origin, but little is known about the interaction of *E* with *S* and *E* with *C* (Fucharoen, 2001). It results from the substitution of lysine for glutamic acid in codon 26 (β26 Glu→Lys and GAG→AAG). Besides being a structural variant, the *E* variant also causes the production of an abnormal mRNA with less β-globin being synthesized. As a result, there is a mild thalassemia phenotype in *EE* homozygotes, but it does not appear to have an effect in *AE* heterozygotes (however, there is a much lower fitness of compound heterozygotes with both the *E* and the α- and β-thalassemia haplotypes; Fucharoen, 2001). *AE* heterozygotes appear to have protection from invasion into erythrocytes by *P. falciparum* malaria (Chotivanich *et al.*, 2002). Where *E* is in high frequency, some other red cell disorders, such as α-thalassemia, are also in high frequency.

Although extensive sequence analysis has not been carried out, from examination of the haplotypic backgrounds for *E* alleles in northeast Thailand, more than one mutational origin (Fucharoen *et al.*, 2002) is suggested. However, the *E* allele found in China is on the same haplotype as that found in Thailand (Ohashi *et al.*, 2005), suggesting that it does not have a different origin. From forward simulations, which used the amount of linkage disequilibrium close to the variant and the observed allele frequency, the age of the *E* allele was estimated to be about 100.3 generations (Ohashi *et al.*, 2004), or 2508 years, and the estimated selection coefficient between genotypes *AA* and *AE* was 0.079.

α–thalassemia

As I mentioned above, there are two linked genes with identical amino-acid sequences, *HBA1* and *HBA2*, on chromosome 16 that code for α-globin. There are two primary types of α-thalassemia: α$^+$ thalassemia, in which one pair of the genes is deleted or inactivated by a point mutation, and α^0-thalassemia, in which both pairs of genes are deleted or inactivated. The deleted forms appear to be generated at a relatively high rate by unequal crossing over and Lam and Jeffreys (2006) estimated the rate of deletion generation in sperm from two men as 0.42×10^{-4}. The haplotypes or gametes for the deleted forms with 2, 1 and 0 genes are indicated by αα, −α and −−, and the homozygous forms of α$^+$ thalassemia and α^0 thalassemia are indicated by −α/−α and −−/−−. The α$^+$ thalassemia is the most common monogenic disease in the world; however, as a homozygote, it results in an extremely mild anemia, −−/αα heterozygotes also have a mild anemia, similar to homozygous α$^+$ anemia, that is, the phenotypic effect appears to be the same whether the deletions are on the same chromosome −−/αα or on different chromosomes −α/−α. For heterozygotes with three α genes, or −α/αα, there is no anemia. When individuals are homozygous for deletions of both genes, −−/−−, or α^0 thalassemia, it is lethal and they are stillborn.

The frequency of α$^+$ thalassemia is generally >10% in regions where malaria is prevalent and in some populations, such as in Nepal, parts of India and Papua New Guinea, it is over 80% (Flint *et al.*, 1998). In fact, α$^+$ thalassemia appears similar to a transient polymorphism found in Nepal and Melanesia and it might eventually reach fixation given that selection continues to favor the −α haplotype. However, in sub-Saharan African populations, α$^+$ thalassemia frequencies do not exceed 50% despite intense malaria selection and Williams *et al.* (2005a) suggested that this might occur because of negative epistasis with the *S* allele (see discussion below).

The four different deletions that commonly cause α$^+$ thalassemia (out of about 80 in total) have different somewhat non-overlapping distributions. For example, deletion −α$^{3.7}$ is found nearly worldwide, but is most common in Africa, deletion −α$^{4.2}$ has a high frequency in Southeast Asia and −α$^{3.7I}$ is found most commonly in African, Indian and Mediterranean populations (Flint *et al.*, 1998). In addition, there are many different local forms of these single-gene deletions, as well as of the deletions that eliminate both genes. Although the age of these variants has not been estimated, their localized presence suggests that they are recent, have arisen in different malarial areas and are increasing because of their selective advantage.

The distribution of both α and β thalassemia variants correspond closely to the regions that have historically had high rates of malaria (Weatherall and Clegg, 2001) and the local distribution of these variants also corresponds to endemic malaria (Siniscalco *et al.*, 1961; Flint *et al.*, 1986; Hill *et al.*, 1988). In addition, several studies have shown protection from severe malaria for individuals with α$^+$ thalassemia, compared with individuals without thalassemia (Allen *et al.*, 1997; Mockenhaupt *et al.*, 2004; Williams *et al.*, 2005b). In a large case–control study in Kenya (Williams *et al.*, 2005b), the *OR* values, standardized to the normal αα/αα genotype, for α$^+$ heterozygotes and homozygotes in Kenya are 0.623 and 0.522 for severe malaria, respectively (Hedrick, 2011b). In the other case–control study in Papua New Guinea

Table 3 The different genotypes for α-thalassemia with the number of α-globin genes and the estimated relative fitnesses of genotypes αα/αα, −α/αα and −α/−α from samples from Kenya and Papua New Guinea (Hedrick, 2011b)

Genotype	αα/αα	−α/αα	−α/−α or −−/αα	−−/−α	−−/−−
Number of genes	4	3	2	1	0
Relative fitness					
Kenya	0.945	0.986	1	w_1	0
Papua new guinea	0.922	0.938	1	w_1	0

(Allen *et al*. 1997), the standardized *OR* values of α^+ thalassemia heterozygotes and homozygotes are 0.781 and 0.362, respectively, for severe malaria. Overall, it appears that many haplotypes that reduce expression of α-globin provide a selective advantage in resistance to severe malaria. Such a general advantage to an entire class of variants that reduces expression appears to be also true for β-globin variants and for G6PD-deficiency variants that reduce enzyme expression (see below).

Using the *OR* values mentioned above and equation (3), and assuming that $m = 0.1$, the estimated relative fitness levels of individuals with different genotypes with different numbers of functional α gene copies are given in Table 3. Here, the fitness of −−/−− is 0, as discussed above and the fitness of −−/−α with one gene is w_1 and the fitness of −−/αα and −α/−α individuals are the same.

From these values, I intuitively expected the −α gamete to increase in frequency to fixation, except for unequal recombination generating other gametes, and for there to be strong selection against −− gametes. This is true for the estimated Kenyan relative fitness values mentioned above for all values of w_1 (Hedrick, 2011b). However, for the Papua New Guinea fitness values −α goes to fixation only when genotype −−/−α has a fitness of 0.79 or above. When w_1 is 0.78 or less, haplotype −α does not increase in frequency and haplotypes −− and αα go to a heterozygote advantage polymorphic equilibrium, not substantially different from that observed for some Asian populations (Hedrick, 2011b).

β–thalassemia

Remember that the 'malaria hypothesis' of Haldane (1949b) was proposed to explain the very high level of β-thalassemia in some Mediterranean populations. There is ordinarily only one copy of the *HBB* gene and β^+ and β^0 thalassemia indicate the reduction and loss, respectively, of the production of functional protein. Individuals with β-thalassemia major are often homozygous for loss of production of β-chain mutants (although they may be compound heterozygotes for two mutants), have profound anemia and, if they are not treated with blood transfusions, die in their first year. Heterozygotes typically have mild anemia (β-thalassemia minor), but symptoms can vary greatly in severity from having severe anemia to being a symptomless carrier. In general, β-thalassemia is more of a public health problem because of this higher morbidity than α-thalassemia.

Most of the β-thalassemias are the result of single-nucleotide substitutions or small insertions or deletions.

Around 200 are known worldwide with many different local forms; therefore, the fact that they are the result of different mutations appears clear. Also, because of the local distributions, it appears that the mutants are recent in age. The frequency of carriers of β-thalassemia variants is from 5 to 20% in some areas, although not as high as the frequency of α-thalassemia variants (Weatherall and Clegg, 2001).

Worldwide distribution of β-thalassemia generally coincides with historic distribution of malaria. Malaria is not present today in most areas with higher frequencies of β-thalassemia, such as around the Mediterranean; therefore, case–control studies are not possible. However, Willcox *et al*. (1983) carried out a case–control study in an endemic malaria area of Liberia and found evidence for malaria protection in β-thalassemia heterozygotes.

β-thalassemia variants exist in areas where there are structural β-globin variants and as a result, compound heterozygotes (heterozygotes with two different mutants) occur. For example, in areas of Africa and around the Mediterranean, there are *S* β-thalassemia heterozygotes. The severity of disease in these individuals varies depending on the nature of β-thalassemia allele, but at its most extreme can resemble sickle-cell anemia. In Southeast Asia, where the structural variant *E* is common, the compound heterozygote *E* β-thalassemia is also common (O'Donnell *et al*., 2009) and causes about 1/3 of the cases of severe thalassemia in Sri Lanka.

Enzymopathies

G6PD deficiency

Glucose-6-phosphate dehydrogenase, G6PD, is an important enzyme that catalyzes the first reaction in the pentose-phosphate pathway, part of glycolysis (Cappellini and Fiorelli, 2008). In the erythrocyte, G6PD is the sole source of enzyme activity that protects from the buildup of super-radicals and to withstand oxidative stress. G6PD deficiency is the most common enzymopathy in humans, affecting around 400 million people worldwide with a global prevalence of 4.9%, although there is substantial variation among populations (Nkhoma *et al*., 2009). Although G6PD-deficient individuals are generally asymptomatic throughout their life, factors that cause oxidative stress, such as particular drugs (including several antimalarials), foods or infections, can cause hemolytic anemia and neonatal jaundice. For example, digestion of fava beans can cause severe hemolytic anemia in individuals with the *Med* G6PD deficiency, a condition traditionally known as favism. Low mortality and low morbidity are associated with G6PD deficiency and it does not seem to generally affect life expectancy.

The *G6PD* gene has 13 exons, is 18 kb in length, is on the X chromosome close to the genes for hemophilia and color blindness and codes for a molecule that is 515 amino acids long. The active enzyme is composed of two or four identical 515 amino-acid subunits. The *G6PD* gene is one of the most polymorphic loci in the human genome and about 140 different molecular variants have been identified (Cappellini and Fiorelli, 2008). Most of these have non-synonymous single-nucleotide replacements, a number of which are polymorphic. These variants range from ones that cause severe deficiency and chronic anemia to some that actually increase G6PD

Table 4 G6PD, red blood cell and HLA variants associated with malaria resistance and their nucleotide and amino acid changes, relative enzyme activity, and their distribution and frequency (much of the G6PD information is from Tishkoff and Verrelli 2004)

Gene (chromosome) allele	Mutation	Number of mutations	Enzyme activity or expression	Disease	Distribution	Frequency
G6PD (Xq28)						
A	Asn126Asp A376G		85%	No	Africa	15–40%
A-	Asn126Asp A376G Val68Met G202A	1	12%	Yes	Africa	0–25%
Med	Ser188Phe C563G	2	3%	Yes	Middle east, Mediterranean	2–20%
Mahidol	Gly160Ser G487A	1	5–32%	Yes	Southeast Asia	24%
Duffy (1q21-q22)						
'Null'	T to C substitution at −33 in promoter	2	Not on erthyrocytes in homozygotes, intermediate in heterozygotes	No	Africa, Papua New Guinea	Near 100% in Africa, 2% in Papua New Guinea
ABO (9q34)						
O	Deletion of nucleotide 261 in exon 6	2 major	None in homozygotes	No	South America, Africa, western Europe	Near 100% in native South Americans to about 0.3% in some Asians
Band 3 (17)						
SAO	Deletions in nucleotides for codons 400–408	1	Abnormal structure in heterozygotes	Lethal in homozygotes	Papua New Guinea	20%
HLA-B (6p21.3)						
Bw53	7 amino-acid gene conversion	1		No	Gambia	40%

activity. G6PD deficiency can be caused by a reduction in the number of enzyme molecules or a structural change causing enzyme instability. Because G6PD is on the X chromosome, males with a single deficient G6PD allele are affected. For females to show deficiency, they generally need to be homozygous for the deficient variant, although some heterozygous females can exhibit G6PD deficiency and even develop hemolytic anemia.

The ancestral normal activity G6PD allele *B* is present worldwide, but the variants causing G6PD deficiency are generally more localized to specific geographic areas. Similar to the sickle-cell allele, the worldwide historic distribution of G6PD-deficient mutations reflects the historic distribution of malaria endemicity. For example, much of Africa has three main polymorphic alleles, the ancestral *B* allele with a frequency of about 0.6–0.8, the *A* allele, which differs from *B* by one amino acid (resulting from a difference at one nucleotide) and has a frequency of about 0.15–0.4, and the *A-* allele, which differs from *A* by one amino acid (also resulting from a difference at one nucleotide) and has a frequency ranging from 0.0 to 0.25 (Table 4).

The *A* allele has 85% of normal G6PD activity and has not been demonstrated to provide resistance to malaria in contemporary populations. On the other hand, the *A-* allele has only 12% of normal activity and is thought to provide resistance to malaria (the *A-* allele is always found on the *A* background, so this combination may be necessary for malaria protection). The *Mediterranean* or *Med* allele is the second most common allele found in countries surrounding the Mediterranean Sea, India and Indonesia. It differs from allele *B* by one amino acid

(resulting from a one-nucleotide change) at another site from that of *A*, has only 3% normal activity and is in frequencies ranging from 0.02 to 0.2. The *Mahidol* allele also differs from *B* by one amino acid (also from a one-nucleotide change) at another site, has 5–32% the activity of the normal enzyme and reaches a frequency of 0.24 in some populations in Southeast Asia.

Overall, it appears in general that a reduction in G6PD enzyme activity, due to these variants and others, results in an increase in resistance to severe malaria. This is not unlike the general effect from number of variants at the α-globin and β-globin genes that result both in thalassemia from lowered gene expression and consequently an increase in resistance to severe malaria. The exact mechanism of increased resistance to malaria from G6PD deficiency is not known, but appears to be related to the inability of the malaria parasite to survive and reproduce in stressed G6PD-deficient cells.

To understand the effect of X-linkage on G6PD frequencies, let me assume that the frequency of a G6PD deficiency allele, say *A-* is q (and the frequency of the non-deficiency allele is $1-q$). Because males are hemizygous for genes on the X-chromosome, the frequency of males with G6PD deficiency is equal to the frequency of the *A-* allele, or q. Assuming Hardy–Weinberg proportions, the three female genotypes, *BB*, *BA-* and *A-A-*, have frequencies of $(1-q)^2$, $2q(1-q)$ and q^2, respectively. All female *A-A-* individuals have G6PD deficiency, but at a rate that is only q as much as males. However, some percentage of females heterozygous for G6PD-deficient alleles also show deficient G6PD activity, making this reduction somewhat less. On the other hand,

if DNA analysis is used to detect the deficiency allele in heterozygotes, regardless of whether they show G6PD deficiency, then the proportion of females with the deficiency allele, either as heterozygotes or homozygotes, is $2q(1-q)+q^2 = 2q-q^2$, a value that is nearly twice as high as the frequency in males.

The results of studies examining the risk of malaria for various G6PD-deficient genotypes are not consistent. For example, Ruwende *et al.* (1995) found in Gambia and Kenya that the reduction in risk (1—*OR*) of severe malaria in male hemizygotes (*A*-) was 58% and that the reduction in risk for heterozygous females (*BA*-) was 46%. On the other hand, Guindo *et al.* (2007) found that in two populations in Mali, the reduction in risk of severe malaria in male hemizygotes (*A*-) was also 58%, but in female heterozygotes (*BA*-) there was no reduction in risk. Further, when Guindo *et al.* (2007) examined only those individuals who were wild-type *AA* at the β-globin locus, the effect in males was increased (*OR* = 0.28) whereas there was still no effect in females. Guindo *et al.* (2007) suggested that the difference in their results and that of Ruwende *et al.* (1995) might be based in the control groups used. In addition, Johnson *et al.* (2009) found no protective effect for either male hemizygotes (*A*-) or female heterozygotes (*BA*-) in a Uganda population, but did find a protective effect for females that were found to be G6PD deficient. This perplexing finding appears to be based on the incomplete correlation of genotype and phenotype for G6PD deficiency in female heterozygotes due to variable inactivation of the two X chromosomes (Mason *et al.*, 2007).

If we use the *OR* values from Ruwende *et al.* (1995) and the approach mentioned above to estimate selection coefficients in females (s_f) and in males (s_m), then they are 0.046 and 0.058, respectively. The three female genotypes, *BB*, *BA* and *A-A-* have fitnesses of $1-s_f$, 1 and 1, respectively, assuming that the heterozygote and the mutant homozygote have the same fitnesses (see Saunders *et al.*, 2002). The two male genotypes, *B* and *A*-, have fitnesses $1-s_m$ and 1, respectively. Therefore, if we assume that the initial frequency of *A*- is 0.01, and using traditional equations for change in allele frequency for X-linked genes (Hedrick, 2011a, p. 154), *A*- increases quickly and reaches a frequency near 1 by 200 generations (solid line in Figure 1). Interestingly, the examples considered in all three figures here (G6PD deficiency, *HBC* and α^+ thalassemia) include situations in which a malaria-resistant variant could go to fixation. If there is a symmetrical heterozygote advantage in females because the *A-A-* homozygotes have a lowered fitness due to G6PD deficiency, and no differential selection in males, then the frequency approaches its equilibrium value of 0.5 in about 300 generations (broken line in Figure 1. When the *A-A-* females and the *A*- males have a selective disadvantage, but there still is a heterozygote advantage in females, then the frequency of *A*- slowly approaches its equilibrium value of 0.126 (dotted line in Figure 1), a value not too different from that observed in many populations.

The age of G6PD mutants have been estimated using several different approaches, some of which assume selection (assuming $s_f = s_m = s$) and others do not (Table 2). Tishkoff *et al.* (2001) jointly estimated the age and selection for both *A*- and *Med* alleles using forward simulations and found that the estimated age of *A*- was
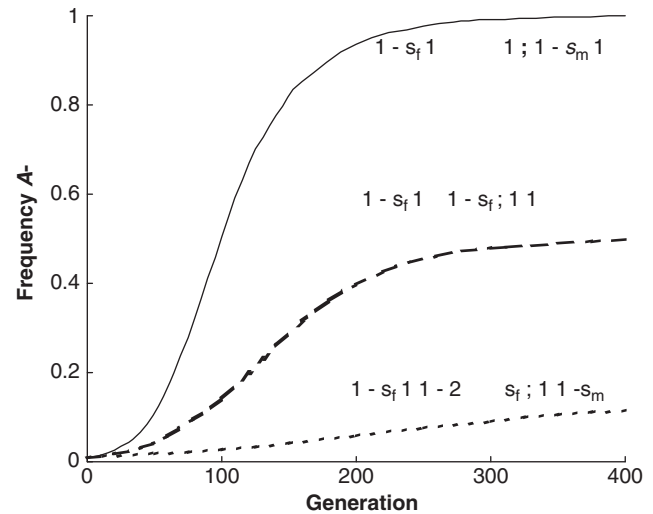

Figure 1 The increase in frequency of allele *A*- at the G6PD locus when it begins at a frequency of 0.01 and fitness of females *BB*, *BA*- and *A-A*- and males *B* and *A*- are for females $1-s_f$, 1, and 1 and for males $1-s_m$ and 1 (solid line), for females $1-s_f$, 1, and $1-s_f$ and for males 1 and 1 (broken line), and for females $1-s_f$, 1, and $1-2s_f$ and for males 1 and 1 -s_m (dotted line), where $s_f = 0.046$ and $s_m = 0.058$ (estimated from the data of Ruwende *et al.*, 1995).

6357 years (254 generations) and the selection coefficient was 0.044, and that the estimated age of *Med* was 3330 years (132 generations) and the analogous selection coefficient was 0.034. From joint estimations using Bayesian methods, Slatkin (2008) estimated that the age of *A*- was only 1000 years (40 generations) and the selection coefficient was 0.25 and Louicharoen *et al.* (2009) estimated that the age of allele *Mahidol* was 1575 years (63 generations) and the selection coefficient was 0.23. The difference in these estimates may be because they were based on different data, with that of Tishkoff *et al.* (2001) based on microsatellite data and that of Slatkin (2008) based on sequence data. The selection coefficients and consequent times in Figure 1 are somewhat more consistent with that estimated by Tishkoff *et al.* (2001) than the estimates of Slatkin (2008) and Louicharoen *et al.* (2009).

Carter and Mendis (2002) suggested that the original selective force for G6PD deficiency could have been either for resistance to *P. falciparum* or *P. vivax*. As evidence for the importance of *P. vivax*, they cite the high frequency of G6PD deficiency in a population in northern Holland, an area where *P. vivax* was prevalent for centuries and *P. falciparum* is thought to have never existed. Recently, evidence has suggested that G6PD deficiency caused by the *Mahidol* mutation in Southeast Asians provides parasite resistance to infection by *P. vivax*, but not to that by *P. falciparum*. Further, G6PD deficiency *Med* in Pakistan has been shown to provide protection from *P. vivax* malaria (Leslie *et al.*, 2010).

Red blood cell surface loci

Two red blood cell loci, Duffy and ABO, that appear to provide protection from malaria, are similar in one sense that they both have two common functional alleles and a defective, loss-of-function allele that consequently

provides protection against malaria. On the other hand, the Duffy variant protects against *P. vivax*, whereas the O phenotype of the ABO system appears to provide protection against *P. falciparum*. The gene for band 3 ovalocytosis, the major transmembrane protein of red blood cells, also has a deletion, loss-of-function, mutant. This mutation provides protection from cerebral malaria, is lethal in homozygotes and appears to provide protection from both *P. vivax* and *P. falciparum*.

Duffy

The Duffy blood group antigen was first observed in 1950 and has two common antigens A and B produced by alleles *FY*A* and *FY*B* (for a review of the Duffy gene and malaria, see Zimmerman, 2004). The Duffy gene is symbolized either by *FY* or *DARC* (Duffy antigen receptor for chemokines). The *FY*B* allele is in highest frequency in Europe whereas the *FY*A* allele is nearly fixed in some Asian populations. Surveys showed that many African-Americans and Africans did not possess either the A or B antigens and had another allele, a Duffy-negative allele that did not produce antigens on erythrocytes and is known generally as a Duffy 'null' allele (*FY*A* and *FY*B* are also called Duffy positive). This allele, *FY*B^{ES}*, now termed erythrocyte silent (ES), is near fixation in sub-Saharan Africa (it is not a true null allele because it is expressed on cell types besides erythrocytes). From analysis of the sequences of these alleles, it appears that *FY*A* is ancestral with amino acid Gly (TGG) at codon 42. Allele *FY*B* differs from *FY*A* with Asp (TAG) at position 42. The *ES* allele in Africa differs from *FY*B* with a T to C SNP mutation, 33 nucleotides upstream from the transcription starting point (−33). This mutation disrupts a binding site for transcription factor GATA-1 on the gene promoter and blocks expression of the gene in erythrocytes.

Several experiments showed that Duffy *ES* homozygotes (*FY*B^{ES}/FY*B^{ES}*) were resistant to infection by two malarial species, *P. knowlesi* and *P. vivax*, but were susceptible to infection by *P. falciparum* (Miller *et al.*, 1976). From these and other studies, it was shown that the Duffy antigen was the obligatory binding receptor used by *P. knowlesi* and *P. vivax* to enter red blood cells and that in homozygous *ES* individuals, no antigen was present on red blood cells and these malaria species could not infect them. In fact, it has been proposed (Hamblin and Di Rienzo, 2000) that the near absence of *P. vivax* in Africa, which is thought to have originated in Asia, occurs because there were no suitable hosts in Africa even though the African environment appears otherwise suitable for *P. vivax*. That is, nearly all Africans were *FY*B^{ES}/FY*B^{ES}*, because of an earlier selective sweep due to some previous selective force when *P. vivax* moved west to Africa from Asia. On the other hand, selection for resistance to *P. vivax* in Africans could have increased the frequency of the *FY*B^{ES}* allele and subsequently eliminated *P. vivax* from Africa (Carter and Mendis, 2002).

There does not appear to be any cost (or disease) associated with the presence of *FY*B^{ES}* homozygotes or heterozygotes. If there were a cost, then because there is no *P. vivax* in Africa and therefore no selection for the *FY*B^{ES}* allele, it would be expected that the frequency of *FY*B^{ES}* would decline. Similarly, the *FY*B^{ES}* allele

frequency in African-Americans would not be expected to decline, even in the absence of *P. vivax* malaria, unless there was a pleiotropic cost to maintaining the allele.

Analysis of the population genetic aspects of the Duffy alleles in Africa is consistent with strong past selection. For example, the F_{ST} value for the *FY*B^{ES}* allele is one of the highest observed for any alleles in humans (Hamblin and Di Rienzo, 2000; Hamblin *et al.*, 2002). Further, the level of variation around the *FY*B^{ES}* allele in Africa is significantly lower than that observed in an Italian sample for the Duffy region. Interestingly, the *FY*B^{ES}* allele in Africa is found on two different haplotypes in three of the five African populations examined by Hamblin and DiRienzo (2000), suggesting either two mutational origins of the *FY*B^{ES}* in Africa or recombination/gene conversion between these haplotypes.

On the basis of phylogenetic relationships of the *FY* sequences, Hamblin and Di Rienzo (2000) estimated the age of the *FY*B^{ES}* allele around 33 000 years ago. As *P. vivax* may have been a significant cause of morbidity before human populations increased in density because of the development of agriculture around 5000 to 10 000 years ago, they suggested that this earlier origin may be consistent with the suggested earlier impact of *P. vivax*. Using a microsatellite locus linked to *FY*, Seixas *et al.* (2002) estimated that the origin of the *FY*B^{ES}* allele was more recent, either 7750 or 12 250 years ago, depending upon the haplotype they used (this assumes a generation length of 25 years).

Surprisingly, a Duffy *ES* allele found in low frequency (0.022) in Papua New Guinea, differed from *FY*A* in exactly the same way at the −33 nucleotide in the promoter region as does the *FY*B^{ES}* allele found in Africa (Zimmerman *et al.*, 1999) and is now called *FY*A^{ES}*. Although this allele appeared to be on a *FY*A* background from analysis of restriction fragment patterns, more detailed sequence analysis could definitively exclude African ancestry for this allele. Zimmerman *et al.* (1999) also documented that there is half as much expression in these heterozygotes *FY*A/FY*A^{ES}* as that in normal homozygotes *FY*A/FY*A*. Further, these heterozygotes were protected from *P. vivax* infection more than normal homozygotes whereas they were not protected from *P. falciparum* infection (Kasehagen *et al.*, 2007). The low frequency of *FY*A^{ES}* appears to be the result of its recent origin, which is reflected in its association with specific alleles at two microsatellite loci about 3 cM from *FY* (Zimmerman *et al.*, 1999). Recently, Ménard *et al.* (2010) also showed that there is about half as much expression in a sample from Madagascar for *FY*A/FY*B^{ES}* heterozygotes as that for for normal homozygotes (*FY*A/FY*A*) or heterozygotes (*FY*A/ FY*B*). In other words, it does not appear that either the expression or the protection from *ES* alleles is recessive, but is intermediate between *ES* homozygotes and genotypes with two normal alleles.

Recently, there have been reports of *P. vivax* in *FY*B^{ES}* individuals from Brazil, Kenya and Madagascar. Perhaps the most definitive study is that of Ménard *et al.* (2010) in Madagascar, a nation where endemic *P. vivax* is prevalent and a diverse ethnic population with both Duffy-negative (*FY*B^{ES}/FY*B^{ES}*) and Duffy-positive (having at least one allele that is *FY*A* or *FY*B*) individuals. In the sample examined of 476 asymptomatic Duffy-negative

individuals from eight locations, *P. vivax* was found in 8.8%. However, they still found a substantial reduction in clinical *P. vivax* malaria among Duffy-negative individuals compared with Duffy-positive individuals. It appears that this population with both Duffy-negative and -positive individuals provided an environment for strains of *P. vivax* to evolve the ability to invade Duffy-negative individuals, giving an example of evolution for utilization of an alternate receptor besides the Duffy antigen for erythrocyte invasion.

SLC4A1 or band 3 ovalocytosis

Gene *SLC4A1* (also known as erythrocyte Band 3 protein gene) codes for the major transmembrane protein of red blood cells and is an anion exchanger in the erythrocyte membrane. One mutation in this gene, found in Papua New Guinea and Malaysia, is due to a 27-nucleotide (9-amino-acid) deletion in a region that is highly conserved throughout species (Jarolim *et al.*, 1991). This deletion results in an abnormal protein structure and function and causes ovalocytosis, an abnormal shape of red blood cells (it is known as Southeast Asian ovalocytosis, or SAO), and mild hemolytic anemia in the heterozygote. The homozygote appears to be lethal because it has never been observed even in progeny from heterozygous matings (Genton *et al.*, 1995).

In populations from lowland Melanesia with high malaria rates, the frequency of this deletion can be as high as 15–20%, but it is not present in the malaria-free highlands. Neither *S* nor *E* hemoglobin variant is present in these populations and only some thalassemias and G6PD deficiency are known to be present to reduce the effects of malaria. Because most individuals in many of these areas are infected with malaria, and often multiple malaria species (Mueller *et al.*, 2009), other unidentified resistance variants appear to be present. There is substantial evidence that deletion heterozygotes are highly protected against malaria. For example, the frequency of the deletion was 0.146 in a sample of healthy Papua New Guinea children whereas it was not present in any of the children with cerebral malaria (Genton *et al.*, 1995 see also; Allen *et al.*, 1999). Carriers also appear to be at lower risk of infection by both *P. vivax* and *P. falciparum* (Cattani *et al.*, 1987; however, see Kimura *et al.*, 2002).

The lethality of homozygotes and the high resistance of heterozygotes to malaria suggest that this deletion variant is somewhat similar to the sickle-cell allele in that the benefit to heterozygotes is high enough to counter the large fitness cost in homozygotes. However, it appears to be a relatively recent mutation because of its limited distribution and a detailed examination of variation at closely linked sites may provide insight as to its age.

ABO

For many decades, the ABO blood group system has been suggested to be associated with diseases and infections, including malaria. The gene consists of seven exons, is more than 18 kb in length and genomic analysis has found over 70 alleles at this locus, suggesting that it is one of the polymorphic genes in humans (Calafell *et al.*, 2008). The ABO glycosyltransferase performs the final step in the production of the ABO molecule by adding sugars to the precursor H antigen. This enzyme either adds the sugars N-acetylgalactosamine or galactose to form the A and B antigens or does not add a sugar, resulting in a functionless H (O) antigen. The three main antigenic classes (A, B and O) are all comprised of numerous alleles determined by both coding and non-coding sequences (Calafell *et al.*, 2008). The silent *O* alleles share a one-nucleotide deletion in codon 87 (nucleotide 261) of exon 6 that results in a frameshift mutation and premature termination of the polypeptide (Yamamoto *et al.*, 1990). The *O* alleles are the most common of the three allelic classes (about 0.6 worldwide) and have frequencies between 0.3 and 0.7 in most populations, the *A* alleles generally have frequencies between 0.2 and 0.3 and the *B* alleles have frequencies between 0.1 and 0.2. Native American or Australian aboriginal populations have *O* allele frequencies of nearly 1.0 (*A* and *B* do not appear to have not been present historically in these populations).

Cserti and Dzik (2007) suggested that there is substantial evidence supporting the importance of allele *O* providing protection against malaria. Their conclusion is based primarily on the consistency between the worldwide distribution of ABO variants and historic presence of malaria, with the *O* allele more common in areas with historic malaria and the association of *O* genotypes with higher resistance to malaria and *A* and *B* genotypes with lower resistance in a number of studies (Uneke, 2007). In a large recent study, Fry *et al.* (2008) showed in three African populations a strong association of O individuals with resistance to severe malaria and found that this effect was recessive, that is, *AO* and *BO* individuals were as likely to be susceptible as *AA* and *BB* individuals and that *AB* genotypes were the most susceptible. In addition, it appears that the *O* allele can protect against severe malaria by a mechanism of reduced rosetting (spontaneous binding of infected erythrocytes to uninfected erythrocytes) (Rowe *et al.*, 2007).

Fry *et al.* (2008) calculated F_{ST} values for the ABO genomic region and found that, even with its polymorphism within populations, it was an outlier for low F_{ST} over populations. They suggested that this similar variation over different populations is the result of similar long-standing balancing selection over populations, perhaps caused by infectious pathogens, including *P. falciparum*. However, Calafell *et al.* (2008) found a more complicated pattern of differentiation over the ABO genomic region. Although the major *O* alleles share a one-nucleotide deletion, they differ in a number of nucleotide substitutions in both exons and introns. Unlike the other recent malaria-resistant alleles discussed here, the *O* human alleles, although different from the *O* chimpanzee allele (Kermarrec *et al.*, 1999), are much older. Calafell *et al.* (2008) estimates that the most common *O* alleles, *O01* and *O02*, are the result of separate mutations and are ∼1.15 and 2.5 million years old, respectively. In other words, assuming that the *O* allele provides protection from malaria, this protection may be present because of selection originally favoring *O* alleles for some other reason.

Immune genes

The genes I have discussed above, such as the hemoglobin mutants and G6PD deficiency, provide innate

resistance to malaria. In addition, although human immunity to malaria is complex (Langhorne *et al.*, 2008), variants in the HLA complex have been shown to provide adaptive immunity to malaria. In general, innate resistance is more important in early childhood survival from malaria whereas adaptive immune response is more important in older children and adults. It appears that malaria protection from particular HLA variants is local and not universal (Hill *et al.*, 1994; Ghosh, 2008). Various factors, such as variation over different populations in malaria strains (and species), in red blood cell polymorphisms, and in HLA polymorphism, may contribute to this heterogeneous response to malaria (Ghosh, 2008).

HLA-Bw53 and DQw5

HLA genes are in the human MHC (major histocompatibility complex) and are the most variable genes in the human genome. They have been shown to be, and have been, under balancing selection, using a number of different approaches (Garrigan and Hedrick, 2003). It is widely accepted that the mechanism of balancing selection at HLA, and for MHC genes in other vertebrates, which retains genetic variants at these genes over species (trans-species polymorphism), is related to the protection that they provide against infectious disease (Hedrick and Kim, 2000). HLA (and MHC) genes can be divided into class I and class II genes in which class I genes recognize intracellular antigens, as from viruses or intracellular stages of parasites, and class II genes recognize extracellular antigens, as from bacteria and parasites. One of the first conclusive demonstrations that HLA variation provided resistance to contemporary infectious disease was the large study showing protection from both specific classes I and II HLA types to malaria in Gambia (Hill *et al.*, 1991).

As there are so many variant forms at many HLA genes (class I genes *HLA-A* and *HLA-B* have over 250 and 500 alleles, respectively; Garrigan and Hedrick, 2003), particular associations may have occurred by chance and after correction for multiple comparisons, statistical significance for particular associations may be lost. As a result, Hill *et al.* (1991) used a two-stage strategy in which they first examined all 45 possible class I HLA-malaria associations in approximately half their samples serologically, and then in stage two in an independent sample examined explicitly the strongest association found in the first group (Bw53), using PCR. As shown in Table 5, the proportions of young children heterozygous for Bw53 were nearly identical in the two independent samples for both the severe malaria and the mild control categories (the mild control group consists of children attending the same hospitals or clinics as those with severe malaria matched for age and area of residence). The calculated *OR* values were nearly identical in the two samples and indicated that Bw53 confers substantial protection from severe malaria in this population. In addition, a class II haplotype (DQw5), consisting of both DQ and DR genes, also showed a strong protective effect in this population (Table 5).

The frequency of individuals with Bw53 (either homozygotes or heterozygotes) is highest in sub-Saharan Africans, reaching 40% in Nigeria and 25% in Gambia, whereas it is very rare (<0–1%) in Asians and Caucasians. Therefore, not only does there appear to be

Table 5 The frequency of HLA class I antigen Bw53 typed in two independent samples (sample sizes in parentheses) of young children by either serotyping or PCR and class II haplotype DQw5 in a population from Gambia for individuals with severe malaria or mild controls (Hill *et al.*, 1991)

	Bw53 (class I)		DQw5 (class II)
	Serotype	PCR	
Severe malaria	0.157 (306)	0.169 (307)	0.057 (436)
Mild controls	0.243 (144)	0.254 (364)	0.105 (958)
OR	0.580	0.597	0.515
s	0.041		0.046

Below are given the odds ratio (*OR*) and the selection coefficient (assuming $m = 0.1$) for Bw53 and for DQw5.

a strong contemporary protective effect, but also its geographical distribution is not unlike that for the sickle-cell allele and indicates past selection favoring Bw53 in malarial environments. Similarly, haplotype DQw5 showed a novel arrangement of genes and it appears that this haplotype is not found, or is very rare, in Caucasians. Because of the high variation in the HLA region, it is not possible to age Bw53 in the way discussed above, but we can estimate the amount of selection and determine the number of generations it would take the genotype to increase in frequency to the present level.

For example, for the Bw53 genotype if $OR = 0.59$ (the average of the values in Table 5) and $m = 0.1$ (WHO, 1998), then $s = 0.041$ from equation (3). Using this amount of selection (and standardizing such that the relative fitness of genotype with the highest fitness is 1) and assuming that the Bw53 allele is dominant, the relative fitnesses of genotypes Bw53/Bw53, Bw53/B(not w53) and B(not w53)/B(not w53) are 1, 1 and 0.959, respectively, where B(not w53) indicates all other alleles at gene *HLA-B*. If we assume that the Bw53 was at an initial frequency of 0.005, as it is now in parts of Europe (Hill, 1991), then using a traditional population genetics selection equation (Hedrick, 2011a, p. 118), it would take ∼86 generations or 2150 years for the frequency of individuals with Bw53 to reach 0.25 (Hill, 1991). This time is substantially less than the time that was expected for strong selection to have occurred for malaria resistance in Africa.

It appears that Bw53 was generated by a gene conversion event in which only seven adjacent amino acids of the most common allele in Gambia, Bw35, were converted to another common motif (Allsopp *et al.*, 1991). Because of the presence of potential donor sequences and the recipient sequence (Bw35) in Gambia, it appears that Bw53 arose locally. As the Bw35 allele does not appear to confer protection from malaria, it appears that this small converted region in its new background provides protection. Further examination of DQw5 by Davenport *et al.* (1995) showed that the higher resistance of this haplotype appeared to be the result of a one-amino-acid difference at the *DRB1* gene and proposed that the resistance from the DQw5 haplotype occurs because this haplotype recognizes an epitope from the malaria parasite.

Multiple resistance variants in the same population

As we find out more information about genes conferring resistance to malaria, we can apply population genetics approaches to predict the eventual outcome of selection when there are two or more variants undergoing selection in a given population. Potentially, different resistance variants may be different alleles at a given locus in which selection may function along with the normal ancestral allele to increase, decrease or stabilize genetic change. Different resistance variants at different loci may be statistically associated (in linkage disequilibrium), resulting in change that is not independent over loci. Moreover, resistance variants at different loci may have fitness values that are not independent over loci (epistasis) and influence genetic change.

There are a number of known genes that could contribute independently (without epistasis) to malaria resistance in areas with high rate of infection; for example, G6PD deficiency and S in sub-Saharan Africa, and Duffy, α-thalassemia and ovalocytosis in Papua New Guinea. All of these genes are unlinked and are present on different chromosomes; therefore, no linkage disequilibrium between the resistance variants at different loci is expected, although there may be strong epistasis between them. However, as more resistance genes become known, some may be linked, so that both linkage disequilibrium and epistasis may become important factors influencing genetic change and overall resistance.

Here, I give two examples, one in which the genetic variants are different alleles at the same gene (S and C) and the other in which the genetic variants are different alleles at different genes, but there is epistasis between them (S and α-thalassemia). Remember that the α-globin and β-globin molecules each contribute two subunits to adult hemoglobin molecules, suggesting that this essential connection could provide potential for gene–gene interaction, or epistasis, as far as fitness. The geographical distributions of α and β thalassemia overlap and in individuals with both, the more severe forms of β$^+$ thalassemia may be changed to a milder condition. Unbalanced production of α and β chains is a major cause of thalassemia, but if both are low, then the severity of anemia can be less. As a result, α-thalassemia can modify the phenotype of β-thalassemia such that it is not as extreme, a potential example of positive epistasis (Penman et al., 2009). In addition, the sickle-cell anemia can be ameliorated by an increase in the production of fetal hemoglobin (Flint et al., 1998), suggesting positive interaction between these genes.

Before I discuss the specific examples, let me provide a general perspective of resistance when several genes are involved. When several different resistance genes are present in the same population, assuming that the resistance variants function independently, the overall expected mean proportion of the population that is resistant to malaria $\bar{R}$ can be expressed as

$$\bar{R} = 1 - \prod_i \left(1 - \sum_j f_{ij}R_{ij}\right) \tag{5}$$

where f_{ij} and R_{ij} are the frequency and resistance, respectively, of genotype j at the ith locus (notice that different loci may have different levels of dominance in this general equation). For example, let me assume that there are two resistant genotypes at locus 1 with frequencies $f_{11} = 0.1$ and $f_{12} = 0.5$ and resistances $R_{11} = 0.5$ and $R_{12} = 0.3$, and one resistant genotype at locus 2 with $f_{21} = 0.3$ and $R_{21} = 0.5$. Therefore, the mean proportion of resistance in the population is

$$\bar{R} = 1 - [1 - (0.1)(0.5) + (0.5)(0.3)][1 - (0.3)(0.5)]$$
$$= 0.32$$

or 32% of the population would be expected to be resistant to malaria.

S and C

The polymorphic structural variants at the β-globin locus, S, C and E, confer resistance to malaria. Variant C is more localized in sub-Saharan Africa than S, but does co-occur with S in many populations. For example, Livingstone (1967) complied survey data on 72 populations from West Africa (about 33 000 individuals) and found a negative correlation between the frequencies of S and C when both genes were present. In general, there were more populations with frequencies of C between 0.075 and 0.15 and $S < 0.05$ and with frequencies of S between 0.075 and 0.15 and $C < 0.05$ than expected. Although some early estimates of fitness suggested that there might be a stable equilibrium for the three alleles A, S and C, a very large data set from Burkina Faso from Modiano et al. (2001) suggested that the C allele would eventually become fixed in regions of malaria (see also Hedrick, 2004; Modiano et al., 2008). Below I will discuss this example and estimate relative fitness levels of the six genotypes when there is simultaneous segregation for S and C in the same population and then use these values to predict change in allele frequencies in this population.

Assume that genotype i provides protection from malaria because the frequency of genotype i in the diseased group is less than that in the control group, or that $f_{id} < f_{ic}$, as I did above. Further, if s_i is the selection coefficient for genotype i, it can be estimated as in equation (3). Because selection for multiple genotypes at the locus is considered here (and relative fitness needs to scaled symmetrically around 1), w_i, the relative fitness of genotype i, is

$$w_i = \frac{1}{1 - s_i} \tag{6a}$$

For example, if $f_{id} = 0.2$ and $f_{ic} = 0.4$ and $m = 0.5$, then $s_i = 0.3125$ and $w_i = 1.4545$.

Now assume that $f_{id} > f_{ic}$, that is, genotype i provides relative susceptibility to malaria. In this case, the selection coefficient can be defined as

$$s_i = m\left(1 - \frac{1}{OR}\right) \tag{6b}$$

Because selection gives a disadvantage to the genotype, the relative fitness of genotype i is

$$w_i = 1 - s_i \tag{6c}$$

For example, if $f_{id} = 0.8$ and $f_{ic} = 0.6$ so that $OR = 2.667$. Further, if $m = 0.5$, then $s_i = 0.3125$ and $w_{i} = 0.6875$. Note

Table 6 The frequency of different genotypes in healthy (control) subjects (f_c) and in malaria (diseased) patients (f_d) and the *OR* for the four genotypes where the odds ratio is statistically significant (Modiano *et al.*, 2001)

	Genotype					
	AA	*AC*	*AS*	*CC*	*SC*	*SS*
f_c	0.6641	0.2172	0.0954	0.0165	—	—
f_d	0.8036	0.1641	0.0275	0.0012		
OR	2.070	0.7075	0.2681	0.0715	—	—
Relative fitness	0.948	1.030	1.079	1.102	—	—
	0.861	0.935	0.979	1	0.498	0.109

Below are the estimated relative fitnesses for the four genotypes assuming the mortality from malaria is $m = 0.1$. In the bottom row, these values are standardized by the fitness of genotype *CC*, the genotype with the highest fitness (Hedrick, 2004).

that $1/0.6875 = 1.4545$, or that the fitness levels in these examples are symmetrical around 1.

In the study by Modiano *et al.* (2001), there were statistically significant effects for the hemoglobin genotypes *AA*, *AC*, *AS* and *CC* (*AC*, *AS* and *CC* showed relative resistance and *AA* showed relative susceptibility). Table 6 lists the frequencies of these four genotypes in the control (healthy subjects) and disease (malaria patients) groups and the resulting *OR* values. The sample sizes for genotypes *SC* and *SS* were very small and the fitness levels used here were based on deviations from Hardy–Weinberg proportions (Hedrick, 2004). Assuming $m = 0.1$, Table 6 lists the relative fitness for these genotypes and also the fitness levels relative to the genotype with highest fitness, *CC*. Using these fitness levels, there are no stable or unstable three-allele equilibria, unlike earlier data (Hedrick, 2004).

With this fitness level, if *S* is introduced by mutation or gene flow into a monomorphic *A* population, it will quickly go to a stable two-allele equilibrium with the equilibrium frequency of *S* equal to 0.12. If *C* is introduced by mutation or gene flow into a monomorphic *A* population, it will increase in frequency to fixation in slightly over 100 generations (Figure 2)). When the *S* and *A* alleles are at their stable equilibrium, *C* can enter the population and eventually increase to fixation after about 200 generations. When both the *C* allele and the *S* allele are introduced simultaneously at low frequencies (0.01 here), the *S* allele will be eliminated (in about 100 generations in the example in Figure 2).

These outcomes occur primarily because genotype *CC* has the highest estimated relative fitness of any genotype. It may also explain particular situations as, for example, why the Dogon people of Mali have a much lower frequency of *S* than most other West Africans and have a high frequency of *C* (Agarwal *et al.*, 2000). Of public health significance, the mean fitness of the population will be 11.2% higher when the population is fixed for *C* than when it is polymorphic for *A* and *S* because of higher average resistance to malaria and the absence of sickle-cell anemia.

Modiano *et al.* (2008) suggested that *C* is not as widespread as *S* because it is recessive in its advantage for malaria resistance. To estimate dominance for the favorable *C* allele, we can standardize the fitness levels so that genotypes *AA*, *AC* and *CC* have fitness levels of 1, $1 + hs$ and $1 + s$, respectively, where h is the level of
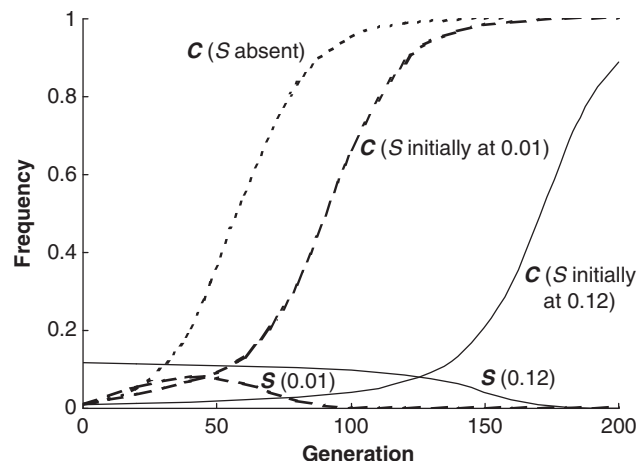

Figure 2 The increase in frequency of allele *C* when it begins at a frequency of 0.01 (short, broken line), when *S* also begins at a frequency of 0.01 (solid lines), and when *S* begins at its equilibrium frequency of 0.0895 (long, broken line) (after Hedrick, 2004). The change in frequency of *S* is also given for the last two situations.

Table 7 The fitness model used to estimate the level of dominance h for the *C* allele and the relative fitnesses standardized against *CC* and then *AA*

	Genotype			s	h
	AA	*AC*	*CC*		
General	1	$1+hs$	$1+s$		
Hedrick, 2004	0.861	0.935	1		
$w_{AA} = 1$	1	1.086	1.166	0.166	0.518

The estimated values of s and h are given for the data in the bottom row.

dominance for the favorable *C* allele (Table 7). Given the estimated fitness levels and assuming that the fitness level of *AA*, *AC* and *CC* are w_{AA}, w_{AC} and w_{CC}, respectively, then

$$s = w_{CC} - w_{AA} \tag{7a}$$

and

$$h = \frac{w_{AC} - w_{AA}}{w_{CC} - w_{AA}} \tag{7b}$$

Using the data in Table 7, the estimates are of $s = 0.166$ and of $h = 0.518$. In other words, it appears that the advantageous effect of the *C* allele is very close to additivity ($h = 0.5$) and is not recessive in its advantageous effect. Interestingly, if the *C* allele is more recessive, that is, the fitness of *AC* is lower, then the *C* allele cannot invade a population in which the *S* allele is at equilibrium.

S and α-thalassemia
Two hemoglobinopathies, sickle cell and α^+-thalassemia, occur in high frequency in sub-Saharan Africa. Williams *et al.* (2005a) examined these disorders in Kenya in a population that was segregating for both variants. In this population, they found that the protection from malaria

Table 8 The fitness values for β-globin genotypes are given on the left column when the normal genotype αα/αα is present at the α-globin locus and the fitness values for the α-globin genotypes are given across the top when the normal genotype AA is present at α-globin locus

β-globin genotype	α-globin genotype		
	αα/αα (1)	−α/αα (1+ht)	−α/−α (1+t)
AA (1−s)	1−s	(1−s)(1+ht)	(1−s)(1+t)
AS (1)	1	1+ht	1+t
SS (0)	0	0	0

The values in the table are the expected fitnesses, assuming no epistasis, for the two-locus genotypes obtained by multiplying the values on the left column by the values across the top.

Table 9 (a) The estimated annual malaria mortality rates for two-locus genotypes at the α- and β-globin loci (Williams et al., 2005a) and (b) the estimated relative survival values for the two-locus genotypes are given

β-globin genotype	α-globin genotype		
	αα/αα	−α/αα	−α/−α
(a) AA	0.02757	0.01578	0.01433
AS	0	0.00340	0.02733
(b) AA	0.552	0.683	0.704
AS	1	0.909	0.554
SS	0	0	0

The survival values are obtained by adding a non-genotype specific annual mortality of 0.034 to each of the mortality rates, taking the inverse, and then standardizing the resulting values so the maximum value occurs for genotype AS αα/αα.

given by each variant was lost when the two disorders were inherited together in the same individual (see also May *et al.*, 2007). They suggested that this negative epistasis could be the reason why α^{+}-thalassemia has not been fixed in any population in sub-Saharan Africa. This situation has been examined theoretically by Williams *et al.* (2005a) and Penman *et al.* (2009), using an epidemiological approach. Here, I will use a traditional population genetics approach and show that the results are generally consistent with their findings, using an epidemiological model. An important caveat is that although most general properties discussed here are enlightening and significant, some conclusions can depend on rather small differences in estimated values (Williams *et al.*, 2005a; Penman *et al.*, 2009; Hedrick, 2011b).

As background for discussion of the fitness array for these two loci, let us examine what the expectation is when there are independent effects on viability for the two loci (no epistasis on a multiplicative scale Hedrick, 2011a, p.554). Table 8 lists on the left margin, the relative fitness values for the β-globin genotypes AA, AS and SS of 1−s, 1 and 0, respectively, and on the top, the relative fitness values for α-globin genotypes αα/αα, −α/αα and −α/−α of 1, 1 + ht and 1 + t, respectively. The product of these marginal values yields the expected fitness levels in the table.

The expected relative viabilities for two-locus genotypes without epistasis can be calculated in the following manner. First, the difference in the relative values for genotypes AS αα/αα and AA αα/αα yields an estimate of s. The fitness of genotype AA αα/αα divided by the fitness of genotype AA −α/−α yields an estimate of 1 + t. Therefore, for the example in Table 9b that I will discuss below, s = 0.448 and t = 0.275, resulting in a value of 1.275 for the expected fitness without epistasis of genotype AS −α/−α. The estimated value for this genotype of 0.554 is substantially lower than this expectation and illustrates negative epistasis.

To calculate the relative viabilities of the different two-locus genotypes (the approach used here was suggested by B. Penman), genotype-specific annual mortality rates were used based on the relative rates of hospital admissions in children with severe malaria (Table 9a; from legend of Figure 2 in Williams *et al.* (2005a)). To each of these mortality rates, an annual non-genotype specific mortality rate was added (in the example here, a rate of 0.034 was added) to yield a total annual mortality rate for each genotype. The reciprocal of these annual

mortality rates provides an estimate of the relative life expectancies. When these values are standardized by the largest relative life expectancy (for these data the genotype AS αα/αα has the maximum value of 29.41), the relative viability values in Table 9b are obtained.

When there is a wild-type genotype at both loci, the estimated relative survival is only 0.552. Notice that both the presence of S in AS and the presence of −α in both genotypes −α/αα and −α/−α increases the estimated relative survival, given normal genotypes at the other locus. However, when there is an AS genotype and homozygosity for −α, the relative survival is reduced to only 0.554, essentially the same relative survival for the wild-type genotype.

There are several things to note about this array of relative viability values. First, when only the normal αα haplotype at the α-globin gene is present (first column), there is a stable polymorphism with S having a frequency of 0.309. Second, when only the normal A allele is present at the β-globin locus (first row), the −α/−α genotype has the highest fitness, predicting that this variant haplotype would be fixed (unless there is a negative effect of anemia). Overall then, the normal AA αα/αα genotype has a low relative viability in this malarial environment.

We can intuit some two-locus behavior by examining these viability values further. For example, if the population is polymorphic for A and S, it appears that −α will be able to increase from a low frequency in the population (enter) because the relative viability of AA −α/αα is greater than that of AA αα/αα (the most frequent genotype at this point; first row of Table 9b) and AS −α/αα is only slightly less viable than AS αα/αα (second row). On the other hand, if the population is fixed for −α−α, it does not appear that S can enter the population because AA −α/−α > AS −α/−α (last column of Table 9b).

However, the genetic behavior of this two-locus fitness array has some further interesting properties that are observed if the two-locus gamete frequency equations with selection (Hedrick, 2011a, p. 559) are iterated. First, if the population is polymorphic at the equilibrium for A and S and −α haplotype enters the population, the system will increase to a two-locus stable equilibrium with −α having a frequency of 0.272 and S having a frequency of 0.256, with no linkage disequilibrium.
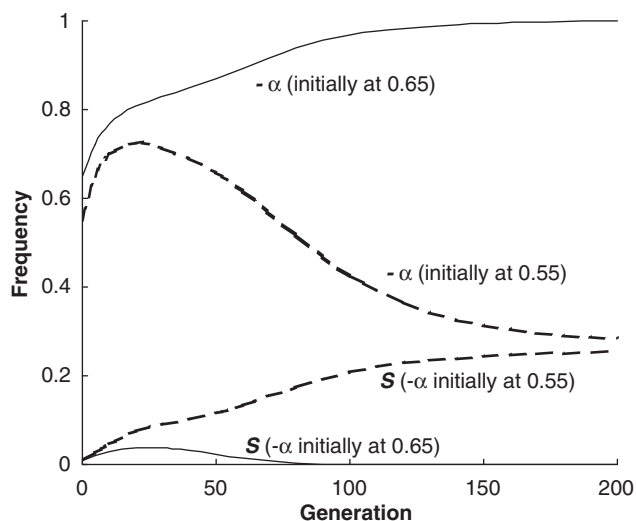
Figure 3 The change in the frequency of S where the initial frequency is 0.01 and the change in the frequency of $-\alpha$ where the initial frequency is either 0.55 or 0.65 using the relative fitness levels in Table 9b.

Second, if the $-\alpha$ is polymorphic, there is an unstable equilibrium such that $-\alpha$ will, depending upon its initial frequency, either continue to increase or go to the two-locus equilibrium discussed above, if S is introduced into the population. Figure 3 presents this situation in which the initial frequency of S is 0.01 and the initial frequency of $-\alpha$ is either 0.55 or 0.65. For 0.55, the frequency of $-\alpha$ initially increases and then approaches the two-locus equilibrium at 0.272. On the other hand, when the initial frequency is 0.65, it continues to increase toward unity whereas S initially increases but then decreases to 0. In other words, the presence of S maintains the frequency of $-\alpha$ lower when the fitness of AS $-\alpha/-\alpha$ is low (high negative epistasis). If fitness of AS $-\alpha/-\alpha$ is somewhat higher (less negative epistasis), then $-\alpha$ can be fixed for all initial frequencies.

Conclusions

Genetic resistance to malaria in humans provides a rich source of data and examples of many aspects of population genetics, particularly cases of strong selection. For example, maintenance of the sickle-cell hemoglobin variant is the classic example of heterozygote advantage, G6PD deficiency is an example of very strong selection at an X-linked locus, variants at the β-globin gene (S, C and E) and causing G6PD deficiency (A-, Med and $Mahidol$) illustrate selective advantage from a single-nucleotide change, HLA-$B53$ provides an example of a selective advantage of a gene conversion product and α-thalassemia provides an example of selection on duplicated loci. In addition, understanding the geographic patterns of these variants includes consideration of selection as well as gene flow, mutation, genetic drift and the interaction of these factors. For example, to illustrate the importance of gene flow, most Mediterranean S alleles appear to descend from the African Benin haplotype (Flint et $al.$, 1998) and the Southeast Asian deletion of both α- globin genes, $--^{SEA}$, is widespread in a number of Asian populations (Chui and Waye 1998; Bernini 2001). Estimating the age during which these

variants have been strongly selected also includes incorporating information about recombination in the genomic region surrounding the selected variants.

An aspect of population genetics that will probably become even more important is the joint impact of multiple malaria resistance variants. Here, I discuss future changes in populations segregating for the two β-globin variants S and C and predict that, given the continued presence of malaria, selection will result in the elimination of S and fixation of C. In populations segregating for α-thalassemia and S, variants at two different loci, because of negative epistasis, either the α-thalassemia haplotype is predicted to be fixed and the S allele eliminated or a stable equilibrium of both variants is predicted depending upon the starting frequencies. Although these results are initially somewhat counterintuitive, one needs to realize that they are dependent upon specific estimated selection coefficients with large confidence intervals and which may vary over time (in the past or in the future) and may vary in different populations.

Obviously, Haldane (1949b) was correct to point out the potential importance of genetic variants in humans for resistance to malaria. Genes conferring resistance to malaria provide some of the best-known case studies of strong positive selection in modern humans and the sickle-cell hemoglobin variant remains the classic example of heterozygote advantage. Overall, the original 'malaria hypothesis' of Haldane that diseases such as thalassemia are polymorphisms with an advantage to heterozygotes in malarial environments has been proven correct; however, much is still to be learned about actual mechanisms of protection, other genes that confer resistance and the population genetics of this variation.

Conflict of interest

The author declares no conflict of interest.

Acknowledgements

References

Ackerman H, Usen S, Jallow M, Sisay-Joof F, Pinder M, Kwiatkowski DP (2005). A comparison of case-control and family-based association methods: the example of sickle-cell and malaria. Ann Hum $Genet$ **69**: 559–565.

Agarwal A, Guindo A, Cissoko Y, Taylor JG, Coulibaly D et $al.$ (2000). Hemoglobin C associated with protection from severe malaria in the Dogon of Mali, a West African population with a low prevalence of hemoglobin S. $Blood$ **96**: 2358–2363.

Agresti A (2007). An $Introduction$ to $Categorical$ $Data$ $Analysis$, 2nd ed. John Wiley & Sons: New York.

Allen SJ, O'Donnell A, Alexander ND, Alpers MP, Peto TE et $al.$ (1997). Alpha+thalassemia protects children against disease caused by other infections as well as malaria. $Proc$ Nat $Acad$ Sci USA **94**: 14736–14741.

Allen SJ, O'Donnell A, Alexander NDE, Mgone CS, Peto TEA et $al.$ (1999). Prevention of cerebral malaria in children in

Papua New Guinea by Southeast Asian ovalocytosis band 3. *Am J Trop Med Hyg* **60**: 1056–1060.

Allison AC (1954). Protection afforded by the sickle-cell trait against subtertian malaria infection. *Br Med J* **1**: 290–294.

Allison AC (1955). Aspects of polymorphism in man. *Cold Spring Harbor Symp Quant Biol* **20**: 239–251.

Allison AC (1956). The sickle-cell and haemoglobin C genes in some African populations. *Ann Hum Genet* **21**: 67–89.

Allison AC (1964). Polymorphism and natural selection in human populations. Aspects of polymorphism in man. *Cold Spring Harbor Symp Quant Biol* **29**: 137–149.

Allison AC (2004). Two lessons from the interface of genetics and medicine. *Genetics* **166**: 1591–1599.

Allsopp CEM, Hill AVS, Kwiatkowski D, Hughes A, Bunce M et al. (1991). Sequence analysis of HLA-Bw53, a common west African allele, suggests and origin by gene conversion of HLA-B53. *Hum Immunol* **30**: 105–109.

Ayodo G, Price AL, Keinon A, Ajwang A, Otieno MF et al. (2007). Combining evidence of natural selection with association analysis increases power to detect malaria-resistance variants. *Am J Hum Genet* **81**: 234–242.

Beet EA (1946). Sickle cell disease in the Balovale district in Northern Rhodesia. *East Afr Med J* **23**: 75–86.

Bernini LK (2001). Geographic distribution of α thalassemia. In: Steinberg MH, Forget BG, Higgs DR, Nagel RL (eds). *Disorders of Hemoglobin: Genetics, Pathophysiology and Clinical Management*. Cambridge University Press: Cambridge. pp 878–894.

Bonham VL, Dover GJ, Brody LC (2010). Screening student athletes for sickle cell trait-a social and clinical experiment. *New Eng J Med* **363**: 997–999.

Calafell F, Roubinet F, Ramirez-Soriano A, Saitou N Bertranpetit J, Blancher A (2008). Evolutionary dynamics of the human *ABO* gene. *Hum Genet* **124**: 123–135.

Cappellini MD, Fiorelli G (2008). Glucose-6-phosphate dehydrogenase deficiency. *Lancet* **371**: 64–74.

Carter R, Mendis KN (2002). Evolutionary and historical aspects of the burden of malaria. *Clin Microbiol Rev* **15**: 564–594.

Cattani JA, Gibon FD, Alpers MP, Crane GG (1987). Hereditary ovalocytosis and reduced susceptibility to malaria in Papua New Guinea. *Trans Roy Soc Trop Med Hyg* **81**: 705–709.

Chotivanich K, Udomsangpetch R, Pattanapanyasat K, Chierakul W, Simpson J et al. (2002). Hemoglobin E: a balanced polymorphism protective against high parasitemias and thus severe *P. falciparum* malaria *Blood* **100**: 1172–1176.

Chui DHK, Waye JS (1998). Hydrops felatis caused by α-thalassemia: an emerging health care problem. *Blood* **91**: 2213–2222.

Clark TG, Diakite M, Auburn S, Campino S, Fry AE et al. (2009). Tumor necrosis factor and lymphotoxin-α polymorphisms and severe malaria in African populations. *J Infect Dis* **199**: 569–575.

Coluzzi M, Sabatini A, Torre A, Di Deco MA, Petrarca V. (2002). A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* **298**: 1415–1418.

Connes P, Reid H, Hardy-Dessources M-D, Morrison E, Hue O (2008). Physiological responses of sickle cell trait carriers during exercise. *Sports Med* **38**: 931–946.

Cox-Singh J, Davis TME, Lee K-S, Shamsul SSG, Matusop A (2008). *Plasmodium knowlesi* malaria in human is widely distributed and potentially life threatening. *Clin Inf Dis* **46**: 165–171.

Crow JF (2004). J. B. S. Haldane's ideas in biology with special reference to disease and evolution. In: Dronamraju KR (ed). *Infectious Disease and Host-Pathogen Evolution*. Cambridge University Press: Cambridge. pp 11–17.

Cserti CM, Dzik WH (2007). The ABO blood group system and *Plasmodium falciparum* malaria. *Blood* **110**: 2250–2258.

Currat M, Trabucher G, Rees D, Perrin P, Harding RM et al. (2002). Molecular analysis of the β-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the β^s Senegal mutation. *Am J Hum Genet* **70**: 207–223.

Curtin PD (1989). *Death by Immigration: Europe's Encounter with the Tropics in the Nineteenth Century*. Cambridge Univ Press: New York.

Davenport MP, Quinn CL, Chicz RM, Green BN, Willis AC et al. (1995). Naturally processed peptides from two disease-resistance associated HLA-DR13 alleles show related sequence motifs and the effects of the dimorphism at position 86 of the HLA DRβ chain. *Proc Nat Acad Sci USA* **92**: 6567–6571.

Dolo A, Modiano D, Maiga B, Daou M, Dolo G et al. (2005). Difference in susceptibility to malaria between two sympatric ethnic groups in Mali. *Am J Trop Med Hyg* **72**: 243–248.

Dronamraju KR (ed) (2004). *Infectious Disease and Host-Pathogen Evolution*. Cambridge University Press: Cambridge.

Eid NA, Hussein AA, Elzein AM, Mohamed HS, Rockett K et al. (2010). Candidate malaria susceptibility/protective SNPs in hospital and population-based studies: the effect of substructuring. *Malaria J* **9**: 119.

Ferwerda B, McCall MBB, Alonso S, Giamarellos-Bourboulis EJ, Mouktaroudi M et al. (2007). TLR4 polymorphisms, infectious diseases, and evolutionary pressure during migration of modern humans. *Proc Nat Acad Sci USA* **42**: 16645–16650.

Flint J, Harding RM, Boyce AJ, Clegg JB (1998). The population genetics of the haemoglobinopathies. *Bailliere Clin Haematol* **11**: 1–51.

Flint J, Hill AV, Bowden DK, Oppenheimer SJ, Sill PR et al. (1986). High frequencies of alpha-thalassemia are the result of natural selection by malaria. *Nature* **321**: 744–750.

Fry AE, Ghansa A, Small KS, Palma A, Auburn S et al. (2009). Positive selection of a *CD36* nonsense variant in sub-Saharan Africa, but no association with severe malaria phenotypes. *Hum Molec Genet* **18**: 2683–2693.

Fry AE, Griffiths MJ, Auburn S, Diakite M, Forton JT et al. (2008). Common variation in the ABO glycosyltransferase is associated with susceptibility to severe *Plasmodium falciparum* malaria. *Hum Molec Genet* **17**: 567–576.

Fucharoen S (2001). Hemoglobin E disorders. In: Steinberg MH, Forget BG, Higgs DR, Nagel RL (eds). *Disorders of Hemoglobin: Genetics, Pathophysiology and Clinical Management*. Cambridge: Cambridge Univ. Press:Cambridge. pp. 1139–1154.

Fucharoen G, Fucharoen S, Sanchaisuriya K, Sae-ung N, Suyasunanond U et al. (2002). Frequency distribution and haplotypic heterogeneity of β^E-globin gene among eight minority groups of northeast Thailand. *Hum Hered* **53**: 18–22.

Garrigan D, Hedrick PW (2003). Detecting adaptive molecular evolution, lessons from the MHC. *Evolution* **57**: 1707–1722.

Genton B, Al-Yaman F, Mgone CS, Alexander N, Paniu MM, Aplers MP (1995). Ovalocytosis and cerebral malaria. *Nature* **378**: 564–565.

Ghosh K (2008). Evolution and selection of human leukocyte antigen alleles by *Plasmodium falciparum* infection. *Hum Immunol* **69**: 856–860.

Giardine B, van Baal S, Kaimakis P, Riemer C, Miller W et al. (2007). HbVar database of human hemoglobin variants and thalassemia mutations: 2007 update. *Hum Mut* **28**: 206.

Gouagna LC, Bancone G, Yao F, Yameogo B, Dabiré KR et al. (2010). Genetic variation in human *HBB* is associated with *Plasmodium falciparum* transmission. *Nat Genet* **42**: 328–331.

Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP et al. (2010). The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Neglected Trop Dis* **4**: e774.

Guindo A, Fairhurst RM, Doumbo OK, Wellems TE Diallo DA (2007). X-linked G6PD deficiency protects hemizygous males but not heterozygous females against severe malaria. *PLoS Med* **4**: e66.

Haldane JBS (1949a). Disease and evolution. *Ric Sci Suppl* **19**: 68–76.

Haldane JBS (1949b). The rate of mutation of human genes. *Hereditas* **35** (S1): 267–273.

Hamblin MT, Di Rienzo A (2000). Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am J Hum Genet* **66**: 1669–1679.

Hamblin MT, Thompson EE, Di Rienzo A (2002). Complex signatures of natural selection at the Duffy blood group locus. *Am J Hum Genet* **70**: 369–383.

Hanchard N, Elzein A, Trafford C, Rockett K, Pinder M. *et al.* (2007). Classical sickle beta-globin haplotypes exhibit a high degree of long-range haplotype similarity in African and Afro-Caribbean populations. *BMC Genet* **8**: 52.

Hedrick PW (2004). Estimation of relative fitnesses from relative risk data and the predicted future of haemoglobin alleles *S* and *C*. *J Evol Biol* **17**: 221–224.

Hedrick PW (2011a). *Genetics of Populations*. 4th ed. Jones and Bartlett: Boston.

Hedrick PW (2011b). Selection and mutation for α thalassemia in non-malarial and malarial environments. *Ann Hum Genet* (in press).

Hedrick PW, Kim T (2000). Genetics of complex polymorphisms: parasites and maintenance of MHC variation. In: Singh R. Krimbas C (eds). *Evolutionary Genetics: From Molecules to Morphology*. Cambridge University Press: Cambridge. pp 204–234.

Hill AVS (1991). HLA associations with malaria in Africa: some implications for MHC evolution. In: Klein J, Klein D (eds). *Molecular Evolution of the Major Histocompatibility Complex*. Springer-Verlag, Berlin. pp 403–434.

Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P *et al.* (1991). Common west African HLA antigens are associated with protection from severe malaria. *Nature* **352**: 595–600.

Hill AVS, Bowden DK, O'Shaughnessy DF, Weatherall DJ, Clegg JB (1988). β thalassemia in Melanesia: association with malaria and characterization of a common variant (IVS1 nt 5 G-C). *Blood* **72**: 9–14.

Hill AVS, Yates SNR, Allsopp CEM, Gupta A, Gilbert SC *et al.* (1994). Human leukocyte antigens and natural selection by malaria. *Phil Trans Roy Soc Lond B* **346**: 379–385.

Ingram VM (2004). Sickle-cell anemia hemoglobin: the molecular biology of the first 'molecular disease'- the crucial importance of serendipity. *Genetics* **167**: 1–7.

Jallow M, Teo YY, Small KS, Rockett KA, Deloukas P *et al.* (2009). Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nat Genet* **41**: 657–665.

Jarolim P, Palek J, Amato D, Hassan K, Sapak P, Nurse GT *et al.* (1991). Deletion in erythrocyte band 3 gene in malaria–resistant Southeast Asian ovalocytosis. *Proc Nat Acad Sci USA* **88**: 11022–11026.

Johnson MK, Clark TD, Njama-Meya D, Rosenthal PJ, Parik (2009). Impact of the method of G6PD deficiency assessment on genetic association studies of malaria susceptibility. *PLoS ONE* **4**: 9 (e7246).

Joy DA, Feng X, Mu J, Furuya T, Chotivanich K *et al.* (2003). Early origin and recent expansion of *Plasmodium falciparum*. *Science* **300**: 318–321.

Karasov T, Messer PW, Petrov DA (2010). Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. *PLoS Genet* **6** (6): e1000924.

Kasehagen LJ, Mueller I Kiniboro B, Bockarie MJ, Reeder JC *et al.* (2007). Reduced *Plasmodium vivax* erythrocyte infection in PNG Duffy-negative heterozygotes. *PLoS ONE* **2** (3): e336.

Kaslow RA, McNicholl JM, Hill AVS (2008). *Genetic Susceptibility to Infectious Diseases*. Oxford University Press: Oxford.

Kermarrec N, Roubinet F, Apoil PA, Blancher A (1999). Comparison of allele O sequences of the human and non-human primate ABO system. *Immunogenetics* **49**: 517–526.

Kimura M, Soemantri A, Ishida T (2002). Malaria species and Southeast Asian ovalocytosis defined by a 27-bp deletion in the erythrocyte band 3 gene. *Southeast Asian J Trop Med Pub Health* **33**: 4–6.

Kwiatkowski DP (2005). How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet* **77**: 171–192.

Kwiatkowski DP, Luoni G (2008). Malaria. In: Kaslow RA, McNicholl JM, Hill AVS (eds). *Genetic Susceptibility to infections Diseases*. Oxford University Press. pp 372–386.

Lam K-WG, Jeffreys AJ (2006). Processes of copy-number change in human DNA: the dynamics of α-globin gene deletion. *Proc Nat Acad Sci USA* **103**: 8921–8927.

Langhorne J, Ndungu FM, Sponaas A-M, Marsh K (2008). Immunity to malaria: more questions than answers. *Nat Immunol* **9**: 723–732.

Lawniczak MKN, Emrich SJ, Holloway AK, Regier AP, Olson M *et al.* (2010). Widespread divergence between incipient *Anophoeles gambiae* species revealed by whole genome sequence. *Science* **330**: 512–514.

Leslie T, Briceño, Mayan I, Mohammed N, Klinkenberg E *et al.* (2010). The impact of phenotypic and genotypic G6PD deficiency on risk of *Plasmodium vivax* infection: A case-control study amongst Afghan refugees in Pakistan. *PLoS Med* **7**: e1000283.

Liu W, Li V, Rudicell RS, Robertson JD, Keele BF *et al.* (2010). Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* **467**: 420–425.

Livingstone FB (1958). Anthropological implications of sickle cell gene distribution in west Africa. *Am Anthrop* **60**: 533–562.

Livingstone FB (1967). *Abnormal Hemoglobins in Human Populations*. Aldine: Chicago.

Livingstone FB (1985). *Frequencies of Haemoglobin Variants*. Oxford University Press: Oxford.

López C, Saravia C, Gomez A, Hoebeke J, Patarroyo MA (2010). Mechanisms of genetically-based resistance to malaria. *Gene* **467**: 1–12.

Louicharoen C, Patin E, Paul R, Nuchprayoon I, Witoonpanich B *et al.* (2009). Positively selected *G6PD-Mahidol* mutation reduces *Plasmodium vivax* density in southeast Asian. *Science* **326**: 1546–1549.

MacFie TS, Nerrienet E, Bontrop RE, Mundy NI (2009). The action of *falciparum* malaria on the human and chimpanzee genomes compared: absence of evidence for a genomic signature of malaria at *HBB* and *G6PD* in three subspecies of chimpanzee. *Infec Genet Evol* **9**: 1248–1252.

Mackinnon MJ, Mwangi TW, Snow RW, Marsh K, Williams TN (2005). Heritability of malaria in Africa. *PLoS Med* **2**: 1253–1259. (e340).

Mason PJ, Bautista JM, Gilsanz F (2007). G6PD deficiency: the genotype-phenotype association. *Blood Rev* **21**: 267–283.

May J, Evans JA, Timmann C, Ehmen C, Busch W *et al.* (2007). Hemoglobin variants and disease manifestations in severe falciparum malaria. *J Am Med Assoc* **297**: 2220–2226.

Ménard D, Barnakas C, Bouchier C, Henry-Halldin C, Gray LR *et al.* (2010). *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proc Nat Acad Sci USA* **107**: 5967–5971.

Miller LH, Mason SJ, Clyde DF, McGinnis MH (1976). The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype FyFy. *New Engl J Med* **295**: 302–304.

Mockenhaupt FP, Ehrhardt S, Gellert S, Otchwemah RN, Dietz E *et al.* (2004). Alpha (+)- thalassemia protects African children from severe malaria. *Blood* **104**: 2003–2006.

Modiano G, Morpurgo G, Terrenato L, Novelletto A, Di Rienzo A *et al.* (1991). Protection against malaria morbidity: near-fixation of the α-thalassemia gene in a Nepalese population. *Am J Hum Genet* **48**: 390–397.

Modiano D, Luoni G, Sirima BS, Simpore J, Verra F et al. (2001). Haemoglobin C protects against clinical Plasmodium falciparum malaria. Nature 414: 305–308.

Modiano D, Bancone G, Ciminelli BM, Pompei F, Blot I (2008). Haemoglobin S and haemoglobin C: 'quick but costly' versus 'slow but gratis' genetic adaptations to Plasmodium falciparum malaria. Hum Molec Genet 17: 789–799.

Mueller I, Widmer S, Michel D, Maraga S, McNamara DT et al. (2009). High sensitivity detection of Plasmodium species reveals positive correlations between infections of different species, shifts in age distribution and reduced local variation in Papua New Guinea. Malaria J 8: 41.

Neafsey DE, Lawniczak MKN, Park DJ, Redmond SN, Coulibaly MB et al. (2010). SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. Science 330: 514–517.

Neel JV, Valentine WN (1947). Further studies on the genetics of thalassemia. Genetics 32: 38–63.

Nkhoma ET, Poole C, Vannappagari V, Hall SA, Beutler E (2009). The global prevalence of glucose-6-phosphate dehydrogenase deficiency: A systematic review and meta-analysis. Blood Cells Molec Dis 42: 267–278.

O'Donnell A, Premawardhena A, Arambepola M, Samaranayake R, Allen SJ et al. (2009). Interaction of malaria with a common form of severe thalassemia in an Asian population. Proc Nat Acad Sci USA 106: 18716–18721.

Ohashi J, Naka I, Patarapotikul J, Hananantachai H, Brittenham G. et al. (2004). Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection. Am J Hum Genet 74: 1198–1208.

Ohashi J, Naka I, Patarapotikul J, Hananantachai H, Brittenham G. et al. (2005). Strong linkage disequilibrium of a HbE variant with the $(AT)_9(T)_5$ repeat in the BP1 binding site upstream of the β-globin gene in the Thai population. J Hum Genet 50: 7–11.

O'Shaughnessy DF, Hill AVS, Bowden DK, Weatherall DJ, Clegg JB (1990). Globin genes in Micronesia: origins and affinities of Pacific Island peoples. Am J Hum Genet 46: 144–155.

Pauling L, Itano HA, Singer SJ, Wells IC (1949). Sickle cell anemia a molecular disease. Science 110: 543–548.

Penman BS, Pybus OG, Weatherall DJ, Gupta S (2009). Epistatic interactions between genetic disorders of hemoglobin can explain why the sickle-cell gene is uncommon in the Mediterranean. Proc Nat Acad Sci USA 106: 21242–21246.

Platt OS, Brambilla DJ, Rosse WF, Milner PF, Castro O et al. (1994). Mortality in sickle cell disease. New Eng J Med 330: 1639–1644.

Ralph P, Coop G (2010). Parallel adaptation: one or many waves of advance of an advantageous allele? Genetics 186: 647–668.

Rowe AK, Rowe SY, Snow RW, Korenromp EL, Schellenberg JR et al. (2006). The burden of malaria mortality among African children in the year 2000. Int J Epidemiol 35: 691–704.

Rowe JA, Handel IG, Thera MA, Deans A-M, Lyke KE et al. (2007). Blood group O protects against severe Plasmodium falciparum malaria through the mechanism of reduced rosetting. Proc Nat Acad Sci USA 104: 17471–17476.

Ruwende C, Khoo SC, Snow RW, Yates SNR, Kwiatkowski D et al. (1995). Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. Nature 376: 246–249.

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. Nature 419: 832–837.

Sanchaisuriya K, Fucharoen G, Sae-ung N, Siriratmanawong N, Surapot S, Fucharoen S. (2001). Molecular characterization of hemoglobin C in Thailand. Am J Hematol 67: 189–193.

Saunders MA, Hammer MF, Nachman MW (2002). Nucleotide variability at G6pd and the signature of malarial selection in humans. Genetics 162: 1849–1861.

Seixas S, Ferrand N, Rocha J (2002). Microsatellite variation and evolution of the human Duffy blood group polymorphism. Molec Biol Evol 19: 1802–1806.

Sheng A-M, Chertow DS, Morens DM, Taubenberger JK (2010). Fatal 1918 pneumonia case complicated by erythrocyte sickling. Emerg Infect Dis 16: 2000–2001.

Siniscalco M, Bernini L, Latte B, Motulsky AG (1961). Favism and thalassemia in Sardinia and their relationship to malaria. Nature 190: 1179–1180.

Slatkin M (2008). A Bayesian method for jointly estimating allele age and selection intensity. Genet Res 90: 129–137.

The Malaria Genomic Epidemiology Network (2008). A global network for investigation the genomic epidemiology of malaria. Nature 456: 732–737.

Timmann C, Evan JA, König IR, Kleensang A Rüschendorf F et al. (2007). Genome-wide linkage analysis of malaria infection intensity and mild disease. PLoS Genet 3: e48.

Tishkoff SA, Varkonyl R, Cahinhinan N, Abbes S, Argyropoulos G et al. (2001). Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science 293: 455–45. 462.

Tishkoff SA, Verrelli BC (2004). G6PD deficiency and malarial resistance in humans: insight from evolutionary genetic analysis. In: Dronamraju KR (ed). Infectious Disease and Host-Pathogen Evolution. Cambridge University Press: Cambridge. pp 113–140.

Tsaras G, Owusu-Ansah A, Boateng FO, Amoateng-Adjepong Y (2009). Complications associated with sickle cell trait: a brief narrative review. Amer J Med 122: 507–512.

Tung J, Primau A, Bouley AJ, Severson TF, Alberts SC, Wray GA (2009). Evolution of a malaria resistance gene in wild primates. Nature 460: 388–391.

Uneke CJ (2007). Plasmodium falciparum malaria and ABO blood group: is there any relationship. Parasitol Res 100: 769–765.

Verra F, Mangang VD, Modiano D (2009). Genetics of susceptibility to Plasmodium flaciparum: from classical malaria resistance genes towards genome-wide association studies. Parasite Immunol 31: 234–253.

Verrelli BC, McDonald JH, Argyropoulos G, Destro-Bisol G, Froment A et al. (2002). Evidence for balancing selection from nucleotide sequence analyses of human G6PD. Am J Hum Genet 71: 1112–1128.

Verrelli BC, Tishkoff SA, Stone AC, Touchman JW (2006). Contrasting histories of G6PD molecular evolution and malarial resistance in humans and chimpanzees. Mol Biol Evol 23: 1592–1601.

Weatherall DJ (2004). J. B. S. Haldane and the malaria hypothesis. In: Dronamraju KR (ed). Infectious Disease and Host-Pathogen Evolution. Cambridge University Press. pp 18–36.

Weatherall DJ (2008). Genetic variation and susceptibility to infection: the red cell and malaria. Br J Haematol 141: 276–286.

Weatherall DJ (2010). The inherited diseases of hemoglobin are an emerging global health burden. Blood 115: 4331–4336.

Weatherall DJ, Clegg JB (2001). The Thalassemia Syndromes, 4th ed. Blackwell Scientific, Oxford.

WHO (1998). Demographic data for health situation assessment and projections. WHO HST/HSP: Geneva.

WHO (2009). World Malaria Report. World Health Organization, Geneva.

Wierenga KJJ, Hambleton IR, Lewis NA (2001). Survival estimates for patients with homozygous sickle-cell disease in Jamaica: a clinic-based population study. Lancet 357: 680–683.

Williams TN, Mwangi TW, Wambua S, Peto TEA, Weatherall DJ. (2005a). Negative epistasis between the malaria-protective effects of α+-thalassemia and the sickle cell trait. Nat Genet 37: 1253–1262.

Williams TN, Wambua S, Uyoga S, Macharia A, Mwacharo JK et al. (2005b). Both heterozygous and homozygous α+

304

thalassemias protect against severe and fatal *Plasmodium falciparum* malaria on the coast of Kenya. *Blood* **106**: 368–371.

Willcox M, Björkman A, Brohult J (1983). Falciparum malaria and β-thalassemia trait in northern Liberia. *Ann Trop Med Parasit* **77**: 335–347.

Wood ET, Stover DA, Slatkin M, Nachman MW, Hammer MF 2005)). The β-globin recombinational hotspot reduces the effects of strong selection around HbC, a recently arisen mutation providing resistance to malaria. *Am J Hum Genet* **77**: 637–642.

Yamamoto F, Clausen H, White T, Marker J, Hakomori S (1990). Molecular genetic basis of the histo-blood group ABO system. *Nature* **345**: 229–233.

Zimmerman PA (2004). The enigma of *Plasmodium vivax* malaria and erythrocyte Duffy negativity. In: Dronamraju KR (ed.). *Infectious Disease and Host-pathogen Evolution*. Cambridge University Press: Cambridge. pp 141–172.

Zimmerman PA, Woolley I, Masinde GL, Miller SM, McNamara D *et al.* (1999). Emergence of *FY*null* in a *Plasmodium vivax*-endemic region of Papua New Guinea. *Proc Nat Acad Sci USA* **96**: 13973–13977.