SHORT COMMUNICATION

Estimation of relative fitnesses from relative risk data and the predicted future of haemoglobin alleles S and C

P. HEDRICK

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

Keywords:
malaria; polymorphism; relative risk; sickle-cell.

Abstract

Epidemiological studies of genetic differences in disease susceptibility often estimate the relative risks (RR) of different genotypes. Here I provide an approach to calculate the relative fitnesses of different genotypes based on RR data so that population genetic approaches may be utilized with these data. Using recent RR data on human haemoglobin β genotypes from Burkina Faso, this approach is used to predict changes in the frequency of the haemoglobin sickle-cell S and C alleles. Overall, it generally appears that allele C will quickly replace the S allele in malarial environments. Explicit population genetic predictions suggest that this replacement may occur within the next 50 generations in Burkina Faso.

Introduction

The normal (A) and the sickle-cell (S) alleles at the human haemoglobin β locus (which differ by a single amino acid at position 6, Glu to Val) have long provided a classic example of balanced polymorphism because AS heterozygotes have the greater resistance to malaria than the two homozygotes (SS homozygotes also suffer from sickle-cell anaemia). It has been known that a third allele C at this locus, which has a different single amino acid substitution at position 6, Glu to Lys, is in substantial frequency in several west African populations. Recently, allele C has been shown to confer higher protection to malaria than allele S from Burkina Faso, West Africa (Modiano et al., 2001). This epidemiological finding supports earlier research suggesting that red blood cells in CC individuals are ‘unsuitable hosts for the malarial parasite primarily because of their inability to lyse and release merozoiotes at the appropriate stage’ (Olson & Nagel, 1986). In addition, it appears that CC homozygotes have only limited, if any, costs from anaemia (Smith & Krevis, 1959). As a consequence, the replacement of the sickle-cell allele S by allele C has been predicted (Modiano et al., 2001; Pennisi, 2001; Pasvol, 2002). However, to explicitly predict genetic change at this locus, it is necessary to parameterize these relative risk (RR) data into relative fitness values so that traditional population genetic approaches can be used.

Various procedures for estimation of the relative fitness of genotypes at the haemoglobin β locus in populations polymorphic for alleles A, S and C have been used (e.g. Allison, 1956; Livingstone, 1967; Cavalli-Sforza & Bodmer, 1971). Estimates in the past of the relative fitnesses for genotypes with the C allele, i.e. AC, CC and CS have not been very accurate, mainly because the sample sizes for these genotypes were often small. The new very large data set from Burkina Faso, which gives both the frequencies of the genotypes of healthy individuals and those with malaria, overcomes this problem. However, these data were not translated into relative fitnesses so that population genetics methodology can be used to predict genetic changes. Below, I first give an approach to estimate relative fitnesses from RR data and then predict the potential impact of these fitnesses on genetic change at the human haemoglobin β locus.

Estimation of relative fitness from relative risk

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

\[ RR = \frac{f_d(1 - f_c)}{f_c(1 - f_d)} \]  

(1)

where \( f_c \) and \( f_d \) are the frequencies of the genotype in control and diseased groups, respectively. If a genotype reduces susceptibility to a disease, then frequency of the genotype in individuals with the disease is lower than in the control group, i.e. \( RR < 1 \). On the other contrary, if the genotype increases susceptibility to the disease, then the frequency of the genotype in the disease group is higher than in the control group and \( RR > 1 \).

In order to utilize RR values for different genotypes in population genetics, they need to be translated into relative fitnesses for the different genotypes. Hill (1991) (see also Hedrick & Kim, 2000) suggested that the selective effect for individuals with a given genotype that gave resistance to malaria could be calculated as

\[ s = m(1 - RR) \]  

(2)

where \( m \) is mortality rate for the subjects infected with malaria, independent of genotype. In this case, the relative fitnesses of the genotype being examined and that in the rest of the population are 1 and 1 - \( s \), respectively. However, to obtain the positive selective effect of the genotype compared with the rest of the population, the relative fitnesses need to be scaled so that the fitnesses of the genotype and the rest of the population become 1/(1 - \( s \)) and 1, respectively. For example, if \( RR = 0 \) (complete resistance of the genotype) and \( m = 0.5 \) (50% mortality of infected individuals), then \( s = 0.5 \) and the relative fitness of the resistant genotype is 2.0.

To determine the relative fitness of genotypes when \( RR > 1 \), the selective effect (which by definition is bounded by 0 and 1) needs to be calculated in a different way. A straightforward approach is to define RR in this case as its reciprocal \( 1/RR \) so that

\[ s = m \left( 1 - \frac{1}{RR} \right) \]  

(3)

and then the fitnesses of the susceptible genotype and the rest of the population become 1 - \( s \) and 1, respectively. For example, if \( 1/RR = 0 \) (complete susceptibility of the genotype) and \( m = 0.5 \), then \( s = 0.5 \) and the fitness is 0.5. These two numerical examples show that a 50% mortality translates into a twofold difference in relative fitness when there is either complete resistance or complete susceptibility of the genotype.

### Estimation of haemoglobin fitness in Burkina Faso

In the study in Burkina Faso by Modiano et al. (2001), the RR values were statistically significant for the haemoglobin genotypes AA, AC, AS and CC (AC, AS and CC showed relative resistance and AA showed relative susceptibility). Table 1 gives the frequencies of these four genotypes in the control (healthy subjects) and disease (malaria patients) groups and the resulting RR values. The RR values for the two other genotypes, SC and SS, were not statistically significant, primarily because there were few individuals of these genotypes in either the control or disease groups.

The low observed frequencies of SC and SS appear to be the result of high mortality (short-life expectancy) because of anaemia and related complications in individuals with these genotypes (Modiano et al., 2001). In this case, the relative fitness can be estimated as

\[ 1 - s = \frac{N_O}{N_E} \]  

(4)

where \( N_O \) and \( N_E \) are the observed and expected numbers (using Hardy–Weinberg proportions) of the genotype (Hedrick, 2000). Modiano et al. (2001) stated that in the control group for genotype SC, \( N_O \) and \( N_E \) were 23 and 46.2, respectively, and for SS, \( N_O \) and \( N_E \) were 1 and 9.2, respectively. The observed values for both these genotypes are statistically significantly less than that expected.

Using the approach in eqn 2 for genotypes AC, AS and CC, the approach in eqn 3 for AA, and approach in eqn 4 for SC and SS. Table 2 gives the relative fitness for the six genotypes.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AA</td>
</tr>
<tr>
<td>Healthy subjects ( f_c )</td>
<td>0.6641</td>
</tr>
<tr>
<td>Malaria patients ( f_d )</td>
<td>0.8036</td>
</tr>
<tr>
<td>Relative risk (RR)</td>
<td>2.070</td>
</tr>
</tbody>
</table>

### Table 1 Genotype frequencies in healthy (control) subjects \( f_c \), malaria (diseased) patients \( f_d \), and the relative risk of malaria for the four genotypes where the relative risk is statistically significant (Modiano et al., 2001).

<table>
<thead>
<tr>
<th>Mortality (m)</th>
<th>AA</th>
<th>AC</th>
<th>AS</th>
<th>CC</th>
<th>SC</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.07</td>
<td>0.964</td>
<td>1.021</td>
<td>1.054</td>
<td>1.070</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>0.901</td>
<td>0.954</td>
<td>0.985</td>
<td>1</td>
<td>0.498</td>
<td>0.109</td>
</tr>
<tr>
<td>0.10</td>
<td>0.948</td>
<td>1.030</td>
<td>1.079</td>
<td>1.102</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>0.881</td>
<td>0.935</td>
<td>0.979</td>
<td>1</td>
<td>0.498</td>
<td>0.109</td>
</tr>
<tr>
<td>0.20</td>
<td>0.897</td>
<td>1.062</td>
<td>1.172</td>
<td>1.228</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>0.730</td>
<td>0.865</td>
<td>0.964</td>
<td>1</td>
<td>0.498</td>
<td>0.109</td>
</tr>
</tbody>
</table>

### Table 2 The estimated relative fitnesses for three levels of mortality (m) from malaria. The fitness in the second row for each value of m is standardized by the fitness of genotype CC, the genotype with the highest fitness.
genotypes. The fitnesses relative to the genotype with highest fitness, CC is also given. Data from the World Health Organization (WHO, 1998) shows that the mortality rates for hospitalized patients may range up to 0.1 and Hill (1991) suggested that mortality rate was 0.07. The mortality rate may have been higher than these estimates in past generations when medical care was not as good so relative fitnesses are given for \( m = 0.07, 0.1 \) and 0.2 in Table 2.

**Population genetic predictions for haemoglobin alleles S and C**

Fitness values for the six possible genotypes from earlier data (e.g. Allison, 1956; Livingstone, 1967; Cavalli-Sforza & Bodmer, 1971) have predicted either stable or unstable equilibria for the three alleles. Using the fitnesses in Table 2, unlike previous fitness arrays, there are no stable or unstable three-allele equilibria (see approach used in Hedrick, 2000). If only alleles A and S are present, there is a two-allele equilibrium, with S at a much lower frequency than A. If C is introduced by mutation or gene flow, whether S is present or not, then C will always increase and eventually the population will become fixed for C, as suggested by Modiano et al. (2001). This outcome occurs primarily because genotype CC has the highest estimated relative fitness of any genotype. Of public health significance, the mean fitness of the population will be 14.3% 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 anaemia.

However, the rate of increase of C is a function of the frequency of S (iterations of the equations provided on p. 123 in Hedrick, 2000 can be used). If S is not present, C is introduced by mutation or gene flow at a low frequency, and \( m = 0.1 \), then C quickly increases to a substantial frequency in a few generations (from 0.01 to 0.5 in approximately 60 generations) (Fig. 1). If \( m = 0.07 \) or 0.2, then the increase from 0.01 to 0.5 takes about 140 and 50% as long. If both C and S are introduced simultaneously at low frequencies, they both initially increase (S actually faster than C). Then, the frequency of C continues to increase quickly and S is reduced to a low frequency and eliminated. If A and S are at their predicted equilibrium frequencies, then the initial increase of C from a low frequency is greatly slowed. Again, as soon as C reaches a frequency greater than 0.05, it begins to quickly increase, as in the other initial conditions.

These numerical examples demonstrate that C will indeed eventually dominate the variation at the human haemoglobin \( \beta \) locus in malarial environments. However, the rate at which C comes to predominate is a function of the initial frequency of S. If S is already at equilibrium, then the increase in frequency is delayed by approximately 100 generations (2000 years if a 20-year generation time is assumed) with the fitness data used here. This interference at low C frequencies, where most C alleles are in heterozygotes, appears to be the result of the low fitness of SC genotypes and the somewhat lower fitness of AC than AS.

An important caveat to the prediction that C will eventually replace S in polymorphic populations is that if the relative fitness of AC is somewhat lower than estimated here, then an unstable equilibrium may result and C may not easily invade. For example, if \( m = 0.1 \) and the fitness of AC is 0.922 (only 1.4% lower than in Table 2), then for C to invade it must be introduced at a frequency greater than the initial frequency of 0.01 used in the numerical examples here. With these fitnesses and A and S at equilibrium, there is an unstable equilibrium just below 0.013 and to invade, the initial frequency of C must be 0.013 or greater. Therefore, introduction by mutation or only by a few migrants, which would probably result in an initial frequency less than this unstable equilibrium would not result in the invasion by C in a population at equilibrium for A and S.

Modiano et al. (2001) estimated that the frequencies of A, S and C in their sample from Burkina Faso were 0.8204, 0.0512 and 0.1284, respectively. In the simulations in Fig. 1, frequencies of S and C very similar to those observed (indicated by solid squares) were seen in generation 59 (when both alleles were simultaneously introduced at low frequencies) and in generation 139 (when S was initially at equilibrium with A and C was introduced at low frequency). Therefore, it appears that C has been slowly increasing in this population because of the presence of S and that now it is poised to rapidly increase to a high frequency.
within the next 50 generations, eliminating allele S and sickle-cell anaemia, and going to fixation. As future generations pass, it will be fascinating to see if changes in the frequency of the S and C alleles are consistent with these predictions.

The relative rate of increase is a function of the differential fitness among the genotypes. Assuming that the relative risks of contracting malaria are correct, then the rate of allele-frequency change is primarily a function of the rate of mortality (m) from malaria. The assumed value of 10% mortality may be a low estimate for selection in past generations and in some areas at present. Although if malaria prevention programmes have a substantial effect, then m may become lower than 10%. As a result, the pattern of change may be quicker or slower if, on average, m is higher or lower.

Acknowledgments

I appreciate comments from D. Garrigan, E. Wood, and two anonymous reviewers. This research was supported by funding from the Ullman Professor.

References


Received 14 January 2003; revised 17 July 2003; accepted 18 July 2003