Reviewer’s report

Title: Loss of Balancing selection in the BetaS Globin locus

Version: 1 Date: 23 June 2009

Reviewer: Deborah Charlesworth

Reviewer’s report:

This is a potentially somewhat interesting study, which tests whether there is a detectably lower advantage to the HbS allele in a population in which malaria infections, and illness due to them, are low, but which has a moderate frequency of the HbS allele, so that selection on the three genotypes can be monitored and its strength estimated. The approach used was to genotype alleles at this locus in individuals in two human populations, from two fairly isolated villages with different ancestry, but both recently moved to within a geographically fairly small region. In the Hausa village, the HbS allele frequency was lower in younger age groups, compared with older ones, consistent with a decrease over a single generation, as might be predicted since this village appears to show low infection rates, and mostly mild disease. In the Massalit village, infection rates and serious symptoms were commoner, and the two generations showed less change in allele frequencies. The relevant page (8) does not actually give the allele frequencies. It would be easiest to give them in a table, together with the frequencies of the 3 genotypes. The table could then show these results broken down by the year or age groups, as relevant.

MAJOR COMPULSORY REVISIONS

As written, the ms is hard to understand in many places, and many important things are not explained, or appear to be inconsistent. Serious revisions of the English are needed, plus (more importantly) revision of the analysis of the data. I have noted many corrections of the English on the ms. Below, I outline some of the problems.

For instance, on p. 7, it says that ‘little or no clinical malaria’ was seen in the 2 villages, but Figure 1 shows % of malaria often > 20%. Perhaps the authors are trying to say that malaria infections were detected, but not illness — certainly, they need to write more clearly. On the same page, anemia is mentioned (p. 7), but it is not explained why this is relevant (presumably, low Hb levels are a sign of infection?).

The first sentence on p. 8 should be omitted, as it is confusing. The age groups used should be explained, and I am not clear what the authors mean by saying that people < or > 30 years old were treated as 2 generations. Clearly, they may be 2 generations, but also, of course, age may affect the kinds of observations, so that one cannot examine allele frequency changes between 2 generations unless one compares individuals of the same age. Standard population genetic analyses compare the youngest age, or age at maturity, say, to test whether
selection has changed an allele frequency. Alternatively, selection can be detected by comparing the allele frequency in young individuals (before selection) with mature individuals (after selection). If such a comparison is made, the strength of selection can be estimated using simple population genetic equations and approaches.

The authors need to distinguish two things: (1) Detecting selection (which they do by testing for Hardy-Weinberg genotype frequencies), and (2) Estimating the selection coefficient to see whether it is detectably lower in the Hausa than the Massalit village.

Testing for Hardy-Weinberg genotype frequencies is described on p. 8 (the abbreviation DHWE isn’t explained). In the Massalit village (but not in the Hausa one), this test suggests some selection, but the ms should explain how the results relate to the selection — is there an excess of heterozygotes in adults that is not seen in young children, for instance, which would suggest that selection favours heterozygotes? It is not enough just to say that there is a significant deviation from HWE. As far as I can see, the Massalit village data have a significant deviation (in an unspecified direction) only in the < 30 age group, which is not what one would expect if selection on HbS is occurring.

As for estimating the selection coefficient, this is very badly explained, and only odds ratios are mentioned, and the use in the analysis isn’t clearly explained on p. 9). If such estimates have been obtained, they could be used to predict future allele frequency changes, but clearly this should be described AFTER the estimates are explained (at present, these predictions are described on p. 8 — BEFORE the estimates). I think the equations used are incorrect (p. 6) because they don’t include the selection coefficient or fitness. Also the term called h is not defined; presumably it is the dominance coefficient, but it is not explained how this was estimated.

**Level of interest:** An article of importance in its field

**Quality of written English:** Not suitable for publication unless extensively edited

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests