Int J Legal Med (2006) 120: 188–189
DOI 10.1007/s00414-005-0032-3

LETTER TO THE EDITOR

A. Salas · C. Phillips · A. Carracedo

Ancestry vs physical traits: the search for ancestry informative markers (AIMs)

Received: 8 March 2005 / Accepted: 8 July 2005 / Published online: 25 August 2005

Keywords SNPs · AIMs · Ancestry · Physical traits · Skin colour

We read with interest the study of Yuasa et al. [1] on the analysis of polymorphisms of the membrane-associated transporter protein (MATP) gene in Germans and Japanese. The main conclusion of their work was that the mutation L374F "is valuable as a population and ancestry informative marker (AIM) for Caucasoids". We wish to outline certain points in order to properly evaluate the contribution of this finding to the field.

We feel it is important from the very beginning to make a distinction between two different aspects of the reported work: (a) the association of these polymorphisms with skin colour and oculo-cutaneous albinism types II and IV, and (b) the power of the polymorphisms to infer ancestry. In our opinion, these two aspects are conceptually mixed in the paper and this could create a confused message. There is no doubt that skin colour inference using genetic polymorphisms might be helpful in medical investigations [2] or even in the forensic field (e.g. police investigations). However, Yuasa and co-authors focused mainly on the second issue, i.e. the power of the described polymorphisms to infer ancestry.

A single locus alone provides very limited information on the ancestral origin of an individual since it can never be representative of the entire genome (e.g. [3]). A familiar parallel for forensic geneticists can be made with mitochondrial DNA (mtDNA) and (non-recombining) Y-chromosome polymorphisms, both of which have very well detailed worldwide phylogeographic distributions. Thus, the presence of an L1b mtDNA or an E3a Y-chromosome profile, for example, although suggestive of sub-Saharan origin, is also found to be widespread in America due to the recent Atlantic slave trade (e.g. [4, 5]) and sporadically in other worldwide populations. It is clear from this that a single SNP test alone is even less appropriate for the inference of ancestry when considered in the context of recent population admixture (e.g. due to the Atlantic slave trade). It is possible, however, to estimate the most probable geographical origin of a particular polymorphism variant or a sequence profile using appropriate algorithms and databases (e.g. [6]); the power of these approaches depends mainly on how informative the marker under study is, as well as the particular phylogeographic features. It must be stressed that this probability is not synonymous with the most probable population origin of a single person; some correlation could exist, and it is this correlation that may be useful in some forensic contexts. Not only do we need to be particularly cautious when using skin pigmentation as a proxy to infer ancestry [2], but also when using genetic markers linked to skin pigmentation or any other gene where varying selection regimes may have acted in different geographical regions. As stated in Tishkoff and Kidd [7], people with dark skin are found in New Guinea, Southern India and Africa, and even within these regions, there can be a very broad range of variation in skin colour; therefore, this complex trait is not always a reliable indicator of ethnicity. As stated above, pigmentation genes such as MAPT could possibly help in the context of police investigations (especially in cases of admixture), but this is not the issue here or in the paper of Yuasa et al. [1].

This begs the question of how many SNPs are necessary to confidently infer ancestry. However, this is a more complex issue dependant on numerous population and genetic assumptions that are generally difficult to gauge and must be defined empirically [8]. The inference of ancestry needs to be made within for example a Bayesian

A. Salas (✉) · C. Phillips · A. Carracedo
Unidade de Xenética, Instituto de Medicina Legal,
Facultad de Medicina, C/San Francisco s/n,
15782 Santiago de Compostela, Galicia, Spain
e-mail: apimlase@usc.es
Tel.: +34-981-582327
Fax: +34-981-580336

A. Salas · C. Phillips · A. Carracedo
Centro Nacional de Genotipado (CeGen),
Hospital Clínico Universitario,
15706 Galicia, Spain

framework making use of a broadly based panel of SNPs (see e.g. [9]). Thus, Shriver et al. [10] stated that "it is possible to estimate the individual ancestry of a person based on DNA analysis with a reasonable number of well-defined genetic markers". A good combination of complementary skewed markers that include such examples as the one described for L374F by Yuasa and co-workers [1] will generally be more effective than a very large collection of markers that individually have limited indicative properties. Unfortunately, the omission from this report of other population samples, outside of Germany and Japan, for this potentially useful SNP means it cannot be properly assessed for use as a reliable AIM in other populations [1].

An additional important point to make is that, since AIMs are by definition those SNPs that show alleles with large frequency differences between populations [10], any neutral marker may be useful for the estimation of ancestry, not just those associated with phenotypic traits. Therefore, it is not particularly relevant that L374F is associated with the MATP gene if the main message of this study relates to the power of the markers described to infer ancestry.

Tools are now freely available to assist in the search for AIMs. The use of the Hapmap database (http://www.hapmap.org) or the Frequency Finder search tool (http://bluegenes.bsd.uchicago.edu/frequencyfinder/), to name two examples, allows AIMs to be isolated with relative ease. For instance, in a quick search, we found 143 SNPs alone within the MATP gene in dbSNP (build 124: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=snp), most of them (~63%) validated. These include SNPs such as rs28117 (frequency of allele A/G: 0/1 CEU in Utah residents with ancestry from northern and western Europe; 0.636/0.364 HCB in Han Chinese in Beijing, China; 0.693/0.307 JPT in Japanese in Tokyo, Japan; and 0.517/0.483 YRI in Yoruba of Ibadan, Nigeria) and rs26722 (frequency of allele C/T: 1/0 CEU, 0.611/0.389 HCB, 0.591/0.409 JPT, 0.950/0.050 YRI) with population informative allele frequency distributions similar to the polymorphisms described in Yuasa et al. [1]; the F374F SNP (rs16891982) described by Yuasa et al. [1] seems to be the marker showing the most skewed distribution (at least for European and Asian populations), indicating promising potential as an AIM in other closely related populations.

In summary, the analysis of SNPs associated with traits such as skin colour could be a useful way to detect AIMs, but mining genomic information is undoubtedly a more economic, rapid and practical strategy, yielding many more loci in the long term. A worldwide population screening should be carried out in order to evaluate the potential of specific SNPs to infer ancestry; in the absence of such a study, special care must be taken with those SNPs that have potentially been subject to selection. Without doubt, a good AIM does not need the prerequisite of association with a population-linked physical characteristic.

References

1. Yuasa I, Umetsu K, Watanabe G, Nakamura H, Endoh M, Irizawa Y (2004) MATP polymorphisms in Germans and Japanese: the L374F mutation as a population marker for Caucasoids. Int J Legal Med 118:364–366
2. Parra EJ, Kittles RA, Shriver MD (2004) Implications of correlations between skin color and genetic ancestry for biomedical research. Nat Genet 36(11 Suppl):S54–S60
3. Pääbo S (2003) The mosaic that is our genome. Nature 421: 409–412
4. Salas A, Richards M, De la Fé T et al (2002) The making of the African mtDNA landscape. Am J Hum Genet 71:1082–1111
5. Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A (2004) The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74:454–465
6. Egeland T, Bøvelstad HM, Storvik GO, Salas A (2004) Inferring the most likely geographical origin of mtDNA sequence profile. Ann Hum Genet 68:461–471
7. Tishkoff SA, Kidd KK (2004) Implications of biogeography of human populations for 'race' and medicine. Nat Genet 36(11 Suppl):S21–S27
8. Bamshad M, Wooding S, Salisbury BA, Stephens JC (2004) Deconstructing the relationship between genetics and race. Nat Rev Genet 5:598–609
9. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 20:2381–2385
10. Shriver MD, Parra EJ, Dios S et al (2003) Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet 112(4):387–399