European Journal of Human Genetics (2005) 13, 131–135

www.nature.com/ejhg

NEWS AND COMMENTARIES

Population genetics

Female migration rate might not be greater than male rate

Mark D Shriver

European Journal of Human Genetics (2005) **13,** 131–132.
doi:10.1038/sj.ejhg.5201329
Published online 3 November 2004

In a recent paper, which appeared in *Nature Genetics*, Jason Wilder[1] and colleagues at the University of Arizona, showed that measures of differentiation among populations for male and female lineages are much more similar than commonly thought.

Until now, the generally accepted view was that there is substantially more geographic differentiation in Y-chromosomal genetic variation than mtDNA variation. Since the Y-chromosomal DNA is inherited only by sons from their fathers and conversely mtDNA is only transmitted through the female lineage (males also have mtDNA, but they do not pass it on to their children) these differences have been taken to reflect sex-specific differences in population structure and/or history. There have been a number of explanations for these higher levels of geographic diversity (which Wilder *et al* estimated as Φ_{ST}, a haploid analog to the diploid F_{ST} statistic), namely greater reproductive variability for men relative to women and higher rates of migration for women relative to men. If these new results hold up, the reasons that earlier analyses gave a different answer should be carefully considered and future studies should be adjusted accordingly.

One of the major reasons that perhaps we should favor the conclusions of this new work over previous studies is that Wilder and colleagues have made a fundamental improvement in their study design compared to these previous efforts. Specifically they were able to sequence completely all of the 389 subjects they studied for the Y-chromosome regions under investigation. The Y-chromosome typically has very low levels of genetic diversity, so it has not generally proven feasible to design studies where all subjects are sequenced. The norm has been to genotype markers previously shown to be polymorphic and this fact alone introduces a form of ascertainment bias leading to higher levels of differentiation (Φ_{ST}). Markers that are discovered in small panels of individuals are likely to be common in one or more of the populations in the discovery panel and even if all populations are represented in this smaller panel, the differentiation will be higher than if all subjects are sequenced. This ascertainment bias effect is even greater if the discovered polymorphisms are used to study populations not represented in the discovery panel.

To avoid such ascertainment bias, Wilder *et al* hit upon the ingenious solution of focusing their sequencing on a special selection of recently inserted human retrotransposons, Y family *Alu* elements, which are hypermutable by virtue of a higher proportion of CpG dinucleotides. In short, this approach allowed them to use complete sequencing to collect enough data to test whether there was more or less geographic differentiation in the Y-chromosome compared to the mtDNA, for which a segment was also sequenced in each of the 389 subjects. The subjects under investigation were sampled from 10 geographically diverse populations; four African, two Asian, two European, and two Oceanic. With such a diverse selection of populations the authors were able to test for differences between these two loci, and the sexes they represent, across a long span of evolutionary time and geographic space, tens of thousands of years and thousands of miles. Overall they found that the levels of mtDNA differentiation ($\Phi_{ST} = 0.382$) were greater than the levels of differentiation on the Y-chromosome ($\Phi_{ST} = 0.334$).

One of the reasons for all of the interest in these patterns of mtDNA and Y-chromosomal diversity is because they can provide clues to the important demographic parameters of populations through time. Have women generally moved to the house and village of the men who fathered their children? Has variability in reproductive success been much higher for men compared to women, with some men having many children and others few or none? Has the generation time for women been substantially shorter, with women becoming mothers younger than when men become fathers (see Helgason *et al*[2])?

As each of these factors, migration rate, reproductive variability, and generation time, would affect mtDNA and Y-chromosomal relative differentiation levels, it is clearly a complex issue. Wilder and colleagues have shown that ascertainment bias, caused by genotyping polymorphic sites and not sequencing each subject, can have an important effect on studies such as these. It is critically important to understand the types of experimental and statistical biases that can affect results, and to take these into account. These authors are to be commended on this account and workers in evolutionary genomics should heed this warning, as it is unlikely that these are the only statistics to be affected by such ascertainment bias.

Nonetheless, given the number of demographic forces that affect patterns of genetic differentiation, the implications of these results should be interpreted cautiously. The authors conclude, 'that the role of female migration is no more important than that of males at the continental- and global scales.' And, while this explanation is clearly consistent with

their data, if generation times are on average shorter for females (ie mothers are generally younger than the babies fathers), which would imply a smaller global effective population size for females compared to males, then greater female migration or greater male reproductive variability might also be consistent with this pattern■

Mark D Shriver is at The Pennsylvania State University, University Park, PA, USA.

References

1 Wilder JA *et al: Nat. Genet* 2004; **36**: 1122–1125.
2 Helgason A *et al: Am. J. Hum. Genet.* 2003; **72**: 1370–1388.

Evolutionary Genetics

Evolutionary path to the heart

Shu Ye

European Journal of Human Genetics (2005) **13,** 132–133.
doi:10.1038/sj.ejhg.5201330
Published online 10 November 2004

The idea that evolutionary analyses can be a valuable complementary approach for assessing functional importance of genetic variants in not new,[1–3] but an elegant recent paper is *Current Biology*, which shows that positive selection might have a significant role in shaping population variation in a gene promoter polymorphism implicated in heart disease, illustrates just how useful these studies can be.[4]

Inherited differences in DNA sequences contribute to interindividual variability in anthropometric characteristics, risk of disease, and response to the environment and medication. Such variants occur roughly every 300–500 bp throughout the human genome, which has about 3 billion base pairs in total. However, only a small percentage of the estimated 10 million polymorphisms result in differences in protein amino-acid sequence or in the level of gene expression. Moreover, it is likely that only some of these functional variants are relevant to disease risk and other phenotypic traits. Molecular genetic epidemiological analyses and *in vitro* functional studies are the most common approaches used to assess the significance of particular genetic variants, but this new study illustrates that evolutionary analyses can be similarly effective.

Matthew Rockman and co-workers investigated the evolutionary history of a polymorphism that has been found to be associated with alternative cardiovascular phenotypes in a number of studies. The polymorphism is located in the promoter region of the matrix metalloproteinase-3 (MMP3) gene, with two alleles identified in humans, one having a run of five thymidines (5T) and the other six thymidines (6T), from nucleotide position-1608 relative to the transcriptional initiation site of the gene. Previous studies have indicated that this poly-T track forms part of two overlapping transcription factor-binding sites. Although the two alleles have similar affinity with the transcription factor ZNF148 (also named ZBP89), which acts as a transcription enhancer, the 5T allele has a lower affinity than the 6T allele for another transcription factor (apparently an NF-κB P50/P50 dimer) that acts as a transcription repressor.[5] As a result, MMP3 transcript and protein levels in *ex vivo* tissues are highest in 5T homozygotes, intermediate in heterozygotes, and lowest in 6T homozygotes.[6]

MMP3 (also known as stromelysin) enzymatically degrades various extracellular matrix proteins in the blood vessel wall and elsewhere. So the expression levels of this gene seem to influence the balance between matrix protein synthesis and degradation, which, in turn, affects the amounts of matrix proteins in tissues. Genetic association studies of the MMP3 gene and atherosclerosis (hardening of the arteries) have shown a genotype–phenotype relationship, such that atheromas (plaques that forms within the walls lining the arteries) in individuals of the 6T/6T genotype tend to be larger, whereas those in individuals of the 5T/6T or 5T/5T genotype are generally smaller but prone to rupture.[7] Individuals who carry the 5T allele also seem to have greater arterial elasticity and a predisposition to developing coronary artery aneurysm.[8] These observations are consistent with a model in which there is an imbalance favouring matrix protein degradation in 5T allele carriers, whereas in 6T/6T individuals matrix protein accumulation is favoured.

The authors data[4] indicate that the poly-T tract mentioned above might lie within a mutational hot spot that has undergone relatively rapid evolution for tens of million years. They compared the human MMP3 gene promoter sequence with those in nine non-human primates, and found that they all contain the poly-T tract but its length differs among the different species. In addition, intraspecies polymorphism was observed in seven of the primate species studied, further supporting that this region is a mutational hot spot.

The study also suggests that natural selection has caused the frequency of the 5T allele in northern Europe to increase. There are considerable differences in 5T allele frequency among human populations: these frequencies range from 0.01 in Cameroon to 0.54 in Sweden. Differences in allele frequency among populations can arise from natural selection and/or genetic drift (change in gene allele frequency due to chance). The former is unique to each locus, whereas the latter affects all autosomal loci equally. To investigate whether the 5T/6T polymorphism has been a target of natural selection, the authors compared the patterns of genetic differentiation at the 5T/6T site among several populations, with reference to patterns at 18 unlinked, neutral polymorphisms that selection is unlikely to affect. The analyses showed