

NEWS AND COMMENTARY

Local, rural and British

A British approach to sampling

Chris Tyler-Smith and Yali Xue

European Journal of Human Genetics (2012) **20**, 129–130; doi:10.1038/ejhg.2011.153;
published online 10 August 2011

The acronym ‘PoBI’ may not yet be familiar to human geneticists in the way that ‘HGDP’, ‘WTCCC’ or ‘HapMap’ are, but a paper in this issue of *EJHG*¹ that introduces the ‘People of the British Isles’ project to the scientific community aims to change this. The PoBI project will collect up to 5000 DNA samples from diverse regions of the British Isles, taking great care to sample individuals with several generations of ancestry in rural locations. These samples are intended to serve as controls for future medical genetic studies, and to provide insights into the peopling of the British Isles over the last few millennia. Many have already been genotyped on standard SNP chips, and 100 have been sequenced genome-wide by the 1000 Genomes Project. Cell lines and sequence data from the latter are already available (<http://www.1000genomes.org/home>). Although readers will have to wait for future publications to discover the insights from these large-scale genetic analyses, the current paper describes the sampling strategy and initial 3865 samples in some detail, outlines an approach to investigating fine-scale population structure using surnames, and presents some preliminary genetic analyses of a handful of chosen loci.

Over the last few years, PoBI project scientists have travelled to multiple parts of England, Wales, Scotland, and Northern Ireland, to meet volunteers who have responded to their advertisements (<http://www.peopleofthebritishisles.org/>). Volunteers whose four grandparents were born in the same local area (<60 km apart) donated 20 ml blood, half of which was used for DNA extraction. With admirable

foresight, the other half was preserved for the establishment of immortal cell lines, and so far, 531 of these have been established.

How well did the project achieve its aim of collecting from donors with all four grandparents born in the same rural area? The median distance between grandparental birthplaces was 16 km, and 75% of distances were less than 45 km, so their strategy was very successful in collecting multigenerational residents from local areas. It was more difficult to assess how many were rural, because what does ‘rural’ mean? If it implies that the grandparents were all born >10 km from even a small town of 20 000 inhabitants, the proportion is 37%. But using the authors’ preferred cut-off of 125 000 inhabitants (which would classify Oxford as urban, but Cambridge as rural: http://en.wikipedia.org/wiki/List_of_largest_United_Kingdom_settlements_by_population), 73% qualify. Perhaps the most important aspect of the dataset here is that ‘rural’ can be treated as a quantitative variable and used quantitatively in subsequent analyses.

The genetic markers chosen were those that must have seemed good candidates for detecting regional differences when the project was conceived: the Y chromosome, HLA, SNPs in the pigmentation gene *MC1R*, and the SNP that distinguishes the A and B blood groups. The results of typing them in around 1000 samples were either reassuring or disappointing, depending on your point of view. Standard analyses using principal components, the programme STRUCTURE² or the classic measure of population differentiation F_{ST} ,³ detected no significant regional differences except a hint of distinction in the people in the Orkney Islands in the north of Scotland, perhaps linked to their Viking ancestry. So the authors developed another approach to search for population structure.

In addition to collecting blood, the project recorded surnames. Using data from a census

performed in 1881, these were classified as ‘local’ or ‘non-local’, and the two classes examined separately. The authors then modelled a population such as that from central England as a mixture between south-western (taken to represent Ancient Britons) and eastern (Anglo Saxon) populations, and estimated the contribution of each population to the central England autosomal genotypes. These contributions differed between the local surname class (mostly eastern) and the non-local class (half and half), which the authors take as evidence of subtle population structure. Published genetic analyses using much larger numbers of markers have already detected low, but significant levels of genetic structure within Britain in more straightforward ways,^{4,5} even with less stringently ascertained samples (Figure 1): Europe-wide south-east to north-west gradients extend into the British Isles. We can look forward to deeper insights into genetic differentiation and its causes when large-scale genetic analyses of the PoBI samples are available.

It is easy to criticise any sampling scheme, and the near-absence of participants from the Republic of Ireland is a glaring omission here, reflecting political rather than genetic realities; even within the UK, sampling is still patchy. More importantly, no medical-genetic study uses such a recruitment scheme, so Winney *et al.*¹ sampling choices mean that naïve use of PoBI samples as controls in case-control comparisons, where the cases come from multi-ethnic cities, would lead to false positives due to the different ascertainment.⁶ But anthropological and evolutionary geneticists should rejoice in the assembly of this resource, the foresight of The Wellcome Trust in funding the project over a decade or so, and hope that resources are available for establishing more cell lines and performing more genome-wide sequencing, so that both the full set of samples and their sequences can be made widely available.

It is obvious why British people interested in their ancestry, and medical geneticists working with British subjects should welcome PoBI, but why should others pay attention? PoBI will not provide information about global genetic diversity in the way that HGDP⁷ and HapMap⁸ do, but its microcosmic survey of genetic variation in a set of small islands off the western coast of the Eurasian continent is revealing the level of differentiation that builds up over millennia via events well documented by archaeology and history, so these alternative data sets can be compared to address questions about the initial peopling of the area, and its subsequent reshaping by internal and external forces. And

Dr C Tyler-Smith and Y Xue are at the Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambs. CB10 1SA, UK.
Tel: +44 (0) 1223 495376; Fax: +44 (0) 1223 494919;
E-mail: cts@sanger.ac.uk

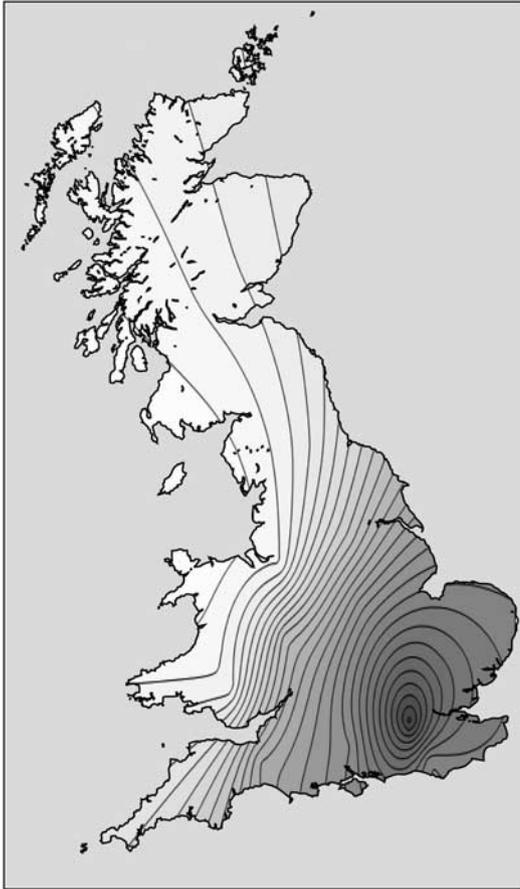


Figure 1 Genetic differentiation within Britain. The WTCCC1 Study⁴ identified rs1042712 in the lactase gene as the most highly differentiated of the ~0.5 million SNPs typed, possibly reflecting the influence of natural selection on the lactase persistence phenotype. Its minor allele frequency ranges from 9.4% (light grey) to 14.6% (dark grey) (redrawn from WTCCC1 data⁴). Future genetic analyses of the PoBI samples should provide further insights into genetic differentiation within Britain.

if the characteristics of the British – politeness, eccentricity, or drunken loutishness, according to your viewpoint and experience – have any genetic basis, perhaps PoBI can provide a starting point for identifying it! ■

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank The Wellcome Trust for support.

- 1 Winney B, Boumertit A, Day T *et al*: People of the British Isles: preliminary analysis of genotypes and surnames in a UK-control population. *Eur J Hum Genet* 2012; **20**: 203–210.
- 2 Hubisz MJ, Falush D, Stephens M, Pritchard JK: Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* 2009; **9**: 1322–1332.
- 3 Weir BS, Cockerham CC: Estimating F-statistics for the analysis of population structure. *Evolution* 1984; **38**: 1358–1370.
- 4 The Wellcome Trust Case Control Consortium: Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* 2007; **447**: 661–678.
- 5 O'Dushlaine CT, Morris D, Moskvina V *et al*: Population structure and genome-wide patterns of variation in Ireland and Britain. *Eur J Hum Genet* 2010; **18**: 1248–1254.
- 6 Price AL, Zaitlen NA, Reich D, Patterson N: New approaches to population stratification in genome-wide association studies. *Nat Rev Genet* 2010; **11**: 459–463.
- 7 Li JZ, Absher DM, Tang H *et al*: Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 2008; **319**: 1100–1104.
- 8 Altshuler DM, Gibbs RA, Peltonen L *et al*: Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; **467**: 52–58.