Diversity of Lactase Persistence Alleles in Ethiopia
Signature of a Soft Selective Sweep

Jones, BL; Raga, TO; Liebert, Anke; Olsen, Anders Krüger; Danielsen, Erik Thomas; Bekele, E; Troelsen, Jesper; Bradman, N; Swallow, Dallas M

Published in:
American Journal of Human Genetics

DOI:
10.1016/j.ajhg.2013.07.008

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact rucforsk@ruc.dk providing details, and we will remove access to the work immediately and investigate your claim.
Diversity of Lactase Persistence Alleles in Ethiopia: Signature of a Soft Selective Sweep

Bryony L. Jones,1 Tamiru O. Raga,2 Anke Liebert,1 Pawel Zmarz,1 Endashaw Bekele,2 E. Thomas Danielsen,3 Anders Krüger Olsen,3,5 Neil Bradman,1,4 Jesper T. Troelsen,3 and Dallas M. Swallow1,*

The persistent expression of lactase into adulthood in humans is a recent genetic adaptation that allows the consumption of milk from other mammals after weaning. In Europe, a single allele (−13910*T, rs4988235) in an upstream region that acts as an enhancer to the expression of the lactase gene LCT is responsible for lactase persistence and appears to have been under strong directional selection in the last 5,000 years, evidenced by the widespread occurrence of this allele on an extended haplotype. In Africa and the Middle East, the situation is more complicated and at least three other alleles (−13907*G, rs41525747; −13915*G, rs41380347; −14010*C, rs145946881) in the same LCT enhancer region can cause continued lactase expression. Here we examine the LCT enhancer sequence in a large lactose-tolerance-tested Ethiopian cohort of more than 350 individuals. We show that a further SNP, −14009T>G (ss 820486563), is significantly associated with lactose-digestor status, and in vitro functional tests confirm that the −14009*G allele also increases expression of an LCT promoter construct. The derived alleles in the LCT enhancer region are spread through several ethnic groups, and we report a greater genetic diversity in lactose digesters than in nondigesters. By examining flanking markers to control for the effects of mutation and demography, we further describe, from empirical evidence, the signature of a soft selective sweep.

Lactase, the enzyme that digests the milk sugar lactose, persists into adult life in approximately 35% of the world’s population.1 This genetic trait of adult lactase persistence (LP [MIM 223100]) is a recent human adaptation permitting those who carry it to use animal milk more readily as a source of nutrition. This LP phenotype is in contrast to the ancestral mammalian phenotype shared by most of the human population where lactase is downregulated before adulthood.

LP is attributable to nucleotide changes in a regulatory region that acts as an enhancer of the expression of the gene encoding lactase, LCT (MIM 603202). This LCT enhancer is located in intron 13 of the neighboring gene, MCM6 (MIM 601806), immediately upstream of LCT (for review see Ingram et al.1). LP in Europe is generally attributable to a single allele (−13910*T, rs4988235) that seems to have been under strong directional selection in the last 5,000–10,000 years.2 The evidence for selection comes from the fact that the allele lies on an extended haplotype3,4 with low microsatellite diversity and is at a significantly higher frequency than expected for the age of the allele.1,4–6 Tests of haplotype homogeneity and population differentiation in genome-wide studies that focus on European samples show that the region of chromosome 2 containing LCT has one of the highest “signatures” of selection.7 In Africa and the Middle East, the situation is more complicated and three additional alleles (−13907*G, rs41525747; −13915*G, rs41380347; −14010*C, rs145946881) in the same LCT enhancer region have been reported to cause lactase persistence.8–16

In Tanzania and Kenya, one particular allele, −14010*C, is at high frequency and gives a significant signal of positive selection in tests of haplotype homozygosity, with expansion of the −14010*C allele dated to approximately 3,000–7,000 years ago.16 In some cases, however, including the Jaali from Sudan and the Somali camel herders from Ethiopia, several different LP alleles are associated with lactase persistence in the same ethnic group.15 Indeed, we observed that in these groups, the sequence of the enhancer region was much more diverse in lactose digesters than in nondigesters. Analysis of mitochondrial DNA, Y chromosome, and autosomal microsatellite variations demonstrated that this difference in diversity was not attributable to hidden population stratification.15

Although it is possible that not all of the described enhancer alleles are functional, we speculated that this difference in diversity had been influenced by what has been described as a soft selective sweep—the phenomenon by which several different variants of similar function rise in frequency simultaneously.17–19 Such soft selective sweeps might not be detectable by published methods of genome-wide detection of selection.20,21

Here, we examine genetic diversity of the LCT enhancer in a larger lactose-tolerance-tested cohort (>350 individuals), which consists of volunteers from several ethnic groups in Ethiopia. We aim to determine whether any

1Research Department of Genetics Evolution and Environment, University College London, Darwin Building, London WC1E 6BT, UK; 2Department of Biology, University of Addis Ababa, PO Box 32597, Ethiopia; 3Department of Science, Systems and Models, Roskilde University, 4000 Roskilde, Denmark; 4Henry Stewart Group, 28/30 Little Russell Street, London WC1A 2HN, UK
5Present address: Department of Pathology, Hospital South, 4700 Naestved, Denmark
*Correspondence: d.swallow@ucl.ac.uk
http://dx.doi.org/10.1016/j.ajhg.2013.07.008. ©2013 The Authors
This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
Table 1. Allele Counts of Variants in Intron 13 of MCM6 and Association with Lactose-Digester Status

<table>
<thead>
<tr>
<th>rs/ss Number</th>
<th>H2NP</th>
<th>Intermediate</th>
<th>Digester</th>
<th>Nondigester</th>
<th>p value</th>
<th>p value (2 allele carriers excluded)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs498823</td>
<td>14</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0.076</td>
<td>0.00008*</td>
</tr>
<tr>
<td>rs145946881</td>
<td></td>
<td></td>
<td>3</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>ss820486563</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>4.63 x 10^-22*</td>
<td></td>
</tr>
<tr>
<td>ss820496514</td>
<td></td>
<td></td>
<td>7</td>
<td>0</td>
<td>0.313</td>
<td>0.00242*</td>
</tr>
<tr>
<td>ss831883152</td>
<td></td>
<td></td>
<td>6</td>
<td>0</td>
<td>0.0264</td>
<td>0.00264</td>
</tr>
<tr>
<td>rs41380347</td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs41456145</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1.91 x 10^-16*</td>
<td></td>
</tr>
<tr>
<td>ss820496565</td>
<td></td>
<td></td>
<td>10</td>
<td>0</td>
<td>0.801</td>
<td>0.801</td>
</tr>
<tr>
<td>rs144412793</td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs4954492</td>
<td></td>
<td></td>
<td>17</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs56348046</td>
<td></td>
<td></td>
<td>2</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs4988235</td>
<td></td>
<td></td>
<td>8</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs41525747</td>
<td></td>
<td></td>
<td>10</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>ss820496514</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs56348046</td>
<td></td>
<td></td>
<td>14</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs41380347</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs41456145</td>
<td></td>
<td></td>
<td>2</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>ss820496565</td>
<td></td>
<td></td>
<td>10</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs144412793</td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs4954492</td>
<td></td>
<td></td>
<td>17</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs56348046</td>
<td></td>
<td></td>
<td>2</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs4988235</td>
<td></td>
<td></td>
<td>8</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs41525747</td>
<td></td>
<td></td>
<td>10</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>ss820496514</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs144412793</td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs4954492</td>
<td></td>
<td></td>
<td>17</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs56348046</td>
<td></td>
<td></td>
<td>2</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

The results reported here are from a total of 356 volunteers (712 chromosomes). To test association of the alleles independently, counts were also made after exclusion of all individuals who carry two different derived alleles (numbers in parentheses). Fisher’s exact test two-sided p values are shown for the comparison of digesters and nondigesters. Asterisks (*) indicate statistically significant association after Bonferroni correction for 12 tests (threshold p = 0.004). Notes that – 14011*T is present in one digester, but the second carrier, despite being a nondigester, also carried – 13907*G, so was presumably suffering from secondary loss of lactase. The location of – 14011*T immediately next to – 14010*C suggests the possibility of function.
enhancer variation and confirmed hydrogen production are more difficult to explain and might be real, suggesting that a different causal mechanism could be at work. A mutation in a different cis-acting region could be responsible for LP in these individuals, or perhaps even a trans-acting factor (although trans-acting variants are less likely to be causal because such mutations are more likely than cis-acting variants to have a pleiotropic effect\textsuperscript{24}). Epigenetic influences could also play a role in the persistence of lactase in some of these individuals.

In an initial analysis of the diversity of the \(\text{LCT}\) enhancer region, we found greater sequence diversity in the digesters than in the nondigesters (Nei’s \(H\) for digesters 0.73 versus 0.023 for nondigesters; \(\pi\) 0.002 for digesters versus 0.0005 for nondigesters), which is consistent with our previous findings in Ethiopian Somali camel herders.\textsuperscript{15} To further explore the causes of this greater diversity in digesters, we sought to control for the possible effects of demography and mutation rate. To do this, we sequenced two control regions either side of the \(\text{LCT}\) enhancer that were unlikely in most cases to have been separated from the enhancer by recombination since the onset of the spread of these alleles (for primers, cycling, and sequencing conditions see Table S1).

We selected a 500 bp fragment 16 kb upstream of the \(\text{LCT}\) enhancer in intron 4 of \(\text{MCM6}\) and a 361 bp fragment 13 kb downstream of the enhancer, approximately 1 kb upstream of \(\text{LCT}\) exon 1. Assuming a recombination rate of 0.5 cM/Mb observed in families for that region (UCSC Genome Browser), 5\% or fewer of the chromosomes are likely to have recombined during the last 300 generations (5,000 years). Intron 4 of \(\text{MCM6}\) is likely to have a similar chromatin state to that of intron 13 and, therefore, a similar vulnerability to germline mutation, thus doubling as a control region with similar mutation rate.

The subsequent analyses were conducted on samples with complete sequence data for all three regions, irrespective of phenotype classification. From the previously collected cohort of Ethiopian Somali camel herders,\textsuperscript{15} 81 individuals were included for comparison. The frequencies of sequence variants in the two control regions for digesters and nondigesters are shown in Table S3.

Haplotypes were inferred with the computer program PHASE\textsuperscript{25} and the results were then checked by visual inspection of the data. The haplotypes that were present in this cohort, excluding those haplotypes with fewer than three occurrences, are shown in Figure 2 in relation to our previous designations.\textsuperscript{27} Figure S1 details the full set of haplotypes and indicates nonrecombinant and recombinant chromosomes. Only 7\% of chromosomes show evidence of historic recombination between the flanking regions and the frequency of detectable recombinants is lower for the chromosomes carrying the derived enhancer alleles (3.9\%). As previously reported, \(-13915^*G\) is found on a C, or a closely related haplotype (of which 85 out of 88 are identical across all three regions), and \(-13907^*G\) is on an extended A, or closely related

![Figure 1. Transfection Experiment Showing the Effect of the \(-14009^*G\) Variant on the Enhancer Activity in Caco-2 Cells](image-url)
haplotype, in 62 out of 63 cases (Figures 2 and S1). The
−14009*G allele occurred in 34 out of 36 cases on a haplo-
type that has a deletion at −942, which, according to our
previous haplotype designations, could be an O, S, T, U,
or X type haplotype. Typing position 5579 of the LCT
cDNA sequence (rs2278544) of a number of the chromo-
somes with the value 0.0002).

As an additional check that the enhancer region in
intron 13 is not intrinsically more mutable than the other
regions that we have examined, we compared the diver-
gence of the three regions from primate sequences by
using the sequences available from the UCSC Human
Genome Browser. We observed that the divergence of the
intron 13 region is in fact less than that of the other two
regions. For example, in comparison with the rhesus
macaque, the values obtained were 5.9%, 5.2%, and
6.3% for control region 1, the enhancer sequence, and con-
trol region 2, respectively.

We then compared haplotype diversity (H) and nucleo-
tide diversity (π) of all three sequence regions (Table 2) in
the digesters and nondigesters from the Amhara, Tigre,
Oromo, and Wolayita and the previously collected Somalis.
Individuals of mixed ancestry or those belonging to groups
with fewer than 20 individuals in our data set, for the pur-
poses of this analysis were grouped together, irrespective
of language family, as “all other Ethiopians.”

Figure 3 shows the distribution of
corrected p value was computed as the
proportion of permutations with an absolute π difference
equal to or larger than the observed π difference be-
tween digesters and nondigesters. For the enhancer, there
was a significant difference in π between digesters and
nondigesters for all ethnic groups (p < 0.005) (Table S4).
Similar comparison of the control sequences showed just
one ethnic group with a significant π difference (the
Oromo for the control region in intron 4, p = 0.0002).

Haplotype diversity of the enhancer region was also
significantly greater in digesters than nondigesters for all
ethnic groups (as determined by the test Hdiff;26 Table
S4). This distribution is depicted in Figure S2. As for the
π analysis, the only significant difference in the control
sequences was the intron 4 region in Oromo (p

The American Journal of Human Genetics 93, 538–544, September 5, 2013 541

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>B,F,G,L</td>
<td>278</td>
</tr>
<tr>
<td>A,H,J,K,P,Y,Z</td>
<td>122</td>
</tr>
<tr>
<td>C,E,M</td>
<td>62</td>
</tr>
<tr>
<td>C.E.M</td>
<td>62</td>
</tr>
<tr>
<td>A,H,J,K,P,Y,Z</td>
<td>41</td>
</tr>
<tr>
<td>A,H,J,K,P,Y,Z</td>
<td>39</td>
</tr>
<tr>
<td>O,S,T,U,X</td>
<td>34</td>
</tr>
<tr>
<td>A,H,J,K,P,Y,Z</td>
<td>18</td>
</tr>
<tr>
<td>C,E,M</td>
<td>10</td>
</tr>
<tr>
<td>B,F,G,L</td>
<td>7</td>
</tr>
<tr>
<td>A,H,J,K,P,Y,Z</td>
<td>4</td>
</tr>
<tr>
<td>A,H,J,K,P,Y,Z</td>
<td>4</td>
</tr>
<tr>
<td>A,H,J,K,P,Y,Z</td>
<td>3</td>
</tr>
<tr>
<td>A,H,J,K,P,Y,Z</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 2. Phased Haplotypes for the Lactase Enhancer and Two Flanking Re-
regions, in Intron 4 of MCM6 and 1 kb Up-
stream of LCT in a Cohort of 422 Ethiopian
Individuals with No Known Shared
Ancestry to the Grandparental Level, for
whom Full Sequencing Data Were
Available

Haplotypes that occurred <3 times (n = 21)
are omitted. The variable sites are
numbered and refer to the following chromosome
positions relative LCT; 1, −30210*C; 2, −30203; 3, −30182*G;
4, −29949*C; 5, −14010*C; 6, −14009*G;
7, −13957*G; 8, −13915*G; 9, −13913*C;
10, −13910*G; 11, −13907*G; 12,
−13806*G; 13, −13730*G; 14, −13603*G;
15, −958*A; 16, −942*del; 17, −875*4; 18,
−678*G. n indicates the number of chro-
mosomes for which that haplotype was
inferred. The lettered haplotypes that are
shown refer to the haplotypes previously
reported by Hollox et al.27 Those that are
shown in bold are the most likely haplo-
type according to previously reported distri-
butions and examination of additional
alleles (unpublished data).

Figure 3
Thus we confirm our previous observation on the diversity of the LCT enhancer sequence region and show with a cohort of independent samples that there is significantly greater nucleotide and haplotype diversity in the enhancer region in digesters than in nondigesters. By sequencing two flanking regions, which have followed the same demographic history as the enhancer over the last 300 generations, both control sequences have rather high diversity in nondigesters as well as in digesters. Indeed, it is the lack of diversity in the enhancer sequence of nondigesters that is noteworthy and this seems to reflect the general conservation of this sequence across primates. Between positions −14028 and −13800, the sequences are approximately 93.5% identical across humans, chimpanzee, gorilla, orangutan, gibbon, baboon, and rhesus macaque, and percentage identity declines on either side of this (see Figure S4). Thus, this sequence region may have been under evolutionary constraint because of its function as a regulatory element for lactase expression in infant mammals. This constraint appears to have been overcome in some human populations, presumably because of the benefit of allowing adult lactase expression.

This study brings the total of confirmed LP causal variants to five (−14010*C, −14009*G, −13915*G, −13910*T, and −13907*G). Their close location, all within the most conserved part of the 450 bp segment known to have

<table>
<thead>
<tr>
<th>Ethnic Group</th>
<th>Digestor Status</th>
<th>Seg Sites</th>
<th>N</th>
<th>Haplotype</th>
<th>H</th>
<th>π</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amhara</td>
<td>nondigester</td>
<td>62</td>
<td>3</td>
<td>4</td>
<td>0.672</td>
<td>0.0019</td>
</tr>
<tr>
<td></td>
<td>digester</td>
<td>32</td>
<td>2</td>
<td>3</td>
<td>0.669</td>
<td>0.0017</td>
</tr>
<tr>
<td>Oromo</td>
<td>nondigester</td>
<td>78</td>
<td>3</td>
<td>4</td>
<td>0.556</td>
<td>0.0013</td>
</tr>
<tr>
<td></td>
<td>digester</td>
<td>54</td>
<td>3</td>
<td>5</td>
<td>0.706</td>
<td>0.0021</td>
</tr>
<tr>
<td>Tigre</td>
<td>nondigester</td>
<td>30</td>
<td>3</td>
<td>4</td>
<td>0.522</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>digester</td>
<td>48</td>
<td>2</td>
<td>3</td>
<td>0.635</td>
<td>0.0015</td>
</tr>
<tr>
<td>Wolayita</td>
<td>nondigester</td>
<td>24</td>
<td>2</td>
<td>3</td>
<td>0.565</td>
<td>0.0013</td>
</tr>
<tr>
<td></td>
<td>digester</td>
<td>16</td>
<td>2</td>
<td>3</td>
<td>0.677</td>
<td>0.0016</td>
</tr>
<tr>
<td>Somali</td>
<td>nondigester</td>
<td>98</td>
<td>3</td>
<td>4</td>
<td>0.559</td>
<td>0.0013</td>
</tr>
<tr>
<td></td>
<td>digester</td>
<td>32</td>
<td>3</td>
<td>4</td>
<td>0.688</td>
<td>0.0017</td>
</tr>
<tr>
<td>Other</td>
<td>nondigester</td>
<td>162</td>
<td>4</td>
<td>5</td>
<td>0.621</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>digester</td>
<td>132</td>
<td>2</td>
<td>4</td>
<td>0.652</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

Using the fully phased data set, measures of genetic diversity were calculated with DnaSP software, using those samples with a clear diagnosis of either lactose-digester or -nondigester status. Note that for the enhancer region, the numbers of segregating sites, numbers of haplotypes, and calculated values for H (haplotype heterozygosity) and π (nucleotide diversity) are in each case larger in the digesters than nondigesters. These analyses were conducted on samples for which uniform ancestry was recorded and a single additional group was composed of all others. Two simple indel polymorphisms (one in intron 4 and one in the upstream LCT region) are considered as SNPs in the calculations. N = number of chromosomes. Detailed statistics are shown in Table S4.

Figure 3. Comparison, in Nondigesters and Digesters, of Nucleotide Diversity π Measured across the Three Sequence Regions

Data points show the five ethnic groups tested: Amhara, Tigre, Oromo, Wolayita, Somali, and “other Ethiopians” as a single group. Red horizontal bars show median values. See Table 2 for n values. Abbreviations are as follows: D, digester; ND, nondigester.
enhancer function, suggests that the functional region is smaller than the sequence that we have tested and this will be explored further in future studies.

The coexistence of all five alleles in Ethiopia goes some way to explaining the greater nucleotide diversity within the LCT enhancer in intron 13 of MCM6 in digesters. Ethiopia has been a crossroads of human migrations in the last 5,000 years since the LP alleles are likely to have come under selection, and studies on other African and Middle eastern populations (B.L.J., D.M.S., and colleagues, unpublished data) show quite different geographic distributions, with overlap in Ethiopia, suggesting that their origins are all different, but determining where these were and how they spread is likely to be difficult. The combination of mutation, large effective population size, migration, and selection has been shown to be important in generating this kind of pattern of diversity, namely parallel selection of multiple alleles of similar function, a so-called soft selective sweep. Here we confirm this unusual pattern of diversity in the LCT enhancer region in Ethiopia, and by testing flanking sequences we control, at a chromosomal level, for possible differences in migrational history, effective population size, and mutation rate between the digesters and nondigesters. Because increased genetic diversity in digesters is localized to the LCT enhancer, we can infer that recent selection is acting on this small relatively conserved functional sequence region. Selection has the effect of increasing the frequency of the background haplotypes on which the derived alleles occur, though the impact on diversity of the flanking sequences is not large. This pattern contrasts with that found with this article online at http://www.cell.com/AJHG/.

Supplemental Data

Supplemental Data include four figures and four tables and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments

This work was funded by the MRC UK (MRC DTA studentship for B.L.J.), the European Union (Marie Curie ITN FP7 Framework Programme grant, LeCHE, Grant ref 215362-2 to A.L.), the Annals of Human Genetics (B.L.J. and A.L.), and Melford Charitable Trust (studentship for T.O.R.). N.B. is the settlor and senior trustee of Melford Charitable Trust. Neither N.B. nor the charitable trust has any intellectual property or other rights with respect to the results of the study. We thank Mari Wyn Burley and Ranji Araseretnam for technical help, Mirna Kovacevic and Ripudaman Bains for their help with data handling, and Ed Hollox for helpful discussion.

Web Resources

The URLs for data presented herein are as follows:

Dnasp, http://www.ub.edu/dnasp/

Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/

UCSC Genome Browser, http://genome.ucsc.edu

References


