

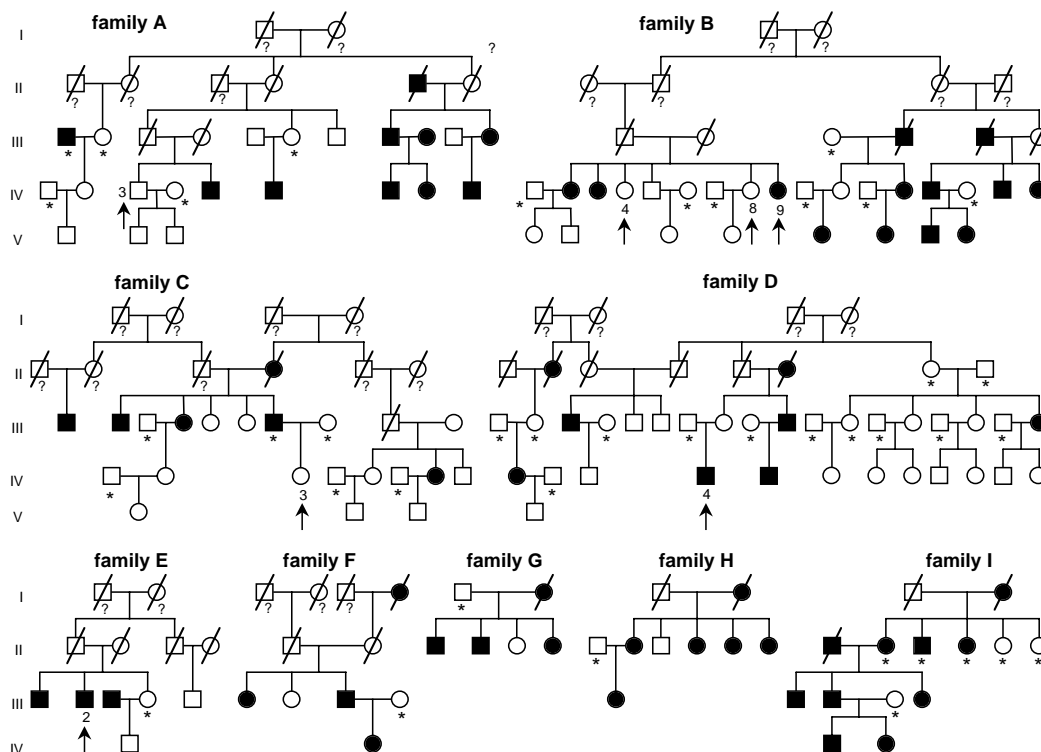
# Identification of a variant associated with adult-type hypolactasia

Nabil Sabri Enattah<sup>1</sup>, Timo Sahi<sup>2</sup>, Erkki Savilahti<sup>3</sup>, Joseph D. Terwilliger<sup>4</sup>, Leena Peltonen<sup>1,5\*</sup> & Irma Järvelä<sup>1,6</sup>

Published online: 14 January 2002, DOI: 10.1038/ng826

Adult-type hypolactasia, also known as lactase non-persistence (lactose intolerance), is a common autosomal recessive condition resulting from the physiological decline in activity of the lactase-phlorizin hydrolase (LPH) in intestinal cells after weaning. LPH hydrolyzes lactose into glucose and galactose. Sequence analyses of the coding and promoter regions of *LCT*, the gene encoding LPH, has revealed no DNA variations correlating with lactase non-persistence<sup>1,2</sup>. An associated haplotype spanning *LCT*, as well as a distinct difference in the transcript levels of 'non-persistence' and 'persistence' alleles in heterozygotes, suggest that a *cis*-acting element contributes to the lactase non-persistence phenotype<sup>3,4</sup>. Using linkage disequilibrium (LD) and haplotype analysis of nine

extended Finnish families, we restricted the locus to a 47-kb interval on 2q21. Sequence analysis of the complete region and subsequent association analyses revealed that a DNA variant, C/T<sub>-13910</sub>, roughly 14 kb upstream from the *LCT* locus, completely associates with biochemically verified lactase non-persistence in Finnish families and a sample set of 236 individuals from four different populations. A second variant, G/A<sub>-22018</sub>, 8 kb telomeric to C/T<sub>-13910</sub>, is also associated with the trait in 229 of 236 cases. Prevalence of the C/T<sub>-13910</sub> variant in 1,047 DNA samples is consistent with the reported prevalence of adult-type hypolactasia in four different populations. That the variant (C/T<sub>-13910</sub>) occurs in distantly related populations indicates that it is very old.



**Fig. 1** Finnish families with lactase non-persistence (hypolactasia). Filled symbols represent biochemically verified hypolactasic individuals with non-persistence; asterisks indicate that no sample was available; question marks represent unknown affection status and arrows and numbers indicate the individuals whose DNA was used for sequencing (see Table 2).

<sup>1</sup>Department of Molecular Medicine, National Public Health Institute, Haartmaninkatu 8, PO Box 104, FIN-00251 Helsinki; and Department of Medical Genetics, University of Helsinki, Finland. <sup>2</sup>Department of Public Health, University of Helsinki. <sup>3</sup>Hospital for Children and Adolescents, University of Helsinki, Helsinki, Finland. <sup>4</sup>Department of Psychiatry and Columbia Genome Center, Columbia University, New York, New York, USA; and Division of Medical Genetics, New York State Psychiatric Institute, New York, New York, USA. <sup>5</sup>Department of Human Genetics, UCLA School of Medicine, Los Angeles, California, USA. <sup>6</sup>HUCH-Laboratory Diagnostics, Laboratory of Molecular Genetics, Haartmanninkatu 2, PO Box 340, FIN-00290 Helsinki, Finland. Correspondence should be addressed to L.P. (e-mail: lpeltonen@mednet.ucla.edu).

**Table 1 • Linkage and linkage disequilibrium in families with adult lactase non-persistence (the fine mapping markers are *D2S3010*–*D2S3018*)**

Marker	Lod score(Z) at $\theta$					LD (P value) <sup>a</sup>
	0.0	0.1	0.2	0.3	0.4	
<i>D2S114</i>	–∞	2.44	1.92	1.13	0.41	0.87195
P6112	2.76	2.20	1.45	0.75	0.22	0.66207
<i>D2S1334</i>	3.15	2.45	1.61	0.84	0.25	0.91039
<i>D2S3010</i>	2.26	1.99	1.36	0.71	0.21	0.53670
<i>D2S3011</i>	3.67	2.94	1.96	1.03	0.31	4 × 10 <sup>–6</sup>
<i>D2S3012</i>	4.09	3.07	2.00	1.00	0.26	5.7 × 10 <sup>–7</sup>
<i>D2S3013</i>	5.91	4.52	2.96	1.53	0.46	5 × 10 <sup>–6</sup>
<i>D2S3015</i>	3.63	2.60	1.66	0.83	0.23	0.03471
<i>D2S3014</i>	6.63	4.88	3.16	1.61	0.44	3.2 × 10 <sup>–8</sup>
<i>D2S3016</i>	3.07	2.22	1.42	0.71	0.19	4 × 10 <sup>–5</sup>
<i>D2S3017</i>	5.33	4.10	2.72	1.39	0.39	0.02166
<i>D2S3018</i>	6.60	4.99	3.25	1.65	0.46	1 × 10 <sup>–5</sup>
<i>D2S2196</i>	7.67	5.62	3.62	1.85	0.54	0.00010
<i>D2S442</i>	3.81	3.08	2.08	1.03	0.27	0.22805
<i>D2S314</i>	4.22	3.61	2.50	1.37	0.45	0.27535
<i>D2S2385</i>	–∞	2.79	1.92	1.01	0.28	0.46457

<sup>a</sup>P values produced using linkage disequilibrium test, given linkage. LD was monitored conditional on the detected linkage, treating the allele frequencies and the recombination fraction as nuisance parameters<sup>23,24</sup>.

Lactase non-persistence limits the use of fresh milk among adults. The age of onset of lactase non-persistence ranges from 1–2 years among the Thai population to 10–20 years among Finns<sup>5,6</sup>. In some populations, LPH activity persists throughout life in the majority of adults, a condition known as adult lactase persistence<sup>7</sup>. To identify the underlying DNA variation, we analyzed seven polymorphic microsatellite markers flanking *LCT* on 2q21 in nine extended Finnish families with hypolactasia (Fig. 1). We found significant evidence of linkage with markers *D2S314*, *D2S442*, *D2S2196* and *D2S1334*; we obtained the highest lod score (7.67) at  $\theta=0$  ( $\theta$  = recombination fraction) with marker *D2S2196* (Table 1). Recombination events defined the centromeric and telomeric boundary for the lactase persistence/non-persistence locus at markers *D2S114* and *D2S2385*, respectively (Fig. 1 and Table 1).

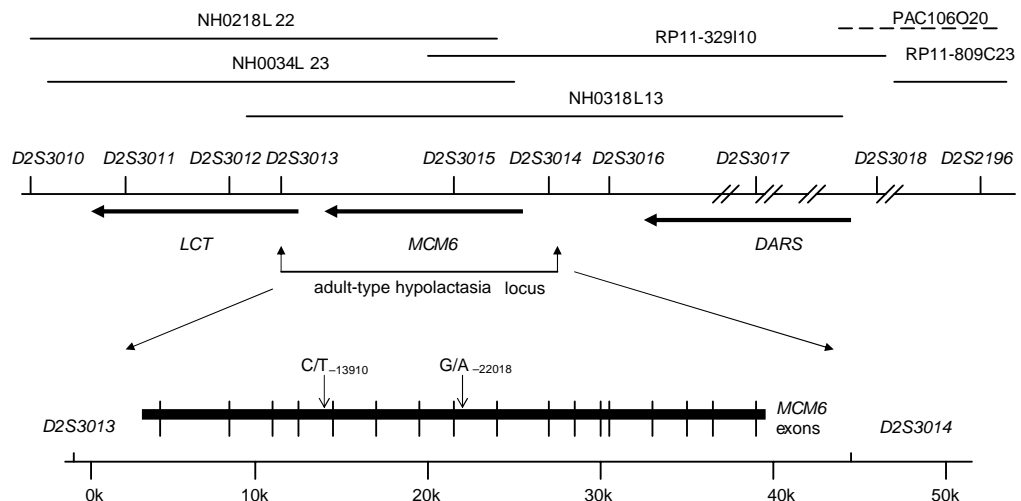
We assembled the draft sequence of the four BAC clones NH0034L23, NH0218L22, NH0318L13 and 329I10 (Fig. 2) into one uninterrupted 222.5-kb contig (Fig. 2). We used nine polymorphic markers in this region for the fine-mapping (Table 1). Six markers spanning a 200-kb region showed highly significant evi-

dence of LD ( $P < 10^{-4}$ ), whereas markers 3' of *LCT* showed no evidence of LD with lactase persistence alleles (Table 1). We constructed marker haplotypes over the critical interval for all alleles that were not identical by descent in the nine Finnish families, a total of 54 non-persistence alleles and 33 persistence alleles. One major haplotype was present in 20 persistence alleles (60%) versus 3 (5%) of the non-persistence alleles, whereas a wide diversity of haplotypes was seen in non-persistence alleles. The remaining 13 persistence alleles (40%) were detected in the ancestral haplotype, based on ancient recombinations (Fig. 3). The conserved haplotype restricted the critical region for lactase persistence to a 47-kb interval between mark-

ers *D2S3013* and *D2S3014*. Consistent with our data, one major persistence haplotype has been reported in other Caucasian populations, such as British and French<sup>4,8</sup>. Earlier efforts to monitor the common haplotype, however, focused solely on *LCT* and sequence in its immediate vicinity<sup>4</sup>.

We analyzed the 47-kb region in DNA of seven family members, four with lactase persistence and three with non-persistence. The region contains only one known gene, *MCM6* (ref. 9), that covers 36 kb of the critical 47 kb (Fig. 2). Although we did not find any variations in the coding region of *MCM6*, we identified a total of 52 non-coding variants: 43 SNPs and 9 deletion/insertion polymorphisms (Table 2). In the Finnish families, only two of the variants (C/T<sub>-13910</sub>, G/A<sub>-22018</sub>), 8 kb apart, showed complete cosegregation with the non-persistence trait (Tables 2 and 3). All family members with non-persistence were homozygous with respect to both C<sub>-13910</sub> and G<sub>-22018</sub> (Table 3). The C/T<sub>-13910</sub> variant is 13,910 bp 5' of the initiation codon of *LCT*, in intron 13 of *MCM6*. The G/A<sub>-22018</sub> variant is 22,018 bp upstream of the first ATG codon of *LCT*, in intron 9 of *MCM6* (Fig. 2).

**Fig. 2** Physical map of adult lactase non-persistence locus. BAC clones are shown above the horizontal line. The three genes *LCT*, *MCM6* and *DARS* are shown by thick black arrows indicating 5'→3' direction. The positions of ten polymorphic microsatellite markers used for fine mapping of the locus are given. The backslashes in the horizontal line denote gaps in the contig sequence. The position of marker *D2S2169* was confirmed by bridging the gap with PAC 106O20 isolated from the PAC library. The organization of *MCM6* is shown below, including the position of the lactase persistence-associated variants in introns 9 and 13, located 13.9 kb and 22 kb 5' of the first ATG of *LCT*.



We analyzed these variants in Finnish DNA samples isolated from a total of 196 intestinal biopsy specimens with biochemically determined disaccharidase (LPH) activity. All 59 samples showing primary lactase deficiency were homozygous with respect to the C allele of C/T<sub>-13910</sub>, 6 were heterozygous with

respect to the G/A<sub>-22018</sub> variant and the remaining 53 were homozygous with respect to the G allele. Of the 137 cases showing lactase persistence, none were homozygous with respect to alleles C and G, at C/T<sub>-13910</sub> and G/A<sub>-22018</sub>, respectively; 74 were homozygous for alleles T and A, with 63 being heterozygous at both positions. All 40 non-Finnish cases with verified disaccharidase deficiency (23 from South Korea, 9 from Italy and 8 from Germany) were homozygous (CC) with respect to C/T<sub>-13910</sub>. One Italian sample was heterozygous (GA) for G/A<sub>-22018</sub>, whereas the remaining 39 cases were homozygous (GG) at this position (Table 3).

To determine the prevalence of the lactase non-persistence-associated variant, we used the solid-phase minisequencing method<sup>10</sup> to screen DNA samples of 938 anonymous Finnish blood donors originating either from the Western, early settlement or Eastern, late settlement regions of Finland. The overall prevalence of the lactase non-persistence genotype CC<sub>-13910</sub> (170 cases) was 18.1%, with a lower prevalence in the sample from the western region (16.8%) than in the eastern region (18.9%;  $P=0.02$ , 1 df; Web Table A). These values are consistent with the results of an epidemiological study<sup>5</sup>.

The prevalence of lactase non-persistence in different populations varies greatly, from less than 5% in northern Europe to almost 100% in southeast Asia<sup>7</sup>. We determined the distribution of the genotype CC<sub>-13910</sub> in 17 French people, 92 white North Americans and 96 African Americans. We observed the genotype CC<sub>-13910</sub> in 41.2% of the French, in 7.6% of the white North Americans and in 79.2% of African Americans, in agreement with epidemiological data (Web Table A)<sup>11,12</sup>.

The variant C/T<sub>-13910</sub>, completely associating with the non-persistence trait in both Finnish and non-Finnish samples, represents the most likely major variant underlying adult lactase non-persistence. The second variant, G/A<sub>-22018</sub>, is also fully associated with lactase persistence in the Finnish families, whereas in the case-control study sample, seven individuals with biochemically verified lactase non-persistence did not carry the associated genotype for this variant. However, we cannot exclude the possibility that both variants in the 8-kb haplotype would participate in the regulation of LPH activity.

Two hypolactasia-associated DNA variants reside at a considerable dis-

**Table 2 • Variations identified within Finnish families with a 47-kb adult lactase non-persistence locus**

Position <sup>a</sup>	Variant	Lactase persistence (homozygous)		Lactase persistence (heterozygous)		Lactase non-persistence		
		BIV4	AIV3	BIV8	CIV3	BIV9	DIV4	EIII2 <sup>b</sup>
-694	A→G	AA	AA	AG	AA	GG	N <sup>c</sup>	AA
-1640/50	T <sub>13</sub> →T <sub>12</sub>	T <sub>13/13</sub>	T <sub>13/13</sub>	T <sub>13/13</sub>	T <sub>13/13</sub>	T <sub>13/13</sub>	T <sub>12/12</sub>	T <sub>12/12</sub>
-2131	C→T	CC	CC	CT	CC	TT	CT*	TT
-3058/ 72	T <sub>15</sub> →T <sub>16</sub>	T <sub>15/15</sub>	T <sub>15/15</sub>	T <sub>15/15</sub>	T <sub>15/15</sub>	T <sub>15/15</sub>	T <sub>16/16</sub>	T <sub>16/16</sub>
-3075	G→T	GG	GG	GG	GG	GG	GG	TT
-4480	T→A	TT	TT	TA	TT	AA	TT	TT
-5440	C→T	CC	CC	CT	CC	TT	CC	CC
-5926	A→T	AA	AA	AA	AA	AA	TA	TT
-8540	G→A	GG	GG	GA	GA	AA	AG	AA
-8630	C→G	CC	CC	CG	CG	GG	GC	GG
-13495	T→C	TT	TT	TC	TT	CC	CT	CC
-13910	T→C	TT	TT	TC	TC	CC	CC	CC
-15239	G→A	GG	GG	GA	GG	AA	AG	AA
-15862	T→C	CC	CC	CT	CC	TT	TC	TT
-16568/79	T <sub>11</sub> →T <sub>12</sub>	T <sub>11/11</sub>	T <sub>11/11</sub>	T <sub>11/12</sub>	T <sub>11/11</sub>	T <sub>12/12</sub>	T <sub>11/11</sub>	T <sub>12/12</sub>
-16888	A→G	AA	AA	GA	AA	GG	GA	GG
-17300	C→T	CC	CC	CC	CC	CC	CT	TT
-19044	T→C	TT	TT	TC	TT	CC	CT	CC
-19519	T→C	TT	TT	TC	TT	CC	TT	TT
-20077	C→G	CC	CC	CG	CC	GG	GC	GG
-20486	G→A	GG	GG	GA	GG	AA	GG	GG
-21721/28	A <sub>7</sub> →A <sub>6</sub>	A <sub>7/7</sub>	A <sub>7/7</sub>	A <sub>7/7</sub>	A <sub>7/7</sub>	A <sub>7/7</sub>	A <sub>7/A<sub>6</sub></sub>	A <sub>7/7</sub>
-21731	A→C	AA	AA	AA	AA	AA	CC	AA
-21736/43	A <sub>9</sub> →A <sub>8</sub>	A <sub>9/9</sub>	A <sub>9/9</sub>	A <sub>9/A<sub>8</sub></sub>	A <sub>9/9</sub>	A <sub>8/8</sub>	A <sub>8/8</sub>	A <sub>8/8</sub>
-22018	G→A	AA	AA	AG	AG	GG	GG	GG
-22741	C→T	CC	CC	CC	CC	CC	N	TT
-22788	A→G	AA	AA	AG	AA	GG	N	GG
-23069	A→G	AA	AA	AG	AA	GG	N	GG
-23442	A→G	AA	AA	AA	AA	AA	N	GG
-23771	T→C	TT	TT	TT	TT	TT	N	CC
-25093/23	Δ30 bp	ΔΔ	ΔΔ	ΔΔ	ΔΔ	ΔΔ	N	II
-27310	A→G	AA	AA	AG	AA	GG	GA	GG
-27480	G→A	GG	GG	GA	GG	AA	AG	AA
-27807	A→C	AA	AA	AA	AA	AA	AC	CC
-30183	A→G	AA	AA	AG	AA	GG	AA	AA
-31268	A→G	AA	AA	AG	AA	GG	AA	AA
-31342	T→C	TT	TT	TT	TT	TT	CT	CC
-33645	C→T	CC	CC	CT	CC	TT	CC	CC
-35176	T→C	TT	TT	TC	TT	CC	CT	CC
-36254	C→T	CC	CC	CT	CC	TT	TC	TT
-36296	G→T	TT	TT	TG	TT	GG	TG	N
-36501	A→T	AA	AA	AT	AA	TT	AT	N
-36506/14	Δ9 bp	ΔΔ	ΔΔ	ΔI	ΔΔ	II	ΔI	N
-36671/77	T <sub>7</sub> →T <sub>6</sub>	T <sub>7/7</sub>	T <sub>7/7</sub>	T <sub>7/6</sub>	T <sub>7/7</sub>	T <sub>6/6</sub>	T <sub>7/7</sub>	T <sub>7/7</sub>
-37565	T→G	TT	TT	TG	TT	GG	GG	TG
-38276	G→C	GG	GG	GC	GG	CC	GG	GG
-39036	G→C	GG	N	GC	N	CC	N	N
-40608	G→C	GG	GG	GG	GG	GG	GC	CC
-41590	T→C	TT	TT	TC	TT	CC	CT	CC
-42081/82	ΔAG	AG	AG	AG/Δ	AG	ΔΔ	AG	AG
-42618	T→C	TT	TT	TC	TT	CC	TT	TT
-42893	G→A	GG	GG	GA	GG	AA	GG	GG

<sup>a</sup>Position number is determined from initiation translation codon (ATG) of *LCT* using the compiled genomic sequence of the BACs NH034L23, NH0218L22, NH0318L13 and RP11-329110. <sup>b</sup>Individuals sequenced from the Finnish families and shown by arrow and number in Fig. 1. <sup>c</sup>N, not determined.

Marker	Haplotype									Total
D2S3011	6	6	6	6	6	6	6	6	6	
D2S3012	4	4	4	4	4	4	4	4	4	
D2S3013	3	2	2	2	1	1	4	4	4	
D2S3015	5	5	5	5	5	5	5	5	5	
D2S3014	2	2	2	4	2	2	2	2	2	
D2S3016	5	5	5	5	5	5	5	5	5	
D2S3017	6	6	5	6	6	3	6	5	3	
Lactase persistence alleles	20	4	1	2	2	1	1	1	1	33
Lactase non-persistence alleles	3	2	0	0	0	0	2	0	2	54

**Fig. 3** The 150-kb haplotypes constructed with seven microsatellite markers in 33 Finnish lactase persistence alleles. The shared haplotype in lactase persistence chromosomes is shaded. Nine of fifty-four different non-persistence chromosomes show some sharing of this haplotype; these haplotypes are also shown. The remaining 45 non-persistence chromosomes carry a wide range of different haplotypes and are not shown.

tance from *LCT*, and we can only speculate on their significance in the regulation of LPH activity. The C/T<sub>-13910</sub> variant affects a binding site of transcription factor AP-2. The C allele, associated with lactase non-persistence, is in the consensus binding motif, whereas the T variant disrupts this motif. It is possible that the AP-2 element has a long-range *cis*-transcriptional effect on developmental stage-specific regulation of *LCT*. To establish the functional significance of this variant, further studies must control for transcriptional effects of all the other sequence variants (including the G/A<sub>-22018</sub>) in the relatively long 14-kb DNA segment 5' of *LCT*, and probably include the use of specialized cell lines, similar to the epithelial cells of the intestine.

We previously mapped<sup>13</sup> the locus for congenital lactase deficiency (CLD) to an interval of approximately 2 cM 5' of *LCT* that spans the variants we describe here. We found no association between these variants and CLD in 19 Finnish families (data not shown).

A shared haplotype, the extent of LD in Finnish alleles and the presence of the same DNA variants in non-persistence alleles in different, distantly related populations together suggest that the persistence variant is old, occurring long before the differentiation of these populations. This is consistent with the hypothesis that adult lactase persistence has become more prevalent since the introduction of dairy culture in around 10,000–8,000 BC<sup>14</sup>. A selection power of 3–5% would be sufficient to explain the present frequency of the lactase persistence allele in northern Europe<sup>15</sup>, assuming it arose around the advent of dairy culture.

The unreliability of the indirect diagnostic methods of adult-type hypolactasia, such as the lactose tolerance test<sup>16</sup>, and the

tedious measurement of LPH activity in jejunal biopsy sample specimens highlights the usefulness of a reliable DNA test. Large-scale epidemiological samples from different populations ascertained for biochemically verified hypolactasia should now be analyzed for the two DNA variants identified here. The outcome of these studies will determine the applicability of these variants as a diagnostic test for adult lactase non-persistence.

**Methods**

**DNA samples analyzed.** The nine extended Finnish pedigrees have been described<sup>17</sup>. We tested 145 of 194 family members for lactase persistence, using lactose-tolerance tests with ethanol (LTTE)<sup>18</sup>. Gluten enteropathy was excluded in all lactose non-persistence cases by measurement of the serum IgA anti-tissue transglutaminase<sup>18</sup>. After obtaining informed consent, we extracted DNA from blood samples of all participating family members in accordance with standard protocols<sup>19</sup>. The case-control study sample represented 196 individuals with biochemically verified disaccharidase activities from jejunal biopsy specimens measured<sup>20</sup> at the Helsinki University Hospital, all of them aged more than 20 y. In addition, DNA samples were available from nine Italian individuals (samples provided by M. Rossi, University of Naples), aged 35–65 y, nine German individuals (provided by M. Lentze, University of Bonn) aged 25–62 y and 22 South Korean individuals (provided by J.K. Seo, Seoul National University), aged 6–18 y, all of which had confirmed lactase non-persistence based on infestinal biopsy samples. One of the individuals from Germany originated from South Korea. Diagnosis was based on the measurement of disaccharidase activities in all but the South Korean samples, for which the diagnosis was confirmed by clinical symptoms. We determined the frequency of the C/T<sub>-13910</sub> variant in 938 anonymous Finnish blood donors from Eastern and Western Finland and in 109 parents of families from Utah and France whose samples are kept by the Centre d'Etude Polymorphisme Humaine. The study was approved by the Ethical Committees of the Helsinki University Hospital and the Finnish Red Cross Blood Transfusion Service.

**Table 3 • Distribution of genotypes in individuals with biochemically verified lactase persistence or non-persistence**

	Genotype	C/T <sub>-13910</sub>				G/A <sub>-22018</sub>		Total
		CC	CT	TT	GG	GA	AA	
Family members	lactase non-persistence	45	0	0	45	0	0	45
	lactase persistence	0	32	13	0	32	13	45
Case-control samples	lactase non-persistence	59	0	0	53	6	0	59
	lactase persistence	0	63	74	0	63	74	137
Non-Finnish*	lactase non-persistence	40	0	0	39	1	0	40
	lactase persistence	0	5	0	0	5	0	5
Total	lactase non-persistence							144
	lactase persistence							187

\*Non-Finnish samples consist of 23 South Korean, 9 Italian and 8 German individuals.

**Genotyping.** We analyzed microsatellite markers flanking *LCT* on 2q21 as described elsewhere<sup>14</sup>. Microsatellite markers within the contig constructed over *LCT* were identified from the published genomic sequence of the BACs (NH034L23, NH0318L13, NH0218L22 and RP11-329I1) using the Repeat Masker program. We synthesized primers flanking the repeats and used PCR conditions as described elsewhere<sup>14</sup>. The amplified fragments were separated on 6% polyacrylamide gel.

**Linkage and allelic association analyses.** We calculated pairwise lod scores using the MLINK option of the LINKAGE program package<sup>21,22</sup>. We assumed autosomal recessive inheritance for adult-type hypolactasia with complete penetrance, no sex difference in recombination fractions and a disease allele frequency of 0.4. Only individuals older than 20 y were included in the study, as the condition is manifested by that age in the Finnish population<sup>7,18</sup>. The affection status for individuals not confirmed by LTTE was regarded as unknown. We estimated allele frequencies and heterozygosity for the markers from family material, using the Downfreq program, to use in the parametric linkage analysis<sup>23</sup>. In addition, we carried out pseudomarker linkage and linkage disequilibrium analyses, assuming an autosomal recessive mode of inheritance, in which the frequencies of both disease and marker alleles are treated as nuisance parameters<sup>24</sup>. We carried out a test of LD conditional on the detected linkage, treating the allele frequencies and the recombination fraction as nuisance parameters<sup>23,24</sup>. *P* values from these analyses are shown in Table 1. We constructed haplotypes using the program GENEHUNTER<sup>25</sup>.

**Detection of the DNA variants.** We sequenced, for the critical 47-kb region, genomic DNA of three individuals with lactase non-persistence and four individuals with lactase persistence, representing phase-known chromosomal haplotypes in Finnish families. Using the published draft genomic sequence, we assembled to one contig the BACs (NH0034L23, NH0218L22, NH0318L23, RP-329I10) that covered the critical region of adult-type hypolactasia, using Sequencher 4 software (Gene Codes). We designed oligonucleotide primers spanning the critical region between them (a list of oligonucleotide primers are available on request). We carried out PCR amplification in a 50- $\mu$ l volume with genomic DNA (100 ng), primers (20 ng each), dNTPs (200  $\mu$ M), 0.5 U of *Taq* polymerase (Dynazyme, Finnzymes) in a standard buffer. Most PCR were amplified using the following PCR cycle conditions: an initial round of denaturation at 94 °C for 3 min, then 35 cycle at 94 °C at 30 s, 55 °C for 30 s, 72 °C for 1.25 min and a final extension of 72 °C for 10 min. In cases where the size of the PCR products was more than 1 kb, we used the Dynazyme extend kit. Purified PCR products (15–40 ng) were cycle-sequenced using BigDye terminator chemistry (PE Biosystems). We analyzed data using ABI Sequencing Analysis 3.3 (PE Biosystems) and Sequencher 4.1 (Gene Codes).

**Screening of the lactase persistence/non-persistence variants.** We amplified the DNA fragment spanning the C/T<sub>-13910</sub> variant using one biotinylated and one unbiotinylated primer. For G/A<sub>-22018</sub> we used one biotinylated and one unbiotinylated primer under the conditions described above. Primer sequences are available upon request. We captured 10  $\mu$ l of the PCR product in a streptavidin-coated microtitre well (Labsystem, Finland) and carried out the minisequencing as described<sup>26</sup>. The minisequencing primers are available upon request. Two parallel minisequencing reactions were carried out for each PCR product.

**URL.** Repeat Masker program, <http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>.

**GenBank accession numbers.** Markers used in the fine-mapping, AF395607–AF395615; BACs: NH0218L22, AC012551; N0034L34, AC011893; NH0318L13, AC011999; RP11-329I10, AC016516.

**Note:** Supplementary information is available on the Nature Genetics web site ([http://genetics.nature.com/supplementary\\_info/](http://genetics.nature.com/supplementary_info/)).

#### Acknowledgments

We are grateful to the families who participated in this study. We thank M. Levander for assistance with minisequencing. This work was supported by The Academy of Finland, The Paediatric Research Foundation (Ulla Hjelt Fond), The Sigrid Jusélius Foundation and Helsinki University Hospital Research Funding and The National Institutes of Health (grant to J.D.T.).

Received 6 September; accepted 18 December 2001.

- Boll, W., Wagner, P. & Mantei, N. Structure of the chromosomal gene and cDNAs coding for lactase-phlorizin hydrolase in human with adult-type hypolactasia or persistence of lactase. *Am. J. Hum. Genet.* **48**, 889–902 (1991).
- Mantei, N. *et al.* Complete primary structure of human and rabbit lactase-phlorizin hydrolase: implications for biosynthesis, membrane anchoring and evolution of the enzyme. *EMBO J.* **7**, 2705–2713 (1988).
- Wang, Y. *et al.* The lactase persistence/non-persistence polymorphism is controlled by a *cis*-acting element. *Hum. Mol. Genet.* **4**, 657–662 (1995).
- Harvey, C.B., Pratt, W.S., Islam, I., Whitehouse, D.B. & Swallow, D.M. DNA polymorphisms in the lactase gene: linkage disequilibrium across the 70 kb region. *Eur. J. Hum. Genet.* **3**, 27–41 (1995).
- Sahi, T., Isokoski, M., Jussila, J. & Launiala, K. Lactose malabsorption in Finnish children of school age. *Acta Paediatr. Scand.* **61**, 11–16 (1972).
- Wang, Y. *et al.* The genetically programmed down-regulation of lactase in children. *Gastroenterology* **114**, 1230–1236 (1998).
- Sahi, T. Genetics and epidemiology of adult-type hypolactasia. *Scand. J. Gastroenterol. Suppl.* **202**, 7–20 (1994).
- Harvey C.B. *et al.* Lactase haplotype frequencies in Caucasians: association with the lactase persistence/non persistence polymorphism. *Ann. Hum. Genet.* **62**, 215–223 (1998).
- Harvey, C.B. *et al.* Characterization of a human homologue of a yeast cell division cycle gene, *MCM6*, located adjacent to the 5' end of the lactase gene on chromosome 2q21. *FEBS Lett.* **398**, 135–140 (1996).
- Syvänen, A.-C. & Landegren, U. Detection of point mutations by solid-phase methods. *Hum. Mutat.* **3**, 172–179 (1994).
- Cuddebecq, Y., Delbrück, H. & Flatz, G. Distribution of the adult lactase phenotypes lactose absorber and malabsorber in a group of 131 Army recruits. *Gastroenterol. Clin. Biol.* **6**, 776–779 (1982).
- McLellan, T., Jorde, L.B. & Skolnick, M.H. Genetic distance between the Utah Mormons and related populations. *Am. J. Hum. Genet.* **36**, 836–857 (1984).
- Järvelä, I. *et al.* Assignment of the locus for congenital lactase deficiency to 2q21, in the vicinity of but separate from the lactase-phlorizin hydrolase gene. *Am. J. Hum. Genet.* **63**, 1078–1085 (1998).
- Clare, H. & Ruth, M. Phylogenetic analysis of the evolution of lactose digestion in adults. *Hum. Biol.* **69**, 605–628 (1997).
- McCracken, R.D. Lactase deficiency: an example of dietary evolution. *Curr. Anthropol.* **12**, 479–517 (1971).
- Arola, H. *et al.* Diagnosis of hypolactasia and lactose malabsorption. *Scand. J. Gastroenterol. Suppl.* **202**, 26–35 (1994).
- Sahi, T. The inheritance of selective adult-type lactose malabsorption. *Scand. J. Gastroenterol. Suppl.* **30**, 1–73 (1974).
- Sulkanen, S. *et al.* Tissue transglutaminase autoantibody enzyme-linked immunosorbent assay in detecting celiac disease. *Gastroenterology* **115**, 1322–1328 (1998).
- Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989).
- Messer, M. & Dahlqvist, A. A one-step ultramicro method for the assay of intestinal disaccharidases. *Anal. Biochem.* **14**, 376–392 (1966).
- Cottingham, R.W. Jr, Idury, R.M. & Schaffer, A.A. Faster sequential genetic linkage computations. *Am. J. Hum. Genet.* **53**, 252–263 (1993).
- Lathrop, G.M., Lalouel, J.-M., Julier, C. & Ott, J. Strategies for multilocus linkage analysis in humans. *Proc. Natl Acad. Sci. USA* **81**, 3443–3446 (1984).
- Göring, H.H.H. & Terwilliger, J.D. Linkage analysis in the presence of errors III: marker loci and their map as nuisance parameters. *Am. J. Hum. Genet.* **66**, 1298–1309 (2000).
- Göring, H.H.H. & Terwilliger, J.D. Linkage analysis in the presence of errors IV: joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when mode of inheritance cannot be accurately specified. *Am. J. Hum. Genet.* **66**, 1310–1327 (2000).
- Krugliak, L., Daly, M.J., Reeve-Daly, M.P. & Lander, E. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.* **58**, 1347–1363 (1996).
- Ihalainen, J., Siitari, H., Laine, S., Syvänen, A.-C. & Palotie, A. Towards automatic detection of point mutations: use of scintillating microplates in solid-phase minisequencing. *Biotechniques* **16**, 938–943 (1994).