

## The T Allele of a Single-Nucleotide Polymorphism 13.9 kb Upstream of the Lactase Gene (*LCT*) (*C – 13.9kbT*) Does Not Predict or Cause the Lactase-Persistence Phenotype in Africans

Charlotte A. Mulcare,<sup>1,2</sup> Michael E. Weale,<sup>1,\*</sup> Abigail L. Jones,<sup>1</sup> Bruce Connell,<sup>3</sup> David Zeitlyn,<sup>4</sup> Ayele Tarekegn,<sup>1</sup> Dallas M. Swallow,<sup>2</sup> Neil Bradman,<sup>1</sup> and Mark G. Thomas<sup>1</sup>

<sup>1</sup>The Centre for Genetic Anthropology (TCGA) and <sup>2</sup>Galton Laboratory, University College London, London; <sup>3</sup>Department of Languages, Literatures, and Linguistics, York University, Toronto; and <sup>4</sup>Centre for Social Anthropology and Computing, Department of Anthropology, University of Kent, Canterbury, United Kingdom

The ability to digest the milk sugar lactose as an adult (lactase persistence) is a variable genetic trait in human populations. The lactase-persistence phenotype is found at low frequencies in the majority of populations in sub-Saharan Africa that have been tested, but, in some populations, particularly pastoral groups, it is significantly more frequent. Recently, a CT polymorphism located 13.9 kb upstream of exon 1 of the lactase gene (*LCT*) was shown in a Finnish population to be closely associated with the lactase-persistence phenotype (Enattah et al. 2002). We typed this polymorphism in 1,671 individuals from 20 distinct cultural groups in seven African countries. It was possible to match seven of the groups tested with groups from the literature for whom phenotypic information is available. In five of these groups, the published frequencies of lactase persistence are  $\geq 25\%$ . We found the T allele to be so rare that it cannot explain the frequency of the lactase-persistence phenotype throughout Africa. By use of a statistical procedure to take phenotyping and sampling errors into account, the T-allele frequency was shown to be significantly different from that predicted in five of the African groups. Only the Fulbe and Hausa from Cameroon possessed the T allele at a level consistent with phenotypic observations (as well as an Irish sample used for comparison). We conclude that the *C – 13.9kbT* polymorphism is not a predictor of lactase persistence in sub-Saharan Africans. We also present Y-chromosome data that are consistent with previously reported evidence for a back-migration event into Cameroon, and we comment on the implications for the introgression of the *– 13.9kb\*T* allele.

### Introduction

Lactose, a disaccharide, is the principal calorific component of milk. To be absorbed, it must be hydrolyzed, a reaction mediated by the enzyme lactase (lactase-phlorizin hydrolase). In most mammals that have been studied, the level of the lactase enzyme is severely reduced some time after weaning, so adults cannot digest lactose effectively (reviewed in Swallow and Hollox 2000). However, in humans, the ability to digest lactose throughout adulthood (lactase persistence [MIM 223100]) exists as a dominant Mendelian polymorphic trait (Sahi 1974; Swallow and Harvey 1993).

Received December 2, 2003; accepted for publication March 10, 2004; electronically published April 20, 2004.

Address for correspondence and reprints: Dr. Dallas Swallow, Galton Laboratory, Department of Biology, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE. E-mail: dswallow@hgmp.mrc.ac.uk.

\* This author devised the statistical analysis method used in this study; the program is available on the TCGA Web site (see the “Electronic-Database Information” section).

© 2004 by The American Society of Human Genetics. All rights reserved. 0002-9297/2004/7406-0004\$15.00

Lactase persistence varies widely in frequency among different human populations, both between and within continents. It is generally found at high frequencies in populations of European descent, in which, for example, Dutch and Swedish studies recorded frequencies of 100% and 99%, respectively (reviewed in Swallow and Hollox 2000). Lactase-nonpersistent individuals (lactose nondigesters) may suffer adverse symptoms from milk ingestion resulting from the breakdown of lactose by bacteria in the gut, varying from mild flatulence to severe abdominal pains and diarrhea.

Although the structure and full exonic sequence of the lactase gene (*LCT* [MIM 603202]) has been known since 1991 (Boll et al. 1991), the causative mechanism for lactase persistence has proved more elusive. Recently, Enattah and colleagues (2002) showed that the T allele of a C/T transition 13.9 kb upstream from exon 1 of *LCT* in intron 13 of the gene *MCM6*, here referred to as *–13.9kb\*T*, was completely associated with lactase persistence in a sample of 196 unrelated Finnish individuals, whose diagnoses were made from intestinal biopsy specimens (lactase persistence:  $n = 137$ ; lactase nonpersistence:  $n = 59$ ). Furthermore, 40 lactase-non-

persistent individuals from various populations (Germany, Italy, South Korea) were all homozygous for the C allele. Finally, DNA from individuals of unknown phenotype collected in Finland ( $n = 938$ ), France ( $n = 17$ ), and the United States (two data sets: African descent [ $n = 96$ ] and European descent [ $n = 92$ ]) had frequencies of the CC genotype that appeared to be consistent with the frequencies of lactase nonpersistence reported for those groups. Although no functional mechanism was shown at that time, presence of  $-13.9kb^*T$  was proposed as a robust marker for lactase persistence. Very recent studies have suggested that this SNP is located in an enhancer element and that the two alleles show some difference in function (Olds and Sibley 2003; Troelsen et al. 2003).  $C-13.9kbT$  typing is now being offered as a genetic test for lactase persistence in Finland (Medix Laboratory), where the strong correlation between the T allele and lactase persistence was first reported, and such testing is being considered for use elsewhere (Buning et al. 2003; Hoegenauer et al. 2003).

To date, there have been no reports of allele frequencies for the  $C-13.9kbT$  polymorphism in populations living in Africa. Although the  $-13.9kb^*T$  allele frequency in Americans with African ancestry is consistent with their lactase-persistence frequency (Enattah et al. 2002), there is known to be substantial admixture between African Americans and European Americans (Parra et al. 1998). Previous studies of African populations showed variation in the frequency of lactase persistence among population groups, as well as a complex pattern of distribution (reviewed in Flatz 1987; Holden and Mace 1997; Swallow and Hollox 2000). Pastoralists, such as the Fulbe in Nigeria, typically have higher frequencies of lactase persistence than nonpastoralists in the same country—for example, the Yoruba and Igbo (Kretchmer et al. 1971; Olatunbosun and Adadevoh 1971; Ransome-Kuti et al. 1972, 1975; Flatz 1987; Holden and Mace 1997). The lactase-persistence phenotype is usually observed at low frequencies in Bantu- and Khoisan-speaking groups (<20%) (Cook and Kajubi 1966; Cook et al. 1967, 1973; Cox and Elliott 1974; Nurse and Jenkins 1974; O'Keefe and Adam 1983; Segal et al. 1983; O'Keefe et al. 1984).

We have typed  $C-13.9kbT$  in 1,671 individuals from 20 different African populations, including both milk-drinking and non-milk-drinking groups. Since phenotype data was not available for these samples, we performed an ethnologically matched group study to determine whether  $-13.9kb^*T$  was associated with lactase persistence in seven African samples and in one northern European sample (to provide a comparative group), using a statistical procedure to take both sampling and phenotyping error into account. We show that, in most cases

in Africa, the frequency of  $-13.9kb^*T$  is too low to explain the observed frequency of lactase persistence.

## Material and Methods

### Samples

DNA was extracted from buccal swabs collected from males belonging to different groups living in various regions of Africa, including populations that have a history of pastoralism and milk drinking and others that do not (table 1). DNA was also obtained from a sample of unrelated Irish individuals for comparative purposes. Informed consent was obtained from all donors. Ethical approval was obtained from University College Hospitals and University College London Joint Committee on the Ethics of Human Research (reference number 99/0196). Appropriate permissions were obtained in each of the collection countries (reference number for Cameroon 0093MINREST/B00/D00/D10/D12). Each donor provided biographical details, such as self-defined ethnic identity, first and second language, and place of birth, with similar information on his mother, father, maternal grandmother, and paternal grandfather. Individuals were classified as belonging to a given cultural group if their own self-declared identity concurred with that they ascribed to both mother and father. Where there were <10 individuals of the same declared cultural identity, they were classified as "other." Individuals with partially unknown or mixed ancestry at the parental level were also classified as "other." The Ethnographic Atlas (Murdock 1967) and a summary table of pastoralists (Blench 1999) were used to obtain information about pastoralism and milk practice, and the Ethnologue Web site was used to check possible linguistic relationships between groups.

### Selection of Matched Populations for Which Phenotypic Data Were Available

For many of the population groups, a matching population sample with lactose-tolerance (digestion) data could be found in the literature. Matching samples were selected to fulfill the following criteria: (1) same declared cultural identity and (2) residency in the same country or in a neighboring country. Sources for data on lactase persistence are listed in table 2.

### Typing the $C-13.9kbT$ Polymorphism

PCR primers LAC-C-M-U (5'-GCTGGCAATACAG-ATAAGATAATGGA-3') and LAC-C-L2 (5'-CTGCTT-TGGTTGAAGCGAAGAT-3') were designed to amplify the region containing the C/T polymorphism (Enattah et al. 2002). The penultimate base of the LAC-C-M-U primer (G) introduces a base change such that the PCR product will be cut by *HinfI* when the T allele is present, giving digestion product sizes of 177 bp and 24 bp, but

not when the C allele is present, giving a digestion product size of 201 bp. PCR reactions were performed in a total volume of 10  $\mu$ l containing 200  $\mu$ M dNTPs, 10 mM Tris-HCl (pH 9.0), 0.1% Triton X-100, 0.01% gelatin, 50 mM KCl, 2.0 mM MgCl<sub>2</sub>, 0.13 U *Taq* polymerase enzyme (HT Biotech), 9.3 nM *TaqStart* monoclonal antibody (BD Biosciences, Clontech), and 0.3  $\mu$ M primers. The *Taq* and *TaqStart* monoclonal antibodies were premixed prior to being added to the other reagents. Thermal cycling conditions were an initial denaturation stage at 95°C for 5 min, then 35 cycles of 95°C for 1 min, 59°C for 1 min, and 72°C for 1 min, followed by a final elongation stage at 72°C for 5 min. Digestions were performed at 37°C overnight in the original PCR plate in a total volume of 25  $\mu$ l. Each reaction contained the entire PCR product, 0.25 U of *Hinf*I, 0.01  $\mu$ g/ $\mu$ l acetylated BSA, and New England Biolabs Buffer 2, as recommended by the manufacturer. The digestion products were run on a 3% agarose gel, and DNA bands were visualized by use of ethidium bromide staining. Gel phenotypes showing a single band of 201 bp (C) or 177 bp (T) were interpreted as genotypes  $-13.9kb*CC$  and  $-13.9kb*TT$ , respectively.

#### Genotype Error Checking

One positive control (a known CT heterozygote) and a blank were included in every 96-well plate. In addition, a set of 50 randomly selected samples were retyped “blind.” All controls and the 50 retyped samples matched the initial typing. As a further control that the results for the SNP matched those reported by Enattah et al. (2002), our SNP protocol was used for typing phenotyped Finnish individuals. There was a high, although incomplete, correlation between the T allele and lactose-tolerance test results (Poulter et al. 2003). The level of discrepancy was attributable to inaccuracies of the tolerance testing and was consistent with the error rates we use in our statistical procedure described below.

#### Statistical Analysis

Our genotype-error-checking procedure (see above) suggested that the genotyping error rate was zero or negligible. However, phenotyping error in the determination of lactose digestion as an indirect test for lactase-persistence status (e.g., by measurement of breath hydrogen or blood glucose) is known to occur at appreciable levels. It is the convention in the medical literature to describe lactose digesters as “negative” and nondigesters as “positive”—that is, giving a positive diagnosis in a lactose-tolerance test. We obtained information on the error rates of false-negative (FN) results (i.e., when a nonpersistent person appears to be a digester) and false-positive (FP) results (i.e., when a persistent person appears to be a nondigester) from three

studies in which the correct lactase phenotype was ascertained by peroral jejunal biopsy and from two studies in which the “correct” phenotype was determined by the “gold standard” method. When the “gold standard” method is used, at least two of three separate noninvasive tests (namely, blood glucose, breath hydrogen, and urine galactose) must concur. Error rates for the three studies ascertained by biopsy were as follows (sample sizes in the denominator): (1) blood glucose FN = 6/25 and FP = 1/25, breath hydrogen FN = 0/25 and FP = 0/25 (Newcomer et al. 1975); (2) blood glucose FN = 0/7 and FP = 0/8, breath hydrogen FN = 0/7 and FP = 0/8 (Howell et al. 1981); and (3) blood glucose not measured, breath hydrogen FN = 5/16 and FP = 2/47 (Arola et al. 1988). Error rates for the two studies ascertained by the “gold standard” method were as follows (sample sizes in the denominator): (1) blood glucose FN = 3/35 and FP = 3/35, breath hydrogen FN = 2/35 and FP = 2/35 (Puehkuri 2000); and (2) blood glucose FN = 1/49 and FP = 1/5, breath hydrogen FN = 2/49 and FP = 1/5 (Kurt et al. 2003). Exact protocols for the blood glucose and breath hydrogen tests varied among the five studies, as they did among the studies on matching populations reported later. However, all protocols involved the measurement of changes in plasma glucose or exhaled hydrogen at one or more time intervals between 30 min and 4 h after administration of at least 50 g lactose. We combined the results from the above five studies to perform a rough averaging over the differences in protocols used (blood glucose FN = 10/116 and FP = 5/73, breath hydrogen FN = 9/132 and FP = 5/120). The combined results suggest that, even if a population has no lactase-persistent individuals, we would expect, by use of one of these two methods, to find between 5% and 10% FNs (i.e., apparent lactose digesters).

Given the above and assuming that the underlying phenotyping error rates acting in these independent studies are applicable to other studies that use the same measurement techniques, we devised a statistical procedure that allowed us to test whether the frequency of lactose digesters predicted by the  $C-13.9kbT$  genotype data was sufficient to explain the observed frequency found in the phenotyped group. We took both phenotyping error and four possible sources of sampling uncertainty into account: (1) sampling uncertainty in  $p$ , the frequency of the  $-13.9kb*T$  in the genotyped group; (2) sampling uncertainty in  $f_n$ , the frequency of false negatives according to the phenotyping method used; (3) sampling uncertainty in  $f_p$ , the frequency of false positives according to the phenotyping method used; and (4) sampling uncertainty in  $L_{app}$ , the frequency of apparent lactase persistence in the phenotyped group.

The procedure was performed as follows:

1. A value for  $p$  was drawn from a Beta( $T+1$ ,  $C+1$ )

**Table 1**

**Published Information on Pastoralism and the Observed Frequency of  $-13.9kb^*T$ , by Population Group**

COUNTRY AND GROUP	NOMADIC PASTORALIST <sup>a</sup> STATUS	% DEPENDENT ON ANIMAL HUSBANDRY (MILKING STATUS) <sup>b</sup>	NO. OF INDIVIDUALS	NO. OF INDIVIDUALS WITH GENOTYPE			OBSERVED FREQUENCY OF $-13.9kb^*T$
				CC	CT	TT	
Cameroon:							
Fulbe <sup>c</sup>	Yes	46–55% (Yes)	49	39	9	1	.112
Hausa <sup>d</sup>	No	16–35% (No)	18	14	3	1	.139
Kwanja <sup>c</sup>	No	NA	70	70	0	0	0
Mambila <sup>c</sup>	No	16–25% (No)	122	121	1	0	.004
Nso <sup>e</sup> (Nsaw)	No	6–15% (No)	126	126	0	0	0
Yamba <sup>c</sup>	No	NA	21	21	0	0	0
Other	NA	NA	128	118	9	1	.043
Nigeria:							
Ibibio	No	6–15% (No)	110	110	0	0	0
Oron	No	6–15% (No) <sup>i</sup>	44	44	0	0	0
Other	No	NA	22	22	0	0	0
Malawi:							
Chewa <sup>c</sup>	No	6–15% (Yes)	84	84	0	0	0
Ngoni <sup>c</sup>	No	6–15% (Yes)	14	14	0	0	0
Tumbuka <sup>c</sup>	No	6–15% (No)	58	58	0	0	0
Yao <sup>c</sup>	No	6–15%	49	49	0	0	0
Other <sup>c</sup>	No	NA	58	58	0	0	0
Senegal:							
Wolof <sup>f</sup>	No	26–35% (Yes)	69	69	0	0	0
Manjak	No	NA	93	93	0	0	0
Other	NA	NA	19	18	1	0	.026
Sudan (North):							
Ga'ali	No	NA	30	30	0	0	0
Shaigi	No	NA	11	11	0	0	0
Other <sup>g</sup>	Mixed	NA	88	88	0	0	0
Sudan (South):							
Dinka	Yes	46–55% (Yes)	34	34	0	0	0
Nuer	Yes	46–55% (Yes)	13	13	0	0	0
Other <sup>g</sup>	Mixed	NA	73	73	0	0	0
Ethiopia:							
Nuer	Yes	46–55% (Yes)	119	119	0	0	0
Anuak (Anywak)	Yes <sup>h</sup>	6–15% (Yes)	108	108	0	0	0
Other	NA	NA	1	1	0	0	0
Uganda:							
Mussesse <sup>c</sup>	No	NA	22	22	0	0	0
Other <sup>c</sup>	No	NA	18	18	0	0	0
Ireland	NA	36–45% (Yes)	47	1	10	36	.872

NOTE.—NA = not available.  
<sup>a</sup> Pastoralists who migrate with their animals. From table 2.1 in Blench (1999).  
<sup>b</sup> Murdock (1967).  
<sup>c</sup> Fulbe with a sedentary lifestyle.  
<sup>d</sup> Immigrant population from Nigeria.  
<sup>e</sup> Bantoid-language speakers.  
<sup>f</sup> Includes 23 “Lebou” individuals. Lebou is a dialect of Wolof.  
<sup>g</sup> This group includes 12 individuals from traditional milk-drinking peoples of known high frequency for lactase persistence (Beja, Misseri, Gomoeia, and Shilluk [Bayoumi et al. 1981, 1982]).  
<sup>h</sup> Probably do not drink fresh milk (A. Tarekegn, unpublished data).  
<sup>i</sup> Identical to Ibibio in this respect.

distribution, where  $T$  is the number of T alleles and  $C$  is the number of C alleles found in the genotyped group. This beta distribution describes the posterior distribution for  $p$ , given the genotype data, assuming a Uniform(0,1) prior.

2. The predicted frequency of true lactase persistence in the population,  $L_{true}$ , was calculated as  $p^2 + 2p(1-p)$  (i.e., the expected frequency of TT + CT genotypes under Hardy-Weinberg equilibrium).
3. Values for  $f_n$  and  $f_p$  were drawn from Beta(11,107)

**Table 2****Comparisons with Published Lactose-Digester Frequencies in Matching Populations, Taking into Account Sampling and Phenotyping Error**

Group	Country of Genotyped Sample (No.)	Expected Frequency of Lactose Digesters	Country of Phenotyped Sample (No.)	Test Method	Observed Frequency of Lactose Digesters in Phenotyped Sample	Reference	P Value
Fulbe <sup>a</sup>	Cameroon ( <i>n</i> = 49)	.265	Nigeria ( <i>n</i> = 24)	Blood glucose	.292	Kretchmer et al. 1971	1
Hausa	Cameroon ( <i>n</i> = 18)	.305	Nigeria ( <i>n</i> = 17)	Blood glucose	.235	Kretchmer et al. 1971	.749
Wolof	Senegal ( <i>n</i> = 69)	.086	Senegal ( <i>n</i> = 53)	Blood glucose	.509	Arnold et al. 1980	0
Ga'ali (Jaali)	Sudan (North) ( <i>n</i> = 30)	.068	Sudan ( <i>n</i> = 113)	Breath hydrogen	.531	Bayoumi et al. 1981	0
Shaigi	Sudan (North) ( <i>n</i> = 11)	.068	Sudan ( <i>n</i> = 42)	Breath hydrogen	.381	Bayoumi et al. 1981	.025
Nuer	Sudan (South), Ethiopia ( <i>n</i> = 132)	.068	Sudan ( <i>n</i> = 23)	Breath hydrogen	.217	Bayoumi et al. 1982	.030
Dinka	Sudan (South) ( <i>n</i> = 34)	.068	Sudan ( <i>n</i> = 208)	Breath hydrogen	.255	Bayoumi et al. 1982	.001
European	Ireland ( <i>n</i> = 47)	.918	Ireland ( <i>n</i> = 50)	Blood glucose	.900	Fielding et al. 1981 <sup>b</sup>	1

NOTE.—Expected frequency of lactose digesters, taking into account the test error rate by the method used in the matched population =  $L_{\text{true}}(1-f_p) + (1-L_{\text{true}})f_n$ , where  $L_{\text{true}}$  = frequency of CT + TT genotypes assuming Hardy-Weinberg equilibrium,  $(f_n, f_p) = (10/116, 5/73)$  if the blood glucose test method is used and  $(f_n, f_p) = (9/132, 5/120)$  if the breath hydrogen test method is used. *P* value = result of test described in the “Statistical Analysis” section.

<sup>a</sup> Fulbe with a sedentary lifestyle.

<sup>b</sup> Blood glucose results only taken from this source, by use of a rise of >20 mg/dl to define lactose digester.

and Beta(6,69) distributions, respectively, if phenotyping was by the blood glucose method and from Beta(10,124) and Beta(6,116) distributions, respectively, if phenotyping was by the breath hydrogen method. Again, these beta distributions describe the posterior distribution for  $f_n$  and  $f_p$ , given the combined false error rate data reported above and assuming a Uniform(0,1) prior.

4. The predicted frequency of apparent lactose digesters accounting for phenotyping error,  $L_{app}$ , was calculated as  $L_{true}(1-f_p) + (1-L_{true})f_n$ .
5. A simulated value for  $n_L$ , the number of lactose digesters observed in the phenotyped group was drawn from a Binomial( $n, L_{app}$ ) distribution, where  $n$  is the number sampled in the phenotyped group.
6. Steps 1–5 were repeated 100,000 times ( $N = 100,000$ ) to build up a Monte Carlo sampling distribution for  $n_L$  under the null hypothesis that the C/T genotype and phenotyping error alone account for the apparent frequency of lactose digesters.
7. Let  $S_g$  be the sum of simulated  $n_L$  values greater than or equal to the observed  $n_L$  value, and let  $S_l$  be the sum of simulated  $n_L$  values less than or equal to the observed  $n_L$  value. A two-tailed  $P$  value for the observed  $n_L$  under the null hypothesis was found as  $2 \times \min(S_g, S_l)/N$ .

### Y-Chromosome Haplotypes

Cruciani and colleagues (2002) reported the presence of a non-African Y-chromosome lineage (M173-derived haplotype 117, or R1\*, by use of nomenclature of the Y-Chromosome Consortium [2002]) in population groups from northern Cameroon. These authors suggested this was due to a back-migration event from outside sub-Saharan Africa. We typed our samples for a genealogically similar marker, 92R7 (Mathias et al. 1994), which is ancestral to M173 but for which intermediate haplotypes (92R7-derived, M173-ancestral) have not been reported in any study of sub-Saharan African populations to date. We also typed 92R7-derived samples for six Y-chromosome microsatellites (DYS19, DYS388, DYS390, DYS391, DYS392, and DYS393) to investigate the intrahaplogroup diversity. Protocols for the typing of 92R7 and the microsatellites are those described by Thomas et al. (1999). Microsatellite repeat numbers were assigned according to the nomenclature of Kayser and colleagues (1997). Genetic diversity,  $h$ , and its SE were calculated according to the unbiased formulae in the work of Nei (1987). Since Y-chromosome typing was not successful in 72 of the 1,671 samples, we also report separately the different Y-chromosome sample sizes.

## Results

### C-13.9kbT in Africa

The frequency of  $-13.9kb*T$  was low or zero in most of the African groups tested, whereas it was very high in the Irish sample (table 1). In the African populations, the  $-13.9kb*T$  allele was only found in a few individuals; all but one of these individuals were from Cameroon and lived close to the same market town, Mayo Darle. Of these individuals, there were 10 Fulbe, 4 Hausa, 1 Mambila, and 10 “others” (mixed ancestry or from other ethnic groups). In all but two cases (one Hausa and one “other” individual), the  $-13.9kb*T$ -carrying individuals or one or both of their parents spoke Fulfulde, a Fulbe language. This association between possession of the  $-13.9kb*T$  allele and speaking Fulfulde was significant both in the Cameroonian sample as a whole ( $P < .001$ ,  $n = 534$ , Fisher’s exact test) and in the non-Fulbe Cameroonians ( $P = .015$ ,  $n = 485$ , Fisher’s exact test). The one individual from Senegal carrying  $-13.9kb*T$  allele was of mixed Wolof and Toucouleur (Tukulor) ancestry. There were no significant departures from Hardy-Weinberg equilibrium in any of the ascribed ethnic groups where T alleles were observed (by use of the method of Guo and Thompson [1992]).

It is noteworthy that  $-13.9kb*T$  was not found in East Africa at all, even though the data sets included many known pastoralists and groups with a high frequency of lactase persistence (table 1).

### Matched Populations for Which Phenotypic Data Were Available

In some cases, it was possible to find closely matching populations in the literature that had phenotype information (table 2). Comparisons of the predicted frequencies of lactase persistence, deduced from the frequency of  $-13.9kb*TT$  and  $-13.9kb*CT$  genotypes, with the reported frequencies obtained from lactose-tolerance testing, showed these were significantly different in all of the African populations except the Fulbe and the Hausa. Only in these two Cameroonian groups was  $-13.9kb*T$  found at frequencies sufficient to explain the raised incidence of lactase persistence. In contrast to the generally poor correspondence between genotypic and phenotypic data in African populations, our Irish sample shows excellent correspondence between predicted and observed frequencies of lactase persistence in the genotyped and phenotyped groups, consistent with the findings of Enattah et al. (2002).

### Y-Chromosome Data

The 92R7-derived haplogroup was extremely rare in the sub-Saharan African populations sampled. Most

92R7-derived chromosomes were found in Cameroon, with 8/42 in the Fulbe, 1/110 in Mambila, 1/65 Kwanja, and 5/113 “others.” Outside Cameroon, we found five 92R7-derived chromosomes in northern Sudan (3/11 Shaigi, 2/29 Ga’ali) and one in southern Sudan (1/72 “others”). The microsatellite haplotype diversity of 92R7-derived chromosomes in Cameroon was high, with 10 haplotypes observed among 15 individuals ( $b = 0.933$ ,  $SE = 0.0449$ , average repeat size variance = 0.224).

## Discussion

The absence of  $-13.9kb^*T$  in most of the African populations typed, which included several milk-drinking groups (table 1), suggests that it is not a reliable predictor of the lactase-persistence phenotype in populations from this region. This, in turn, indicates either that it is not a causative mutation or that it is not the sole causative mutation in all human populations.

This conclusion is consistent with a previous study (Poulter et al. 2003). In a series of 48 London patients of various ancestry, from whom intestinal biopsies were obtained, the correlations of lactase activity and sucrase/lactase ratio with  $-13.9kb^*CT$  and  $-13.9kb^*TT$  genotype were not as tight as might have been expected for a *cis*-acting causal change. In contrast to this, in a recent Finnish study, the  $13.9kb^*CT$  heterozygotes did have activity intermediate between the  $13.9kb^*CC$  and  $13.9kb^*TT$  homozygotes (Kuokkanen et al. 2003).

Previous studies have shown that, outside Africa, there are very few common *LCT* gene haplotypes (A, B, C, and U) (Hollox et al. 2001), and recent studies in Europeans have shown that linkage disequilibrium extends over at least 1 Mb (Poulter et al. 2003), as demonstrated clearly by Bersaglieri and colleagues in this issue of the *Journal* (Bersaglieri et al. 2004 [in this issue]). The  $-13.9kb^*T$  allele is carried on the background of the extended A haplotype that is most common in northern Europeans and may have reached high frequencies as a result of selection (Poulter et al. 2003). However, many A haplotype chromosomes do not carry T at  $-13.9$  kb. A comparison of the occurrence of this allele, as well as alleles at other recently described loci that subdivide the A haplotype (such as  $G-22kbA$ , [Enattah et al. 2002]), suggests that  $-13.9kb^*T$  is the most recent (Poulter et al. 2003). It is possible that the  $C-13.9kbT$  transition occurred more recently than another (as yet unknown) mutation that is the true causal change both in Africa and Europe. Recent transfection studies do, however, suggest a functional role for  $C-13.9kbT$  (Olds and Sibley 2003; Troelsen et al. 2003).

In a few rare individuals, high expression of the mRNA transcript, encoded by the *LCT* allele of a non-

A haplotype chromosome, has been observed (Poulter et al. 2003). In particular, a single individual in a United Kingdom cohort was interpreted as being heterozygous for the A and B haplotypes, as well as for  $C-13.9kbT$ , and showed high expression of lactase from both transcripts, suggesting that there may be heterogeneity of the cause of lactase persistence in Europe (Poulter et al. 2003).

It seems probable that the C-to-T transition at  $-13.9$  kb occurred in a non-sub-Saharan African population that contributed to the current population of Europe. If this were the case, then its presence in Cameroon, and especially in people of Fulbe cultural identity or with Fulfulde-speaking ancestry, could be explained by introgression from outside sub-Saharan Africa.

Our Y-chromosome data corroborate the results of Cruciani and colleagues (2002) in finding high frequencies in our Cameroonian samples of a haplogroup that is generally absent from sub-Saharan Africa. Phylogeographic arguments suggest that this haplogroup (R1\*, by use of the nomenclature of the Y-Chromosome Consortium [2002]) has a non-African origin. Cruciani and colleagues (2002) found R1\* Y chromosomes at an average frequency of 40% in several northern Cameroonian groups, including one Fulbe group. We found evidence for the same haplotype (typed by use of a marker that appears phylogenetically identical in this part of Africa) in our samples from central Cameroon, with a particularly high frequency (19%) in the Fulbe group that was tested. The Y-chromosome microsatellite diversity we observed indicates that this haplogroup could not have been brought to this part of Africa by a single recent founder. The origins of the Fulbe are the subject of debate, but the group is thought to be from outside Cameroon; on the basis of ethnic traditions and linguistic similarities between Fulbe languages and Tukulor (Toucouleur), an origin in the Futa Toro region of the Senegal river basin has been proposed (Newman 1995). It is possible that the back-migration event that led to the introduction of R1\* into sub-Saharan Africa (Cruciani et al. 2002) also brought the  $-13.9kb^*T$  allele and that the Fulbe of central Cameroon migrated locally from the north. However, haplogroup R1\* is also found at high frequencies in several non-Fulbe groups in the Extreme North Province of Cameroon, where the  $-13.9kb^*T$  allele is found at low frequencies (<3%, data not shown). Thus, the demographic processes leading to the presence of the  $-13.9kb^*T$  allele in Cameroon may be not be the same as those leading to the Y-chromosome introgression but could instead relate more specifically to Fulbe migration history. Further studies on the distribution of the  $-13.9kb^*T$  allele and of other genetic markers in this part of Africa are required to resolve this question.

We have shown that  $-13.9kb^*T$  is not associated

with lactase persistence in a wide range of African populations. It would now be appropriate to undertake more extensive genotyping and phenotype characterization on the same individuals in multiple African groups. It will be of interest to determine whether lactase persistence is associated with the same or a different haplotype and whether such haplotypes have an extended length consistent with a recent selective sweep, as was found for Europeans (Bersaglieri et al. 2004 [in this issue]). African populations display multiple lifestyles, with milk-drinking and non-milk-drinking groups often living in close proximity, and have complex demographic histories. Understanding the genetic determinants of lactase persistence in African populations will help explain the genetic history of the lactase-persistence phenomenon and should ultimately have positive implications for public health.

### Conclusion

The T allele located 13.9 kb upstream of *LCT* has been claimed by Enattah and colleagues (2002) as a predictor of lactase persistence in European populations. It does not fulfill that function in sub-Saharan Africans. Use of the *C-13.9kbT* polymorphism as a diagnostic predictor of adult hypolactasia outside Europe should therefore be approached with caution. Our results show that the *-13.9kb\*T* allele cannot be causal of lactase persistence in most Africans, although it could possibly explain lactase persistence in some Cameroonians. Data presented in this study support the possibility that the presence of the *-13.9kb\*T* allele in Cameroon is due to introgression from outside sub-Saharan Africa. The combined results from *C-13.9kbT* and the Y-chromosome analysis suggest a complex demographic history for this part of Africa, which includes at least one major introduction of genes from outside the region.

### Acknowledgments

We thank Dominic Gormis, Esther William, Tanelli Helenius, Jim Wilson, Pieta Nasanen, John Greenhalgh, Jane Moore, Richard Phillips, Katya Bulgina, Alex Murray, Ali Barwhani, Corine Atton, and Noreen von Cramon-Taubadel, who collected and extracted many of the DNA samples used in the present study and/or tested for Y-chromosome markers. We also thank Dr. Roger Blench and Dr. Clare Holden for helpful discussions. C.A.M. was funded by a BBSRC CASE studentship.

### Electronic-Database Information

The URLs for data presented herein are as follows:

Ethnologue, Languages of the World, <http://www.ethnologue.com/>

Medix Laboratory, [http://www.medix.fi/tiedotteet/tuoteinfo\\_02/03.htm](http://www.medix.fi/tiedotteet/tuoteinfo_02/03.htm) (for SNP testing for lactase-persistence diagnosis)  
 Online Mendelian Inheritance in Man (OMIM): <http://www.ncbi.nlm.nih.gov/Omim/> (for lactase persistence and *LCT*)  
 The Centre for Genetic Anthropology (TCGA) Software Page, <http://www.ucl.ac.uk/tcga/software/> (for the statistical analysis program used for the ethnologically matched group study)

### References

- Arnold J, Diop M, Kodjovi M, Rozier J (1980) L'intolérance au lactose chez l'adulte au Sénégal (Lactose intolerance in adults in Senegal). *C R Seances Soc Biol Fil* 174:983–992 (In French)
- Arola H, Koivula T, Jokela H, Jauhiainen M, Keyrilainen O, Ahola T, Uusitalo A, Isokoski M (1988) Comparison of indirect diagnostic methods for hypolactasia. *Scand J Gastroenterol* 23:351–357
- Bayoumi RAL, Flatz SD, Kuhau W, Flatz G (1982) Beja and Nilotes: nomadic pastoralist groups with opposite distributions of the adult lactase phenotypes. *Am J Phys Anthropol* 58:173–178
- Bayoumi RAL, Saha N, Salih AS, Bakkar AE, Flatz G (1981) Distribution of the lactase phenotypes in the population of the Democratic Republic of the Sudan. *Hum Genet* 57:279–281
- Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN (2004) Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* 74:1111–1120 (in this issue)
- Blench R (1999) Why are there so many pastoral groups in eastern Africa? In: Azarya V, Breedveld A, De Bruijn M, Van Dijk H (eds) *Pastoralists under pressure? Fulbe societies confronting change in west Africa*. Brill Press, Boston
- Boll W, Wagner P, Mantei N (1991) Structure of the chromosomal gene and cDNAs coding for lactase-phlorizin hydrolase in humans with adult-type hypolactasia or persistence of lactase. *Am J Hum Genet* 48:889–902
- Buning C, Jurga J, Fiedler T, Kupferling S, Worm M, Weltrich R, Genschel J, Lochs H, Schmidt H, Ockenga J (2003) Genetic background of lactose intolerance and implications for diagnosis. *Gastroenterology Suppl* 124:A144
- Cook G, Asp N, Dahlqvist A (1973) Lactose absorption kinetics in Zambian African subjects. *Br J Nutr* 30:519–527
- Cook G, Kajubi S (1966) Tribal incidence of lactase deficiency in Uganda. *Lancet* 1:725–729
- Cook G, Lakin A, Whitehead R (1967) Absorption of lactose and its digestion products in the normal and malnourished Ugandan. *Gut* 8:622–627
- Cox J, Elliott F (1974) Primary adult lactose intolerance in the Kivu lake area: Rwanda and the Bushi. *Am J Dig Dis* 19:714–724
- Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R, Underhill PA (2002) A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet* 70:1197–1214

- Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I (2002) Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 30:233–237
- Fielding J, Harrington M, Fottrell P (1981) The incidence of primary hypolactasia amongst the Irish. *Ir J Med Sci* 150:276–277
- Flatz G (1987) Genetics of lactose digestion in humans. *Adv Hum Genet* 16:1–77
- Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361–372
- Hoegenauer C, Hammer HF, Mellitzer K, Renner W, Toplak H (2003) Evaluation of a new genetic test compared to the lactose hydrogen breath test for the diagnosis of acquired primary lactase deficiency. *Gastroenterology Suppl* 124:A64
- Holden C, Mace R (1997) Phylogenetic analysis of the evolution of lactase digestion in adults. *Hum Biol* 69:605–628
- Hollox EJ, Poulter M, Zvarik M, Ferak V, Krause A, Jenkins T, Saha N, Kozlov AI, Swallow DM (2001) Lactase haplotype diversity in the Old World. *Am J Hum Genet* 68:160–172
- Howell JN, Schockenhoff T, Flatz G (1981) Population screening for the human adult lactase phenotypes with a multiple breath version of the breath hydrogen test. *Hum Genet* 57:276–278
- Kayser M, de Knijff P, Dieltjes P, Krawczak M, Nagy M, Zerjal T, Pandya A, Tyler-Smith C, Roewer L (1997) Applications of microsatellite-based Y chromosome haplotyping. *Electrophoresis* 18:1602–1607
- Kretchmer N, Ransome-Kuti O, Hurwitz R, Dungy C, Alakija W (1971) Intestinal absorption of lactose in Nigerian ethnic groups. *Lancet* 2:392–395
- Kuokkanen M, Enattah NS, Oksanen A, Savilahti E, Orpana A, Jarvela I (2003) Transcriptional regulation of the lactase-phlorizin hydrolase gene by polymorphisms associated with adult-type hypolactasia. *Gut* 52:647–652
- Kurt I, Abou Ghoush M, Hasimi A, Serdar M, Kutluay T (2003) Comparison of indirect methods of lactose absorption. *Turk J Med Sci* 33:103–110
- Mathias N, Bayes M, Tyler-Smith C (1994) Highly informative compound haplotypes for the human Y chromosome. *Hum Mol Genet* 3:115–123
- Murdock G (1967) *Ethnographic atlas*. University of Pittsburgh Press, Pittsburgh
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York
- Newcomer A, McGill DB, Thomas P, Hofmann A (1975) Prospective comparison of indirect methods for detecting lactase deficiency. *N Engl J Med* 24:1232–1235
- Newman J (1995) *The peopling of Africa: a geographic interpretation*. Yale University Press, New Haven, CT
- Nurse G, Jenkins T (1974) Lactose intolerance in San populations. *Br Med J* 2:728
- O'Keefe S, Adam J (1983) Primary lactose intolerance in Zulu adults. *S Afr Med J* 63:778–780
- O'Keefe S, Adam J, Cakata E, Epstein S (1984) Nutritional support of malnourished lactose intolerant African patients. *Gut* 25:942–947
- Olatunbosun D, Adadevoh B (1971) Lactase deficiency in Nigerians. *Am J Dig Dis* 16:909–914
- Olds LC, Sibley E (2003) Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a *cis* regulatory element. *Hum Mol Genet* 12:2333–2340
- Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Dekka R, Ferrell RE, Shriver MD (1998) Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 63:1839–1851
- Poulter M, Hollox E, Harvey CB, Mulcare C, Peuhkuri K, Kajander K, Sarner M, Korpela R, Swallow DM (2003) The causal element for the lactase persistence/non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. *Ann Hum Genet* 67:298–311
- Puehkuri K (2000) Lactose, lactase and bowel disorders. PhD thesis, Hakapaino, Helsinki, <http://ethesis.helsinki.fi/julkaisut/laa/biola/vk/peuhkuri/> (accessed April 5, 2004)
- Ransome-Kuti O, Kretchmer N, Johnson J, Gribble J (1972) Family studies of lactose intolerance in Nigerian ethnic groups. *Pediatr Res* 6:359
- (1975) A genetic study of lactose digestion in Nigerian families. *Gastroenterology* 68:431–436
- Sahi T (1974) The inheritance of selective adult-type lactose malabsorption. *Scand J Gastroenterol Suppl*:1–73
- Segal I, Gagjee P, Essop A, Noormohamed A (1983) Lactase deficiency in the South African black population. *Am J Clin Nutr* 38:901–905
- Swallow DM, Harvey CB (1993) Genetics of adult-type hypolactasia. *Dyn Nutr Res* 3:1–7
- Swallow DM, Hollox EJ (2000) The genetic polymorphism of intestinal lactase activity in adult humans. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) *The metabolic and molecular basis of inherited disease*, 8th ed. McGraw-Hill, New York
- Thomas MG, Bradman N, Flinn HM (1999) High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y-chromosome. *Hum Genet* 105:577–581
- Troelsen JT, Olsen J, Moller J, Sjoström H (2003) An upstream polymorphism associated with lactase persistence has increased enhancer activity. *Gastroenterology* 125:1686–1694
- Y-chromosome consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339–348