Screening of variants for lactase persistence/non-persistence in populations from South Africa and Ghana

Suvi Torniainen, M Iqbal Parker, Ville Holmberg, Elisa Lahtela, Collet Dandara and Irma Jarvela*

Address: 1Department of Medical Genetics, University of Helsinki, Helsinki, Finland, 2International Centre for Genetic Engineering and Biotechnology and Division of Medical Biochemistry, University of Cape Town, Anzio Road, Observatory, 7925, Cape Town, South Africa, 3Department of Bacteriology and Immunology, University of Helsinki, Helsinki, Finland and 4Helsinki University Hospital, Laboratory Services, Helsinki, Finland

Email: Suvi Torniainen - suvi.torniainen@helsinki.fi; M Iqbal Parker - iqbal.parker@uct.ac.za; Ville Holmberg - ville.holmberg@helsinki.fi; Elisa Lahtela - laura.lahtela@helsinki.fi; Collet Dandara - collet.dandara@wits.ac.za; Irma Jarvela* - irma.jarvela@kolumbus.fi

* Corresponding author

Abstract

Background: Lactase non-persistence is a condition where lactase activity is decreased in the intestinal wall after weaning. In European derived populations a single nucleotide polymorphism (SNP) C/T-13910 residing 13.9 kb upstream from the lactase gene has been shown to define lactase activity, and several other single nucleotide polymorphisms (G/C-14010, T/G-13915, C/G-13907 and T/C-13913) in the same region have been identified in African and Middle East populations.

Results: The T-13910 allele most common in European populations was present in 21.8% mixed ancestry (N = 62) individuals and it was absent in the Xhosa (N = 109) and Ghana (N = 196) subjects. Five other substitutions were also found in the region covering the previously reported variants in African and Middle East populations. These included the G/C-14010 variant common in Kenyan and Tanzanian populations, which was present in 12.8% of Xhosa population and in 8.1% of mixed ancestry subjects. Two novel substitutions (C/T-14091 and A/C-14176) and one previously reported substitution G/A-13937 (rs4988234) were less common and present only in the Xhosa population. One novel substitution G/A-14107 was present in the Xhosa and Ghanaian populations. None of the other previously reported variants were identified.

Conclusion: Identification of the G/C-14010 variant in the Xhosa population, further confirms their genetic relatedness to other nomadic populations members that belong to the Bantu linguistic group in Tanzania and Kenya. Further studies are needed to confirm the possible relationship of the novel substitutions to the lactase persistence trait.

Background

Lactase non-persistence is a condition where the activity of lactase declines during childhood leading to malabsorption of lactose [1]. Certain DNA-variants, however, enable individuals to never lose their lactase activity. A single nucleotide polymorphism (SNP) C/T-13910 residing in intron 13 of the MCM6 gene 13.9 kb upstream from the lactase gene (LCT) has been found to correlate perfectly with lactase activity [2-5]. Functional studies have shown that the C/T-13910 variant acts as an enhancer [6,7] and it
regulates the LCT gene at the transcriptional level [3,8]. This long distance regulatory region seems to be a binding site for the Oct-1 transcription factor [9]. Recently, other variants close to the original variant C/T-13910 in the same enhancer region have been identified in nomadic pastoralist and non-pastoralist groups from East Africa, Cameroon and Middle East populations [10-12]. The G/C-14010 variant is common in East Africa. The indirect lactose tolerance test (LTT) has been used to show that this variant has a significant association with the lactase persistence trait. Other variants associated with lactase persistence in Africa or the Middle East are T/G-13915, C/G-13907 and T/C-13913. Functional in vitro studies using dual-luciferase reporter assays have shown that the mutant C-14010, G-13907 and G-13915 alleles increase LCT promoter expression [10]. Tissue samples from intestinal biopsies on individuals from the Arabian Peninsula have been used to show a correlation between allele G-13915 and lactase activity [11].

To further investigate genetic variants in African populations we sequenced the 400 bp region covering the previously identified variants from DNA-samples of subjects from Xhosa or mixed ancestry groups from South Africa. As a comparison, we sequenced the same region in the Akan people, a Western African linguistic group, which has a long history of farming and trading, but no dairy culture.

Results

The identified single nucleotide substitutions and their allele frequencies are shown in Table 1. We found a total of six variants in the region covering 400 bp on both sides of the original C/T-13910 variant. The distances between the variants are shown in Figure 1.

Xhosa population

The most common substitution in Xhosa population was allele C-14010 present in 12.8% of the subjects (Table 1). All the Xhosa subjects were homozygous for the genotype C/C-13910 and the T-13910-allele was completely absent. In addition, one previously known G/A-13937 (rs4988234) and three other novel substitutions C/T-14091, G/A-14107, A/C-14176 were found (Table 1, Figure 1). None of the previously reported variants identified in African and Arab populations (C/G-13907, T/C-13913, T/G-13915) [10-12] were found in this study.

Mixed ancestry population

The frequency of the T-13910 allele was 21.8% and the C-14010 allele 8.1% in the mixed ancestry population (Table 1). None of the previously reported mutations (C/G-13907, T/C-13913, T/G-13915, G/A-13937) [10,12] or the novel substitutions presented in this study were identified.

Akan population in Ghana

Both T-13910 and C-14010 alleles were absent in the Ghanaian population. Only one of the novel substitutions found in the Xhosa population, (G/A-14107) in a heterozygous form, was present in (the non-malaria group from) the Ashanti Region in Ghana. None of the other novel substitutions or the previously reported mutations (C/G-13907, T/C-13913, T/G-13915, G/A-13937) [10,12] were present in this study population.

Discussion

In Europe the C/T-13910 variant determines lactase activity in lactase persistence/non-persistence. In pastoralist populations of Africa several other SNPs in the same region upstream of LCT have been identified. Functional evidence for lactase persistence has been obtained for some of these variants [10,12]. We analyzed the distribution of substitutions near C/T-13910 in a South African population that consumes milk products and compared it to substitutions found in the Akan peoples from Ghana, who does not have a dairy culture.

Modern humans have lived in South Africa for at least 77,000 years [13]. The early settlers of South Africa, the KhoiKhoi and San people (collectively called the Khoisan), were stone-age hunter-gatherers. They are often thought to represent distinct ethnic groups [14]. Bantu speaking people started to migrate south 2000 years ago bringing with them smelting technology (the iron-age) and expertise in farming [14-16]. Among the Bantu people who migrated to Southern Africa were the Nguni who included the Xhosas, Zulus and Ndebeles. The Xhosa and

Figure 1

Locations of single nucleotide substitutions found in the study. Variants previously associated with lactase activity are shown in bold. Substitutions are located upstream of the LCT gene. The numbers show the distance in nucleotides between the variant and the beginning of the LCT gene (marked as 0).
Zulu people comprise a large proportion of modern day Black South Africans. Xhosa speaking people include many subgroups with related but distinct ancestries [17].

Most of the Xhosa-speaking peoples in today’s South Africa, originally came from the Eastern Cape province. They consume fermented milk from the many cattle they keep. This is generally true for most of the Bantu-speaking populations and to some extent indicates their ability to tolerate lactose. We did not find the C/T-13910 variant common in populations of European origin, further confirming previously reported findings that this SNP is rare in African populations [10,12,18]. The results of our study show that in the Xhosa population the C-14010 allele of the polymorphic site G/C-14010 is the most common of those alleles previously associated with lactase activity. The C-14010 allele has also been found to be present in Tanzania with a higher prevalence (31.9%) than in the South African Xhosa population [10]. Because of the frequency difference and the fact that the Xhosa population originated from Central Africa and migrated slowly southwards, we hypothesize that the lactase persistence allele has migrated with them and gradually diluted the frequency of C-14010 by intermarriages with the Khoisan people [15,16,19].

Interestingly, this study revealed three novel substitutions C/T-14091, G/A-14107, A/C-14176 in the Xhosa population. The substitutions are not reported in the SNP-database http://www.ncbi.nlm.nih.gov/sites/entrez. The frequencies of the substitutions are low and unlike the previously detected T/G-13915, T/C-13913 and C/G-13907 they are not located close to the C/T-13910 variant in the Oct-1 transcription factor binding site. However, the size of enhancer sites can be in excess of several hundred basepairs and it is therefore possible that all these variants are included in the same enhancer region [20]. In order to investigate their functional role in the regulation of lactase activity, subjects carrying these mutations need to be recruited and to investigate any associations between them and the results of their lactose tolerance test (LTT) or preferably the lactase activity of their intestinal biopsy specimens. It is known that when the Xhosa people migrated south, they assimilated with the earlier inhabitants of Southern Africa, the Khoisan, thus, it would be interesting to investigate the distribution of these novel substitutions in these populations. However, as they are rare, a large number of samples would be required in order to uncover any functional significance for the three novel substitutions from association studies. Another possibility is that these variants are not associated with lactase persistence but represent the natural diversity of African populations.

Among the people of mixed ancestry origin both the T-13910 and the C-14010 allele were found reflecting the genetic influence of both Africans and Europeans in this group. This is explained by the diverse ethnic origin (Africans, Western Europeans, the Khoisan, Indonesians and Malaysians) of these people. Therefore the results of the mixed ancestry population group support their ethnic background.

The Akan in the rural Ashanti region live mainly on subsistence agriculture and rarely keep milk producing livestock. Their consumption of fresh milk products has thus remained very limited. This explains why there has not been any evolutionary pressure towards an increasing frequency of lactose persistence genotypes, although these genotypes are known to be present in the population.

**Conclusion**

Identification of the G/C-14010 variant in the Xhosa population, further confirms their genetic relatedness to other nomadic populations members that belong to the Bantu linguistic group in Tanzania and Kenya. Further studies are needed to confirm the possible relationship of the novel substitutions to the lactase persistence trait.

**Methods**

**Study population**

The study material consisted of 171 samples collected from Southern Africa for the study of genetic risk factors affecting esophageal cancer [21]. Of these 109 were Xhosa and 62 of mixed ancestry. The Black Xhosa-speaking South Africans originated from the Eastern or Western Capes. The mixed ancestry (commonly referred to as “coloured”) subjects are the result of intermarriages between...
races including Black Africans, Western Europeans, the Khoisan, Indonesians and Malaysians who settled in the Cape from the middle of the 17th century. Written consent was obtained before subjects were enrolled in the study. A questionnaire gathering details on demographics (age, sex, and race), smoking habits and alcohol consumption was completed for all subjects by a trained phlebotomist. Blood was collected from participating subjects and processed for DNA isolation [21]. The study was approved by Cape Town University’s Human Research Ethics Committee.

A total of 196 subjects were analyzed from Ghana. Diagnostic procedures and the malarial indices of the women have been described elsewhere [22]. DNA was extracted from the subjects’ peripheral venous blood (QiAmp, Qiagen). All the participants gave their informed consent. The study protocol was reviewed and approved by the Committee on Human Research Publication and Ethics, School of Medical Sciences, University for Science and Technology, Kumasi, Ghana, and the institution’s guidelines were followed in conducting the study.

**Genotyping**

The DNA fragment spanning the C/T-13910 variant was amplified by polymerase chain reaction (PCR) and then sequenced. The primer sequences used were: 5'-CCCTCGTTAATACCCACGTGACC-3' (forward) and 5'-GCTACATTCGATATGAGTGAGGACGCA-3' (reverse). The PCR products were separated on a 1.5% agarose gel containing ethidium bromide for size verification. The excess primers were removed by digestion with 2.5 U of Shrimp Alkaline Phosphatase (USB) and 5 U of Exonuclease I (New England Biolabs) at 37°C for 45 min followed by inactivation at 80°C for 15 min. For sequencing the BigDye 3.1 terminator (Applied Biosystems) was used according to the manufacturer’s instructions. The sequencing reaction was as follows: at 96°C for 1 min, then 25 cycles at 96°C for 10 s, at 55°C for 5 s and at 60°C for 4 min, which was followed by purification with MilliQ Multiscreen plates (Millipore, USA) with Sephadex G-50 Superfine sepharose (Amersham Biosciences, Sweden), electrophoresis with an ABI 3730 DNA Analyzer (Applied Biosystems) and base calling using the Sequence Analysis 5.2 software (Applied Biosystems). The results were analyzed by Sequencher 4.1.4 software (Gene Codes, USA).

**Authors’ contributions**

ST carried out sequencing, analysed sequences, drafted the manuscript and made the figure and the table. CD and IP provided the samples from South Africa and their expertise in drafting the manuscript. VH provided samples from Ghana and description of this study population and participated in drafting of the manuscript. EL participated in the laboratory work and drafting. II got the idea for the study, participated on designing of the study and final drafting of the manuscript. All authors read and approved the final manuscript.

**Acknowledgements**

The authors thank all the participants of this study. This work is based upon research supported by the South African Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation given to IMP. Grants from the Medical Research Council of South Africa (MRC) obtained by IMP, and the Sigrid Jusélius Foundation, Helsinki, Finland given to IJ are gratefully acknowledged.

The foundations had no role in design, in the collection, analysis, and interpretation of data, in the writing of the manuscript; and in the decision to submit the manuscript for publication.

**References**

1. Sahi T: Genetics and epidemiology of adult-type hypolactasia. Scand J Gastroenterol 1994:7-20.
2. Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I: Identification of a variant associated with adult-type hypolactasia. Nat Genet 2002, 30:233-7.
3. Kuokkanen M, Enattah N, Oksanen A, Savilahti E, Orpana A, Jarvela I: Transcriptional regulation of the lactate-phlorizin hydrolase gene by polymorphisms associated with adult-type hypolactasia. Gut 2003, 52:467-52.
4. Rasinpera H, Savilahti E, Enattah NS, Kuokkanen M, Törmänen N, Lindahl J, Jarvela I, Kolho KL: A genetic test which can be used to diagnose adult-type hypolactasia in children. Gut 2004, 53:1571-6.
5. Enattah NS, Kuokkanen M, Forsblom C, Natah S, Oksanen A, Jarvela I, Peltonen L, Savilahti E: Correlation of intestinal disaccharidase activities with the C/T-13910 variant and age. World J Gastroenterol 2007, 13:3508-12.
6. Troelsen JT, Olsen J, Möller J, Sjöström J: An upstream polymorphism associated with lactase persistence has increased enhancer activity. Gastroenterology 2003, 125:1866-94.
7. Olds LC, Sibley E: Lactase persistence DNA variant enhances lactase promote activity in vitro: functional role as a cis regulatory element. Hum Mol Genet 2003, 12:2333-40.
8. Rasinpera H, Kuokkanen M, Kolho K-L, Enattah NS, Lindahl H, Savilahti E, Orpana A, Jarvela I: Transcriptional downregulation of the lactase (LCT) gene during childhood. Gut 2005, 54(11):1660-1661.
9. Lewinsky RH, Jensen TG, Möller J, Steneballe A, Olsen J, Troelsen JT: T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. Hum Mol Genet 2005, 14:3945-53.
10. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, mortensen HM, Hirbo JB, osman M, Ibrahim M, Omar SA, Lema G, Nyambo TB, Ghori J, Bumpstead S, Pritchard JK, Wray GA, Deloukas P: Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 2007, 39:31-40.
11. Imtiaz F, Savilahti E, Sarnesto A, Trabzuni D, Al-Kahtani K, Kagevi I, Rashid NS, Meyer B, Jarvela I: T/G-13915 variant upstream of the lactase gene (LCT) is the founder allele of adult-type hypolactasia in an Arab population(s). J Med Genet 2007, 44(10):e89.
12. Ingram CJ, Elamin FF, Mulcare CA, Weale ME, Tarekegn A, Raga TO, Bekele E, Elamin FM, Thomas MG, Bradman N, Swallow DM: A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? Hum Genet 2007, 120:779-88.
13. Henshilwood C, d’Erricco F, Yates R, Jacobs Z, Tribolo C, Duller G, Mercier N, Sealy J, Valladas H, Watts I, Wintle A: Emergence of modern human behavior: middle stone age engravings from South Africa. Science 2002, 295:1278-80.
14. Behar D, Villetos R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkani H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith C, Wells S, Rosset S: The Geographic Consortium. The dawn of human matrilineal diversity. Am J Hum Genet 2008, 82:1130-40.
15. Oliver R: The problem of the Bantu expansion. *J African History* 1966, 7:361-76.

16. Cavalli-Sforza L, Menozzi P, Piazza A: Demic expansions and human evolution. *Science* 1993, 259:639-46.

17. Lane A, Soodyall H, Arndt S, Rashikhoph A, Jonker E, Freeman C, Young L, Morar B, Toffie L: Genetic substructure in South African Bantu-speakers: evidence from autosomal DNA and Y-chromosome studies. *Am J Phys Anthropol* 2002, 119:175-85.

18. Mulcare CA, Weale ME, Jones AL, Connell B, Zeitlyn D, Tarekegn A, Swallow DA, Bradman N, Thomas MG: 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. *Am J Hum Genet* 2004, 74:1102-10.

19. Pereira L, Gusmao L, Amorim A, Prata MJ: Bantu and European Y-lineages in Sub-Saharan Africa. *Ann Hum Genet* 2002, 66:369-78.

20. Visel A, Bristow J, Pennacchio LA: Enhancer identification through comparative genomics. *Semin Cell Dev Biol* 2008, 18:140-52.

21. Dandara C, Li DP, Walther G, Parker MI: Gene-environment interaction: the role of SULT1A1 and CYP3A5 polymorphisms as risk modifiers for squamous cell carcinoma of the oesophagus. *Carcinogenesis* 2006, 27:791-797.

22. Mockenhaupt FP, Ulmen U, von Gaertner C, Bedu-Addo G, Bienzle U: Diagnosis of placental malaria. *J Clin Microbiol* 2002, 40(1):306-8.