SHORT COMMUNICATION

Lactase persistence may have an independent origin in Tibetan populations from Tibet, China

Min-Sheng Peng[1,3,5], Jun-Dong He[1,2,3,5], Chun-Ling Zhu[1,3], Shi-Fang Wu[1,3], Jie-Qiong Jin[1,3] and Ya-Ping Zhang[1,3,4]

Milk consumption is prevalent in daily diets of Tibetans. To digest the milk sugar lactose, lactase persistence (LP) should be required. However, little is known about the genetic basis of LP in Tibetans. We screened 495 Tibetan individuals for five previously reported single-nucleotide polymorphisms (SNPs): −13907C/G (rs41525747), −13910C/T (rs4988235), −13915T/G (rs41380347), −14010G/C and −22018G/A (rs182549), which are associated with the LP in populations from a vast region surrounding Tibet. The five SNPs were nearly absent in Tibetan populations, suggesting LP likely to have an independent origin in Tibetans rather than to be introduced via gene flow from neighboring populations. We identified three novel SNPs (−13838G/A, −13906T/A and −13908C/T) in Tibetans. In particular, −13838G/A might be functional as it is located in the binding motif for HNF4α that acts as a transcription factor for intestinal gene expression. To investigate the potential association of this variant with LP, further detailed studies are required in the future.

With the arising of agricultural revolution, the evolution of food producing has transformed human global demographics within the Holocene.[1] The bio-cultural adaptations to new diets, particularly those associated with the farming and animal husbandry, have left several signatures in human genomes.[2,3] One of the most famous cases, to digest lactose from milk, human needs continued production of lactase throughout adult life (lactase persistence, LP; OMIM #223100). This trait is caused by genetic differences *cis*-acting to the lactase gene (*LCT*).[4] Four causative single-nucleotide polymorphisms (SNPs) ∼13.9 kb upstream from *LCT*: −13910C/T (rs4988235), −13907C/G (rs41525747), −13915T/G (rs41380347) and −14010G/C, have been subsequently identified as the candidate *cis*-acting elements based on genotype–phenotype association analyses and functional experiments during the past decade.[5] The derived allele of −13910*T is associated with LP in European[6] and some Central Asian[7] populations. This allele is identified in Indian populations and is inferred to make substantial contribution to LP in India.[8] The rest three alleles, −13907*G, −13915*G and −14010*C are responsible for the LP in some African and Middle Eastern populations.[9,10] These results suggest that LP alleles have emerged independently in several geographic/ethnic groups,[2] which

was most likely driven by recent positive selection,[10,11] as well as demographic factors (for example, migration[12]).

In the Tibetan Plateau, milk and milk products (for example, from domestic yaks) are important ingredients of the daily diets for Tibetans, especially for the herders.[13] The lactose tolerance test with measurement of breath hydrogen revealed that the LP prevalence in Tibetans (9/30; 30%) were significantly higher than in Han Chinese (1/30; 3.3%).[14] This result was in accordance with the history of milk consumption for Tibetans,[14] whereas the lack of milk from traditional Han Chinese diets.[15] However, how the culture of milk usage shaped the genetic diversity of Tibetan populations remains an enigma. To investigate genetic variants for LP in Tibetans, we sequenced a region of 321 bp (position −14044 to −13724 upstream *LCT*) covering the previously identified four SNPs (−13907C/G, −13910C/T, −13915T/G and −14010G/C) from DNA samples of 495 Tibetan individuals living in Tibet (Figure 1). The PCR and sequencing protocols were previously reported.[9] All sequences have been deposited in the GenBank (accession numbers: JQ395072–JQ395566). Meanwhile, allele −22018*A[16] but not allele −13910*T[17] was suggested to be associated with LP in populations from Northern China. So we
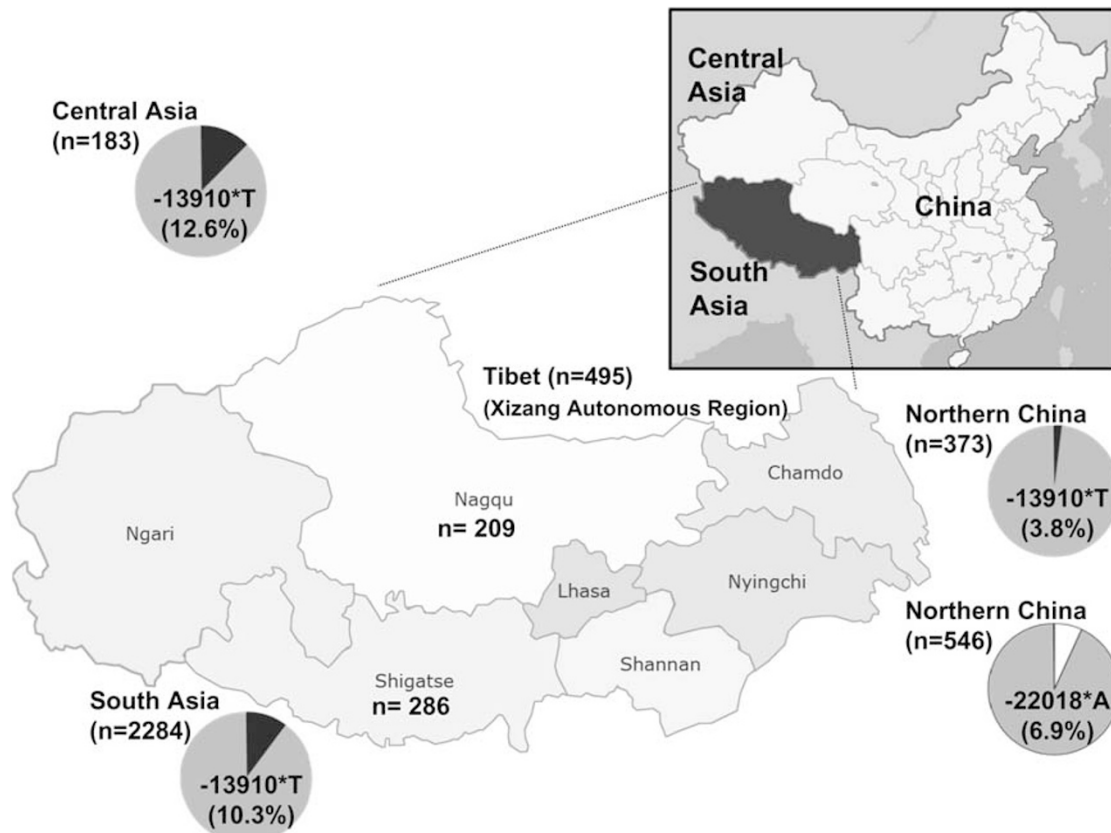
Figure 1 Population frequencies for −13910*T and −22018*A alleles associated with LP surrounding Tibet. The information of frequencies is retrieved from published literatures.[7,8,16,17] The sampling in Tibet is also shown in the map.

further screened −22018G/A (rs182549) in our samples using the RFLP method as described before.[16]

For the four SNPs (−13907C/G, −13910C/T, −13915T/G and −14010G/C), the alleles responsible for LP are completely absent in Tibetan samples (Table 1 and Supplementary Table S1). For −22018G/A, only one heterozygous form is present in the population from Nagqu. The results indicate that the LP in Tibetans may not be explained by these known SNPs. When looking at the published populations with LP around Tibet (Figure 1), −13910*T was responsible for the substantial proportion of LP in Central Asian[7] and Indian[8] populations; −22018*A was likely associated with the LP in populations from Northern China.[16] In terms of previous studies about mitochondrial DNA and Y chromosome variation in Tibetans, some but rare gene flows from Central Asians and South Asians were detected.[18,19] The Neolithic genetic components from Northern China were revealed to contribute substantially to the current gene pool of Tibetans.[18–20] Nevertheless, the susceptible alleles for LP in those surrounding populations were not introgressed into Tibetan populations via the prehistorical and historical migrations. Taken together, we propose that the LP in Tibetans is likely to have an independent origin.

In this study, we have identified three more SNPs: −13838G/A, −13906T/A and −13908C/T in the sequenced region (Table 1 and Supplementary Table S1). Both −13906*A (0.6%; 6/990) and −13908*T alleles (0.1%; 1/990) occurred at low frequencies. The sporadic distribution patterns suggest that the two rare mutations are unlikely to be associated with LP in Tibetan populations, although the possibility that the LP due to multiple low frequency mutations cannot be excluded completely. For −13838G/A, with the exception of absence in Baqên and Nyima (Table 1), its frequency ranges from 1.9 (1/54, Rinbung) to 20.8% (5/24, Baingoin) in Tibetan populations; and it has an overall frequency of 6.6% (65/990). No significant deviations from Hardy–Weinberg equilibrium for −13838G/A were observed in the whole Tibetan population as well as the two geographically defined sub-populations (Table 2). Interestingly, the site of −13838 is located in the binding motif (position −13854 to −13830 upstream LCT) for HNF4α—a transcription factor for intestinal gene expression.[21] The previous functional study has revealed that this motif is important for enhancer activity of −13910*T and co-expression of HNF4α can increase both −13910*T and −13910*C enhancer activities.[21] The above evidence imply that more attention should be paid for SNP −13838G/A. To discern whether the −13838G/A and some unidentified variants are responsible for LP in Tibetans or not, further comprehensive genotype–phenotype association analyses and functional experiments are required in the future.

Table 1 Allele of variants in the *LCT* enhancer in Tibetan populations

Prefecture	Country	Number of people	−22018*A	−14010*C	−13915*G	−13910*T	−13908*T	−13907*G	−13906*A	−13838*A
Nagqu	Amdo	26					1			1
Nagqu	Baqên	12								
Nagqu	Baingoin	12								5
Nagqu	Biru	12								1
Nagqu	Lhari	10								3
Nagqu	Nagqu	81	1							14
Nagqu	Nyima	9								
Nagqu	Nyainrong	29								4
Nagqu	Xainza	9								1
Nagqu	Sog	9								1
Shigatse	Ngamring	13							1	2
Shigatse	Bainang	52							1	10
Shigatse	Tingri	15								1
Shigatse	Gyangzê	22								3
Shigatse	Kangmar	22								3
Shigatse	Lhazê	37								3
Shigatse	Namling	11							1	1
Shigatse	Nyalam	6								1
Shigatse	Rinbung	27							1	1
Shigatse	Xigazê	23							1	4
Shigatse	Saga	11							1	1
Shigatse	Yadong	43								4
Shigatse	Zhongba	4								1
Nagqu	Total	209	1				1			30
Shigatse	Total	286							6	35
All	Total	495	1				1		6	65

Abbreviation: *LCT*, lactase gene.

Table 2 The χ^2 Hardy–Weinberg equilibrium test for −13838G/A

Population	Size	GG	GA	AA	χ^2-value	P-value
Nagqu	209	180	28	1	0.01	>0.05
Shigatse	286	253	31	2	0.91	>0.05
All	495	433	59	3	0.4	>0.05

Note: the test was performed by using the Hardy–Weinberg equilibrium calculator (http://www.oege.org/software/hwe-mr-calc.html).[22]

1 Diamond, J. & Bellwood, P. Farmers and their languages: the first expansions. *Science* **300,** 597–603 (2003).
2 Patin, E. & Quintana-Murci, L. Demeter's legacy: rapid changes to our genome imposed by diet. *Trends Ecol. Evol.* **23,** 56–59 (2008).
3 Laland, K. N., Odling-Smee, J. & Myles, S. How culture shaped the human genome: bringing genetics and the human sciences together. *Nat. Rev. Genet.* **11,** 137–148 (2010).
4 Wang, Y., Harvey, C. B., Pratt, W. S., Sams, V. R., Sarner, M., Rossi, M. *et al.* The lactase persistence/non-persistence polymorphism is controlled by a *cis*-acting element. *Hum. Mol. Genet.* **4,** 657–662 (1995).
5 Ingram, C. J., Mulcare, C. A., Itan, Y., Thomas, M. G. & Swallow, D. M. Lactose digestion and the evolutionary genetics of lactase persistence. *Hum. Genet.* **124,** 579–591 (2009).
6 Enattah, N. S., Sahi, T., Savilahti, E., Terwilliger, J. D., Peltonen, L. & Järvelä, I. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30,** 233–237 (2002).
7 Heyer, E., Brazier, L., Frédéric, L., Hegay, T., Austerlitz, F., Quintana-Murci, L. *et al.* Lactase persistence in Central Asia: phenotype, genotype, and evolution. *Hum. Biol.* **83,** 379–392 (2011).
8 Gallego Romero, I., Basu Mallick, C., Liebert, A., Crivellaro, F., Chaubey, G., Itan, Y. *et al.* Herders of Indian and European cattle share their predominant allele for lactase persistence. *Mol. Biol. Evol.* **29,** 249–260 (2012).
9 Ingram, C. J., Elamin, M. F., Mulcare, C. A., Weale, M. E., Tarekegn, A., Raga, T. O. *et al.* A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Hum. Genet.* **120,** 779–788 (2007).
10 Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S. *et al.* Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39,** 31–40 (2007).
11 Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A. *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* **74,** 1111–1120 (2004).
12 Itan, Y., Powell, A., Beaumont, M. A., Burger, J. & Thomas, M. G. The origins of lactase persistence in Europe. *PLoS Comput. Biol.* **5,** e1000491 (2009).
13 Dong, S. K., Long, R. J. & Kang, M. Y. Milking and milk processing: traditional technologies in the yak farming system of the Qinghai-Tibetan Plateau, China. *Int. J. Dairy. Techol* **56,** 86–93 (2003).
14 Jiang, Z. & Liu, X.-c. Initial study on the lactose malabsorption and lactose intolerance of Tibetan middle school students. *J. Chongqing Med. Univ* **20,** 272–274 (1995).
15 Simoons, F. J. Primary adult lactose intolerance and the milking habit: a problem in biologic and cultural interrelations II. A culture historical hypothesis. *Am. J. Dig. Dis* **15,** 695–710 (1970).
16 Xu, L., Sun, H., Zhang, X., Wang, J., Sun, D., Chen, F. *et al.* The -22018A allele matches the lactase persistence phenotype in northern Chinese populations. *Scand. J. Gastroenterol.* **45,** 168–174 (2010).
17 Sun, H. M., Qiao, Y. D., Chen, F., Xu, L. D., Bai, J. & Fu, S. B. The lactase gene -13910T allele can not predict the lactase-persistence phenotype in north China. *Asia Pac. J. Clin. Nutr.* **16,** 598–601 (2007).
18 Zhao, M., Kong, Q. P., Wang, H. W., Peng, M. S., Xie, X. D., Wang, W. Z. *et al.* Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proc. Natl Acad. Sci. USA* **106,** 21230–21235 (2009).
19 Gayden, T., Cadenas, A. M., Regueiro, M., Singh, N. B., Zhivotovsky, L. A., Underhill, P. A. *et al.* The Himalayas as a directional barrier to gene flow. *Am. J. Hum. Genet.* **80,** 884–894 (2007).

20 Peng, M. S., Palanichamy, M. G., Yao, Y. G., Mitra, B., Cheng, Y. T., Zhao, M. *et al.* Inland post-glacial dispersal in East Asia revealed by mitochondrial haplogroup M9a'b. *BMC Biol.* **9,** 2 (2011).

21 Lewinsky, R. H., Jensen, T. G., Moller, J., Stensballe, A., Olsen, J. & Troelsen, J. T. T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity *in vitro*. *Hum. Mol. Genet.* **14,** 3945–3953 (2005).

22 Rodriguez, S., Gaunt, T. R. & Day, I. N. Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies. *Am. J. Epidemiol.* **169,** 505–514 (2009).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)