Identifying and Characterizing Causal Genetic Variants for Lactose Tolerance in African Populations

Anya Jayanthi (COL 2025), Michael McQuillan PhD, Sarah Tishkoff PhD
Department of Genetics and Biology, Perelman School of Medicine, University of Pennsylvania

Introduction

Cow's milk contains **lactose**, a sugar that is broken down into glucose and galactose by the enzyme lactase in the small intestine[1].

Lactase persistence (LP), or high lactase levels after weaning, is an inherited trait caused by substitution mutations called **single nucleotide polymorphisms (SNPs)** that increase the expression of the lactase-coding gene, *LCT.* In Africa, four causal SNPs for LP have been identified that account for 45% of LP phenotypic variation, suggesting that additional SNPs exist[1].

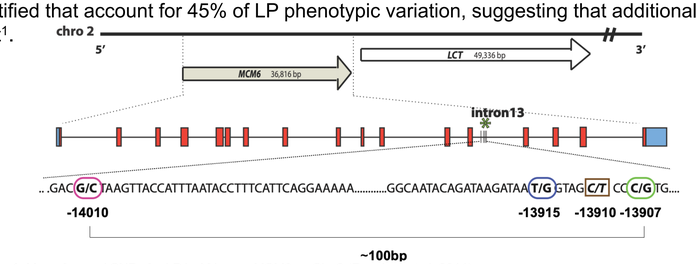

Figure 1. Map of causal SNPs for LP in Africa on *MCM6* on Chr 2 (Ranciaro et al. 2014).

Pastoralist (cattle herding) groups and their descendants are expected to have greater LP prevalence. The first expansion of pastoralism in Africa is correlated with **Afro-Asiatic** language speakers, the second expansion with **Nilo-Saharan** language speakers. Non-pastoralist farmers (Niger-Congo speakers) and foragers (Khoesan speakers e.g. the Hadza) are expected to have less LP prevalence[2,3,4,5].

We aimed to **assess LP phenotypic variation** among diverse African populations and **identify novel causal SNPs** for LP.

Goal: untangle the **evolutionary history** of LP in Africa, the correlation between LP and the spread of pastoralism, and the **genetic basis of human adaptation**.

Methods

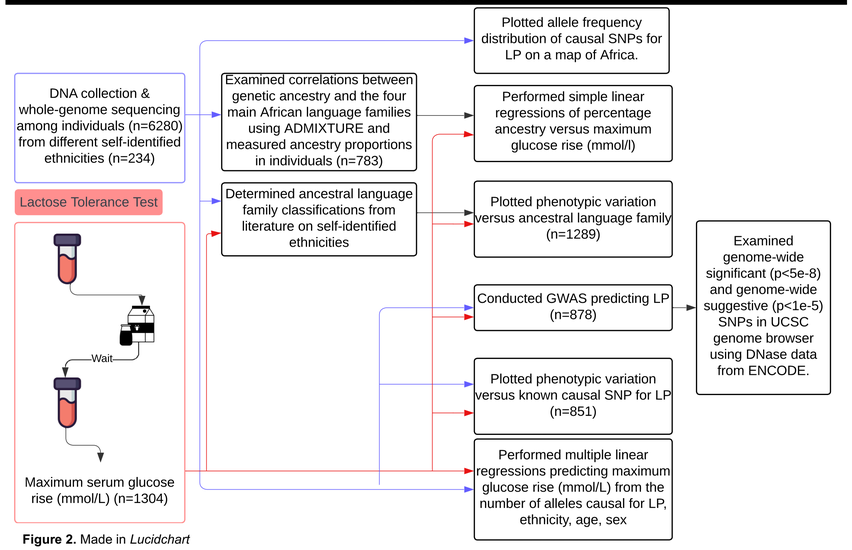

Figure 2. Made in *Lucidchart*

Results

Table 1. Summary of simple linear regressions for percent of an individual's ancestry by maximum glucose rise (mmol/L).

Afro-Asiatic **pastoralist ancestry** is positively correlated with LP, but Nilo-Saharan pastoralist ancestry is not.

Ancestry	Trend	P-Value	R^2 Value
Afro-Asiatic	+	4.325e-08	0.03474
Nilo-Saharan	NA	0.9695	1.726e-06
Hadza	+	0.1689	0. 01654

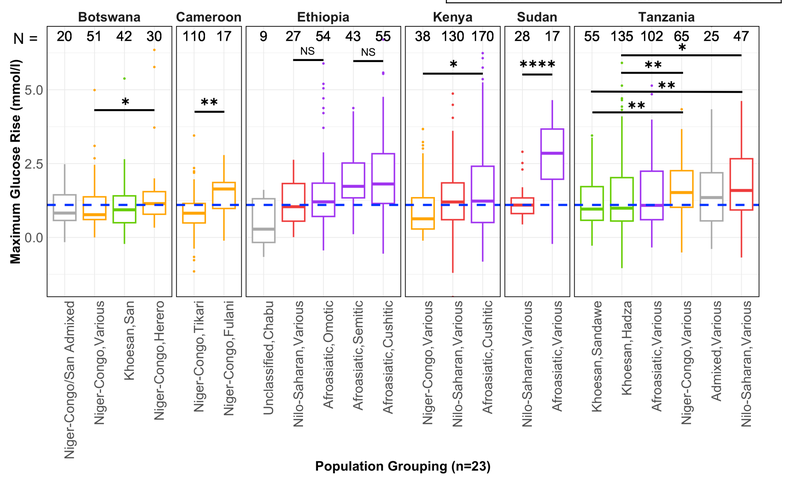

Figure 3. Maximum serum glucose rise (mmol/L) by population categorization boxplots in response to lactose consumption (n=1289). (Kruskal-Wallis: p=1.196e-14. The dotted line indicates 1.1 mmol/L, above which are the LP and lactase intermediate persistence (LIP) phenotypes. Wilcoxon pairwise significant differences are indicated by stars (p-values > 0.05 are "NS" or "non-significant," 0.05 ≥ are "✱," 0.01 ≥ are "✱✱," and each subsequent "✱" indicates an increase by e-01).

Pastoralist groups tend to have a higher prevalence of LP. **Afro-Asiatic** groups tend to have a higher prevalence of LP than **Nilo-Saharan** groups.

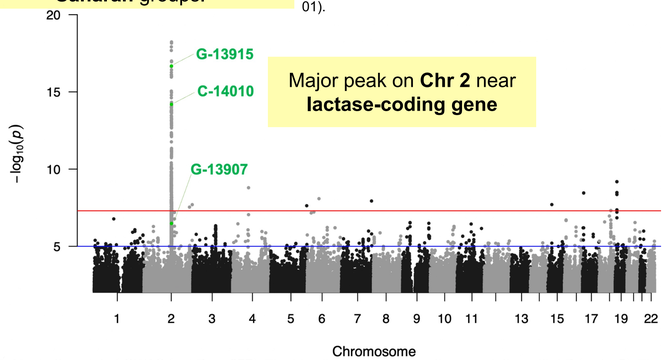

Figure 4. Manhattan plot of GWAS for LP (n=877). The blue line marks suggestive correlation with LP (p=1e-5). The red line marks genome-wide significantly correlation with LP (5e-08). The green dots mark known causal SNPs for LP.

Putative causal SNPs for LP are near genes related to **glucose uptake**.

Table 2. Summary of SNPs from GWAS identified that could be related to LP.

Chr	# of SNPs	Possible Relation to LP
2	2	Near LCT and MCM6. Overlap with small intestine regulatory peaks.
2	8	Near LCT. Overlap with regulatory peak(s).
2	25	Near gene(s) related to glucose uptake. Overlap with regulatory peak(s).
3	16	Overlap with regulatory peak(s).
4	3	Near gene(s) related to glucose uptake.
18	5	Near gene(s) related to glucose uptake.
19	2	Near gene(s) related to glucose uptake and/or overlap with regulatory peak(s).

Table 3. Summary of Wilcoxon pairwise significant differences in maximum glucose rise values between groups in Figure 4.

Variant(s) 1	Variant(s) 2	P-Value
13915	None	5.73E-15
None	14010	8.82E-14
None	13907	4.40E-09
13915	14010	4.30E-06
13907; 13915	None	2.76E-04
13915; 14010	None	4.74E-04
13907; 13915	14010	1.81E-03
14010	13907	2.06E-03
13907; 13915	13910	3.04E-03
13915	13910	3.82E-03
None	13910	4.57E-03
13910	13907	8.69E-03
13915; 14010	14010	1.17E-02
13915; 14010	13910	1.35E-02
13907; 13915	13907	4.04E-02

G-13915 and G-13907 tend to have greater maximum glucose rise values. Carriers of >1 different causal SNPs for LP tend to have **greater** glucose rises than carriers of just one or none

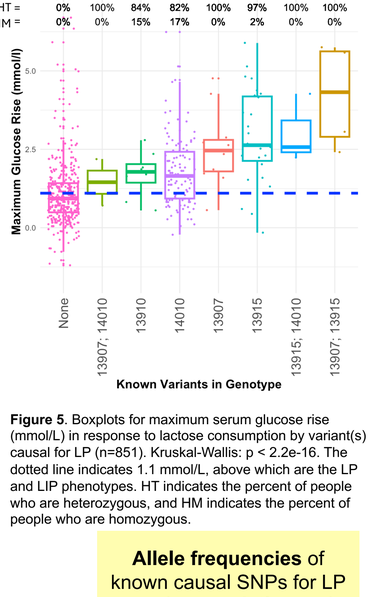

Figure 5. Boxplots for maximum serum glucose rise (mmol/L) in response to lactose consumption by variant(s) causal for LP (n=851). Kruskal-Wallis: p < 2.2e-16. The dotted line indicates 1.1 mmol/L, above which are the LP and LIP phenotypes. HT indicates the percent of people who are heterozygous, and HM indicates the percent of people who are homozygous.

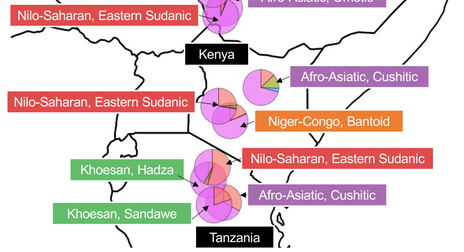

Figure 6. Pie-charts for distribution of SNPs in East Africa mapped by population grouping.

Allele frequencies of known causal SNPs for LP vary by **population** and **language**.

Afro-Asiatic groups tend to have higher frequencies of **G-13907** and **G-13915** SNPs than **Nilo-Saharan** groups.

Conclusions and Future Directions

Our study findings largely supported our expectations that LP is more prevalent in pastoralist populations; however, findings suggest that LP is more prevalent among the Hadza than expected, and less prevalent among Nilo-Saharan groups than expected

Our findings also suggested a potential additive effect among known causal SNPs for LP, and/or greater expression of lactase among G-13915 and G-13907

We identified 58 SNPs associated with greater increases in blood glucose after lactose consumption; 35 were near loci related to glucose metabolism. These SNPs should be further evaluated for association with LP or changes in glucose metabolism using functional genomic analyses, HiC data analyses, and genomic scans of selection.

Acknowledgements

Thank you to Dr. Michael McQuillan and Dr. Sarah Tishkoff for their patience, guidance, and support; Alessia Ranciaro for her earlier work in this area; our collaborators Gurja Belay, Dawit Wolde Meskel, Charles Fokunang, Muntaser Ibrahim, Thomas Nyambo, Sabah Omar, Sununguko Wata Mpoloka, Charles Wambebe, Gaonyadiwe George Mokone, and Alfred K. Njamnshi; and, especially, the African participants who donated their time, DNA, and phenotype information to make this research possible. This work was funded by the NIH grants 1R35GM134957, R01AR076241, and ADA 1-19-VSN-02, R35-GM134922, USSP at the University of Pennsylvania Perelman School of Medicine, and PURM and GfFMUR at the University of Pennsylvania.

Contact
Anya Jayanthi
University of Pennsylvania
Email: anyaj@sas.upenn.edu

References

1. Ranciaro A, Campbell MC, Hirbo JB, Ko WY, Froment A, Anagnostou P, Kotze MJ, Ibrahim M, Nyambo T, Omar SA, Tishkoff SA. Genetic origins of lactase persistence and the spread of pastoralism in Africa. Am J Hum Genet. 2014 Apr 3;94(4):496-510. doi: 10.1016/j.ajhg.2014.02.009. PMID: 24630847; PMCID: PMC3980415.
2. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet.* 2008;9:403-433. Prendergast ME, Lipson M, Sawchuk EA, et al. Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa. *Science*. 2019;365(6448):eaaw6275. doi:10.1126/science.aaw6275
3. Veyrieras JB, Kudaravalli S, Kim SY, et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet. 2008;4(10):e1000214. doi:10.1371/journal.pgen.1000214
4. Ségurel L, Bon C. On the Evolution of Lactase Persistence in Humans. Annu Rev Genomics Hum Genet. 2017;18:297-319. doi:10.1146/annurev-genom-091416-035340.