GENETIC TESTING OF ADULT-TYPE HYPOLACTASIA IN PRESENT-DAY AND ANCIENT SAMPLES

Summary of Ph.D. thesis

DÓRA NAGY M.D.

Paediatric Department

Faculty of Medicine of the University of Szeged

Szeged, Hungary

ABBREVIATIONS

ATH	adult-type hypolactasia
bp	base pair
dCAPS	derived cleaved amplified polymorphic sequence method
dNTP	deoxynucleotide-triphosphate
EDTA	ethylenediaminetetraacetic acid
HBT	hydrogen breath test
HVSI	hypervariable region I
HVSII	hypervariable region II
kb	kilo base
LNP	lactase non-persistence
LP	lactase persistence
MCM6	minichromosome maintenance-6 gene
mIPEP	modified improved primer extension preamplification
mtDNA	mitochondrial DNA
PCR	polymerase chain reaction
RFLP	restriction fragment length polymorphism
SNP	single nucleotide polymorphism
UV	ultraviolet

INTRODUCTION

Adult-type hypolactasia (ATH) also known as lactase non-persistence (LNP) is a common trait worldwide which varies in the prevalence and in the age of onset among populations both ethnically and geographically. The enzyme lactase on the apical surface of the brush border enterocytes facilitates lactose digestion in fresh milk. Its activity decreases after weaning in most humans, but persists in others. The lactase persistent (LP) phenotype is very frequent among Northern Europeans but declines towards the south and west. The trait is largely absent in Asians. It has been strongly associated with a single nucleotide polymorphism (SNP) 13910 base pairs (bp) upstream of the lactase gene in one of the introns of the MCM6 gene. The C/T_{-13910} polymorphism, as a *cis*-acting element, enhances the expression of the lactase gene. The CC_{-13910} genotype has been observed to be associated with LNP, CT_{-13910} and TT_{-13910} genotypes with LP phenotypes.

The ability to digest lactose into adulthood is one of the traits which developed in mankind under strong selection pressure. The selection forces could have been the nutritional benefit, the water and electrolyte contents of the milk or the improved calcium absorption after milk consumption. The most commonly acknowledged view is the *gene-culture co-evolution hypothesis*, which supposes that LP most likely originated as a mutation in populations which used milk as an important source of adult nutrition. A strong positive selection occurred for the LP phenotype from the ancestral LNP phenotype, around 5 000-10 000 years ago.

The Hungarian population belongs in the Finno-Ugric branch of the Uralic linguistic family, a diverse group of people. After a migratory period, the Hungarian pastoralist nomads entered Europe as seven major tribes that invaded the Danubian Basin from across the Carpathians around 895 AD. The Carpathian Basin had been settled for thousands of years before the arrival of Hungarians, by Dacians, Romans, Sarmatians, Goths, Huns, Avars, Slavs, and many others: it is probable that on the eve of the Hungarian conquest the overwhelming majority of the indigenous population were Slavic.

OUR AIMS

We determined the prevalence of LP and LNP genotypes in the Hungarian population, the age of onset of ATH and the applicability of hydrogen breath test (HBT), a clinical standard method, in comparison with genetic screening of the C/T_{-13910} SNP in symptomatic and asymptomatic children.

The other aim of this study was to evaluate the prevalence of C/T_{-13910} genotypes in remains from the Hungarian population of the 10th–11th centuries AD, as compared with present-day Hungarian-speaking populations. Random samples were evaluated, and our results were analyzed relative to other populations, which are believed to have been in contact with Hungarians during the migratory period, and after the settlement in the Carpathian basin.

SUBJECTS AND METHODS

Four groups of 432 present-day Hungarian-speaking individuals and a group of 42 bone samples from the period of the Hungarian conquest were analyzed.

Group 1 consisted of 82 persons with LNP-related symptoms after milk consumption. Beside these symptoms 15% of these subjects exhibited coeliac disease, 9% giardiasis, 3% inflammatory bowel disease and 7% bacterial or viral intestinal infection. ***Group 2*** comprised 104 randomly-selected control individuals with no abdominal symptoms and no previous gastrointestinal disease. The genotype frequencies of C/T_{-13910} variants in present-day Hungarian-speaking populations were established from ***group 3,*** 181 randomly-selected voluntary Hungarians, and ***group 4***, 65 Seklers.

42 bone samples in an excellent state of biomolecular preservation, originating from burials in the period of the Hungarian conquest were included in the analysis. The samples had been excavated in cemeteries from the 10th–11th centuries from different regions of the Carpathian basin. Both the burial sites and the bones were archaeologically and anthropomorphologically well-defined. Classical Hungarian conquerors were those excavated from rich graves. Commoners were found in

graves with impoverished burial remains. All 42 ancient bone samples had previously yielded reproducible mitochondrial DNA (mtDNA). All present-day and ancient samples underwent the molecular genetic screening for the C/T_{-13910} SNP. The individuals of group 1 and 2 also underwent HBT for the clinical evaluation of the results. In group 3 and 4 and in ancient samples the genetic testing was extended for the mtDNA haplogroups.

HBT was performed after the ingestion of 1 g of lactose/body weight kilogram (max. 50 g) in a 10% aqueous solution. HBT results were considered positive ≥20 ppm at 120 min after lactose load.

Genetic testing of present-day samples was based on DNA extraction from the root portion of single hairs or buccal smears, using Chelex-based method. Derived cleaved amplified polymorphic sequence method (dCAPS) was applied to analyse the genotypes, a restriction enzyme recognition site, including the SNP, was introduced into the PCR product by the forward primer containing a mismatch to template DNA (forward primer: 5'-GGCAATACAGATAAGATAATGGAG-3'; (the mismatching nucleotide is underlined); and reverse primer: 5'-CCTATCCTCGTGGAATGCAGG-3'). Polymerase chain reaction (PCR) amplification reactions accorded to the AmpliTaq Gold DNA polymerase protocol (Applied Biosystems, California, USA). The amplification protocol was: 6 min at 94°C, 35 cycles of 93°C for 45 sec, 54°C for 45 sec, 72°C for 60 sec and final extension at 72°C for 5 min. In restriction fragment length polymorphism (RFLP) the 119 bp PCR products were cleaved separately by NlaIV or HinfI endonucleases (Fermentas, Ontario, Canada), which resulted in the case of the C_{-13910} allele in 96-, 23-bp-long, and in that of the T_{-13910} allele in 97-, 22-bp-long fragments, and were run on an 8% native polyacrylamide gel and visualized by UV transillumination (UVP BioImaging System, Upland, CA, USA). 33 PCR products of the subjects with discrepant genotype and phenotype results were sequenced (BigDyeTM Terminator Version 3.1) to confirm or disprove the results of the PCR-RFLP.

Mitochondrial DNA (mtDNA) testing was previously carried out, based on the polymorphisms in the hypervariable region I and II and coding region of the human mtDNA. All ancient and present-day Sekler samples and 71 of the 181 present-day Hungarian samples were tested for mtDNA.

Processing of ancient samples was carried out according to the method, previously developed by our research team in the Genetic Institute of the Biological Research Center of the Hungarian Academy of Sciences. Bone powdering process included the cleaning, irradiation, the removal of the surfaces of the bones (femurs) and then the grinding of small portions of the epiphysis into fine powder. The DNA extraction procedure (suspension of large quantities of bone powder in ethylenediaminetetraacetic acid (EDTA), incubation overnight at 37°C with continuous vertical rotation and centrifugation) was repeated for 3-5 times. The sediment was suspended and incubated overnight at 37°C in extraction buffer. After centrifuging the solution the supernatant, containing the DNA, was stored at -20°C. Standard DNA isolation methods, and alternatively, when needed, a modified method incorporating the DNeasy Blood & Tissue Kit and QIAmp DNA Mini Kit (Qiagen, Valencia, CA, USA) was used.

After the successful and contamination-free mtDNA amplification, C/T_{-13910} genotyping was carried out on the ancient samples. The primers and restriction enzymes were identical to those used in the reactions of the present-day samples. The standard amplification reaction was carried out according to the GoTaq Hot Start Polymerase protocol (Promega, Wisconsin, USA). The amplification protocol was: 94°C for 6 min, 10 cycles of 30 sec at 93°C, 40 sec at 56°C, 40 sec at 72°C, 40 cycles of 30 sec at 93°C, 40 sec at 54°C, 40 sec at 72°C, and a final extension at 72°C for 5 min. 20 of the 42 ancient samples yielded DNA in the PCR reaction. The remaining 22 samples were subjected to a modified improved primer extension preamplification (mIPEP) method to enhance the efficiency of the amplification. 3 further ancient samples were successfully C/T_{-13910} genotyped. The

PCR products were then subjected to RFLP as described in connection with the present-day samples.

For ***contamination prevention*** with modern DNA, strict precautions were taken during each step of the ancient sample preparation. All steps of sample processing (bone powdering, DNA extraction, preamplification, amplification and post-PCR analysis) were carried out wearing appropriate protective clothing in separate rooms dedicated only for ancient DNA work. All workspaces and appliances were cleaned with bleach and subsequently irradiated with UV-C light before use. All solutions used were filtered and subsequently irradiated with UV-C light. During all steps Universal Fit Filter Tips (Corning Incorporated, Lowell, MA, USA) were used for pipetting. PCR and Eppendorf tubes were sterilized before use by autoclaving. The surface of the bone samples was cleaned and removed as described above in order to prevent possible contamination. The 11 persons, who participated in the sample processing or worked in the labs, were mtDNA tested and C/T_{-13910} genotyped. Each steps of bone processing was carried out at least two times by at least two researchers. Those samples were accepted that had consistent results in all successful PCR reactions and had different haplogroup and genotype results from that of the researchers who analyzed the samples. Extraction, preamplification, amplification, and digestion blanks (with no bone powder, template DNA or PCR product) were used as negative controls in each reaction to screen for contaminants entering the process at any stage. Positive controls (CC, CT, and TT_{-13910}) were also used in each digestion.

Due to DNA degradation, primers were designed to amplify short sequences of templates. To prove the authenticity of ancient human DNA further, DNA was isolated from an ancient horse remain, excavated from one of the human burial sites, and amplified with both the horse-specific and the human-specific C/T_{-13910} primers. Only the horse-specific primers resulted in amplification product.

We also made efforts to reduce the possible allele dropout. High-quality and -quantity bone powder was used to optimize DNA extraction; preamplification

method (mIPEP) was applied to increase the quantity of template DNA before PCR reaction; the number of amplification cycles were increased to reach the detection limit of the machine without strong artefacts; and amplification results were concluded from several, consistent PCR reactions.

GraphPad Prism version 4.00 for Windows software package (GraphPad Software, San Diego, California, USA) was used for ***statistical evaluation***. Parametric data were calculated by using variance analysis and unpaired Student's *t*-test with Welch correction (*d*-probe), when the variances between the pairs of groups differed significantly from each other ($p<0,05$ in the *F*-test). Data are expressed as mean ± standard error of the mean (mean ± SEM). Fisher exact test was applied to compare HBT with the results of the C/T_{-13910} genotypes and to compare the genotypes in the ancient samples with present-day populations. A p value $<0,05$ was considered as statistically significant. Deviation from Hardy-Weinberg equilibrium was calculated in present-day and ancient Hungarian populations concerning the C/T_{-13910} genotypes.

RESULTS

The prevalences of the C/C_{-13910}, C/T_{-13910} and T/T_{-13910} genotypes were 73, 26 and 1% in group 1; as compared with 38, 50 and 12% in group 2.

Comparison of HBT and C/T_{-13910} genotype results

The symptoms of LNP did not manifest under the age of five, except in one patient, aged four, where intestinal infection enhanced the symptoms.

The patients with abdominal complaints possessed positive HBT in 89% above the age of 12 and 100% above the age of 16 years, while in controls in the same age-group the prevalence was 69% and 71%, respectively.

With sequencing we could correct the genotypes of 5 samples that had discrepancy between the genotype by RFLP and the phenotype. In all, 12 patients had coeliac disease.

In case of CC_{-13910} genotype significant difference was measured between group 1 and 2 in the exhaled H_2 levels at 120 min (52±5,9 ppM versus 27±4,7 ppM, p<0,01) in the subgroup of 5-12 years; and at 60 min (32±8,9 ppM versus 12±2,8 ppM, p<0,05), at 90 min (54±11,1 ppM versus 21±4,5 ppM, p<0,01) and at 120 min (63±11,3 ppM versus 31±5,9 ppM, p<0,05) in the >12-year subgroup. In case of CT_{-13910} and TT_{-13910} genotypes, between group 1 and 2, at 120 min significant difference (22±7,2 ppM versus 3±1,6 ppM, p<0,05) was observed in the subgroup of 5-12 years and no significant difference (19±10,2 ppM versus 4±1,3 ppM, p=0,20) was found above the age of 12 years.

The evaluation of the results of HBT compared to the genetic screening was carried out by Fisher exact test in groups 1 and 2. The 120-min sampling point of the HBT results proved to be the most reliable both in the 5–12-year (group 1: p<0.01, group 2: p <0.001) and in the >12-year subgroup (group 1: p <0.001, group 2: p <0.001). The positive and negative predictive values, and the sensitivity and specificity of the HBT at 120 min in groups 1 and 2, and the results of the genetic testing are similar to the findings of previous studies.

Results of ancient samples

The genotyping of the C/T_{-13910} autosomal SNP was successful in 23 ancient bone samples (13 classical conquerors, nine commoners, and one not determined). The prevalence of the CC_{-13910}, CT_{-13910}, and TT_{-13910} genotypes among the 23 ancient Hungarians was 87%, 4%, and 9%; as compared with 39%, 50%, and 11% among 181 present-day Hungarians; and 29%, 62%, and 9% among 65 present-day Seklers. The allele frequencies associated with lactase persistence (T_{-13910}) in the groups of ancient, present-day Hungarians and present-day Seklers were 10.9%, 35.9%, and 40%, respectively. Although all 13 classical conquerors had CC_{-13910} genotype, three of the commoners displayed CT_{-13910} (11%) and TT_{-13910} genotypes (22%). The T_{-13910} allele frequency was 28% among the commoners.

Significant difference was found in the C/T_{-13910} genotypes and allele frequencies between the ancient Hungarian conquerors and the present-day Hungarian-

speaking populations, but no significant difference was found between the present-day Hungarian-speaking populations and ancient Hungarian commoners.

Present-day Hungarian-speaking populations exhibit a similar prevalence of the T_{-13910} allele to those in neighbouring countries, such as Austria, the Czech Republic, Slovenia, and Germany. In contrast, the group of all ancient Hungarians displayed a significantly lower prevalence. Although the prevalence of the T_{-13910} allele in the subgroup of ancient commoners was similar to that of the present-day Hungarians, North-west Russians, Austrians, Slovenians, Czechs, and Germans; in the subgroup of classical conquerors it corresponded well with the prevalence of Ob-Ugric present-day populations, such as Khantys, or Maris, and certain Central-Asian and Turkish populations.

The prior mtDNA testing identified six major mtDNA haplogroups (H, U, T, N1a, JT, X) among Hungarian conquerors, six among commoners from the time of the conquest (H, HV, M, R, T, U) and 13 (H, HV, I, J, K, JT, M, R, T, U, V, W, X) among present-day Hungarian-speaking populations. The three ancient samples with LP genotype were all commoners and all displayed haplogroup H, the most common in Europe. Haplogroup N1a and M (indicative of a Near Eastern, Asian origin) were present in 9% and 4% of the ancient Hungarians, but absent or very rare in the present-day Hungarian and Sekler populations. The high prevalence of haplogroup U (23% of the ancient samples), and especially the haplogroup U4, is characteristic of Finno-Ugric populations and populations in south-eastern Europe and western Siberia.

Significant deviation from Hardy Weinberg Equilibrium was calculated in the group of ancient Hungarians ($p<0.001$) and in the subgroups of commoners ($p<0.05$). No significant deviation was found in the present-day Hungarian population ($p>0.10$).

DISUSSION and CONCLUSIONS

ATH is a common genetic enzyme deficiency in humans. Its prevalence and the age at onset vary both geographically and ethnically.

1. Our study has demonstrated that the prevalence of ATH in present-day Hungarian-speaking populations corresponds well with the overall level in Europe. The prevalence of the ATH genotype is 73% in children with ATH-related symptoms, 38% in healthy control children, 39% in present-day Hungarians and 29% in present-day Seklers.

2. The age at onset of ATH in the Hungarian population ranges between 5 and 16 years. Under the age of 5 years ATH is not manifested, except in the case of gastrointestinal infection, where it is manifested above the age of 16 years in all symptomatic individuals and in the majority (71%) of those who are asymptomatic.

3. Our results prove that the C/T_{-13910} variants are strongly associated with ATH phenotypes in Hungarians and correlate well with the results of the HBT. The observed discrepancies between the genotypes and phenotypes may be explained by physiological and pathophysiological factors.

4. We suggest a supplementary use of the genetic and HBT tests for the correct diagnosis and first-stage screening of ATH.

The ability to digest lactose into adulthood is an example of gene-culture co-evolution. It appeared around 5 000−10 000 years ago in those populations that had domesticated their animals previously and used their milk as part of their nutrition. LP developed under strong selection pressure since it meant nutritional, water and electrolyte benefits for the digesters.

1. We found that in bones from the Carpathian Basin from the 10th−11th centuries, the LP allele was present in only 10.9%, and exclusively in bones of the commoners with European mitochondrial haplogroups, who may have

been of pre-Hungarian indigenous ancestry, but not in bones of the classical conquerors.

2. This is despite animal domestication and dairy products having been introduced into the Carpathian Basin early in the Neolithic Age. This anomaly may be explained by the Hungarian use of fermented milk products, their greater consumption of ruminant meat than milk, hierarchical and cultural differences, or by their having other lactase-regulating genetic polymorphisms than C/T_{-13910}.

3. Our results on LP frequency, together with previous studies on mitochondrial DNA haplogroups, prove that the Hungarian conquerors comprised only a small proportion of the inhabitants of the Carpathian Basin at the time of the Hungarian Conquest, though they dominated the region culturally.

4. While the low LP prevalence in ancient Hungarians is similar to that in present-day populations of the Uralic linguistic family and certain Central-Asian and Turkish populations, the significantly higher LP prevalence in present-day Hungarian populations corresponds well with that in neighbouring European populations. This provides additional information on the Asian origin of Hungarians and their genetic assimilation to the European community during 1000 years of turbulent history.

ACKNOWLEDGEMENTS

First I wish to express my special thanks to Professor István Raskó, who encouraged and supported me with his recommendation that I should perform my scientific studies on this topic. His enthusiastic guidance and advice have always been extremely valuable.

I am truly grateful to my colleagues, Erika Bogácsi-Szabó, Gyöngyvér Tömöry and Bernadett Csányi, for introducing me into the laboratory work, for their great help in allowing me to overcome the difficulties and for their discussions of the problems and questions that arose during my work.

My special thanks are due to Professor Ágnes Várkonyi for providing samples from paediatric patients, and to Mrs. Mária Radó and Mrs. Gabriella Lehőcz for their skilled technical assistance in the laboratory work.

I would like to express my gratitude to all of my colleagues at the Genetic Institute of the Biological Research Centre and at the Paediatric Department of the University of Szeged who have contributed to my work in so many ways.

Last, but not least, I thank my mother for her encouraging support, her wise and professional advice and her love throughout my studies and work; and my family for their constant support. Without them I would not have been able to carry out my work.

LIST OF PUBLICATIONS RELATED TO THE Ph.D. THESIS

1. **Prevalence of adult-type hypolactasia as diagnosed with genetic and lactose hydrogen breath tests in Hungarians.**
   Nagy D, Bogácsi-Szabó E, Várkonyi Á, Csányi B, Czibula Á, Bede O, Tari B, Raskó I.
   European Journal of Clinical Nutrition 63:909-912. 2009. IF: 3.072

2. **Comparison of lactase persistence polymorphism in ancient and present-day Hungarian populations.**
   Nagy D, Tömöry G, Csányi B, Bogácsi-Szabó E, Czibula Á, Priskin K, Bede O, Bartosiewicz L, Downes CS, Raskó I.
   American Journal of Physical Anthropology 145:262-269. 2011. IF: 2.756

3. **A felnőtt-típusú laktóz intolerancia előfordulási gyakoriságának diagnosztikai célú molekuláris genetikai vizsgálata gyermek és felnőtt magyar populációban.**
   Bogácsi-Szabó E, Nagy D, Tari B, Várkonyi Á, Bede O, Raskó I.
   Gyermekgyógyászat 57(3):279-286. 2006.

IN EXTENSO LIST OF PUBLICATIONS

1. **Effects of magnesium supplementation on the glutathione redox system in atopic asthmatic children.**
   O Bede, D Nagy, A Surányi, I Horváth, K Gyurkovits.
   Inflammation Research 57:279-286. 2008. **IF:** 1,457

2. **Toxocariasis associated with chronic cough in childhood: a longitudinal study in Hungary.**
   O Bede, Zs Szénási, K Gyurkovits, D Nagy.
   Journal of Helminthology 82:357-363. 2008. **IF:** 1,229

3. **Cystas fibrosis molekuláris biológiai pathomechanizmusa.**
   O Bede, D Nagy. *Medicina Thoracalis* 61(4):166-172. 2008.

4. **Egy beteg - két betegség: cystas fibrosis és coeliakia.**
   O Bede, D Nagy. *Medicina Thoracalis* 61(4):181-184. 2008.

5. **Sóvesztő szindróma vagy cisztás fibrózis?**
   O Bede, D Nagy, C Lakatos, S Túri.
   Gyermekgyógyászat 60(3): 170-173. 2009.

6. **Az asthma bronchiale más megvilágításban.**
   O Bede, D Nagy. *Medicina Thoracalis* 62(5):309-314. 2009.

7. **Az újszülöttkor immunológiája és az eosinophil sejtek.**
   O Bede, D Nagy. *Medicina Thoracalis* 62(5):315-320. 2009.

8. **A diffúz parenchymás tüdőbetegségek patomechanizmusa.**
   O Bede, D Nagy. *Medicina Thoracalis* 63(5):321-335. 2010.

9. **Légúti vírusfertőzések: pathomechanizmus, immunológia.**
   Bede O, Nagy D. *Gyermekgyógyászat* 62(5):321-335. 2011.

10. **Légúti vírusfertőzések klinikai jellegzetességei és epidemiológiája.**
    Nagy D, Bede O. *Gyermekgyógyászat* 63(2):80-82. 2012.