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

Ph.D. thesis

Dóra Nagy M.D.

Tutor

Prof. István Raskó M.D., PhD., DSc.

Paediatric Department

Faculty of Medicine of the University of Szeged

Szeged, Hungary

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.**
   D Nagy, E Bogácsi-Szabó, Á Várkonyi, B Csányi, Á Czibula, O Bede, B Tari, I Raskó.
   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.**
   D Nagy, G Tömöry, B Csányi, E Bogácsi-Szabó, Á Czibula, K Priskin, O Bede, L Bartosiewicz, CS Downes, I Raskó.
   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.

CONTENTS

ABBREVIATIONS

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

1. INTRODUCTION

1.1. HISTORY AND EVOLUTION OF LACTOSE TOLERANCE/INTOLERANCE

Genes and cultures are believed to interact. One candidate example that has been put forward is lactose tolerance (Malström *et al*., 2010).

In most parts of the world, the majority of adults are unable to digest lactose, the main carbohydrate in raw milk. Infants can digest lactose when milk is the most important part of their nutrition, but the lactose digestion capacity declines after the weaning period and results in lactose intolerance/lactase non-persistence (LNP). In others it persists into adult life, causing lactose tolerance/lactase persistence (LP) (Gudmand-Hoyer, 1994).

It was long believed that lactose intolerance was the abnormal trait, and this condition was described as lactase deficiency, because most early research was conducted in countries where lactose tolerance was the most frequent phenotype. However, it was soon recognized that this supposed abnormality was the most frequent trait and the normal mammalian developmental pattern worldwide (Swallow, 2003). A phylogeny study of lactase haplotypes suggested that LP mutations arose only in humans (Holden *et al*., 2002) and evolved within a short time period as a result of natural selection which favoured the advantageous LP phenotype (Bersaglieri *et al*., 2004).

Several hypotheses have been proposed to explain the particular distribution of LP.

The ***gene-culture co-evolution***, also known as the ***cultural-historical hypothesis,*** is the most commonly acknowledged hypothesis. It is presumed that LP is most probably a genetic adaptation to the drinking of milk from domestic livestock. The capacity to digest lactose was a strong selective advantage for adults in pastoralist populations because LP adults were able to derive nutritional benefit from milk, while LNP was nutritionally detrimental (Beja-Pereira *et al.* 2003; Holden *et al*., 2002; McCracken, 1971; Simoons, 1970). It has been proposed that the strong positive selection for the LP phenotype occurred around 5 000−10 000 years ago (Bersaglieri *et al*., 2004). Populations which kept livestock but did not milk them, *e.g*. in China, south-east Asia and parts of sub-Saharan Africa, may not have been selected for LP (Holden *et al*., 2002).

The ***calcium assimilation hypothesis*** suggests that carriers of the LP allele are favoured in high-latitude regions, where sunshine is limited and humans are at risk of vitamin D deficiency, rickets and osteomalacia. Proteins and lactose in fresh milk are assumed to act like vitamin D and promote calcium absorption (Holden *et al*., 2002).

The ***arid climate hypothesis*** speculates that fresh milk as a liquid might be a valuable source of water in desert environments where water and food are scarce, thereby helping to

maintain the electrolyte balance of the body. It is hypothesized that milk-related diarrhoea and consequent water depletion in LNP pastoralists would have caused selection against those individuals (Holden *et al*. 2002).

According to the ***reverse cause hypothesis***, dairying was adopted only in populations with preadaptive high frequencies of the LP allele (Burger *et al*., 2007).

1.2. HUNGARIAN HISTORY

The Hungarian population belongs in the Finno-Ugric language family of the Uralic linguistic family, a diverse group of people. Finno-Ugric speakers include the Saamis (Lapps) in the northern Finn-Scandinavian and Kola Peninsulas, the Erzas, Moksas, Maris, Udmurts and Komis in the northern woodland zone of European Russia, and the Mansis (Voguls) and Khantys (Ostyaks) around the river Ob in western Siberia (Vékony, 2002). Distantly related to the Finno-Ugrians are the various Samoyed peoples of Siberia, the Nenets, Enets, Nganassans and Selkups. After a migration period, Hungarians entered central European history as seven major tribes that invaded the Danubian Basin from across the Carpathians around 895 AD. This was the last in a series of migrations. The Carpathian Basin had been settled for thousands of years before the arrival of Hungarians, by Dacians, Romans, Sarmatians, Goths, Huns, Avars, Slavs and others: it is probable that on the eve of the Hungarian Conquest the overwhelming majority of the indigenous population were Slavic (Tömöry *et al*., 2007).

1.3. CLINICAL MANIFESTATION OF LACTOSE INTOLERANCE

The specific enzyme lactase-phlorizin hydrolase (lactase) occurs on the apical surface of the brush border enterocytes in the small intestine, with the highest expression in the mid-jejunum. Lactase is a large glycoprotein, which catalyses the hydrolysis of the β(1-4)-glycosidic bond in the disaccharide lactose, producing two monosaccharides: β-D-glucose and β-D-galactose (Swallow *et al.*, 2000). These monosaccharides are absorbed and used as energy sources. In the event of lactose intolerance, the activity of lactase is low, resulting in the typical abdominal symptoms (distension, bloating, pain, cramps, flatulence, loose stools or diarrhoea) due to the osmotic load of unhydrolysed lactose and its fermentation by the bacterial flora of the colon, producing short-chain fatty acids, methane, hydrogen and carbon dioxide (Adolfsson *et al.*, 2004).

1.4. CLASSIFICATION AND EPIDEMIOLOGY OF LACTOSE INTOLERANCE

Lactose intolerance can be divided into three categories.

1.4.1. Congenital lactase deficiency (CLD)

CLD is a rare autosomal recessive gastrointestinal disorder characterized by watery diarrhoea starting during the first 10 days of life in infants fed with lactose-containing milk. CLD is the most severe form of lactose intolerance, with an almost total lack (0−2%) of lactase activity on jejunal biopsy (Järvelä *et al*., 1998). The prevalence of CLD in the Finnish population is 1:60 000. CLD is considered to belong to the Finnish disease heritage because of the founder effect and genetic drift (Järvelä *et al*., 1998). However, it has recently also been identified in Turkish and Italian families (Torniainen *et al.*, 2009a).

1.4.2. Adult-type hypolactasia (ATH)/LNP

ATH, also known as primary lactose malabsorption or LNP, is the most common genetic deficiency in humans. The activity of lactase gradually decreases after the weaning period to 5−10% of the original functional capacity (Swallow, 2003).

The age at onset of ATH differs in different populations. While the majority of Thai children become hypolactasic by the age of 2−4 years (Keusch *et al*., 1969), ATH is rarely manifested before the age of 5 years in the Caucasian population. In African children it occurs between 1 and 8 years, in Indian children between 3 and 8 years, *i.e*. much earlier than the range 5−20 years in the Finnish population (Kuchay *et al*., 2011; Rasinpera *et al*., 2004; Sahi *et al*., 1972). The prevalence of ATH varies widely between populations, both ethnically and geographically (Sahi *et al*., 1973). The prevalence of ATH among Caucasian populations in Europe generally varies between 3% and 70% (Sahi, 1994; Swallow, 2003). LP is most prevalent in north-western Europe, with the highest frequency among Swedes and Danes, and there is a decline in frequency as one moves south and west. In general, the frequency in the rest of the world is low. The prevalence of LP is lowest in Asia; it is 0−10% among Thai, Chinese and Japanese people, and ~10−30% in African populations (Gerbault *et al*., 2009). However, LP is frequent in the milk-dependent nomads of the Afro-Arabian desert zone. LNP is a common phenotype in native populations of Australia and America (Swallow *et al*., 2000).

1.4.3. Secondary lactose intolerance

This is a temporary condition due to the destruction of the intestinal epithelium by bowel inflammation, infection or nutritive allergy (Swallow, 2003). The activity of lactase is affected by other disaccharidases of the intestinal epithelium.

1.5. DIAGNOSIS OF LACTOSE INTOLERANCE

1.5.1. Direct test

The measurement of intestinal disaccharidase (lactase, sucrase and maltase) activities is the gold standard for the diagnosis of ATH, though, it is not suitable for routine diagnosis because of its invasiveness. The activities of lactase and other disaccharidases are measured by the method developed by Dahlqvist (1964). ATH is considered to exist when the lactase activity is <10 U/g protein and the ratio lactase/sucrase is <0.3 (Rasinperä *et al.*, 2004). This method is suitable for distinguishing between primary and secondary lactose intolerance.

1.5.2. Indirect tests

The clinical standard is the ***hydrogen breath test*** (HBT). The quantity of exhaled hydrogen produced by the bacterial fermentation of lactose is measured before and at various time intervals after lactose ingestion (Metz *et al*., 1975). The *l**actose tolerance test*** measures the extent of the elevation of the blood glucose level after lactose intake in cases of lactose tolerance (Newcomer *et al*., 1975). ***Isotope tests*** are based on the same method, but can detect ^{13}C-labelled glucose in the serum or ^{13}C-labelled carbon dioxide in the exhaled air (Vonk *et al*., 2000). Many nutritional and individual factors may influence lactose digestion, and indirect tests are therefore sometimes inaccurate (Newcomer *et al*., 1975).

1.6. GENETICS OF LACTOSE INTOLERANCE

1.6.1. Lactase gene and protein

Lactase is encoded by a single gene (LCT) of approximately 50 kilo bases (kb), located on chromosome 2q 21, with a 1-kb long promoter region preceding the LCT. The gene has 17 exons and encodes a messenger RNA (mRNA) transcript of 6 274 nucleotides and a preprotein of 1 927 amino acid residues, containing a putative signal peptide, a large proportion and a mature protein with 2 catalytic sites. The cleavage and protein maturation take place in the endoplasmic reticulum and Golgi complex (reviewed by Swallow, 2003).

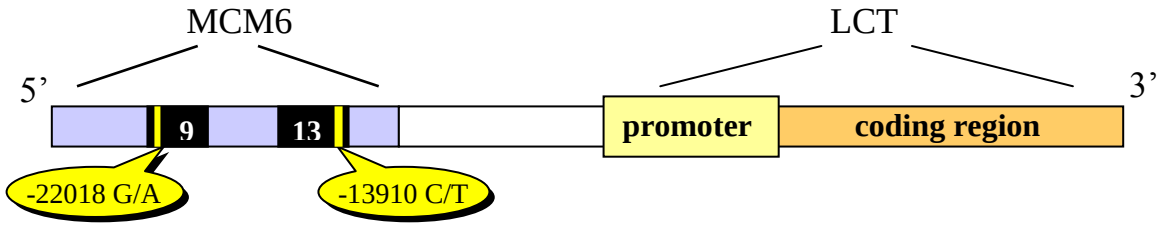
1.6.2. CLD and ATH loci

CLD and ATH are autosomal recessive traits. The CLD locus was previously considered to be at a distance of >2 mega baise pairs (bp) from the LCT on chromosome 2q21 (Järvelä *et al.*, 1998). However, a total of 9 mutations (Y1390X, G1363S, S1666fsX1722, S218fsX224, Q268H, E1612X, R1587H, S688P and V565fsX567) in the coding region of LCT recently found to be associated with CLD in Finnish, Turkish and Italian populations

result in molecular changes in the protein (Kuokkanen *et al*., 2006; Torniainen *et al*., 2009b).

Sequencing of the region immediately upstream of the LCT and also of the exons failed to identify any nucleotide difference that could be causal of the phenotypic polymorphism of ATH (Swallow, 2003). The ATH locus was identified with the aid of linkage disequilibrium and haplotype analysis of Finnish families. Sequencing of a 47 kb region upstream of the human LCT resulted in two single nucleotide polymorphisms (SNPs): a C/T variant (rs4988235) located 13 910 bp upstream and a G/A variant (rs182549) located 22 018 bp upstream of the LCT in introns 13 and 9 of the minichromosome maintenance-6 gene (MCM6) on chromosome 2q21 (Figure 1). The C/T_{-13910} genotypes were completely, while the G/A_{-22018} variants were partially associated with LP/LNP phenotypes (Enattah *et al*., 2002).

FIGURE 1 *Locations of the C/T_{-13910} and G/A_{-22018} variants relative to the LCT*



MCM6: minichromosome maintenance-6 gene; LCT: lactase gene

1.6.3. Enhancer activity of the C/T_{-13910} region

In two independent *in vitro* studies, the functional role of the C/T_{-13910} SNP as a *cis*-acting element in the regulation of LCT expression was characterized (Olds *et al*., 2003; Troelsen *et al*., 2003). Those transfection experiments involved the use of lactase promoter-driven reporter gene constructs combined with short fragments of either the human C_{-13910} or T_{-13910} region in differentiated and undifferentiated intestinal cell lines. The region containing the C/T_{-13910} SNP possessed transcriptional enhancer activity. The T_{-13910} variant activated the lactase promoter-driven expression 2−4-fold in the undifferentiated cell line, and up to 40-fold in the differentiated cells. The C_{-13910} variant exhibited a significantly lower enhancer activity than that of the T_{-13910} variant, with the result that LP was observed in individuals with the TT_{-13910} and CT_{-13910} variants, whereas in those with the CC_{-13910} variant LNP may develop. In these two functional studies the regulatory role of the G/A_{-22018} variants was also examined. The -22 018 region was a weak silencer of the -13 910 enhancer activity, but no difference could be measured between the G_{-22018} and A_{-22018}

variants. This indicated that the G/A_{-22018} SNP might not be functionally involved in the LP/LNP phenotypes.

A transcription factor, Oct-1, was identified to bind more strongly to the T_{-13910}-containing region than to the C_{-13910} region (Lewinsky *et al*., 2005). The relacement of cytosine by thymidine at the -13 910 site seemed to create a stronger Oct-1-binding site. Moreover, binding sites of intestinal transcription factors (GATA-6, HNF4 α, Fox and Cdx-2) were also identified in the C/T_{-13910} region, as a further indication that this region underlies the regulation of lactase expression (Lewinsky *et al*., 2005).

The developmental down-regulation of LCT appears to be regulated at a transcriptional level in humans. However, post-transcriptional and post-translational control mechanisms are also believed to be involved in the decline of lactase activity (Rossi *et al*., 1997).

1.6.4. New mutations in association with ATH

The C/T_{-13910} variant has demonstrated a complete association with LP/LNP phenotypes in Finland (Enattah *et al*., 2002) and strong associations in Europe and in some Asian and American populations (Kuchay et al., 2011; Global Lactase Persistence Association Database - GLAD), but failed to correlate with LP/LNP phenotypes in certain African and Middle Eastern populations (Mulcare *et al*., 2004). Other polymorphisms (G/C_{-14010}; T/G_{-13915} - rs41380347; T/C_{-13913} - rs41456145; C/G_{-13907} - rs41525747; and T/C_{-3712} - rs55660827) in the close vicinity of the C/T_{-13910} variant ha been identified as associated with LP/LNP phenotypes in Sudanese, Tanzanian, Kenyan, Somali, Xhosa and Saudi populations (Enattah *et al*., 2008; Imitiaz *et al*., 2007; Ingram *et al*., 2007; Tishkoff *et al*., 2006; Toriainen *et al*., 2009b). Most of these SNP-containing regions act as enhancers and bind transcriptional factors such as HNF1α or Oct-1 (Ingram *et al*., 2007; Jensen *et al*., 2011; Olds *et al*., 2011; Sibley *et al*., 2011).

1.7. OUR AIMS

The aimes of this study include determinations of the prevalence of LP and LNP genotypes in the Hungarian population, the age at onset of ATH, and the applicability of the HBT in comparison with genetic screening of the C/T_{-13910} SNP in symptomatic and asymptomatic children.

A further aim was an evaluation of 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 the results were analysed relative to those on other populations believed to have been in contact with Hungarians during the migratory period, and after the settlement in the Carpathian Basin.

2. SUBJECTS AND METHODS

2.1. SAMPLE CHARACTERISTICS

Four groups comprising a total of 432 present-day Hungarian-speaking individuals and a group of 42 bone samples from the period of the Hungarian Conquest were analysed.

2.1.1. Present-day samples

Group 1 consisted of 82 persons (46 females, 36 males; mean age 11 years; range: 2-19 years) with LNP-related symptoms after milk consumption (abdominal pain in 57%, diarrhoea in 48%, bloating and flatulence in 53%). 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 (58 females, 46 males; mean age: 11 years; range: 2–20 years) with no abdominal symptoms and no previous gastrointestinal disease.

The subjects in groups 1 and 2 were patients undergoing treatment at the Department of Paediatrics and Child Health Centre at the Faculty of Medicine of the University of Szeged.

The genotype frequencies of C/T_{-13910} variants in present-day Hungarian-speaking populations were established from ***group 3*** (a DNA bank obtained from 181 randomly-selected voluntary Hungarian participants living in different parts of Hungary) and ***group 4*** (65 Seklers, an outlying minority Hungarian-speaking population in eastern Transylvania, supposedly descendants of one of the seven invading tribes).

The subjects and/or their parents gave their informed consent to participation in the study, which was approved by the local ethical committee.

2.1.2. Ancient samples

Forty-two 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, provided by the Archaeological Institute of the Hungarian Academy of Sciences, had been excavated in cemeteries dating from the 10th–11th centuries in different regions of the Carpathian Basin (samples that yielded a genotyping result are shown in Table 1). Both the burial sites and the bones were archaeologically and anthropomorphologically well-defined. Ancient bone samples were classified on the basis of the grave findings. The ancient remains excavated from rich graves, containing a horse skull, harness, arrow- or spear-heads, mounted belts, braided ornaments and earrings, were identified as the bones

TABLE 1 *Anthropomorphologic data on bone samples from the Carpathian Basin*

Sample code	Sample location (country)	Estimated age (century)	Sex	Social status
1	Aldebrő-Mocsáros (HU)	late-10th	F	classical conqueror
2	Besenyőtelek-Szőrhát (HU)	10th	M	classical conqueror
3	Csekej (SK)	10th	nd	nd
4	Eger-Szépasszonyvölgy (HU)	10th	nd	classical conqueror
5	Fadd-Jegeshegy (HU)	late-10th–early-11th	M	commoner
6	Fadd-Jegeshegy (HU)	late-10th	F	commoner
7	Harta-Freifelt (HU)	10th	M	classical conqueror
8	Harta-Freifelt (HU)	10th	F	classical conqueror
9	Harta-Freifelt (HU)	early-10th	F	classical conqueror
10	Kolozsvár (RO)	10th	nd	classical conqueror
11	Magyarhomoróg (HU)	middle-10th	M	commoner
12	Mözs-Szárazdomb (HU)	middle-10th	M	commoner
13	Mözs-Szárazdomb (HU)	middle-10th	F	commoner
14	Mözs-Szárazdomb (HU)	middle-10th	F	commoner
15	Orosháza-Görbics (HU)	10th	F	classical conqueror
16	Őrménykút (HU)	late-10th	M	classical conqueror
17	Őrménykút (HU)	late-10th	F	classical conqueror
18	Sárrétudvar-Hízóföld (HU)	middle-10th	F	classical conqueror
19	Szabadkígyós-Pálliget (HU)	middle-10th	M	classical conqueror
20	Szabadkígyós-Pálliget (HU)	middle-10th	M	classical conqueror
21	Szegvár-Oromdűlő (HU)	early-11th	F	commoner
22	Szegvár-Oromdűlő (HU)	late-11th	F	commoner
23	Zalavár-Kápolna (HU)	11th–12th	M	commoner

HU = Hungary, SK = Slovakia, RO = Romania, F = female, M = male, nd = not determined.

of classical conquerors; while ancient bone samples from graves involving simple burials were identified as the remains of ancient commoners. Based on this archaeological classification, the ancient samples from the 10th−11th centuries are mentioned later as ancient Hungarians, classical conquerors or commoners. All 42 ancient bone samples had previously yielded reproducible mitochondrial DNA (mtDNA). The bone samples were identical to those analyzed by Tömöry *et al*. (2007) (Table 1).

2.2. METHODS

All present-day and ancient samples were subjected to molecular genetic screening for the C/T_{-13910} SNP. The individuals in groups 1 and 2 also took part in the HBT for clinical evaluation. In groups 3 and 4 and the ancient samples, the genetic testing extended to the mtDNA haplogroups.

2.2.1. Lactose HBT and supplementary clinical tests in present-day subjects

Testing took place on an empty stomach early in the morning. The HBT was performed after the oral ingestion of 1 g lactose/body weight kg (max. 50 g) in 10% aqueous solution by the standard method by Metz *et al*. (1975). In the event of a positive test result (≥20 ppm at 120 min and an increase of ≥10 ppm above the basal value within 60-120 min after the lactose load), the HBT was repeated after 2 months.

If a secondary lactase deficiency, such as coeliac disease, was suspected in the background of a positive HBT result, despite the presence of LP genotypes, small intestinal biopsy was obtained to establish the correct diagnosis. Coeliac disease was diagnosed histologically from the intestinal samples of 12 patients who displayed seropositivity for IgA anti-endomysial antibodies and/or IgA anti-gliadin antibodies and tissue transglutaminase. The histological staging was carried out in accordance with the Marsh system: type 1, 2, 3a, 3b or 3c (Oberhuber, 2000).

2.2.2. DNA testing of present-day samples

DNA extraction

Genomic DNA was extracted from the root portion of single hairs or from buccal smears, through use of the Chelex-based method (Walsh *et al*., 1991).

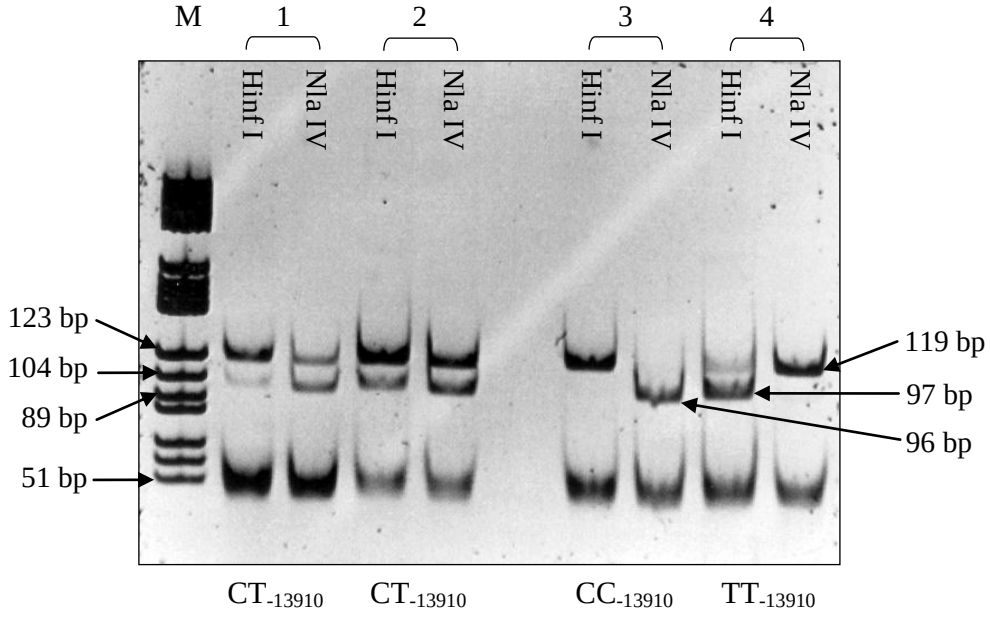
C/T_{-13910} genotyping

The derived cleaved amplified polymorphic sequence method (dCAPS) was applied to analyse the genotypes by using the dCAPS Finder 2.0 program

(http://helix.wustl.edu/dcaps/dcaps.html). 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').

The polymerase chain reaction (PCR) was carried out in a volume of 40 µl containing 1× Ampli Taq Gold Buffer, 6 µM of each of the primers, 200 µM of each of the deoxynucleotide-triphosphates (dNTPs), 2.5 mM $MgCl_2$, 20 ng of DNA extract and 1 U of AmpliTaq Gold Polymerase (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 and 72 °C for 60 sec, and final extension at 72 °C for 5 min.

FIGURE 2 *HinfI and NlaIV enzymatic cleavage of 119 bp PCR products in individual reactions in 4 present-day samples, with CC_{-13910}, TT_{-13910} or CT_{-13910} genotypes*



M: DNA marker (pBR322/HaeIII – 300 ng)

7 µl of 119 bp PCR product was subjected to treatment with the individual enzymes NlaIV and HinfI (Fermentas, Ontario, Canada) in a restriction fragment length polymorphism (RFLP) method, with 5 U of endonuclease and 1× reaction buffer in a reaction volume of 20 µl. The NlaIV cleavage resulted in two fragments (96 bp and 23 bp) in the case of the

C_{-13910} allele, while HinfI digestion in the presence of the T_{-13910} allele resulted in a 97 bp and a 22 bp fragment (Figure 2). For each reaction, a negative control (no template) and three types of positive controls (CC_{-13910}, CT_{-13910} and TT_{-13910}) were included. The enzyme-cleaved PCR products were run on 8% native polyacrylamide gel and visualized by UV transillumination (UVP BioImaging System, Upland, CA, USA) after ethidium bromide staining.

The PCR products of 33 subjects with discrepant genotype and phenotype results were sequenced to confirm or disprove the results of RFLP digestions and to discover other possible mutations in the background of the discrepancy. Sequencing was conducted under BigDyeTM Terminator Version 3.1 cycling conditions. The products were purified by ethanol precipitation and analysed on an Automatic Sequencer 3730xl (Applied Biosystems, California, USA).

Mitochondrial DNA testing

The hypervariable region I (HVSI) of the control region and, if necessary, HVSII and coding regions of the mtDNA were analysed by Tömöry *et al*. (2007), either by sequencing or by PCR-RFLP to elicit the mtDNA haplogroup of the sample. Polymorph positions of the mtDNA were identified by using the revised Cambridge Reference Sequence (Andrews *et al*., 1999). The samples were mtDNA haplogrouped on the basis of the mutational patterns reported by Tömöry *et al*. (2007). All present-day Sekler samples were mitochondrially haplogrouped. Seventy-one of the 181 present-day Hungarian samples were randomly selected for mtDNA testing.

2.2.3. Genetic testing of ancient samples

Ancient samples were processed in accordance with the results of the studies of Kalmár *et al*. (2000), Tömöry *et al*. (2007) and Csányi *et al*. (2008).

Bone powdering and DNA extraction

The surface of the bones (femurs) was washed with diluted bleach and distilled water and then subjected to UV-C irradiation at 1 J/cm^2 for 30 min. A 2 x 3 cm portion was cut from each bone epiphysis and the surface of the portion was first washed with bleach and then removed (at least 2–3 mm deep) with a UV-C-treated sterilized sand disk. The bone portion was next irradiated on each side with UV-C light at 1 J/cm^2 for 30 min, ground into a fine powder by using a mineralogy mill (Retsch MM301; Haan, Germany) and stored in a sterile tube at 4 °C. A bone powder sample (1.3–1.5 g) was suspended in 10 ml of ethylenediaminetetraacetic acid (EDTA) and incubated overnight at 37 °C with continuous

vertical rotation. The samples were centrifuged (2000g for 15 min), EDTA was removed and the sediment was resuspended in 10 ml of EDTA daily for 3–5 days. The sediment was suspended in 1.8 ml of extraction buffer, incubated overnight at 37 °C with continuous vertical rotation, and centrifuged at 12 000 rpm for 10 min. The supernatant, containing the DNA, was stored at -20 °C.

DNA isolation

Standard isolation methods were used as described by Kalmár *et al*. (2000). Alternatively, when needed, a modified method incorporating the DNeasy Blood & Tissue Kit and QIAmp DNA Mini Kit (Qiagen, Valencia, CA, USA) was applied. In this modified method, DNA was isolated from 350 μl of bone extract by treatment with 350 μl of 4 M NH_4-acetate and 700 μl of 96% ethanol at -70 °C for 10 min. The mixture was transferred into a DNeasy Mini spin column and centrifuged at 6 000 g for 1 min. The column was washed twice and the DNA was eluted in a final volume of 40 μl (Tömöry *et al*., 2007).

Mitochondrial DNA testing

All ancient samples had previously been tested for mtDNA haplogroups (Tömöry *et al*., 2007; Tables 1 and 2).

C/T_{-13910} genotyping

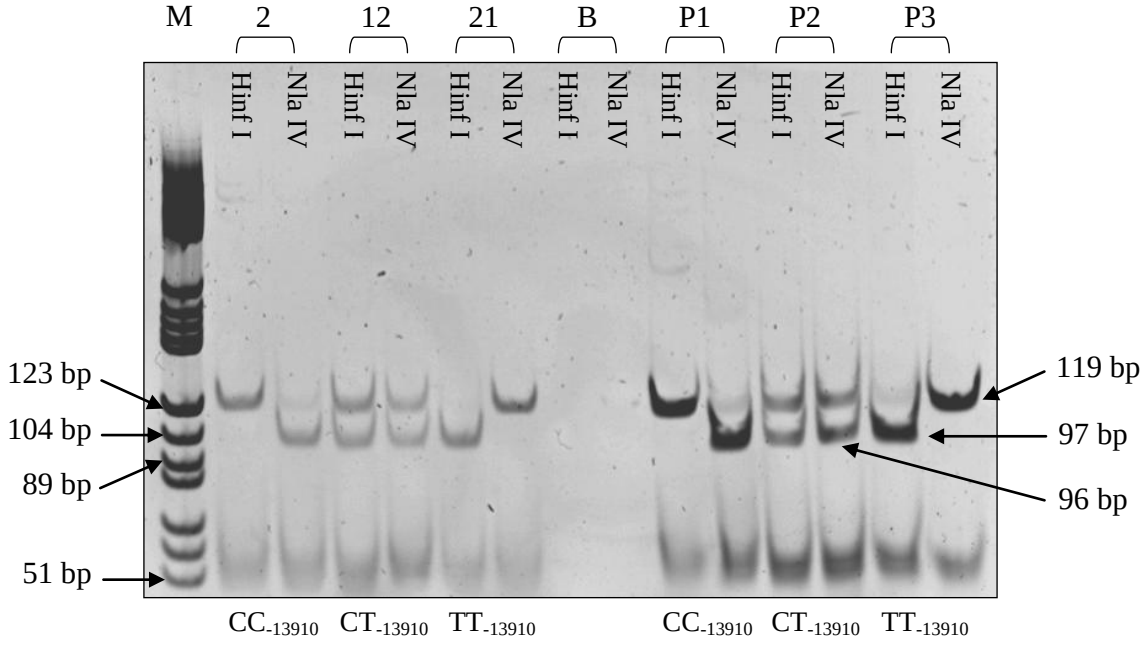
After successful and contamination-free mtDNA amplification, C/T_{-13910} genotyping was carried out on the ancient samples. The primers and restriction enzymes were identical to those used in the reactions of the present-day samples. The standard amplification reaction of the ancient samples contained 10 μl of bone isolate, 200 μM of each of the dNTPs, 25 pmol of each primer, 2.5 mM $MgCl_2$, 1 x Colorless GoTaq Flexi Buffer, and 1.25 U of GoTaq Hot Start Polymerase (Promega, Wisconsin, USA) in a total reaction mixture volume of 50 μl. The amplification protocol was 94 °C for 6 min, 10 cycles of 30 sec at 93 °C, 40 sec at 56 °C and 40 sec at 72 °C, 40 cycles of 30 sec at 93 °C, 40 sec at 54 °C and 40 sec at 72 °C, and a final extension at 72 °C for 5 min. Twenty 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 (Csányi *et al*., 2008; Hanson *et al*., 2005) to enhance the efficiency of the amplification. Ten-μl aliquots of mIPEP products were used in the subsequent C/T_{-13910} genotyping. Three further ancient samples were successfully typed (Table 2). Ten μl of the PCR product was subjected to restriction enzyme digestions as described in connection with the present-day samples (Figure 3).

TABLE 2 *Genetic data on ancient bone samples from the Carpathian Basin*

Sample code	C/T_{-13910}	mtDNA[a]		
	genotype	Haplogroup	Mutations in HVSI	Mutations in HVSII and coding region
1	CC	U4	356C	12308G
2	CC	N1a	147A 172C 183C 189C 223T 248T 320T 355T	+10238HphI
3	CC	J2	069T 093C 126C 193T 265T	+10871MnlI
4	CC[b]	U	259A 311C	-73ApaLI, +7025AluI, 12308G, +14766MseI
5	CC	HV	311C	-73ApaLI, -14766MseI, +7025AluI
6	CC	R	CRS	+73ApaLI -7025AluI 11719A 12308A +14766MseI
7	CC[b]	H	093C 221T	-73ApaLI, -7025AluI
8	CC	T3	126C 248T 292T 294T	10873T
9	CC	H	294T 304C	-73ApaLI, -7025AluI, -14766MseI, 10310G
10	CC[b]	H	189C 295T	-73ApaLI, -7025AluI, 12308A, -14766MseI
11	TT	H	CRS	-73ApaLI, -7025AluI, -14766MseI
12	CT	H	093C 366T	-73ApaLI, -7025AluI, -14766MseI
13	CC	M	223T 311C	10400T
14	CC	T2	126C 148T 218T 294T 304C	
15	CC	N1a	147A 172C 183C 189C 223T 320T 355T	+10238HphI
16	CC	H	CRS	-73ApaLI, -7025AluI, +14766MseI
17	CC	X	183C 189C 223T 278T	+10871MnlI, 14470C
18	CC	T	126C 182C 183C 189C 294T 296T 298C	9 bp del
19	CC	U4	223T 356C	10400C 12308G 12705C
20	CC	JT	126C 163G 186T 189C	+10871MnlI
21	TT	H	CRS	-73ApaLI, -7025AluI, +14766MseI
22	CC	U5a1	114A 192T 256T 270T 294T	12308G
23	CC	U3	311C 343G	12308G

mtDNA = mitochondrial DNA, HVSI = hypervariable region I, HVSII = hypervariable region II, CRS = Cambridge Reference Sequence.
[a] The analysis and classification of the mtDNA results are discussed in detail by Tömöry *et al.* (2007). [b]Only the miPEP+PCR reaction yielded DNA for the C/T_{-13910} reaction.

FIGURE 3 *Gel photo of three ancient and three present-day samples*



M: DNA marker (pBR322/HaeIII – 300 ng); 2, 12 , 21: ancient samples (the ancient sample codes are the same as used in Tables 1 and 2); B: digestion blanks; P1-P3: present-day samples.

Contamination prevention and authentication

To prevent any possible contamination with modern DNA, strict precautions were taken during each step of the ancient sample preparation, as described by Tömöry *et al* (2007). The 11 persons who participated in the sample processing or worked in the laboratories, were mtDNA-tested and C/T$_{-13910}$-genotyped (Table 3). The number of persons involved in the processing was minimized in order to prevent contamination. All steps of sample processing (bone powdering, DNA extraction, preamplification, amplification and post-PCR analysis) were carried out by staff wearing appropriate protective clothing (gloves, face mask, hair net, glasses and laboratory coats) in separate rooms dedicated to ancient DNA work and no other molecular work. All workspaces and appliances were cleaned with bleach and subsequently irradiated with 1 J/cm^2 UV-C light for 2 h before use. All solutions used were filtered and subsequently irradiated with UV-C light for 30 min. During all steps, Universal Fit Filter Tips (Corning Incorporated, Lowell, MA, USA) were used for pipetting. PCR and Eppendorf tubes were sterilized by autoclaving before use. The surface of the bone samples was cleaned and removed as described above, in order to prevent possible contamination.

TABLE 3 *C/T_{-13910} genotypes and mtDNA haplogroups of coworkers*

	C/T_{-13910}	mtDNA[a]		
		HG	Mutations of HVS1	Mutation of HVSII & coding region
R1[b]	TT	H	354T	-73ApaLI, -7025AluI, -14766MseI
R2[b]	CC	K	224C 287T 311C	12308G
R3[b]	CT	J	069T 126C	+10871MnlI
R4[b]	CT	V	298C	72C, -73ApaLI, -4580NheI, -14766MseI
R5[b]	CC	U	146G 180G 342C	+73ApaLI, +10871MnlI, 12308G, 12705C, +14766MseI
Ar1[b]	TT	V	240G, 298C	72C, -73ApaLI, -4580NheI, -14766MseI
Ar2[b]	CT	W	172C 223T 231C 292T	+10871MnlI
Lw1[c]	CC	J	069T 126C	
Lw2[c]	CT	W	172C 223T 231C 292T	+10871MnlI
Lw3[c]	TT	M	223T 234T 288C 298C 327T 359C	-10871MnlI
A[c]	CT	H	274A	-73ApaLI, -7025AluI, -14766MseI

mtDNA=mitochondrial DNA, HG=haplogroup, HVSI=hypervariable region I, HVSII=hypervariable region II, R=researcher, Ar=archaeologist, Lw=lab worker, A=assistant
[a] The analysis and classification of the mtDNA results are discussed in detail by Tömöry *et al*. (2007).
[b] These individuals handled the samples. [c] Present in the labs when the samples were processed.

Bone samples were powdered independently by at least two researchers with different mtDNA haplogroups and C/T_{-13910} genotypes, at least twice each. In each case, two independent DNA extractions were carried out, and at least two successful mIPEP and/or PCR amplifications were performed on each extract in order to assess the reproducibility and authenticity of the results. Only those findings were accepted that involved consistent results in all successful PCR reactions and where the haplogroup and genotype results differed from those of the researchers who analysed 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 the possibility of contaminants entering the process at any stage. Positive controls (CC, CT and TT_{-13910}) were also included in each digestion.

In consequence of DNA degradation, primers were designed to amplify short sequences of templates during mtDNA testing (Tömöry *et al*., 2007) and C/T_{-13910} genotyping (119-bp PCR product).

To confirm the authenticity of the ancient human DNA, DNA was additionally isolated from an ancient horse remain, excavated from one of the human burial sites, and amplified with both horse-specific (forward: 5'-CACCATACCCACCTGACATGCA-3' and reverse: 5'-GCTGATTTCCCGCGGCTTGGTG-3') and the human-specific C/T_{-13910} primers. Only the horse-specific primers yielded amplification product.

It has been proposed (Burger *et al*., 2007) that the allele determination of ancient samples by PCR may be misleading on account of the possibility of allele dropout: the ancient DNA may be so sparse that random distribution of the alleles in the PCR reaction mixture may result in only one allele being present. If this occurs, a heterozygote will be falsely reported as a homozygote. We made efforts to reduce the possible allele dropout. High-quality and -quantity bone powder was used to optimize DNA extraction; preamplification (mIPEP) was applied to increase the quantity of template DNA before the PCR; the number of amplification cycles was increased to reach the limit of detection of the machine without strong artefacts; and the amplification results were derived from several, consistent PCR reactions, as suggested in studies on ancient DNA by Burger *et al*. (1999) and Hummel (2003).

2.2.4. Statistical analysis

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 the unpaired Student's *t*-test with the 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's exact test was applied to compare the HBT findings with the results on the C/T_{-13910} genotypes and to compare the genotypes in the ancient samples with those in the present-day populations. A p value <0.05 was considered to be statistically significant.

Deviation from the Hardy-Weinberg equilibrium was calculated (Rodriguez *et al*., 2009) in the present-day and ancient Hungarian populations as concerns the C/T_{-13910} genotypes.

3. RESULTS

3.1. RESULTS ON GROUPS 1 (PATIENTS) AND 2 (CONTROLS)

3.1.1. C/T_{-13910} genotype 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.

3.1.2. Comparison of HBT and C/T$_{-13910}$ genotype results

The individuals in groups 1 and 2 with different genotypes and HBT results were subdivided into three age groups to evaluate the age of onset of LNP (Table 4). The symptoms of LNP were not manifested before the age of 5 years with the exception of one patient, aged 4, where an intestinal infection enhanced the symptoms.

In case of CC$_{-1310}$ genotypes, among the patients with abdominal complaints, the HBT was positive in 89% of those aged >12 years and in 100% of those aged >16 years; among the controls, the corresponding proportion were 69% and 71%, respectively.

In the event of a discrepancy between the genotype and the phenotype, sequencing was performed. Five of the 33 sequenced samples gave different results from those of RFLP. The correct C/T$_{-13910}$ genotypes, based on sequencing, are presented in Table 4.

In all, 12 patients had type 3 coeliac disease. Five of these 12 patients had the CC$_{-13910}$ genotype and the remaining 7 had the CT$_{-13910}$ genotype. Among the patients with the CC$_{-13910}$ genotype, 1 (20%) had type 3a, 3 (60%) had type 3b and 1 (20%) had type 3c coeliac disease. One CT$_{-13910}$ genotyped patient had type 3a (14%), 4 had type 3b (57%) and 2 had type 3c (29%). Subjects with coeliac disease and negative HBT results were on a gluten-free diet.

In the case of the CC$_{-13910}$ genotype, significant differences in exhaled H_2 levels were measured between groups 1 and 2 at 120 min (52±5.9 ppM vs 27±4.7 ppM, $p<0.01$) in the subgroup aged 5−12 years; and at 60 min (32±8.9 ppM vs 12±2.8 ppM, $p<0.05$), at 90 min (54±11.1 ppM vs 21±4.5 ppM, $p<0.01$) and at 120 min (63±11.3 ppM vs 31±5.9 ppM, $p<0.05$) in the >12-year subgroup. In the cases of the CT$_{-13910}$ and TT$_{-13910}$ genotypes, a significant difference (22±7.2 ppM vs 3±1.6 ppM, $p<0.05$) was observed between groups 1 and 2 at 120 min in the subgroup aged 5−12 years, but no significant difference (19±10.2 ppM vs 4±1.3 ppM, $p=0.20$) was detected above the age of 12 years (Figure 4).

The results of the HBT relative to those of genetic screening in groups 1 and 2 were evaluated by means of Fisher's exact test. The 120-min sampling point of the HBT results proved to be most reliable in both the 5–12-year (group 1: $p<0.01$, group 2: $p<0.001$) and the >12-year subgroup (group 1: $p<0.001$, group 2: $p<0.001$).

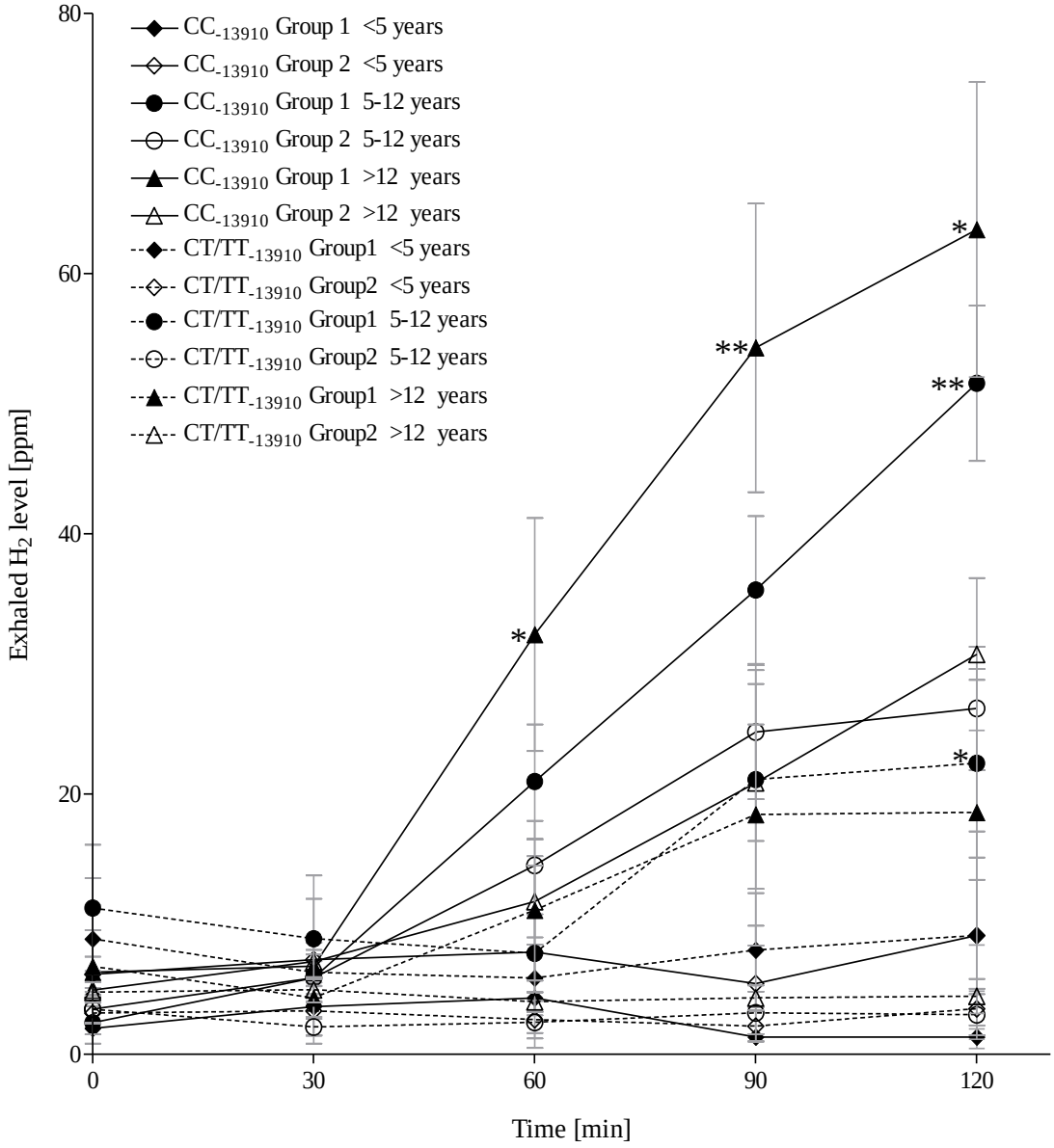
The positive and negative predictive values, and the sensitivity and specificity of the HBT at 120 min in groups 1 and 2, together with the results of previous studies, are presented in Table 5.

TABLE 4 *Comparison of the C/T$_{-13910}$ genotypes with the HBT results in the age-subgroups of group 1 (patients) and 2 (controls)*

	CC$_{-13910}$ genotype				**CT$_{-13910}$ and TT$_{-13910}$ genotypes**			
	Group 1 (n=60)		**Group 2 (n=40)**		**Group 1 (n=22)**		**Group 2 (n=64)**	
	HBT (n (%))				HBT (n (%))			
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Total n	54 (90)[a]	6 (10)	23 (58)	17 (42)	3 (14)[e]	19 (86)[f]	2 (3)[g]	62 (97)
<5 years	0	3 (100)[b]	1 (14)[c]	6 (86)	1 (14)[g]	6 (86)	0	12 (100)
5−12 years	29 (100)	0	11 (65)	6 (35)	1 (20)	4 (80)	1 (5)	21 (95)
>12 years	25 (89)	3 (11)[d]	11 (69)	5 (31)[d]	1 (10)	9 (90)	1 (3)	29 (97)
↳ *12−14 years*	*8 (80)*	*2 (20)*	*6 (86)*	*1 (14)*	*1 (20)*	*4 (80)*	*0*	*11 (100)*
14−16 years	*8 (89)*	*1 (11)*	*0*	*2 (100)*	*0*	*3 (100)*	*0*	*10 (100)*
> 16 years	*9 (100)*	*0*	*5 (71)*	*2 (29)*	*0*	*2 (100)*	*1 (11)*	*8 (89)*

The results for the age-subgroup >12 years are detailed after the arrow. The severity of coeliac disease was classified according to the Marsh system: type 1, 2, 3a, 3b or 3c.
[a] Coeliac disease was diagnosed in 4 subjects aged 6 (type 3c), 7 (type 3b), 16 (type 3b) and 17 (type 3a).
[b] Coeliac disease was diagnosed in 1 subject aged 4 (type 3b).
[c] ATH-related symptoms were enhanced by intestinal infection.
[d] The exhaled H_2 levels were close to the cut-off level (17-19 ppm) in 2 cases in group 1 and in 3 cases in group 2.
[e] Coeliac disease was diagnosed in 1 subject aged 12 (type 3c) and giardiasis in 2 others aged 7 and 14.
[f] Coeliac disease was diagnosed in 6 subjects aged 3 (type 3b), 3 (type 3b), 10 (type 3b), 12 (type 3c), 13 (type 3b) and 17 (type 3a).
[g] Secondary lactose malabsorption developed due to a virus infection.

FIGURE 4 *Mean ± SEM lactose HBT data at 0, 30, 60, 90 and 120 min after lactose intake for the genotypes in groups 1 (patients, n=82) and 2 (controls, n=104)*



Asterisks denote significant differences (* $p<0.05$; ** $p<0.01$) between the group 1 and 2 data.

TABLE 5 *Evaluation of the HBT and genetic screening in groups 1 (patients) and 2 (controls) and in previous studies*

			Number of		Sensitivity	Specificity	Positive PV	Negative PV
			Children	Adults	% (95% CI)		% (95% CI)	
Our study[a]	Total		186	-	77 (68−85)	94 (87−98)	94 (86−98)	78 (69−85)
	Group 1		82	-	90 (79−96)	86 (65−97)	95 (85−99)	76 (55−91)
	Group 2		104	-	58 (41−73)	97 (89−100)	92 (74−99)	78 (69−87)
Newcomer *et al.*[b]	1975	USA	-	50	100	100	-	-
Arola *et al.*[b]	1988	Finland	-	63	69	96	-	-
Högenauer *et al.*[c]	2005	Austria	-	123	75	99	-	-
Büning *et al.*[c]	2005	Germany	-	166	91.4	96.0	98.1	82.8
Schirru *et al.*[c]	2007	Sardinia	-	84	100	96	98	100
Bernardes-Silva *et al.*[c]	2007	Brazil	-	75	100	83	76	100
Krawczyk *et al.*[a]	2008	Germany	-	58	100	95	88	100
Mottes *et a*l.[c]	2008	Italy	43	-	91	55	-	-
Pohl *et al.*[c]	2010	Switzerland	-	194	97	95	90	98
Kuchay *et al.*[d]	2011	India	176	-	97.2	100	100	94.7

PV: predictive value, HBT: H_2 breath test.
[a] HBT results were compared with those of genotyping.
[b] HBT results were compared with the lactase activity of intestinal biopsy samples.
[c] Genotyping results were compared with those of the HBT.
[d] Genotyping results were compared with the lactase activity of intestinal biopsy samples.

3.2. RESULTS ON ANCIENT SAMPLES, AND ON GROUPS 3 (PRESENT-DAY HUNGARIANS) AND 4 (PRESENT-DAY SEKLERS)

3.2.1. C/T$_{-13910}$ genotype results

The genotyping of the C/T$_{-13910}$ autosomal SNP was successful in 23 ancient bone samples (13 classical conquerors, 9 commoners, and 1 not determined). The C/T$_{-13910}$ genotype and the mtDNA haplogroup results and the features of the bone samples are presented in Tables 1 and 2.

The prevalence of the CC$_{-13910}$, CT$_{-13910}$, and TT$_{-13910}$ genotypes among the 23 ancient Hungarians was 87%, 4% and 9% (Table 6); as compared with 39%, 50% and 11% among 181 present-day Hungarians; and 29%, 62% and 9% among 65 present-day Seklers (Table 7). The allele frequencies associated with LP (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 the CC$_{-13910}$ genotype, 3 commoners displayed CT$_{-13910}$ (11%) and TT$_{-13910}$ genotypes (22%) (Table 6). The T$_{-13910}$ allele frequency among the commoners was 28%.

3.2.2. mtDNA haplogroup results

The prior mtDNA testing identified 6 major mtDNA haplogroups (H, U, T, N1a, JT and X) among the Hungarian conquerors, 6 among the commoners from the time of the conquest (H, HV, M, R, T and U) and 13 (H, HV, I, J, K, JT, M, R, T, U, V, W and X) among the present-day Hungarian-speaking populations (Tables 6 and 7). The 3 ancient samples with LP genotype were all commoners and all displayed haplogroup H, which is the most common in Europe (Richards *et al*., 1998). Two of these samples exhibited the TT$_{-13910}$ genotype. Although they both displayed haplogroup H, their mutations in the HVSII and coding regions of the mtDNA were not identical, which excludes their maternal relationship. They were buried in graves in different locations and at different times, which may further indicate that these individuals were not related to each other (Tables 1 and 2). Haplogroup N1a (indicative of a Near-Eastern, Asian origin—*Haak et al*., 2005) and haplogroup M (indicative of an Asian origin—Maca-Meyer *et al*., 2001) were present in 9% and 4% of the ancient Hungarians, but absent or very rare in the present-day Hungarian and Sekler populations (Tables 6 and 7). The high prevalence of haplogroup U (23% of the ancient samples), and especially haplogroup U4, is characteristic of Finno-Ugric populations and populations in south-eastern Europe and western Siberia (Bermisheva *et al*., 2002; Richards *et al*., 1998) (Table 8).

TABLE 6 *Distribution of mtDNA haplogroups and C/T$_{-13910}$ genotypes among ancient Hungarians*

	Number of classical conquerors (%)				Number of commoners (%)				Total number of ancient Hungarians (%)			
	CC$_{-13910}$	CT$_{-13910}$	TT$_{-13910}$	Total	CC$_{-13910}$	CT$_{-13910}$	TT$_{-13910}$	Total	CC$_{-13910}$	CT$_{-13910}$	TT$_{-13910}$	Total
mtDNA haplogroup	13 (100)	0	0	13 (100)	6 (67)	1 (11)	2 (22)	9 (100)	20 (87)	1 (4)	2 (9)	23 (100)
H	4	0	0	4 (31)	0	1	2	3 (33.5)	4	1	2	7 (31)
U	3	0	0	3 (23)[a]	2	0	0	2 (22.5)[b]	5	0	0	5 (23)
T	2	0	0	2 (15)[c]	1	0	0	1 (11)[d]	3	0	0	3 (13)
N1a	2	0	0	2 (15)	0	0	0	0	2	0	0	2 (9)
JT	1	0	0	1 (8)	0	0	0	0	1	0	0	1 (4)
X	1	0	0	1 (8)	0	0	0	0	1	0	0	1 (4)
HV	0	0	0	0	1	0	0	1 (11)	1	0	0	1 (4)
R	0	0	0	0	1	0	0	1 (11)	1	0	0	1 (4)
M	0	0	0	0	1	0	0	1 (11)	1	0	0	1 (4)
J	0	0	0	0	0	0	0	0	1[e]	0	0	1 (4)

mtDNA = mitochondrial DNA.
Further types were identified within haplogroup U: [a]U4 in 2 cases, [b]U3 and U5a1 in 1 case each;
haplogroup T: [c] T3 in 1 case, [d]T2 in 1 case;
haplogroup J: [e]J2 in 1 case; the social status of this sample was not classified.

TABLE 7 *Distribution of mtDNA haplogroups and C/T$_{-13910}$ genotypes among present-day Hungarian-speaking populations*

	Number of present-day Hungarians (%)				Number of present-day Seklers (%)			
	CC$_{-13910}$	CT$_{-13910}$	TT$_{-13910}$	Total	CC$_{-13910}$	CT$_{-13910}$	TT$_{-13910}$	Total
	71 (39)	90 (50)	20 (11)	181 (100)[a]	19 (29)	40 (62)	6 (9)	65 (100)[a]
mtDNA haplogroup	30 (42)	37 (52)	4 (6)	71 (100)[b]	19 (29)	40 (62)	6 (9)	65 (100)[b]
H	13	15	1	29 (41)	6	16	2	24 (36)
U	0	6	0	6 (8)[c]	2	8	1	11 (16.5)[d]
T	3	2	0	5 (7)[e]	4	6	1	11 (16.5)[f]
J	5	3	0	8 (11)[g]	1	3	1	5 (8)[h]
K	4	3	0	7 (10)	3	4	0	7 (11)
V	2	1	1	4 (6)	0	0	0	0
HV	2	0	1	3 (4.2)	2	0	0	2 (3)
W	0	3	0	3 (4.2)	0	1	1	2 (3)
R	1	2	0	3 (4.2)	0	0	0	0
X	0	2	0	2 (3)	1	0	0	1 (2)
M	0	0	1	1 (1.4)	0	0	0	0
JT	0	0	0	0	0	1	0	1 (2)
I	0	0	0	0	0	1	0	1 (2)

mtDNA = mitochondrial DNA.

[a] Number of all C/T$_{-13910}$ genotyped samples, [b] Number of all mtDNA-tested and C/T$_{-13910}$-genotyped samples. Further types were identified within haplogroup U: [c] U5, U5a and U5b in 1 case each and U4 in 2 cases, [d] U3, U4 and U5a1a in 1 case each and U5a1 in 5 cases; haplogroup T: [e] T1 and T3 in 1 case each and T2 in 2 cases, [f] T2, T2b, T3 and T5 in 1case each and T1a in 5 cases; haplogroup J: [g] J1 and J1a in 1 case each and [h] J2 in 1 case.

TABLE 8 *Characteristics of mitochondrial DNA haplogroups present in our samples*

Haplogroups	Characteristic of	Appearance (years ago)	Overall frequency in Europe (%)	Reference
HV *	South-eastern Europe, Mediterranean			Peričić *et al*., 2005
H *	Europe, Near East	25−30 000	40-60	Richards *et al*., 1998
V *	North-western Europe, Iberian Peninsula	12−13 000		Torroni *et al*., 1998
JT *	Southern Europe, Near East, Russia			Richards *et al*., 1998
T *	Central Mediterranean	46 500	8	Peričić *et al*., 2005
J *	Eastern Mediterranean	28 000	11	Richards *et al*., 1998
J1	Central, northern Europe, Mediterranean (J1a), Near East (J1b)			Richards *et al*., 1998
J2	Turkey, Italy, Iberian Peninsula, Iceland			Richards *et al*., 1998
U *	South-eastern Europe (Balkan), Middle East, India	>50 000	14	Helgason *et al*., 2001
U3	South-eastern Europe			Richards *et al*., 1998
U4	Finno-Ugric populations	16−25 000		Bermisheva *et al*., 2002
U5	North-western (U5a1) and southern Europe (U5a), Lapps (U5b)	50 000	7	Richards *et al*., 1998
K	Balkans	15 000	7	Richards *et al*., 1998
I *	North-western Europe, Near East, Caucasus, Balkans	35 000	2	Richards *et al*., 1998
N1a	Near East/Middle East		0.2	Haak *et al*., 2005
W	Northern and southern Europe	18 500	1	Richards *et al*., 1998
X	Everywhere	24 000	2	Pakendorf *et al*., 2005
M	Asia	60−65 000		Maca-Meyer *et al*., 2001

* Haplogroups developed from haplogroup R.

Haplogroups H, U, J, T, K, V, N1, X and W comprise 95% of the European haplogroups (Haak *et al*., 2006). The European haplogroups correspond well with those of Near-Eastern populations. More than 80% of the European lineage (haplogroups U, J, T and R) already existed in the Palaeolithic Age (Semino *et al*., 2000), while less than 20% (haplogroups J1b, J2a, T1a, R1, R2 and N1a) was introduced from the Near East during the Neolithic Age (Torroni *et al*., 2006).

Significant differences in C/T_{-13910} genotypes and allele frequencies were found between the ancient Hungarian conquerors and the present-day Hungarian-speaking populations, but no significant difference was observed between the present-day Hungarian-speaking populations and ancient Hungarian commoners (Table 9).

TABLE 9 *Comparison of the distributions of C/T_{-13910} genotypes and allele frequencies in present-day and ancient Hungarian-speaking populations*

	Present-day		
	Hungarians	Seklers	Classical conquerors
All ancient Hungarians			
Genotype	$p<0.0001$	$p<0.0001$	-
Allele frequency	$p=0.0004$	$p=0.0002$	-
Classical conquerors			
Genotype	$p<0.0001$	$p<0.0001$	-
Allele frequency	$p<0.0001$	$p<0.0001$	-
Commoners			
Genotype	$p=0.1617$	$p=0.0537$	$p=0.0545$
Allele frequency	$p=0.6169$	$p=0.4399$	$p=0.0079$

$p<0.05$ is considered to be statistically significant.

Table 10 presents the frequencies of the T_{-13910} allele and digesters in present-day populations of the Uralic linguistic family and populations which, according to historical accounts (Vékony, 2002), might well have been in contact with ancient Hungarians during their westward migrations, and after the settlement in the Carpathian Basin. 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. The prevalence of the T_{-13910} allele in the subgroup of ancient commoners was similar to those 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 present-day Ob-Ugric populations, such as the Khantys or Maris, and certain Central-Asian and Turkish populations (Table 10).

TABLE 10 *Frequencies of the T_{-13910} allele, lactose digesters and non-digesters in Eurasian populations*

Population (Country and/or region)	**N**	**T_{-13910} allele (%)**	**N**	**LP* (%)**	**References**
Ancient Hungarians (Carpathian Basin)	23	10.9	23[a]	13	present study
Classical conquerors	13	0	13[a]	0	present study
Commoners	9	28	9[a]	33	present study
Present-day Hungarians (Hungary)	181	35.9	535	63	present study, Czeizel *et al.*, 1983
Present-day Seklers (Romania)	65	40	65[a]	71	present study
Present-day populations of the Finno-Ugric (Uralic) linguistic family					
Ob-Ugric speakers (Russia)	62	3.2	62[a]	6.5	Enattah *et al.*, 2007
Khantys (Russia)	195[b]	10.7	195[c]	19.6	Lember *et al.*, 1995, Kozlov *et al.*, 1998
Mansis (Russia)	81[b]	15	81	28	Kozlov *et al.*, 1998
Maris (Russia)	207[b]	10.1	207[c]	19.9	Kozlov *et al.*, 1998
Komis (Russia)	10	15	168[c]	45.7	Enattah *et al.*, 2007, Kozlov *et al.*, 1998
Erzas (Russia)	30	26.7	46	48	Enattah *et al.*, 2007, Kozlov *et al.*, 1998
Mordvin Mokshas (Russia)	30	28.4	27	48	Enattah *et al.*, 2007, Kozlov *et al.*, 1998
Udmurts (Russia)	30	33.4	105[c]	46.4	Enattah *et al.*, 2007, Kozlov *et al.*, 1998
Nenets (Russia)	99[b]	12	99	22	Kozlov *et al.*, 1998
Finns (Finland)	1900	58.7	638	83	Anthoni *et al.*, 2007, GLAD
Estonians (Estonia, Russia)	314	51.4	749[c]	72.7	Lember *et al.*, 2006, Kozlov *et al.*, 1998
Saamis (Lapps) (Finland, Russia)	30	16.7	50	52	Enattah *et al.*, 2007, Kozlov *et al.*, 1998
North-west (Russia)	231	39.2	47	51.1	Khabarova *et al.*, 2009, GLAD
Present-day populations that came into contact with Hungarians during the migratory period and after the settlement in the Carpathian Basin					
Turks (Turkey)	49	3	470	28.7	GLAD, Flatz *et al.*, 1986
Adygei, Caucasus (Russia)	17	11.8	17[a]	22.2	Bersaglieri *et al.*, 2004
Kazakhs (China)	94	4.79	195	23.6	Sun *et al.*, 2007
Mongols (China)	82	2.44	198	12.1	Sun *et al.*, 2007
Uzbekistanis (Uzbekistan)	18	0	-	0	Gerbault *et al.*, 2009
Uzbeks (Afghanistan)	38	8	16	0	GLAD
Uygurs (China)	10	5	10[a]	9.8	Bersaglieri *et al.*, 2004
Daghestanis, Druss (Russia)	17	11.8	17[a]	23.5	Enattah *et al.*, 2007
Daghestanis, mixed (Russia)	23	11.9	23[a]	17.4	Enattah *et al.*, 2007
Daghestanis, Nogs (Russia)	20	12.5	20[a]	25	Enattah *et al.*, 2007
Austrians (Austria)	123	47.2	633[c]	76.2	Högenauer *et al.*, 2005, GLAD
Plzan Czechs (Czech Republic)	200[b]	40.8	200	65	Gerbault *et al.*, 2009
Former Czechoslovakians	17[b]	57.6	17	82	GLAD
Germans (Germany)	30	56	1865[c]	85.2	GLAD
Slovenians (Slovenia)	153[b]	28.6	153	49	Gerbault *et al.*, 2009
Ukrainians (Ukraine, Russia)	46	22	142	77	GLAD, Kozlov *et al.*, 1998

LP = lactase persistence phenotype (digesters), GLAD = Global Lactase Persistence Association Database.
*Frequency of phenotypes obtained by HBT and/or glucose tolerance test and/or measurement of lactase activity from intestinal biopsy. [a] Frequency of phenotypes predicted from the frequency of C/T_{-13910} alleles. [b] Frequency of T_{-13910} allele calculated from phenotypic data. [c] Weighted average of summarized data.

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

4. DISCUSSION

4.1. CLINICAL EVALUATION OF C/T$_{-13910}$ GENETIC TESTING

The 39% incidence of ATH in the Hungarian population corresponds to the overall level in Europe (Sahi, 1994); it is rarely manifested at <5 years. The coincidence of the CC$_{-13910}$ genotype with a positive HBT increases with age. The symptoms develop at around the age of 12 years; all those aged >16 years in group 1 had ATH-related symptoms, whereas in the control group the coincidence never reached 100%.

The CC$_{-13910}$ genotype displayed a close association with a positive HBT in group 1, whereas the LP genotypes correlated better with a negative HBT in group 2. The patients had significantly higher H_2 levels than the controls from 60 min (Figure 4), presumably as a result of a lactase deficiency enhancement, primarily because of additional diseases that do not exist in group 2.

Despite introducing sequencing as an independent genotyping method, we were not able to clarify the discrepancies between the genotypes and the HBT results, except in 5 cases. The genotyping error rate of the RFLP in our study is 15%, which is concordant with previous results (Hosking *et al*., 2004; Hüebner *et al*., 2007). The discrepancy between the CC$_{-13910}$ genotype and the negative HBT results at >12 years may be explained by an abnormal colon bacterial metabolism or glucose metabolism, slow intestinal transit or hypolactasia later in life. The positive HBT results conflicting with the LP genotypes may be a consequence of accelerated intestinal transit, secondary lactose malabsorption because of inflammation, coeliac disease, giardiasis, bacterial or viral infections, a carrier status of a congenital lactase deficiency (Järvelä *et al*., 1998) or other genetic factors (Enattah *et al*., 2008). Nevertheless, with sequencing we were not able to detect any other SNP responsible for the ATH-related symptoms in the close vicinity of the C/T$_{-13910}$ variant. With regard to the sensitivity, specificity, and positive and negative predictive values of the HBT, our results more or less correspond to those in previous papers. The slight differences between our own and the earlier data might be explained by the differences in the HBT methodology, the number, the nationality or the age of the subjects, and the analysis procedure (Table 5).

The HBT has several disadvantages: it is time-consuming, a lactose intake may cause symptoms, physiological and pathological factors may influence the result, and there can be a lack of distinction between primary and secondary lactose malabsorption. The DNA genotyping is exact, but provides no information on the manifestation of the symptoms after a lactose intake; it is suitable for the prediction of ATH even in childhood and for dietary intervention. We suggest supplementary use of the two tests to attain a correct diagnosis.

4.2. HISTORICAL AND ARCHAEOLOGICAL EVALUATION OF C/T$_{-13910}$ GENETIC TESTING

We succeeded in genotyping the C/T$_{-13910}$ autosomal SNP in an unprecedented number of ancient bone samples. Strict precautions were taken to ensure contamination-free ancient DNA work. The steps in the DNA analyses were repeated several times in independent experiments in order to achieve authentic and reproducible results. To interpret our results correctly, we favoured a multidisciplinary approach, including data on the genetic testing of mtDNA and Y chromosomes, Hungarian history, and direct and indirect evidence from archaeology and ethnography.

The ability to digest lactose into adulthood is one of the traits which developed in mankind under strong selection pressure (Bersaglieri *et al*. 2004). 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. While the arid climate hypothesis may be valid in desert areas, it is unlikely to explain the high frequency of LP in northern Europe. In these regions, the calcium assimilation hypothesis is widely accepted (Gerbault *et al*., 2009). The reverse cause hypothesis has been rejected by most studies (Burger *et al*., 2007; Holden *et al*., 2002).

The gene-culture co-evolution hypothesis (Holden *et al.* 2002; McCracken, 1971; Simoons, 1970) postulates that the selection force was the nutritional benefit in those nomad pastoral populations which drank milk prior to the introduction of the LP allele, and that the population migrations led to LP demonstrating a special geographic distribution. This hypothesis was supported by the study of Beja-Pereira *et al*. (2003), who observed a high allelic diversity in cattle milk protein in those regions of Europe where the incidence of LP was high and large herds were kept for the high milk yield.

Ruminants had already been domesticated in the Near East by 8000 BC and this process subsequently developed relatively quickly. The earliest use of milk, assessed via the $\delta^{13}C$

values of fatty acids, was detected in pottery vessels from around 7000 BC in the Near East and south-eastern Europe (Evershed *et al*., 2008). Domestic animals were exploited for milk to different degrees throughout the Neolithic Age, depending on cultural, economic and environmental factors. Though only small-scale dairying existed between 6000 and 4000 BC, it increased in the period 4000−3000 BC (Craig *et al*., 2005). Beja-Pereira *et al.* (2003) conclude that the ability to consume milk co-evolved with cattle dairying. It is debatable, however, whether Early Neolithic farmers had the necessary genetic adaptation to be able to digest lactose in fresh milk. The study by Burger *et al.* (2007) indicated that the LP T_{-13910} allele was not observed in human remains from Neolithic and Mesolithic sites (5840−2267 BC), whereas it was in one Mediaeval sample (400−600 AD).

The haplotype analysis by Enattah *et al*. (2007) identified 9 LP haplotypes in the global populations. The T_{-13910} variant was found in two highly divergent haplotype backgrounds. The T_{-13910} allele arose twice, once ~5000−12000 years ago in the Caucasian region in the haplotype which is now dominant in Europe, and once, ~1400–3000 years ago in a region north of the Caucasus and west of the Urals. This later independent introduction of the T_{-13910} allele is consistent with the low prevalence of this allele among ancient Hungarians.

Itan *et al*. (2009) used a demic computer simulation model to explore the spread of European LP, and concluded that it originated in central Europe, in a region including modern Hungary, about 5500 BC. The molecular evidence then is in favour of gene-culture coevolution, and not of the reverse-cause hypothesis.

How do our results fit into this picture? Present-day Hungarian-speaking populations and ancient commoners exhibit a similar prevalence of the T_{-13910} allele to those in neighbouring countries, such as Austria, The Czech Republic, the Ukraine and Slovenia. In contrast, the ancient classical conquerors displayed a significantly lower prevalence, which corresponds well with those of present-day populations of the Uralic linguistic family, such as the Khantys, Mansis and Maris, and certain Central-Asian and Turkish populations (Table 10). This is consistent with the original Hungarian invaders having roots far to the east of modern Hungary being a minority in the Carpathian Basin, and further diluted in the subsequent turbulent history of that area. This is supported by previous Hungarian studies on mitochondrial haplogroups and the Y chromosome (Csányi *et al*., 2008; Tömöry *et al*., 2007), which showed the genetic assimilation of the present-day Hungarians with their geographical neighbours, but also a significant Asian influence on the genetics of the Hungarian conquerors, from their Siberian origins and from the genetic effects of those

populations with whom the ancient Hungarians came into contact during their westward migrations, such as the Kazars, Petchenegs, Bolgars, Savirs and Iranians.

Among the archaeological remains from the period of the Hungarian conquest, the absence of the T_{-13910} allele in the classical conquerors, but not in the commoners, might suggest a hierarchical difference in nutrition. It might also be due to some of the commoners being descendants of the pre-existing inhabitants of the Carpathian Basin (Romans, Goths, Huns, Avars, Slavs and Germans) who lived a settled life and had a different dairy culture from that of the Hungarian conquerors.

There is some evidence for a change in stock-raising practices at the time of the Hungarian conquest. In archaeological sites in the Carpathian Basin from the Sarmatian (1st–4th centuries AD), Slavic and Avar (5th–9th centuries AD) and Hungarian conquest period (9th–10th centuries AD), cattle remains predominate in animal bone assemblages, but during the Hungarian conquest a small increase occurred in the proportion of horse bones and a small decrease in the proportion of cattle bones. The fragmentation of the bones indicated a butchering process; the change could therefore be due to the differential consumption of horse and cattle meat (Bartosiewicz, 2003). Nevertheless, milk products must have been renewable sources of food for the ancient Hungarian pastoralists. Mobile pastoralists collected milk from their herds, including mares (Bartosiewicz, 2003; Outram *et al*., 2009).

The absence of adult lactose tolerance in the ancient Hungarians we have studied is compatible with their milking their herds, given milk fermentation (Ingram *et al*., 2009; Myles *et al*., 2005). While fresh milk was the basic dairy food in Scandinavia, a situation in accord with the cold climate and good sanitation, in south-eastern Europe and in south-western Asia processed foods prepared from soured milk were preferred. Milk was less appetizing as a fresh warm beverage than as fermented foods, such as a wide range of fresh or ripened cheese, butter, curd, kashk, kishk, doogh, ghee, dahi, kefir or koumiss (Kosikowski, 1981; Outram *et al*., 2009).

The lactose content of fresh milk is 4.42–5.15 g/g% in cattle (Cerbulis *et al*., 1974; Miglior *et al*., 2006), 4.66–4.82 g/g% in goats (Baldi *et al*., 2002; Contreras *et al*., 2009), 4.57–5.40 g/g% in sheep (Addis *et al*., 2005; Fuertes *et al*., 1998), and 6.91–7.04 g/g% in horses (Caroprese *et al*., 2007). Since the lactose content can be reduced by 50–60% by bacterial fermentation (Kilara *et al*., 1975), processed milk products have no or low lactose contents (range 0–3.7 g/g%: 3.7 g/g% in buttermilk; 3.6 g/g% in ropy milk; 3.5 g/g% in kefir; 2.6 g/g% in acidophilus milk; 2.4 g/g% in yogurt; 3 g/g% in cream; 1.4 g/g% in cottage

cheese; and only traces of lactose in all other cheeses). Lactose malabsorbers reported fewer or no symptoms after consuming fermented milk products (Alm, 1982; Savaiano *et al*., 1987).

Such processing of milk is of considerable antiquity, to judge by the archaeological evidence. The presence of degraded milk fats (high abundances of $C_{16:0}$ and $C_{18:0}$ fatty acids) and lipid pyrolysis products (mid- and long-chain ketones) in some archaeological ceramics suggests that the dairy products were heated, perhaps as part of their processing (Craig *et al*., 2005; Evershed *et al*., 2008). This is supported by the fact that a high frequency of ruminant milk lipids has been detected from such ceramics, whereas raw milk lipids (high abundances of $C_{4:0}$ to $C_{12:0}$ fatty acids) are rapidly destroyed by burial (Copley *et al*., 2003; Evershed *et al*., 2008). Hungarian pastoralists, averse to drinking fresh milk, would have had little or no selection for LP in earlier millennia.

Alternatively, the low prevalence of European-type (C/T_{-13910}) LP in ancient Hungarians may be due to their having one or more non-European polymorphisms in the lactase regulatory regions. Recent studies (Enattah *et al*., 2008; Ingram *et al*., 2007; Itan *et al*., 2010) have revealed other alleles that can also produce LP in South and East-African, Saudi and other Middle-Eastern populations. Some cases of undoubted LP have not yet been identified with any known allele in the control regions sequenced (Itan *et al*., 2010). It is therefore possible that ancient Hungarians were lactose tolerant, despite their lack of the characteristic European allele. In view of the fairly high proportion of European Y chromosome and mitochondrial haplogroups among Hungarian conquerors (Csányi *et al*., 2008; Tömöry *et al*., 2007), that is perhaps unlikely; and in view of our ignorance of the non-European allele that might be involved, it is premature to search for it in ancient Hungarian samples.

5. 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.

6. REFERENCES

1. **Addis M**, Cabiddu A, Pinna G, *et al*. Milk and cheese fatty acid composition in sheep fed Mediterranean forages with reference to conjugated linoleic acid *cis*-9, *trans*-11. *J Dairy Sci* 88:3443–3454. 2005.
2. **Adolfsson O**, Meydani SN, Russell RM. Yogurt and gut function. *Am J Clin Nutr* 80:245–256. 2004.
3. **Alm L**. Effect of fermentation on lactose, glucose, and galactose content in milk and suitability of fermented milk products for lactose intolerant individuals. *J Dairy Sci* 65:346–352. 1982.
4. **Andrews RM**, Kubacka I, Chinnery PF, *et al*. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23(2):147. 1999.
5. **Anthoni SR**, Rasinperä HA, Kotamies AJ, *et al*. Molecularly defined adult-type hypolactasia among working age people with reference to milk consumption and gastrointestinal symptoms. *World J Gastroenterol* 13(8):1230–1235. 2007.
6. **Arola H**, Koivula T, Jokela H, *et al*. Comparison of indirect diagnostic methods for hypolactasia. *Scand J Gastroenterol Suppl* 23:351–357. 1988.
7. **Baldi A**, Modina S, Cheli F, *et al*. Bovine somatotropin administration to dairy goats in late lactation: effects on mammary gland function, composition and morphology. *J Dairy Sci* 85:1093–1102. 2002.
8. **Bartosiewicz L**. A millennium of migrations: proto-historic mobile pastoralism in Hungary. *Bull Fla Mus Nat Hist* 44(1):101–130. 2003.
9. **Beja-Pereira A**, Luikart G, England PR *et al*. Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nat Genet* 35(4):311–313. 2003.
10. **Bermisheva M**, Tambets K, Villems R, *et al*. Diversity of mitochondrial DNA haplotypes in ethnic populations of the Volga-Ural region of Russia. *Mol Biol (Mosk)* 36:990–1001. 2002.
11. **Bernardes-Silva CF**, Pereira AC, Fátima Alves da Mota G, *et al*. Lactase persistence/non-persistence variants, C/T_{-13910} and G/A_{-22018}, as a diagnostic tool for lactose intolerance in IBS patients. *Clinica Chimica Acta* 386:7–11. 2007.
12. **Bersaglieri T**, Sabeti PC, Patterson N, *et al*. Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* 74:1111–1120. 2004.

13. **Burger J**, Hummel S, Herrmann B, *et al*. DNA preservation: A microsatellite-DNA study on ancient skeletal remains. *Electrophoresis* 20:1722–1728. 1999.
14. **Burger J**, Kirchner M, Bramanti B, *et al*. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *PNAS* 104(10):3736–3741. 2007.
15. **Büning C**, Genschel J, Jurga J, *et al*. Introducing genetic testing for adult-type hypolactasia. *Digestion* 71(4):245–250. 2005.
16. **Caroprese M**, Albenzio M, Marino R, *et al*. Behavior, Milk yield, and milk composition of machine and hand-milked murgese mares. *J Dairy Sci* 90:2773–2777. 2007.
17. **Cerbulis J**, Farrell Jr. HM. Composition of milks of dairy cattle. I. Protein, lactose, and fat contents and distribution of protein fraction. *J Dairy Sci* 58(6):817–827. 1974.
18. **Contreras A**, Paape MJ, Miller RH, *et al*. Effect of bromelain on milk yield, milk composition and mammary health in dairy goats. *Trop Anim Health Prod* 41:493–498. 2009.
19. **Copley MS**, Berstan R, Dudd SN, *et al*. Direct chemical evidence for widespread dairying in prehistoric Britain. *PNAS* 100(4):1524–1529. 2003.
20. **Craig OE**, Chapman J, Heron C, *et al*. Did the first farmers of central and eastern Europe produce dairy foods? *Antiquity* 79:882–894. 2005.
21. **Czeizel A**, Flatz G, Flatz SD. Prevalence of primary adult lactose malabsorption in Hungary. *Hum Genet* 64:398–401. 1983.
22. **Csányi B**, Bogácsi-Szabó E, Tömöry G, *et al*. Y-chromosome analysis of ancient Hungarian and two modern Hungarian-speaking populations from the Carpathian Basin. *Ann Hum Genet* 72:519–534. 2008.
23. **Dahlqvist A**. Method for assay of intestinal disaccharidases. *Anal Biochem* 57:18–25. 1964.
24. **Enattah NS**, Jensen TG, Nielsen M, *et al*. Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am J Hum Genet* 82:57–72. 2008.
25. **Enattah NS**, Sahi T, Savilahti E, *et al*. Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 30:233–237. 2002.
26. **Enattah NS**, Trudeau A, Pimenoff V, *et al*. Evidence of still-ongoing convergence evolution of the lactase persistence T_{-13910} alleles in humans. *Am J Hum Genet* 81:615–625. 2007.

27. **Evershed RP**, Payne S, Sherratt AG, *et al*. Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature* 455:528–531. 2008.

28. **Flatz G**, Henze HJ, Palabiyikoglu E, *et al*. Distribution of the adult lactase phenotypes in Turkey. *Trop Geogr Med* 38(3):255–258. 1986.

29. **Fuertes JA**, Gonzalo C, Carriedo JA, *et al*. Parameters of test day milk yield and milk components for dairy ewes. *J Dairy Sci* 81:1300–1307. 1998.

30. **Gerbault P**, Moret C, Currat M, *et al*. Impact of selection and demography on the diffusion of lactase persistence. *PLoS ONE* 4(7): e6369. doi:10.1371/ journal.pone.0006369. 2009.

31. **Global Lactase Persistence Association Database**: GLAD web site: http://www.ucl.ac.uk/mace-lab/GLAD

32. **Gudmand-Hoyer E**. The clinical significance of disaccharide maldigestion. *Am J Clin Nutr* 59:735S–741S. 1994.

33. **Haak W**, Forster P, Bramanti B, *et al*. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310:1016–1018. 2005.

34. **Hanson EK**, Ballantyne J. Whole genome amplification strategy for forensic genetic analysis using single or few cell equivalents of genomic DNA. *Anal Biochem* 346:246–257. 2005.

35. **Helgason A**, Hickey E, Goodacre S, *et al*. mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* 68:723–737. 2001.

36. **Holden C**, Mace R. Pastoralism and the evolution of lactase persistence. In: Leonard WR, Crawford MH, editors. The Human Biology of Pastoral Populations. Cambridge: Cambridge University Press pp 280–307. 2002.

37. **Högenauer C**, Hammer HF, Mellitzer K, *et al*. Evaluation of a new DNA test compared with the lactose hydrogen breath test for the diagnosis of lactase non-persistence. *Eur J Gastroenterol Hepatol* 17:371–376. 2005.

38. **Hosking L**, Lumsden S, Lewis K, *et al*. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. *Eur J Hum Genet* 12:395–399. 2004.

39. **Hummel S**. DNA markers. In: Ancient DNA typing: methods, strategies and applications. Berlin, Heidelberg: Springer-Verlag pp 19–41. 2003.

40. **Hüebner C**, Petermann I, Browning BL, *et al*. Triallelic single nucleotide polymorphisms and genotyping error in genetic epidemiology studies: MDR1

(ABCB1) G2677/T/A as an example. *Cancer Epidemiol Biomarkers Prev* 16(6):1185–1192. 2007.

41. **Imtiaz F**, Savilahti E, Sarnesto A, *et al*. The T/G-13915 variant upstream of the lactase gene (LCT) is the founder allele of lactase persistence in an urban Saudi population. *J Med Genet* 44(10):e89. doi:10.1136/jmg.2007.051631. 2007.

42. **Ingram CJ**, Elamin MF, Mulcare CA, *et al*. A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Hum Genet* 120(6):779–788. 2007.

43. **Ingram CJ**, Mulcare CA, Itan Y, *et al*. Lactose digestion and the evolutionary genetics of lactase persistence. *Hum Genet* 124:579–591. 2009.

44. **Itan Y**, Jones BL, Ingram CJ, *et al*. A worldwide correlation of lactase persistence phenotype and genotypes. *BMC Evol Biol* 10:36. doi:10.1186/1471-2148-10-36. 2010.

45. **Itan Y**, Powell A, Beaumont MA, *et al*. The Origins of Lactase Persistence in Europe. *PLoS Comput Biol* 5(8):e1000491. doi:10.1371/journal.pcbi.1000491. 2009.

46. **Järvelä I**, Enattah NS, Kokkonen J, *et al*. Assignment of the locus for congenital lactase deficiency to 2q 21, in the vicinity of but separate from the lactase-phlorizin hydrolase gene. *Am J Hum Genet* 63:1078–1085. 1998.

47. **Jensen TGK**, Liebert A, Lewinsky R, *et al*. The -14010*C variant associated with lactase persistence is located between an Oct-1 and HNF1α binding site and increases lactase promoter activity. *Hum Genet* 130:483–493. 2011.

48. **Kalmár T**, Bachrati CZ, Marcsik A, *et al*. A simple and efficient method for PCR amplifiable DNA extraction from ancient bones. *Nucleic Acids Res* 28(12):E67. 2000.

49. **Keusch GT**, Troncale FJ, Miller LH, *et al*. Acquired lactose malabsorption in Thai children. *Paediatrics* 43:540–545. 1969.

50. **Khabarova Y**, Torniainen S, Nurmi H, *et al*. Prevalence of lactase persistent/non-persistent genotypes and milk consumption in a young population in north-west Russia. *World J Gastroenterol* 15(15):1849–1853. 2009.

51. **Kilara A**, Shahani KM. Lactase activity of cultured and acidified dairy products. *J Dairy Sci*. 59(12):2031–2035. 1975.

52. **Kosikowski FW**. Dairy foods of the world - evolution, expansion and innovation. *J Dairy Sci* 64:996–1007. 1981.

53. **Kozlov AI**, Balanovskaia EV, Nurbaev SD, *et al*. Genogeographic primary hypolactasia in the Old World populations. *Genetika*. 34(4):551–561. 1998.

54. **Krawczyk M**, Wolska M, Schwartz S, *et al*. Concordance of genetic and breath tests for lactose intolerance in a tertiary referral centre. *J Gastrointestin Liver Dis* 17(2):135–139. 2008.

55. **Kuchay RA**, Thapa BR, Mahmood A, *et al*. Effect of C/T-13910 cis-acting regulatory variant on expression and activity on lactase in Indian children and its implication for early genetic screening of adult-type hypolactasia. *Clin Chem Acta* 412:1924–1930. 2011.

56. **Kuokkanen M**, Kokkonen J, Enattah NS, *et al*. Mutations in the translated region of the lactase gene (LCT) underlie congenital lactase deficiency. *Am J Hum Genet* 78:339–344. 2006.

57. **Lember M**, Tamm A, Piirsoo A, *et al.* Lactose malabsorption in Khants in western Siberia. *Scand J Gastroenterol* 30(3):225–227. 1995.

58. **Lember M**, Torniainen S, Kull M, *et al*. Lactase non-persistence and milk consumption in Estonia. *World J Gastroenterol* 12(45):7329–7331. 2006.

59. **Lewinsky RH**, Jensen TGK, Møller J, *et al*. 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.

60. **Maca-Meyer N**, Gonzalez AM, Larruga JM, *et al*. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2:13. 2001.

61. **Malmström H,** Linderholm A, Lidén K, *et al*. High frequency of lactose intolerance in prehistoric hunter-gatherer population in northern Europe. *BMC Evol Biol* 10:89. 2010.

62. **McCracken RD**. Lactase deficiency – example of dietary evolution. *Curr Anthropol* 12:479. 1971.

63. **Metz G**, Jenkins DJ, Peters TJ, *et al*. Breath hydrogen as a diagnostic method for hypolactasia. *Lancet* 1:1155–1157. 1975.

64. **Miglior F**, Sewalem A, Jamrozik J, *et al*. Analysis of milk urea nitrogen and lactose and their effect on longevity in Canadian dairy cattle. *J Dairy Sci* 89:4886–4894. 2006.

65. **Mottes M**, Belpinati F, Milani M, *et al*. Genetic testing for adult-type hypolactasia in Italian families. *Clin Chem Lab Med* 46(7):980–984. 2008. (abstract)

66. **Mulcare CA**, Weale ME, Jones AL, *et al*. 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. *Am J Hum Genet* 74(6):1102–1110. 2004.

67. **Myles S**, Bouzekri N, Haverfield E, *et al*. Genetic evidence in support of a shared Eurasian-North African dairying origin. *Hum Genet* 117:34–42. 2005.

68. **Newcomer AD**, McGill DB, Thomas PJ, *et al*. Prospective comparison of indirect methods for detecting lactase deficiency. *N Engl J Med* 293:1232–1236. 1975.

69. **Olds LC**, Sibley E. Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a cis regulatory element. *Hum Mol Genet* 12:2333–2340. 2003.

70. **Olds LC**, Ahn JK, Sibley E. -13915*G DNA polymorphism associated with lactase persistence in Africa interacts with Oct-1. *Hum Genet* 129:111–113. 2011.

71. **Oberhuber G.** Histopathology of celiac disease. *Biomed Pharmacother* 54:368–372. 2000.

72. **Outram AK**, Stear NA, Bendrey R, *et al*. The earliest horse harnessing and milking. *Science* 323:1332–1335. 2009.

73. **Pakendorf B**, Stoneking M. Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet* 6:165–183. 2005.

74. **Pericić M**, Barać Lauc L, Martinović Klarić I, *et al*. Review of Croatian genetic heritage as revealed by mitochondrial DNA and Y chromosomal lineages. *Croat Med J* 46:502–513. 2005.

75. **Pohl D**, Savarino E, Hersberger M, *et al*. Excellent agreement between genetic and hydrogen breath tests for lactase deficiency and the role of extended symptom assessment. *Br J Nutr* 104:900–907. 2010.

76. **Rasinperä H**, Savilahti E, Enattah NS, *et al*. A genetic test which can be used to diagnose adult-type hypolactasia in children. *Gut* 53:1571–1576. 2004.

77. **Richards MB**, Macaulay VA, Bandelt HJ, *et al*. Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 62:241–260. 1998.

78. **Rodriguez S**, Gaunt TR, Day INM. Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies. *Am J Epid* 169(4):505-514. 2009.

79. **Rossi M**, Maiuri L, Fusco MI, *et al*. Lactase persistence versus decline in human adults: multifactorial events are involved in down-regulation after weaning. *Gastroenterology* 112:1506–1514. 1997.

80. **Sahi T**, Isokoski M, Jussila J, *et al*. Lactose malabsorption in Finnish children of school age. *Acta Paediatr Scand* 61:11–16. 1972.

81. **Sahi T**, Isokoski M, Jussila J, *et al*. Recessive inheritance of adult-type lactose malabsorption. *Lancet* 2:823–826. 1973.

82. **Sahi T**. Genetics and epidemiology of adult-type hypolactasia. *Scand J Gastroenterol* Suppl 202:7–20. 1994.

83. **Savaiano DA**, Levitt MD. Milk intolerance and microbe-containing dairy foods. *J Dairy Sci* 70:397–406. 1987.

84. **Schirru E**, Corona V, Usai-Satta P, *et al*. Genetic testing improves the diagnosis of adult type hypolactasia in the Mediterranean population of Sardinia. *Eur J Clin Nutr* 61(10):1220–1225. 2007.

85. **Semino O**, Passarino G, Oefner PJ, *et al*. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 290:1155–1159. 2000.

86. **Sibley E**, Ahn JK. Theodore E. Woodward Award: Lactase persistence SNPs in African populations regulate promoter activity in intestinal cell culture. *Trans Am Clin Climatol Assoc* 122:155–164. 2011.

87. **Simoons FJ**. Primary lactose intolerance and the milking habit: a problem in biological and cultural interrelations, II. A culture historical hypothesis. *Am J Dig Dis* 45:489–497. 1970.

88. **Sun HM**, Qiao YD, Chen F, *et al*. The lactase gene -13910T allele cannot predict the lactase-persistence phenotype in north China. *Asia Pac J Clin Nutr* 16(4):598–601. 2007.

89. **Swallow DM**. Genetics of lactase persistence and lactose intolerance. *Annu Rev Genet* 37:197–219. 2003.

90. **Swallow DM**, Hollox EJ. The genetic polymorphism of intestinal lactase activity in adult humans. In Scriver CR, Beaudet AL, Sly WS, Valle D (editors): The Metabolic and Molecular Basis of Inherited Disease. New York: McGraw-Hill; pp 1651–1663. 2000.

90. **Tishkoff SA**, Ranciaro RA, Voight BF, *et al*. Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genet* 39:31–40. 2006.

91. **Torniainen S**, Freddara R, Routi T, *et al*. Four novel mutations in the lactase gene (*LCT*) underlying congenital lactase deficiency (CLD). *BMC Gastroenterol* 9:8 doi:10.1186/1471-230X-9-8. 2009a.

92. **Torniainen S**, Parker MI, Holmberg V, *et al*. Screening of variants for lactase persistence/non-persistence in populations from South Africa and Ghana. *BMC Genetics* 10:31. doi:10.1186/1471-2156-10-31. 2009b.

93. **Torroni A**, Bandelt HJ, D'Urbano L, *et al*. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 62:1137–1152. 1998.

94. **Tömöry G**, Csányi B, Bogácsi-Szabó E, *et al*. Comparison of maternal lineage and biogeographic analysis of ancient and modern Hungarian populations. *Am J Phys Anthropol* 134:354–368. 2007.

95. **Troelsen JT**, Olsen J, Moller J, *et al*. An upstream polymorphism associated with lactase persistence has increased enhancer activity. *Gastroenterology* 125:1686–1694. 2003.

96. **Vékony G**. Magyar őstörténet – Magyar honfoglalás. Budapest: Nap Kiadó. 2002.

97. **Vonk RL,** Lin Y, Koetse HA, *et al*. Lactose (mal)digestion evaluated by the ^{13}C-lactose digestion test. *Eur J Clin Invest* 30(2):140–146. 2000.

98. **Walsh PS**, Metzger DA, Higuchi R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10:506–513. 1991.

7. 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.