MOLECULAR GENETICS OF LACTASE PERSISTENCE

Nabil Sabri Enattah

Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland and Department of Medical Genetics, Faculty of Medicine, University of Helsinki, Helsinki, Finland

Academic Dissertation

To be publicly discussed with the permission of the Medical Faculty of the University of Helsinki, in the lecture hall 2 of the Biomedicum Helsinki, Haartmaninkatu 8, on February 3rd, 2005 at 12 noon

Helsinki 2005

Supervised by:

Professor Leena Peltonen-Palotie
Department of Molecular Medicine
National Public Health Institute and
Department of Medical Genetics,
University of Helsinki, Finland

and

Docent Irma Järvelä
Department of Medical Genetics,
University of Helsinki,
Helsinki,
Finland

Reviewed by

Professor Jaakko Ignatius
Department of Clinical Genetics
Oulu University Hosptial
Finland

and

Docent Tarja Ruuska
Department of Pediatric Gastroenterology
Hospital for Pediatric and Adolescence
Tampere, Finland

To be publicly discussed with:

Docent Maija Wessman
Finnish Genome Center &
Folkhälsen Research Center
University of Helsinki, Finland

Julkaisija-Utgivare-Publisher
Kansanterveyslaitos (KTL)
Mannerheimintie 166
00300 Helsinki
Puh.vaihde (09) 47441, telefax (09) 47448408

Folkhälsoinstitutet
Mannerheimvägen 166
00300 Helsingfors
Tel. växel (09) 47441, telefax (09) 47448408

National Public Health Institute
Mannerheimintie 166
FIN-00300 Helsinki, Finland
Telephone +358 9 47441, telefax +358 9 47448408

Publications of the National Public Health Institute, KTL A4/2005
ISBN 951-740-491-3 (Paperback)
ISSN 0359-3584 (printed versions)
ISBN 951-740-492-1 (PDF)
ISSN 1458-6290 (electronic versions)
http://ethesis.helsinki.fi

Yliopistopaino
Helsinki 2005

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original articles, which are referred to in the text by their Roman numerals. In addition, some unpublished data are presented

I Järvelä I*, Enattah NS*, Kokkonen J, Varilo T, Savilahti E, Peltonen L (1998) Assignment of the locus for congenital lactase deficiency to 2q21, in the vicinity of but separate from the lactase-phlorizin hydrolase gene. ***American Journal of Human Genetics* 63: 1078-1085**

II Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen, Järvelä I (2002) Identification of a variant associated with adult-type hypolactasia. ***Nature Genetics* 30 (2): 233- 237**

III Kuokkanen M, Enattah NS, Oksanen A, Savilahti E, Orpana I, Järvelä I (2003) Transcriptional regulation of the lactase-phlorizin hydrolase gene by polymorphisms associated with adult-type hypolactasia. ***Gut* 52: 747-652**

IV Enattah NS, Forsblom C, Rasinperä H, Tuomi T, Groop P-H, Järvelä I and the FinnDiane study group (2004) The genetic variant of lactase persistence C (-13910) T as a risk factor for type I and II diabetes in the Finnish population. ***European Journal of Clinical Nutrition :* 58(9):1319-1322**

V Enattah NS, Trudeau A, Pimenoff V, Maiuri L, Rossi M, Aurrichio S, Creco L, Lenzte M, Seo JK, Rahgozar S,Khalil I, Alifrangis M, Natah S, Shaat N, Groop L, Comas D, Bulaeva K, Mehdi QS, TerwilligerJD, Sahi T, Savilahti E, Perola M, Sajantila A, Järvelä I, Peltonen L (2004) The introduction of lactase persistence mutation into the global population. ***Submitted***

* These authors contributed equally to this work
The original publications have been reproduced with the permission of the copyright holders.

ABBREVIATIONS

aa	amino acids
ASHG	American Society of Human Genetics
ATH	adult-type hypolactasia
BAC	bacterial artificial chromosome
BCM	Baylor college of medicine
BLAST	basic local alignment search tool
BMD	bone mineral density
bp	base pair
cDNA	complementary DNA
Cdx-2	caudal-related protein 2
CEPH	Centre d´Etude polymorphisme Humain
CLD	congenital lactase deficiency
cM	centi Morgan
cR	centi Ray
cSNP	coding single nucleotide polymorphism
DARS	aspartyl-tRNA synthetase
DDGE	denaturing gradient gel electrophoresis
DNA	deoxyribonucleic acid
dNTP	deoxynucleosidetriphosphate
E.C.	Enzyme Commission Number
EMBL	European Molecular Biology Database
ELSI	the ethical, legal, and social issues of human genome project
ER	endoplasmic reticulum
EST	expressed sequence tag
FDH	Finnish disease heritage
FISH	fluorescence in situ hybridization
FREACs	Fork-Head related activators
H	Heterozygosity
HGP	Human genome project
HNF1α	Hepatic Nuclear Factor 1 α
HOX11	Homeo box 11
kb	kilobase
KD	kilodalton
LD	linkage disequilibrium
LNP	lactase non-persistence
LOD	logarithm of odds
LP	lactase persistence
LPH	lactase-phlorizin hydrolase
LTT	lactose tolerance test

LTTE	lactose tolerance test with ethanol
Mb	megabase
MCM6	minichromosome maintenance deficient 6
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
nt	nucleotide
OMIM	Online Mendelian Inheritance in Man
p	short arm of chromosome
PAC	P1-artifical chromosome
PCR	polymerase chain reaction
q	long arm of chromosome
RFLP	restriction fragment length polymorphism
RH	radiation hybrid
RNA	ribonucleic acid
RT	reverse transcriptase
RT-PCR	reverse transcriptase polymerase chain reaction
SNP	single nucleotide polymorphism
STR	short tandem repeat
tRNA	transfer RNA
UTR	untranslated region
YAC	yeast artificial chromosome
Θ	recombination fraction
λ	proportion of excess of allele in chromosomes carrying the disease allele
χ^2	chi-square test

SUMMARY

Two types of lactase deficiency exist in human, congenital lactase deficiency and adult-type hypolactasia. Congenital lactase deficiency (CLD) is an autosomal recessive severe gastrointestinal disorder in newborns characterized by watery diarrhoea after breast fed milk due to osmosis developed by unhydrolyzed lactose. The severe diarrhoea followed by dehydration, acidosis, and weight loss is usually diagnosed during the first weeks or months of life. CLD is considered one of the 36 rare monogenic disorders enriched in Finnish population. In contrast, adult-type hypolactasia is a normal physiological condition, in which the lactase deficiency is a result of the down regulation of the lactase enzyme after weaning in mammals including human. The condition can be clinically presented with a wide diversity of intestinal symptoms such as: meteorism, borborgymi, flatulence, fullness, abdominal colicky pains, loose stools and diarrhoea after ingestion or eating lactose containing foods.

In this study we localized the CLD locus between markers D2S114 and D2S132 on chromosome 2q in 19 CLD families. Further we fine mapped the locus based on linkage disequilibrium (LD) and ancestral haplotype analyses between markers D2S314-D2S2385, about 1 Mb 5′ of the lactase-phlorizin hydrolase (LPH) gene. Further we localized the locus for adult-type hypolactasia on a 6 cM region flanking the LPH gene between markers D2S114-D2S2385 in nine extended Finnish families. Using linkage disequilibrium and haplotype analysis we restricted the region to 47 kb interval between markers D2S3014-D2S3012. Sequence analysis of this region revealed two variants, C/T$_{-13910}$ and G/A$_{-22018}$ that show significant correlation with lactase persistence /non persistence trait in Finnish families and lactase enzyme activity in case-control study materials. Mutational analysis of the variants C/T$_{-13910}$ and G/A$_{-22018}$ associated with lactase persistence revealed no correlation with CLD and provided evidence that two separate genetic loci underlying congenital lactase deficiency and lactase persistence, respectively, are present on 2q21. Both molecular epidemiological studies in different populations and recent functional studies show that C/T$_{-13910}$ variant is most probably the causative variant of lactase persistence trait.

The analysis of disaccharidase activities shows that the mean level of lactase activity among CC$_{-13910}$ genotype was 6.86±0.35 U/g, CT$_{-13910}$ genotype it was 37.8±1.4 U/g, and 57.6±2.4 U/g protein for the TT$_{-13910}$ genotype and age per se has no significant effect on the

disaccharidase activity in adults. Further, relative quantitation of the expressed LPH alleles in the intestinal mucosa showed that the mRNA levels in individuals with the T_{-13910} allele several times higher compared to that found in individuals with the C_{-13910} allele, suggesting regulation of the LPH gene at the transcriptional level. To trace back the age and origin of lactase persistence mutation haplotype analysis was performed using SNPs flanking the associated SNPs and covering 30 kb region in 37 populations. Haplotype analysis revealed that two major haplotypes could be identified as carrying the lactase non persistent variant whereas only one major background haplotype was observed in lactase persistent alleles in all populations studied. Based on haplotype analysis and LD in global populations we propose that the geographic region west of the Ural Mountains represents the most likely origin of the major global lactase persistence mutation. The major lactase persistent haplotype most probably originated in a nomadic population tribes some (4800-6600 years ago) and the mutation then spread with migration of tribes westward towards Europe as well as to the south to Western Asia and Middle East. This would imply that lactase persistence was introduced to Europe by migrations of Indo-European tribes from Asian Steppes, not from Middle East, the region where farming and dairy practice are supposed to originate.

Significant correlation of the C/C_{-13910} genotype with low lactase activity, lactase/sucrase (L/S)-ratio and the similar prevalence figures of C/T_{-13910} with lactase persistence in >30 populations studied facilitated the introduction of a genetic test of adult-type hypolactasia to clinical practice. Identification of the C/T_{-13910} polymorphism has facilitated large-scale population based studies on the effect of lactase persistence/non-persistence on different clinical conditions like diabetes and osteoporosis. For example, the analysis of 1455 patients with type I and 615 with type II diabetes and 446 non diabetic controls in the Finnish population shows no detected differences in the lactase persistence genotype frequencies (C/T_{-13910} and TT_{-13910}) between diabetic and non diabetic subjects. Thus, we conclude that the C/T_{-13910} polymorphism associated with lactase persistence is not a risk factor for type I or type II diabetes in the Finnish population. In addition, the data that emerged from the analysis of the genetic variation of the LPH locus will help to shed light on the history of lactase persistence and provide the basis for analyses of evolutionary forces which have made the variant a predominant allele among some populations.

REVIEW OF THE LITERATURE

1. LACTASE-PHLORIZIN HYDROLASE (LPH)

Lactase-phlorizin hydrolase (LPH) (EC 3.2.1.23/62) is an integral glycoprotein of the microvillus membrane of small intestinal epithelial cells (Mantei et al. 1988). The mature enzyme has two enzymatic activities: Lactase (β-d-Galactoside galactohydrolase) (EC 3.2.1. 23), and the phlorizin hydrolase (glycosyl-N-acyl-sphinosine glycohydrolase) (EC .3.2.1 62) (Schlegel-Haueter et al. 1972; Colombo et al. 1973; Skovbjerg et al. 1981; Skovbjerg et al. 1982). Both catalytic activities are produced by a single polypeptide chain (Mantei et al. 1988).

Lactase is an enzyme responsible for hydrolysing the milk sugar lactose (the main carbohydrate in mammalian milk) to glucose and galactose **(Figure 1)** whereas, Phlorizin hydrolase is responsible for hydrolysing aryl and alkyl β- glycosides to phlorizin and β-glycosylceramides (which are part of diet of most vertebrates) (Lorenz-Meyer et al. 1972; Keller P 1993; Arola and Tamm 1994).

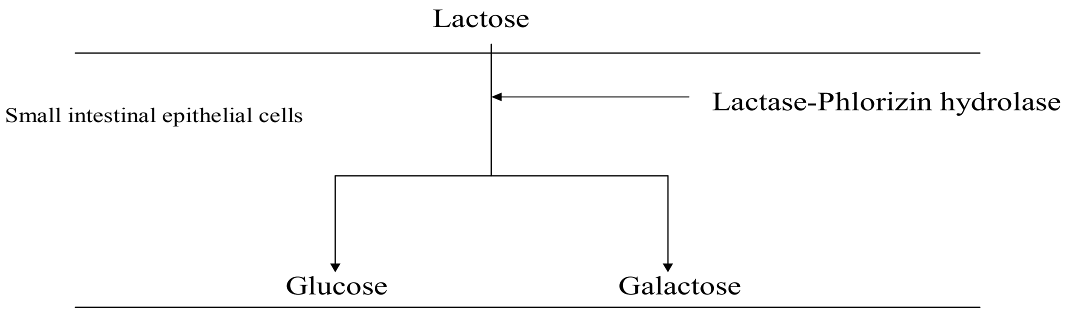

Lactose
Small intestinal epithelial cells
Lactase-Phlorizin hydrolase
Glucose
Galactose

Figure 1. The milk sugar, lactose, is hydrolyzed in the small intestinal epithelial cells to glucose and galacatose, by the enzyme lactase-phlorizin hydrolase.

1.1 Biosynthesis & Structure of lactase-phlorizin hydrolase

Intestinal epithelial cells synthesize LPH as a single- chain large precursor protein (Pro LPH) with a molecular weight of 215 KDa (Danielsen et al. 1984; Skovbjerg et al. 1984; Naim et al. 1987). This precursor post-translationally processed to the mature LPH of molecular weight about 135-160 KDa. The pro-LPH protein consists of five domains: an N-terminal signal sequence (19 amino acids), a LPHα profragment size of 849 amino acids (none of which appear in mature, membrane bound LPH), for which the cellular destination is not known, an extracellular domain of 1104 amino acids, $LPH_{\beta final,}$ (which carry both active sites of the enzymes), a hydrophobic trans-membrane anchor domain (19 amino acids) near the carboxy terminus, and short C-terminal cytosolic domain of 26 amino acids (Mantei et al. 1988) **Figure 2**.

The activity and the structure of the disacchridases are modulated by various mechanisms. Through a multi-step synthesis and the migration of the enzyme to the brush-border membrane, the mature enzyme gets localized to its site of action. This process is influenced by numerous factors, such as age, degree of differentiation of cells along the villus, glycosylation, and enterocyte life span. The glycosylation of the polypeptide is apparently similar to other disaccharidases (Danielsen et al. 1981; Roth 1987) and includes two main steps: the co-translational acquisition of glucan units of a high mannose type at the ER and subsequent trimming and complex glycosylation in the Golgi apparatus. During its passage through the Golgi complex, the intestinal brush-border hydrolases also get glycosylated with O-linked oligosaccharides. There is evidence that lactase is O-glycosylated through serines and threonine as well as N-glycosylated (through aspargine), and this glycosylation probably affects enzymatic activity as well as folding and intracellular transport (Naim and Lentze 1992).

Pro-LPH is glycosylated in the endoplasmic reticulum (ER) by mannose rich N-linked oligosaccharides (Naim and Naim 1996). In ER, two mannose rich pro-LPH homodimers form a dimer that is further transferred to the Golgi apparatus. The O-linked sugars of pro-LPH dimer are glycosylated and N-linked sugars are further processed in cis-Golgi resulting in a glycoprotein with a molecular weight of 230 KDa (Hauri et al. 1985; Naim et al. 1991).

The mature pro-LPH undergoes two proteolytic cleavage steps: The first cleavage occurs intracellularly and removes the large LPHα profragment at Arg_{734}/Leu_{735} resulting in a membrane bound $LPH\beta_{intial}$ (Leu_{735}-Tyr_{1927})(von Heijne 1986; Jacob et al. 2000).

Although details have been disputed, it is now considered that the active site at Glu1273 in domain III is responsible for hydrolysis of glucosides such as phlorizin, whereas the other in domain IV, at Glu1749, catalyzes the hydrolysis of galactosides such as lactose (Arribas et al. 2000). $LPH\beta_{intial}$ is targeted to the intestine brush border membrane where it is cleaved by trypsin at Arg_{868}/Ala_{869} leading to a 160 KDa mature $LPH\beta_{final}$ enzyme (Figure. 2) (Naim and Naim 1996; Wuthrich et al. 1996). LPH is anchored to the intestinal membrane by a hydrophobic region near its carboxy terminus in the C_{in}-N_{out} orientation and the catalytic sites of the enzyme are located in the lumen of the intestine (Mantei et al. 1988).

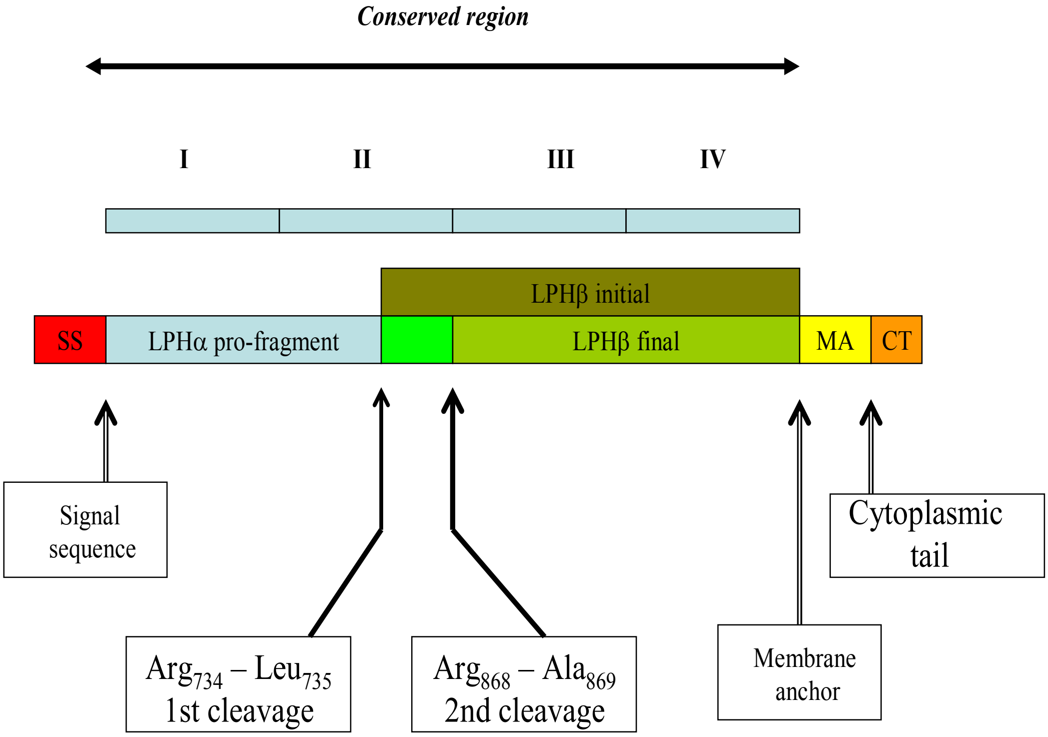


Figure 2. The structure of pro-LPH in human. The pro-LPH contains a cleavable signal sequence from Met1 to Gly19 that guides the Polypeptide to endoplasmic reticulum (ER) (von Heijne 1986; Mantei et al 1988). The region from Ser20 to Thr1882 consists of Four homologous domains (I-IV). The pro-LPH is processed by two proteolytic cleavages: an intracellular cleavage occurs between Arg734 and Leu735 which produces LPH β initial and a cleavage in the intestinal lumen between Arg868-Ala869 generates LPH βfinal the mature enzyme (Jacob et al 1996; Wüthrich et al 1996). Modified from (Jacob et al 2000).

The LPH gene has been mapped to the long arm of chromosome 2q21 (Kruse et al. 1988; Harvey et al. 1993). The gene is approximately 55 kb in size and composed of 17 exons. The size of cDNA is 6274 nt, including the 5781 nucleotide coding for 1927 amino acids of the primary translation product (pro-LPH) (Mantei et al. 1988). There are four (I-IV) conserved structural and functional regions in the pro LPH polypeptide of which two (I-II) are deleted via proteolytic cleavages from the final LPH molecule (Figure 2). Mantei and colleagues (1988) suggest that the regions III-IV are result of gene duplication of the I-II regions of an ancestor gene. Sequence similarities of I-IV regions of the pro-LPH to β-glycosidases found in archaebacteria, eubacteria, and fungi support this hypothesis and thus, LPH probably belongs to the β-glucosidase and β-galactosidase superfamily (Naim 2001).

1.2 Regulation of lactase-phlorizin hydrolase

The regulation of LPH expression in both humans and animals has been studied extensively during past 20 years, greatly stimulated by the interest of the 2 phenotypes of lactase status in human, persistence and non persistence. However, the molecular mechanisms behind the developmental down regulation of LPH expression after infancy have remained unclear (Sahi 1978; Sahi 1994b; Swallow 2003).

There is accumulating evidence that the primary mechanism behind the developmental down-regulation of the LPH expression might be mainly transcriptional (Escher et al. 1992; Lloyd et al. 1992; Fajardo et al. 1994). This has been shown in humans (Fajardo et al. 1994; Wang et al. 1995), rabbits (Keller et al. 1992), sheep (Lacey et al. 1994), pigs (Torp et al. 1993), and rats (Buller et al. 1990; Duluc et al. 1993; Krasinski et al. 1994). In the majority of cases, the LPH mRNA levels have been shown to correlate with lactase activity or the ratio of lactase to sucrase (L/S) activities (Escher et al. 1992; Lloyd et al. 1992; Fajardo et al. 1994). However, a comparison of LPH mRNA levels and lactase activity/LPH mRNA level ratios revealed a heterogeneous pattern in both hypolactasic (lactase non persistence) and lactase persistent individuals (Rossi et al. 1997). Further, biosynthesis of proLPH has been found to correlate with lactase mRNA levels but not with lactase activity in rabbits and rats, suggesting that a significant control of lactase gene expression could take place at the posttranscriptional level (Sebastio et al. 1989). Therefore, it has been suggested that the regulation of LPH gene expression involves both transcriptional and posttranscriptional control (Freund et al. 1991; Maiuri et al. 1994; Rossi et al. 1997).

Both a delayed posttranslational processing or/and reduction of pro-LPH synthesis have been observed in metabolic labelling studies in cells of hypolactasic individuals (Sterchi et al. 1990; Witte et al. 1990; Lloyd et al. 1992), however, the transcriptional regulation is most probably the most important factor affecting the level of LPH. A number of *cis* elements with a putative role in LPH transcription have been identified in the DNA sequences immediately upstream of the TATA box of the promoter. The human, rat, and pig promoter show stretches of homology in this region. The binding sites for transcription factor binding sites are clustered within 100 bp upstream of a TATA box, as is the case for most genes **(Figure 3).**

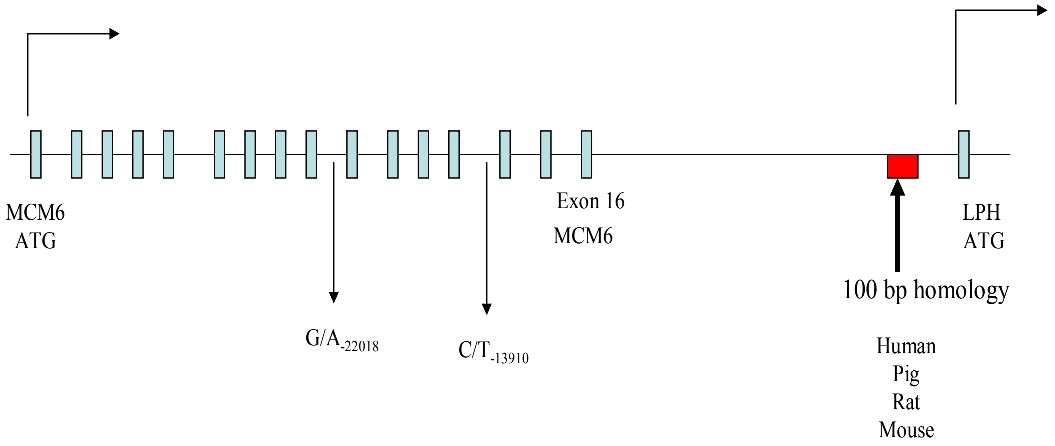


Figure 3. The species comparison shows a short segment (100 bp) have been conserved in different species just 5`LPH gene. This segment have been shown to be involved in interaction with different transcription factors to modulate the LPH Transcription. Whereas the region flanking the SNPs C/T$_{-13910}$ and G/A$_{-22018}$ does not shows any conservation.

The expression of the LPH gene is regulated by multiple transcription factors and their interactions which all influence the decline of the LPH enzyme after childhood. The effects of multiple transcription factors in the activation of the LPH promoter have been studied using transfection of reporter constructs, yeast one-hybrid cloning, and gel shift assays together. Specific antibodies, introduction of specific mutations, and co-transfection experiments with transcription factors, show evidence of the involvement of the caudal homologue Cdx2, HNF1α, HOX11, FREACs and GATA 4, 5 and 6 factors (Troelsen et al. 1992; Troelsen et al. 1994b; Troelsen et al. 1997; Fitzgerald et al. 1998; Hollox et al. 1999; Spodsberg et al. 1999; Fang et al. 2000; Mitchelmore et al. 2000; Fang et al. 2001; Krasinski et al. 2001; van Wering et al. 2002a; van Wering et al. 2002b; Troelsen et al. 2003a).

Analyses of pig and the rat LPH promoters in transgenic mouse models have indicated that approximately 1 kb 5´flanking sequence of the pig promoter and approximately 2 kb 5´flanking sequence of the rat promoter are sufficient to produce the reporter gene expression in a pattern similar to the endogenous LPH expression (i.e. small intestinal-specific expression down regulation after weaning, and a correct expression pattern along the longitudinal axis of the small intestine) (Troelsen et al. 1994a; Krasinski et al. 1997; Lee et al. 2002).

1.3 Terminology and classification of human lactase deficiencies

The types of lactase deficiency (Villako and Maaroos 1994) can be divided into:
a) Primary lactase deficiency: in which the lactase enzyme is selectively deficient.
There are two types of primary lactase deficiency:

i) Congenital lactase deficiency: lactase enzyme is almost nonexistent in the intestine of the newborn.

ii) adult-type hypolactasia (lactase non persistence): lactase enzyme is physiologically decreased (down-regulated) in adulthood to a level about 1/10 of that in newborn.

Although, the two types are considered as primary types of lactase deficiency; we should stress the fact that the first one is a pathological condition, whereas the second one can be considered to represent a physiological condition.

b) Secondary lactase deficiency: the lactase enzyme activity is affected with other disacchridase enzymes of the intestinal epithelial cells.

This type of lactase deficiency is usually due to an injury to intestinal mucosa. Injuries can be the result of diseases like in inflammatory bowel disease, Coeliac disease (Kosnai et al. 1980), acute enteritis (Ulshen and Rollo 1980), Tropical sprue or parasitic infections like Giardia lamblia, and Ascaris lumbricoides (Carrera et al. 1984). Also severe protein deficiency (Brunser et al. 1976), and oral medicines like neomycin, colchicines, or gamma irradiation can result in the severe injury of intestinal mucosa and lead to secondary deficiencies of several intestinal disacchridase enzymes.

Before we go on pause is necessary to clarify the terminology used. Sahi recommend the following accurate terminology (Sahi 1978) to describe lactase activity-related phenotypes: **Hypolactasia** is a very low activity of lactase in jejunal mucosa. **Adult-type hypolactasia** is used to differentiate from congenital lactase deficiency which affects the newborn.

Table 1. Frequently used terms for lactase activity related-phenotypes.

Low lactase activity	Normal lactase activity
Lactase non-persistence	Lactase persistence
Hypolactasia	
Adult-type hypolactasia	
Lactase restriction	
Alactasia	
Lactose malabsorption	
Lactose maldigestion	
Low lactose digestion capacity (low LDC)	High lactose digestion capacity (high LDC)
Lactose intolerance	
Milk intolerance	

Some have recommended the term **lactase restriction** rather than hypolactasia (Flatz 1987). The terms **lactase deficiency and alactasia** imply a total lack of lactase activity, which does happen even in congenital lactase deficiency. As a counter part for hypolactasia, the term **lactase persistence** instead of hyperlactasia has been used, meaning moderate or high lactase activity in intestinal mucosa of adults. Since lactase persistence is the most common term used, then **lactase non persistence** should be used instead of hypolactasia. The terms **lactose malabsorption and lactose maldigestion** are used to describe a poor lactose hydrolyzing capacity which can be determined by lactose tolerance test. It almost always implies hypolactasia, so in practice these terms are often interchangeable. Flatz recommends the terms **low lactose digestion capacity (Low LDC)** and its counterpart **high lactose digestion capacity (high LDC).** Although these are the proper and accurate terms, they are somewhat cumbersome to use. The most common public term used **is lactose intolerance** to mean lactose malabsorption or adult-type hypolactasia with abdominal symptoms. However there are some lactase persistence people who have abdominal symptoms and via vice versa some hypolactasic people who do not have any symptoms. Another confusing term is **milk intolerance** which means that a person suffers from abdominal symptoms after milk ingestion. Finally, we should remember to differentiate between **primary hypolactasia**, mention above, and **secondary hypolactasia** which could appear due to infection or infestation of small intestine, in celiac disease. In these cases the histology of intestine is often abnormal and affects all diasaccharidases.

In the following text the terms lactase nonpersistence/persistence will be used except in some situation where the term adult-type hypolactasia will be used instead of lactase nonpersistence.

2. Congenital lactase deficiency

2.1 Historical background

Congenital lactase deficiency (CLD), (MIM 223000) (http://www.ncbi.nlm.nih.gov/omim) is an autosomal recessive inherited severe gastrointestinal disorder in newborns. Holzel et al (1959) described the first patients, two siblings who had watery diarrhoea from birth and a very low lactase activity in small intestine (Holzel et al. 1959). In 1966 Launiala et al reported a selective absence of lactase activity in duodenal specimens of infants with severe diarrhoea after breast feeding (Launiala et al. 1966). In a clinical study on 16 patients Savilahti et al obtained the first evidence for recessive mode of inheritance for CLD (Savilahti et al. 1983).

CLD is a rare inborn error that is most prevalent in the Finnish population (Asp et al. 1973; Asp and Dahlqvist 1974; Savilahti et al. 1983).The incidence of CLD have been estimated to be 1:60000 in the Finnish population (Savilahti E, personal communication). Subsequently, CLD is considered one of the 36 rare monogenic disorders enriched in Finnish population (Norio et al. 1973; Norio 2003a; Norio 2003b; Norio 2003c). So far, 46 patient in 39 families have been diagnosed in Finland (Savilahti et al. 1983), (personal communication) whereas only 18 cases have been reported elsewhere in the world (Holzel 1967). In 1991, Poggi et al have been reported in an ASHG meeting that complete sequence of LPH gene (the candidate gene) in one CLD patient (from our series) revealed no pathological sequence changes could be detected in LPH gene (Poggi and Sebastio 1991), but no further report has been shown, and the causative gene remained however unclear.

2.2 Clinical presentations of CLD

The hallmark symptom of CLD is watery diarrhoea that the newborn develops soon after the first doses of breast fed milk due to osmosis developed by unhydrolyzed lactose. The severe diarrhoea followed by dehydration, acidosis, and weight loss are usually diagnosed during the first weeks or months of life (Savilahti et al. 1983). Despite the symptoms, CLD infants are vigorous, and hungry. The child may survive for several months and the baby shows delayed growth due to loss of nutrients, dehydration and acidosis. In laboratory investigation, the faeces are strongly acidic, and contain large amount of lactose. Blood cholesterol is below normal. Low lactase activity is pathognomic. The activity of both sucrase and maltase is normal. Lactase activity measured in duodenal biopsy specimen is very low (0-10 U/g protein). After the child is put on lactose -free diet, diarrhoea stops and infant quickly begin to gain weight. In long term follow up studies of 16 cases, normal psychomotor development and growth of the affected children on lactose free diet have been observed (Savilahti et al. 1983). In a series of 11 infant diagnosed of CLD at 6 to 88 days of life, hypercalemia and nephocalcinosis has been reported in which the hypercalcemia has been responded to treatment within one week of the start of lactose free-diet in most patients (Saarela et al. 1995). The mechanism of hypercalcemia is unclear but at the follow up examinations at ages 2 to 10 years of age, one of the patient still had hypercalciuria, three of 11 was still had nephrocalcinosis.

CLD should be differentiated from other very rare condition severe familial lactose intolerance (MIM 150220), where activity of lactase in new born is normal. In this condition symptoms vomiting, lactosuria and aminoaciduria develop during the first days of life (Holzel et al. 1962; Berg et al. 1969; Villako and Maaroos 1994).

2.3 The Finnish disease heritage

Although the prevalence data for CLD is globally not highly reliable, this very rare inborn error seems to be slightly more common in Finland than elsewhere (Asp et al. 1973; Asp and Dahlqvist 1974; Savilahti et al. 1983) and therefore it is considered to belong to the so called Finnish Disease Heritage (FDH). FDH is a group of rare hereditary diseases that are overrepresented in Finland (Norio 2003a). The reason for this is the peculiar history of Finland, small founder populations with long time isolation due to geographical and linguistic factors. This has resulted so far in the enrichment of about 36 monogenic disorders in this population. Characteristically, there is one founder mutation which is responsible for the distinct majority of disease alleles (Norio et al. 1973; Norio 2003a; Norio 2003b; Norio 2003c). The high quality health care system with comprehensive population registers have provided the means for the clinicians and geneticists in Finland to identify these cases and to describe the concept of Finnish Disease Heritage concept (Norio et al. 1973). The incidence of these disorders varies mostly between from 1:10000 to 1:100000. The combined carrier frequency of all tested Finnish mutations in one DNA array based study monitored for the prevalence of 31 rare and common disease mutation varied between 1:11 and 1:6 in regional study populations (Pastinen et al. 2001). The variation of carrier frequency of different mutations within the country provides evidence for the relatively recent population bottlenecks (Pastinen et al. 2000; Pastinen et al. 2001).

When tracing the early Finns, the genetic data implies that the majority of the Finnish gene pool originate from later small founder populations of Indo-European speakers who arrived from the south approximately 2000 years ago (Varilo 1999) . It is actually more probable that small immigration groups arrived in Finland continuously after glacial period (Peltonen et al. 1999). Analyses of the genetic diversity of the Y chromosome and mitochondrial DNA show that Finns differ from other European populations in having an exceptionally reduced amount of Y-chromosomal and mitochondrial diversity. This indicates that relatively few people have contributed to the genetic lineage of today´s Finns (Sajantila et al. 1996). Based on Y chromosomal haplotype studies, Finland has been inhabited in two waves (Kittles et al. 1998). The first migratory wave of Uralic speakers from the east occurred some 4000 years ago and has had a distinct effect on the Finnish gene pool. A review the history of Finland reveals that for hundreds of years Finland remained very sparsely populated. In the 12th century the population was only 50000, by the 16th century the population had expanded to 250000 but was still concentrated in the coastal areas. The internal migration in the 16th century resulted in the foundation of regional sub isolates and expansion of local populations shows existence for several bottle necks. The most severe was the great famine at the end of 17th century that killed one third of the population of 400000 between the years 1690-1730. Since then the population has grown rapidly to today's 5.2 million inhabitant over three centuries. As a consequence of regional expansion most diseases belonging to the FDH present a regional clustering. A distribution equal to the population density indicates that the mutation is old (like diastrphic dysplasia and Meckel syndrome), whereas a tight regional distribution suggests a more recent introduction of the mutation.

So far, out of the 36 diseases, the gene has been mapped in 33 (92%) and characterized in 27 (75%). The founder mutation was responsible for the majority of disease genes. Among the 27 characterized FDH genes, the main mutation is found in 100% of the chromosomes in 8 disorders. In most of the others, one mutation is represented in more than 90% of the disease alleles. In two diseases, the corresponding fraction is 70% (Peltonen et al. 1999; Norio 2003b). Strong evidence of the descent of many Finnish disease genes from a single founding ancestor has been obtained from linkage disequilibrium and haplotype data. Long LD intervals reaching up to 13 cM (Peltonen et al. 1999) around a particular disease mutation, have been observed and used as a very powerful tool by geneticist to tackle the molecular background of FDH. An update of the mutations behind the Finnish disease Heritage can be obtained in references (Peltonen et al. 1999; Norio 2003c) and on the website (www.findis.org).

3. Lactase persistence/nonpersistence

3.1 Historical background

Lactose, the milk sugar, has been found in 1860s to cause diarrhoea in dogs (Sahi 1994b). In the beginning of 20th century it was shown that lactase enzyme was present in intestine of infant animals and greatly diminished in adult animals (Mendel 1907; Kretchmer 1971; Sahi 1994b). The developmental curve for the lactase enzyme was constructed for a number of animals including rat, mouse, dog, pig, and rabbit (Heilskov 1951), in which the activity was high in infancy and then on weaning the lactase activity was down regulated to one tenth the level in newborn in adult animals. In contrast, there were only a few direct observations of human intestinal lactase activity (Sahi 1994b). The activity was established during infancy but it was sharply reduced by severe disease. There is consensus that the small intestinal mucosa is the principal site of lactase activity, which is highest in jejunum. However, controversy existed on the precise site of enzyme activity until Borgstöm et al (Borgstrom et al. 1957) and Dahlqvist et al (Dahlqvist and Borgstrom 1961) showed that the hydrolysis of lactose takes place in the outer membrane (brush border) of the mucosal epithelia cells.

Low lactase activity in healthy adults humans was discovered independently by two groups in 1963 (Auricchio et al. 1963; Dahlqvist et al. 1963). In those early days most of the studies were conducted on subjects from Northern Europe, among whom lactase persistence is common. This led to belief that lactase activity remains throughout the life span in human, and the term adult-type hypolactasia was coined to indicate the low lactase activity in the jejunal mucosa in healthy adults. When the studies expand to other populations it turned out that the hypolactasia is prevalent in adulthood in the most other groups and represents the normal state for humans like in other mammals (Sahi 1994b).

3.2 Diagnosis of lactase nonpersistence (adult-type hypolactasia)

Hypolactasia per se does not give to any disturbance; symptoms appear only after ingestion of lactose containing foods. In nonpersistent subjects foods containing lactose causes abdominal symptoms such as meteorism. borborgymi, flatulence, fullness, abdominal colicky pains, loose stools and diarrhoea (Villako and Maaroos 1994). There is a considerable individual variation in the manifestation of symptoms; this depends on the amount of milk products consumed and the individual sensitivity to stomach pains.

Strictly speaking there are two types of tests for diagnosing hypolactasia, direct and indirect methods (Metz et al. 1975; Newcomer et al. 1975; Flatz and Rotthauwe 1977; Arola 1994).
1) The direct method is an invasive method in which in which intestinal biopsy specimen processed for an assay of mucosal disaccharides. The disaccharidase assay is performed according to the standard method developed mainly by Dahlqvist & Burgess (Dahlqvist 1964). Diagnosis of hypolactasia is suggested when the lactase activity < 10 IU/g protein and lactase/sucrase ratio <0.3 with normal histology. There are different cut-off values used by different laboratories, in the Finnish samples analyzed in this thesis, a cut-off value of < 20 IU/g protein for lactase activity was used.
2) Indirect methods which are based on lactose tolerance tests.
There are many different tests but the two used worldwide are the following:
a) Lactose tolerance test (LTT) is based on the measurements of the increase in blood glucose by serial determinations after oral lactose load. Lactose dose varies in different modifications from physiologic dose of 12.5 g to the usual tolerance dose of 50 g. An increase to values < 1.1 mmol/l has been considered as indicative of hypolactasia whereas an increase to values >1.7 mmol/l is considered to indicate lactase persistence. To increase the reliability of the test a LTT with ethanol (LTTE) to inhibit the conversion of galactose to glucose by liver has been used by Sahi in the diagnosis of the family subjects (Sahi 1974a) involved in the present study. Hypolactasia is diagnosed if the blood galactose concentration is < 0.03 mmol/l at 40 min after lactose and ethanol ingestion.
b) Breath hydrogen test after lactose ingestion (BHT), hydrogen concentration in expired air is determined by gas chromatography after oral lactose load. Samples are taken at zero time point and then at intervals of 15 to 60 min for 2 to 6 hours. An increase in hydrogen

concentration > 20 ppm or > 0.3 ml/min over baseline is interpreted as a diagnostic for hypolactasia (Metz et al. 1975; Arola 1994).

As stated before, the average sensitivity of traditional laboratory tests (lactose tolerance test, breath-hydrogen test) vary between 69% to 100% (Arola 1994). The conflicting reports are complicated by the fact that the correlation between lactose intolerance and lactase non persistence is poor; depending on the test used, 11% to 32% of individuals with lactase non persistence report no symptoms from lactose containing milk products (Carroccio et al. 1998; de Vrese et al. 2001), whereas up to 57% of subjects with self-reported lactose intolerance have normal lactose absorption in breath-hydrogen test (Carroccio et al. 1998; Saltzman et al. 1999). The development of lactose intolerance per se seems not only to be related to the lactase enzyme level i.e. hypolactasia, but other factors interact in complex network to produce the clinical symptoms.

3.3 Genetics of lactase persistence/nonpersistence

The decline in lactase activity to a very low level after infancy remained unclear for a long time until in 1963 isolated intestinal lactase deficiency in adults was reported (Auricchio et al. 1963; Dahlqvist et al. 1963) and an accurate enzymatic method for intestinal disaccaridase determinations was published (Dahlqvist 1964). Since then, two hypotheses underlying the down regulation of lactase in adults have been suggested the adaptive and the genetic hypothesis.

The **adaptive hypothesis** suggests that lactase enzyme activity depends on lactose feeding and they propose that the hypolactasia appears merely because of the lack of milk supply (the major lactose source) (Bolin et al. 1969; Bolin and Davis 1970; Bolin et al. 1971). Animal feeding experiments have been carried out using rats, calves and pigs. Most of studies have fail to show any adaptation (Sriratanaban et al. 1971; Lebenthal et al. 1973; Leichter 1973). In none of the studies lactose feeding able to prevent the normal post weaning decline in lactase activity (Bolin et al. 1969; Lebenthal et al. 1973). Feeding trials in humans also have shown that feeding lactose in diet does not prevent the down regulation, and the lack of lactose in diet does not provoke a decline in lactase activity (Kretchmer 1971; Gilat et al. 1972).

In contrast, family studies have suggested a genetic aetiology behind the decline in lactase activity (Ferguson and Maxwell 1967; Welsh 1970; Gilat et al. 1973). **The genetic hypothesis** suggests that the decline in lactase activity in adult is genetically defined and not related to lactose feeding. Recessive inheritance of lactase decline, adult-type hypolactasia, was supported by family studies (Ferguson and Maxwell 1967; Welsh 1970; Gilat et al. 1973; Sahi et al. 1973; Sahi 1974a), a twin study (Metneki et al. 1984), and the distribution in disaccharidase activity (Ho et al. 1982; Flatz 1984). Strong evidence for recessive inheritance was supported by family studies in Finnish population (Sahi et al. 1973; Sahi 1974a). This study involved 11 probands and 327 family members (on which our study is also based). In this material the most valid indirect diagnostic method was used; the lactose tolerance test with ethanol which allowed the researchers to perform a reliable characterization of the phenotype of all family members and conclusive pedigree analysis.

Once it had been shown that adult-type hypolactasia is genetically determined recessively inherited trait (Ferguson and Maxwell 1967; Sahi et al. 1973; Sahi 1974a), the search for the causing factors intensified. In 1988 the human lactase gene was mapped to chromosome 2 by analysis of Southern blots of DNA from a panel of human-rodent cell hybrids containing characteristic sets of human chromosomes (Kruse et al. 1988). In the same year the complete structure of primary human and rabbit lactase-phlorizin hydrolase translation products deduced from the cDNA sequences was reported (Mantei et al. 1988). Disappointingly in 1991 Boll *et al* (Boll et al. 1991) showed that there were no sequence differences in any of 17 exons of LPH gene, in exon-intron boundaries or in the 1 Kb region 5´of LPH gene, between lactase persistent and lactase non persistent subjects. They concluded that humans with high or low levels of lactase carry intact coding sequences of the LPH gene. In 1995 Wang et al (Wang et al. 1995) had shown that subjects heterozygous for the lactase persistence allele expressed one allele of the LPH transcript at much lower levels than the other allele and these subjects tend to have intermediate lactase activities. This finding supported the idea that the lactase persistence/non-persistence trait is likely to be controlled by cis-acting element(s) residing within or adjacent to the lactase gene rather than by a variation in a trans-acting factor (Wang et al. 1995). Thus, they suggested that expression of two LPH alleles could be independently regulated (Wang et al. 1995). In addition, further support has been observed by strong linkage disequilibrium (LD) across the 70 kb haplotype spanning the lactase gene (Harvey et al. 1995). An interval of LD over 70 KB spanning the LPH gene has been observed in different populations with only 3 common haplotypes (Harvey et al. 1995; Harvey et al. 1998; Hollox et al. 2001). One particular haplotype , called A, seemed to be associated with lactase persistence and was found at much higher frequencies in Northern European than any other populations (Harvey et al. 1998; Hollox et al. 2001). During these years a large number of single nucleotide polymorphism were found but none of them was shown to be the causative variant of lactase persistence (Boll et al. 1991; Lloyd et al. 1992; Harvey et al. 1995; Hollox et al. 1999). After studies by Boll et al (1991) no further sequence of the full length of the LPH gene has been reported before the present work.

3.4 Mechanisms that underlie adult-type hypolactasia

The mechanisms behind the timing of the onset of hypolactasia are poorly understood. Earlier studies have shown a wide ethnic and regional variation in the age of onset of adult-type hypolactasia. The majority of Thai children have shown to become hypolactasic by the age of two years, in black populations adult-type hypolactasia has been shown to manifest between one to eight years, whereas in white populations low lactase levels are rarely seen in children under five years of age (Keusch et al. 1969; Sahi et al. 1972; Welsh et al. 1978; Simoons 1980). Wang et al (1995) studied the onset of lactase non persistence in children aged from 2 months to 11 years. They analyzed subjects heterozygous for the lactase persistence using polymorphism within exons of the LPH gene as they did in previous study in adults, and monitored the asymmetric expression of the LPH transcripts, indicative for the down regulation of an allelic LPH transcription (Wang et al. 1998). Genetically programmed down-regulation of the lactase gene was observed starting from the second year of life, although the extent and onset was not constant. They concluded that a developmentally regulated trans-acting DNA-binding protein could bind to only one kind of lactase allele and influence transcription and/or mRNA stability (Wang et al. 1998). Previous studies of the Finnish population, based on lactose tolerance test, have shown that adult-type hypolactasia can manifest up to 20 years (Sahi et al. 1972). However, later recent data have shown that the majority of Finns have developed hypolactasia by age of 10 (Rasinpera H 2004).

4. Evolution of Lactase persistence

Only little attention has been paid to the fact that is not quite appropriate to state that the prevalence of hypolactasia (instead of lactase persistence) varies considerable from less than 5% to 100%, since this can be considered as the normal physiological condition of humans. Below I will present the prevalence data systematically for the abnormal state which is lactase persistence.

Lactase persistence varies greatly between different and within populations from almost 0 % in South East Asia to 99 % in Northern Europe (Sahi 1994a). To explain these highly geographic variations in the prevalence of lactase persistence, various researchers have produced some hypotheses to explain these variations focused on some selective factors related to milk supply because it is the only source of lactose, the substrate of the lactase enzyme. These hypotheses are

1) *The culture historical hypothesis* is based on genetic selection and correlates the occurrence of lactase persistence with dairy culture. This hypothesis is the most widely accepted and proposed by Simoons (Simoons 1969; Simoons 1970), later by McCracken (McCracken 1970) and refined by others (Johnson et al. 1974; Flatz and Rotthauwe 1977; Simoons 1978). It states that individuals with lactase persistence were able to use all the nutrients of milk, therefore, they were stronger, better equipped to survive and possibly had more children. Thus the proportion of people with lactase persistence would increase in a population.

2) *Calcium absorption hypothesis*: This hypothesis was put forward to explain the prevalence of lactase persistence in Northern Europe (Flatz and Rotthauwe 1973). In this region of the world the nutritional supply of vitamin D was low and it was proposed that lactose could enhance absorption of calcium and thus individuals with lactase persistence will have less rickets and pelvic deformities resulting in a selection in favour of lactase persistence. Much criticism of this hypothesis been presented by Simoons (Simoons 2001) who have shown that lactase nonpersistence people can absorb calcium as lactase persistence people and this hypothesis is not confirmed by historical, osteoarachelogical or biomedical evidence.

3) Selective advantage of lactase persistence to survive cholera and other epidemics have been proposed by Cook and Al Torki (Cook and al-Torki 1975). This hypothesis was put to explain the high frequency of lactase persistence in hot climates such as desert regions.

These three hypotheses were tested by comparative methods and most support was obtained for the historical culture hypothesis. Further, using maximum likelihood analysis it was shown that the evolution of milking seems to precede the evolution of lactase persistence (Holden and Mace 1997). Nowadays, there is a consensus that selection has been responsible for the high prevalence of lactase persistence. Haplotype analyses have suggest that much of the variation in the LPH locus has been affected by genetic drift with recent directional selection for lactase persistence (Hollox et al. 2001). The question remained then how much of the selection power was necessary to produce the observed frequencies in different populations. Some suggest that a selection power of 1% was sufficient to increase the gene from 5% to 60% in 10000 years, whereas others suggest a high selection power of 5-7% to explain the frequency in Northern Europe (Cavalli-Sforza 1973; Aoki 1986; Flatz 1987; Aoki 2001).

5. Lactase persistence/nonpersistence and human diseases

5.1 Bone development and Osteoporosis

The association of lactase non persistence and osteoporosis is one of the most common conditions studied so far. Lactase non persistence can be manifested clinically as lactose intolerance by abdominal bloating, cramping, distention, flatulence, and diarrhea, causing many people to avoid drinking milk, the main source of calcium (Birge et al. 1967). The role of calcium from milk and other dairy products has been shown to be essential for bone mass development. Therefore, lactose intolerance might lead to diminished calcium intake and reduced calcium absorption. Thus lactase non persistence has been considered to be a risk factor for osteoporosis and fractures (Birge et al. 1967; Newcomer et al. 1978; Finkenstedt et al. 1986; Horowitz et al. 1987; Mainguet et al. 1991; Wheadon et al. 1991; Corazza et al. 1995; Di Stefano et al. 2001).

However, there are conflicting reports whether the risk to osteoporosis is the result of lactase nonpersistence alone (Birge et al. 1967), or both lactase non persistence and lactose intolerance together (Finkenstedt et al. 1986; Horowitz et al. 1987; Mainguet et al. 1991; Slemenda et al. 1991). Two studies of Finnish perimenopausal women aged 38-57 years, have shown that lactose intolerant women have a slightly reduced perimenopausal bone mineran density (BMD) and elevated risk of fractures (Honkanen et al. 1996; Honkanen et al. 1997). The main difficulty in studies addressing the impact of lactase nonpersistence on the osteoporosis and fracture risk in humans have been tedious and inaccurate diagnostic laboratory tests so far used for lactase non persistence.

5.2 Diabetes Mellitus & other diseases

In humans, intestinal lactase activity is increased in diabetes and been shown to be normalized with insulin treatment (Tandon et al. 1975; Mahmood et al. 1978; Schedl et al. 1983; Murakami and Ikeda 1998). Furthermore, milk lipids, contain high amount of energy, might contribute to obesity and increased risk of diabetes. A high frequency of lactose absorbers (lactase persistence) was observed among diabetic type I and type II patients in Sardinia (Meloni et al. 2001). They studied 50 control subjects, 50 with diabetes type I, and 50 with diabetes type II. They used breath hydrogen test and found 14% prevalence of lactase persistence compared to 48% in patient with type I diabetes, and 52% in patient with type II diabetes.

Since people with lactase persistence supposedly drink more milk than people with lactase non persistence, high lactose consumption leads to a greater exposure to glucose and galactose. This might have implications for many other disease risks. Some support exist for a role of lactose ingestion and galactose cytotoxicity in the pathogenesis of ovarian cancer (Cramer 1989; Macdonald 1989; Mettlin and Piver 1990; Risch et al. 1994; Herrinton et al. 1995; Webb et al. 1998; Meloni et al. 1999; Britton et al. 2000; Goodman et al. 2002), and in senile cataract (Simoons 1982; Bengtsson et al. 1984; Rinaldi et al. 1984; Spinelli et al. 1987; Lisker et al. 1988; Meloni et al. 1999).

Considering ischemic heart disease ; it is thought that excessive intake of milk may result in abnormal serum profiles of triglycerides, and cholesterol increasing the risk for ischemic heart disease (Segall 1980; Segall 1994; Segall 2003). No correlation was found in males, and in females the sample size was not sufficient for statistical significant conclusions. In an Estonian study it was found that patients with acute myocardial infarction (MI) consumed more milk than control persons (Lember and Tamm 1988). They found that the relative risk of MI for those who consumed 3 or more glasses of milk daily was 4 (95%CI 1.4-13.3) when compared to those who consumed less milk. It should emphasized that all the clinical correlations with lactase persistence/nonpersistence rather reflect the effect on milk drinking and since this habit has many determinants, beyond intestinal lactase activity, the results remain non- conclusive.

6. Human genome Project (HGP) (1990-2003)

6.1 Historical background, paving the way toward HGP

The rules of heredity were established by Mendel, based on breeding experiments on pea plant and advanced by Sutton in his **Chromosomal theory of heredity** (Mendel 1866; Sutton 1903). Hereditary transmission through the **sperm and egg** became known by 1860. It took 20 years, in which the details of **mitosis, meiosis and fertilization** were clarified, to establish that the **chromosomes** were the active players in these processes (Olby 1966).

Recognition the fact that **DNA is the chemical hereditary material** arose from studies of transformation of pneumonia causing bacteria (Griffith 1928; Avery 1944). **Chargaff´s rules stated that** nucleotide pairing **A=T**, and **C=G** in DNA and was based on the fact that the number of adenine residues was always equal to thymine and the number of cytosines equal to guanines (Chargaff 1949; Chargaff 1951). They showed that **3´-5´phosphodiester bonds** regularly link the nucleotides of DNA. Then detailed analysis of high-quality photographs of **X-ray diffraction pattern of DNA** provided the basis for the correct determination of **double helix structure for DNA** by Watson and Crick. Structural studies greatly facilitated the understandings the details of **DNA replication** was very soon recognized (Astbury 1951; Franklin and Gosling 1953; Watson and Crick 1953b; Watson and Crick 1953a). **DNA synthesis** in cell free extracts of bacteria with **DNA polymerase I enzyme** was demonstrated (Kornberg 1960) and the deciphering of the **genetic code** was finally interpreted (Nirenberg and Matthaei 1961; Cold Spring Harbor Publications 1966; Nirenberg 2004). The development of **recombinant DNA technologies** occurred in1970s (Jackson et al. 1972; Cohen et al. 1973; Lobban and Kaiser 1973). The concept that DNA molecule can be cut at defined points by **restriction enzymes (Smith and Wilcox 1970)**, and again joined by **DNA ligase enzyme** (Lobban and Kaiser 1973), as well as the invention of powerful methods **for DNA sequencing** (Sanger and Coulson 1975; Maxam and Gilbert 1977; Smith et al. 1986), in vitro **DNA amplification (polymerase chain reaction, PCR**) (Saiki et al. 1988) set the stage for the official international Human genome Project to begin in 1990.

6.2 Human genome project (HGP) 1990-2003

The Human Genome project was officially launched in 1990 to create publicly accessible databases for high-quality sequences of genomes of human and key model organisms (human genome project information web site, 2004). The initial goals of HGP were to *identify* all genes in human DNA, *determine* the sequences of some 3 billion base pairs that make up human DNA, *store* this information in databases, *improve* tools for data analysis, *transfer* related technologies to the private sector, and *address* the ethical, legal, and social issues (ELSI) that may arise from the project. Although the project was initiated, and funded mainly by the US government, some 18 countries have participated in the worldwide effort, with significant contributions from the Sanger Center in the United Kingdom and genome centers in Germany, France, and Japan. Within the last decade, the program has rapidly progressed from the generation of genetic and physical maps, of chromosomes via the positioning of 42000 ESTs (Deloukas et al. 1998) , to the production of a draft sequence of the human genome (Weissenbach et al. 1992; Cohen et al. 1993; Weissenbach 1993; Hudson et al. 1995; Dib et al. 1996; McPherson et al. 2001). The Human Genome Project reached its major goals in 2003 with the official completion of the human sequence on the 50th anniversary of Watson and Crick's description of the fundamental structure of DNA.

The announcement of a private company Celera, in 1998, that it will finish the DNA sequence of human in a very rapid time frame using whole genome shotgun sequencing strategy (Venter et al. 1998), led the public efforts, which relied on hierarchial sequencing strategy, to modify their goals and respond to the challenge. In June 2000, international leaders of the Human Genome Project (HGP) announced that the rough draft of the entire human genome would be completed a year ahead of schedule. Today this race is viewed as beneficial for the scientific community and it resulted in the publication the first version of the human genome sequence in February 2001. The outcome of public effort was published in Nature (Lander et al. 2001) and that of Celera in Science (Venter et al. 2001).

The ongoing analyses of the working draft have revealed interesting facts: The human genome contains 3164.7 million nucleotide bases and the average gene consists of 3000 bases, the sizes varying greatly. The largest known human gene is dystrophin with 2.4 million bases. The total number of genes is estimated to be 22000, much lower than previous estimates of 80,000 to 140,000 that had been based on extrapolations from gene-rich areas of the genome (Cold Spring Harbor Genome meeting 2004, www.cshl.edu). Less than 2% of the genome sequence encodes for proteins. Repeated sequences, not coding for proteins make up at least 50% of the sequence. The human genome has a much greater portion (50%) of repeat sequences than the mustard weed (11%), the worm (7%), and the fly (3%)(Lander et al. 2001; Venter et al. 2001) (Human genome project information web site, 2004). Chromosome 1 has highest number of genes (2968), and the Y chromosome the lowest (231).

In addition to the human DNA sequence, the HGP also aimed to sequence the genomes of model organisms, which serve as outstanding tools for the identification of the genes and their regulatory elements as well as the functional protein domains. Comparative genomics provides a key informational tool for understanding the functions of human genome as well as defects resulting in human diseases. For example, approximately 80% of the 30000 human genes seem to have one single identifiable ortholog in the mouse genome, whereas the proportions of human genes without any detected homologous in the mouse is less than 1% (Waterston et al. 2002). To date, the genomes of numerous model organisms have been sequenced (NCBI and NHGRI web pages) including the completed genome sequences of *E. coli (Pennisi 1997)*, *S. cerevisiae (Goffeau et al. 1996)*, *C. elegans (1998)*, *D. melanogaster (Adams et al. 2000)*, as well as whole-genome drafts of species like *C. briggsae*, *D. pseudoobscura*, mouse (Mus musculus) (Waterston et al. 2002), and rat (Rattus Norvigicus) (Gibbs et al. 2004). In addition, sequence efforts are on the way for genomes of other mammalians like the oow (Bos taurus), Pig (Sus scrofa), and Dog (boxer breeder). Information of these genomes will add profoundly to our understanding of the evolutionary processes of the last 100 million years, and will greatly facilitate identification of diseases genes.

6.3 Beyond the HGP in the 21st century

The analytical power arising from the reference DNA sequences of entire genomes provides to an era that has been predicted to be the century of biology. One of the greatest impacts of having the genome sequences available lie in our entirely new possibilities in biological research. We have entered the genomic era and the potential of the HGP could be used to improve human health and well-being and fighting disease (Collins et al. 2003). The HGP set the challenges for the future to understand the geography and function of the genome (DNA sequence organization, chromosomal structure and organization, noncoding DNA types; amount; distribution; information content; and functions) and opened avenues to research problems which could not be addressed before: precise gene number and function, gene regulation, the spatiotemporal expression pattern of the human genes, novel metabolic pathways and protein interactions. We can develop genome-based approaches for disease diagnostics and for predictions of normal biological features like individual drug response. These explorations will result in a more comprehensive view to human biology and provide us with profound understanding of these complex systems.

6.4 Genetic Diversity in Humans, HapMap project

Even though the human genome sequence exists, detailed characterization of the heritable variation in the human genome is needed to increase our understanding of traits and diseases, many of them involving the interplay between multiple genetic and environmental factors (Daly et al. 2001; Peltonen and McKusick 2001; Gabriel et al. 2002; Bersaglieri et al. 2004).

The DNA sequence of any two people is 99.9 percent identical, but the remaining 0.1% is important since it contains the genetic variants that influence how people differ in their risk of disease or their response to drugs. Sites in the DNA sequence where individuals differ at a single DNA base are called **single nucleotide polymorphisms (SNPs).** Sets of nearby SNPs on the same chromosome are inherited in blocks. This pattern of SNPs on a block is a haplotype. Blocks may contain a large number of SNPs, but a few SNPs are enough to uniquely identify the haplotypes in a block. A map of these haplotype blocks and the specific SNPs that identify the haplotypes are called tag SNPs will constitute what is known The

HapMap which will describe the common patterns of human DNA sequence variation (Johnson et al. 2001; 2003). This will make genome scan approaches to finding regions with genes that affect diseases much more efficient and comprehensive by reducing the number of SNPs required to examine. The haplotype map, or "HapMap," is hoped to provide the scientific community with a tool that will allow researchers to find genes (Van Den Oord and Neale 2003) and genetic variations that affect: health and disease, response to environmental factors, susceptibility to infection, and the effectiveness of and adverse responses to drugs and vaccines (Cardon and Abecasis 2003; Deloukas and Bentley 2004), (www.hapmap.org).

7. Identification of disease genes

7.1 Principles and Strategies

A few areas of biological research have progressed as fast as human disease gene identification. The choice of strategy depends on what resources are available for the study (family material, cases control, funding ...etc), and how much is known about the pathogenesis of the disease. In summary the disease gene identification studies involve two steps:

1) Initial disease gene identification

2) Verification of the causative role and the population attributable fraction of the identified variants

Disease gene identification can be performed via functional or positional cloning.

When some information exits about the metabolic disturbance behind the disease, this information can be used to identify the mutated gene. Information of the suspect gene products (proteins) can be used to produce gene specific oligonucleotides or raising specific antibodies for screening cDNA libraries to identify clones encoding the genetic elements of interest.

In positional cloning the disease gene is isolated based on its chromosomal mapping location. Generally this includes two approaches:

a) Position-independent candidate gene approach: if the phenotype of the disease resembles another phenotype in animals or humans for which the gene is known, or if pathogenesis suggests that the gene may be a member of a known gene family, then we sequence that gene directly without knowing the chromosomal location. This approach has only rarely been successful.

b) Positional dependent approach: involves the following steps

1. Identification of the chromosomal region

i) Via genome-wide genotyping and linkage analysis: the most common approach used for mapping Mendelian disorders.

ii) Via identification of chromosomal abnormalities: chromosomal staining can implicate the missing or excessive region.

iii) Via comparative hybridization and/or Loss of heterozygosity screening: commonly used in tumor genes due to common deletion reported in tumors.

2. Physical mapping: This step is getting less and less demanding, thanks to the HGP. Once you mapped the locus, you just go to the databases, find the physical map and the complete sequence of the region you need to analyze for mutations (www.ncbi.nih.gov).

3. Fine mapping: by *using Linkage disequilibrium (LD), haplotype sharing*, usually the linkage analysis maps the region to few cM, which is still considered too large for sequence analyses. In order to restrict the disease locus to small region, one can, especially in founder populations use LD and/or allelic association, and haplotype sharing for fine mapping of the locus.

4. Selection of candidate gene in the restricted region: Even after fine mapping the disease locus the region typically remains too large and contains tens of genes. It is useful to collect all information guide to the most probable candidate genes; This approach called ***positional candidate gene approach***.

This includes looking for genes which show ***appropriate expression pattern and for function of the genes*** in the region, or ***homology to a relevant gene or ESTs in human or model organisms***. Sometimes, the restricted region seem not to harbor any known gene, and one has to use ***computational methods for gene identification*****,** a process known as ***cloning in silico.*** There are many computer software tools for the analysis of genomic sequence which have been developed to predict gene structure like ***GeneID (Guigo et al. 1992)*** and ***Genescan*** (Burge and Karlin 1997) **, or** only exon prediction programs like ***MZEF*** (Zhang 1997) ***and Grail II*** (Xu et al. 1994)**.** Although the accuracy of the prediction programs range between 80-90 % (Burge and Karlin 1997; Zhang 1997), the approach is less time consuming than non computational methods like ***Zoo-blottin*** (Claudio et al. 1994), ***CpG island identification*** (Antequera and Bird 1993), ***hybridization to mRNA/cDNA (Northern blot), exon trapping*** (artificial RNA splicing assay) (Church et al. 1994), or ***cDNA selection/capture*** (Lovett 1994) methods used for identification of genes in a restricted chromosomal region.

5. Testing candidate genes

Ultimately, individual candidate genes have to be tested individually to see if there is compelling evidence that mutations in them do cause the disease in question. This can be performed by various means:

I) ***Mutation screening****:* screening for patient-specific DNA-variants is the most immediate task. If the correct gene is tested, identifying mutations in several unrelated patients; including some with an obvious deleterious effect; and absence of the variant from the control samples strongly suggests a causative role.

II) **Confirmation of the mutation**: Formal proof that mutation is indeed the causative one requires additional evidence. This can be conducted by *Restoration of normal phenotype in vitro* by transfection of a cloned normal allele into a model cell line carrying the disease phenotype can provide evidence that that gene is associated with the disease or *production of a mouse model of the disease* to show some resemblance to humans with the disease, although this may not always be the case even if the correct gene has been identified due to differences between human and mouse biology.

In some other circumstances the confirmation of mutations and the interpretation of mutation screening can be difficult due to several reasons including: *Unsuspected locus heterogeneity* (mutations in the candidate gene account only for a small proportion of the cases tested), *Mutations are not ambiguously pathogenic* (difficult to distinguish pathogenic missense mutations from neutral variants), *Mutations may be hard to find* ; this may be due to the large size of the candidate gene, or to difficulties in detecting mutations using the standard PCR methods (e.g. mutations in the *F8C* gene causing severe hemophilia A were initially hard to find because many were large inversions which disrupted the gene) (Strachan 1999).

Most disease genes that have been identified by a positional candidate strategy represent ***Mendelian monogeneic rare disorders***. However, in ***complex disorders*** or ***common disorders*** positional cloning is not an easy way to identify susceptibility genes due to locus heterogeneity, low penetrance and differences between the populations. In most cases no single susceptibility gene will be mutated in every patient with the disease. However, today's sequence information on the web has moved disease gene identification from the map-based gene discovery to sequence-based gene discovery. Precise transcript maps, identification of millions of SNP-variants combined with novel biostatistical strategies will facilitate monitoring of "genome-wide risk profiles" and hopefully expose genetic variants which also predispose individuals to common diseases

7.3 Linkage analysis in disease gene mapping

The aim of genetic mapping is to assign a disease locus to a specific chromosomal region by the use of the genome-wide marker set and linkage analyses. Linkage analysis (typically in families) calculate how often two loci (a marker and unknown disease locus) are separated by meiotic recombination through monitoring the co-segregation of alleles in subjects from families with trait under investigation.

In the early days the first genetic markers used were **restriction fragment length polymorphism (RFLP)** (Botstein et al. 1980), tediously identifiable with restriction enzymes, in the PCR era, a switch to use **short tandem repeats (STR),** also known as **microsatellites** has taken place. STR has numerous advantages over RFLPs; they are highly informative, easy to type, and even dispersed throughout the genome. STR usually consists of 10-50 copies of di, tri, tetra, or pentanucleotides repeats. The most common repeats $(CA)_n$ are distributed evenly at 30-50 kb intervals in the human genome.

The principle of linkage analysis is based on fact that for any two loci on the same chromosome the chance of cross over event between them (recombination) during the meiosis depends on distance between them. The closer two loci are to each other, the smaller is the chance for recombination between them. The measure of genetic linkage is the recombination fraction, theta ($0 \subseteq \Theta \subseteq 0.5$), which is a measure of the distance between two loci defined by the frequency at which a cross over event took place in the observed number of meiotic events. An estimate of $\Theta = 0.5$ is consistent with the two loci being unlinked (this can be considered as two loci on different chromosomes, or located more than 50 cM on the same chromosome). It can be performed in a parametric manner (when mode of inheritance is known, other parameters can be justified, as the case in Mendelian disorders) or nonparametric (when the mode of inheritance is not known, or the trait does not obey Mendelian rules like in complex disorders). In Mendelian monogenic diseases tests of linkage are mostly based on maximum likelihood estimation and likelihood ratio testing, also called parametric lod score analysis. In linkage analysis the overall likelihood of the data on two alternative assumption is calculated in which the likelihood (L) that two loci (i.e. marker and disease locus) are linked with recombination fraction (Θ) is compared to the likelihood that they are not linked (Θ=0.5). The logarithm to the base 10 ($\log_{10}$) of the ratio of these two likelihoods is what is called the lod score (Z) (Morton 1955).

$$Z(\theta) = \log_{10}\left[\frac{L(\theta)}{L(0.5)}\right]$$

The most likely distance between two loci (e.g. marker and a disease locus) is the recombination fraction (Θ) at which the two point lod score peaks. Traditionally, an odd ratio of more than 1000:1 in favour of linkage (expressed on a logarithmic scale as LOD score of 3) is considered a statistically significant demonstration of linkage. An odd ratio of < 1:100 (LOD score <-2) is agreed to demonstrate the exclusion of linkage (a 100 times likely not to be linked) (Morton 1955; Ott 1999). To extract more information about a region, a multipoint lod score analysis, using the information of adjacent marker loci.

Computer based program packages have been developed to facilitate the calculation of linkage such as LINKAGE (Lathrop and Lalouel 1984; Lathrop et al. 1984).

The main disadvantage of parametric linkage analysis is the dependence on the estimates for mode of inheritance, penetrance, and gene frequency- parameters usually well enough characterized for a single gene disorders but not in complex disorders (Goring and Terwilliger 2000a; Terwilliger and Goring 2000). Model free nonparametric methods like Affected Sibpair analysis (ASP), homozygosity mapping, and association analysis have been used for complex disorders to circumvent some of the problems. There are many computer programs have been developed to carry out these analyses. These programs include GENEHUNTER (Kruglyak et al. 1996), MAPMAKER/SIBS (Kruglyak and Lander 1995), and SIMWALK (Sobel and Lange 1996).

7.4 Linkage disequilibrium

Linkage disequilibrium (LD), is defined as the non random association of alleles of linked markers (Terwilliger et al. 1998). LD can arise from a variety of causes including: recent mutation, population founder effects, recent admixture of populations with different allele frequencies, selection in favour of a specific allele, and demographic history of a population (Slatkin 1994; Laan and Paabo 1997). Several factors influence the level of observed LD, such as the chromosomal region under study, the age and mutual distance of the markers (properties of the markers), the age and the history of the population (genetic drift, population growth and structure, admixture or migration) and selection (Ardlie et al. 2002; Nemeth et al. 2003). LD is disrupted by recombination, mutation, and conversion events. Average LD declines with distance, but there is large variation between different chromosomal regions. **Figure** 4 graphically presents the principle used for LD mapping.

Five different measures have been used for LD (Guo 1997). The correlation coefficient Δ, Lewontin´s D´, the robust formulation of the population attributable risk δ, and Kaplan and Weir's proportional difference *d*. They show that under complete LD between marker and disease locus and of no mutation at the marker or the disease locus, δ is the best measure for fine mapping purposes, and D´ is equally best in many realistic settings (Devlin and Risch 1995). The statistical significance can be measured by traditional chi-square (χ^2) test, Fisher's exact test and the likelihood based lambda (λ) test (Terwilliger 1995). In simulation studies they show that the performance of all measures depends on the marker allele frequency and initial incomplete LD could render all measures useless (Guo 1997).

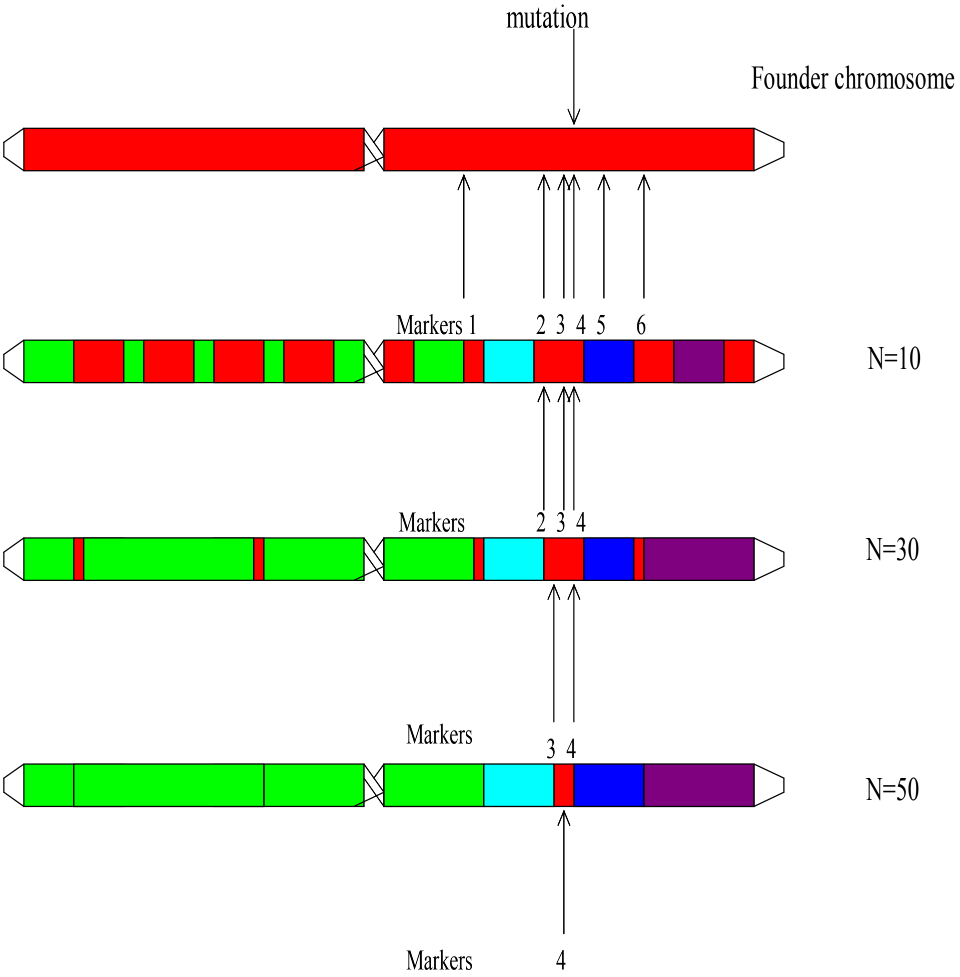


Figure 4. The Principle of linkage disequilibrium (LD) mapping. The conserved region flanking the disease mutation, diminish after generations due to recombination events. In the course of time, only a short chromosomal segment is preserved from the founder chromosome. This occurrence is detected as LD with the typed markers like in the figure only marker 4 will be in LD with disease mutation after long time. The same principle can be used in a similar manner in Haplotype analyses and for evaluating the age of mutation.

LD is proven to be a powerful method for the high resolution mapping of monogenic disorders (Hastbacka et al. 1994; Peltonen et al. 1999). LD and haplotype sharing strategies have provided short cuts to disease gene identification in isolates (Peltonen et al. 2000). The length of LD interval along the disease allele can be utilized to estimate the age of mutations (Varilo 1999). The younger the mutation the more extended is the observable region of LD flanking the disease mutation. This can be seen clearly in the mapping studies of mutations behind Finnish disease heritage (Varilo et al. 1996b; Peltonen et al. 1999).

AIMS OF THE PRESENT STUDY

This study was initiated when no family studies using molecular genetics methods of human lactase deficiencies had been preformed and the genetic loci were unknown and no detailed genetic characterization for the lactase persistence/non persistence variant had been conducted.

The main aim of this work was:

1. To position the locus for the congenital lactase deficiency
2. To identify the molecular variant behind the adult lactase persistence and to address the functional role of the DNA-variant.
3. To study the prevalence of DNA-variant and to analyze the allelic diversity of lactase persistence locus in global populations
4. To address the potential relationship between lactase persistence and diabetes type I and type II in the Finnish population

MATERIALS AND METHODS

1. Study materials

1.1 The samples analyzed in different studies

In total, more than 5000 samples, collected by our clinical collaborators have been analyzed in this project:

I) Congenital lactase deficiency study (CLD) (Study I):

A total of 91 DNA samples comprising of 27 CLD patients belonging into 19 families and 64 healthy family members were enrolled in the linkage study for CLD project. The pedigrees for the 19 families shown into (**Figure** 1, study I). The diagnosis of all patients was based on clinical symptoms in addition an estimation of lactase activity in a jejunal biopsy specimen was performed for all but one patient as described (Savilahti et al. 1983). In addition, in the genealogical study information from 31 families were traced for 3 generations of ancestors on the basis of local church registries (Varilo et al. 1996a).

II) Lactase nonpersistence/persistence studies (studies II, III, IV, V)

a- A total of 145 individual tested by LTTE, from nine extended Finnish pedigrees originally studied by Timo sahi in 1970s for adult-type hypolactasia were enrolled in the linkage study for adult-type hypolactasia. We also enlarged this family study by collecting the DNA samples (19 samples) from family members in the younger generations **(Figure 1**, study II). In total 194 individual >20 years at time of testing were enrolled in the linkage study, in which 90 DNA samples were available for the study.

b- 1037 DNA random samples, collected from Finland, Utah, France, African Americans, Italy, and South Korea were included in the epidemiological study of adult-type hypolactasia (study II).

c- 250 intestinal biopsy sample specimens ascertained for disaccharidase activities were tested for any correlation between the genetic variants identified and adult-type hypolactasia (study II and unpublished data).

d- A total of 52 intestinal biopsy specimens from an independent set of 52 patients with abdominal complaints were collected prospectively to study relationship between the LPH mRNA and C/T$_{-13910}$ and G/A$_{-22018}$ polymorphisms associated with lactase persistence (Study III).

e- A total of 1455 patients with type 1 and 615 with type 2 diabetes in addition to 446 non-diabetic subjects from Western Finland have been studied for the relationship between lactase persistence and diabetes (study IV).

f- A total of 1611 samples from 37 populations of 11 different global regions were collected for the study addressing the age and the origin of lactase persistence mutation (study V) Table 2.

g- In addition, 122 anonymous Finnish families with both parents and one child were analyzed for LD and haplotype block analyses of the LPH locus (study V).

Table 2. The population samples analyzed in the study V.

Region	Population	Population groups
A. East Asia:	South Korea	
	Han Chinese	
B. Trans-Urals:		
I. East Urals:	Ob-Ugric speakers	
II. West Urals:	Komi	
	Udmurts	
	Mokshas	
	Erzas	
	Saami	
	Finns-East region	
	Finns-West region	
C-Caucasus:	Dahgestan:	Druss
		Nog
		Mixed
D-West Asia:	Pakistan:	Somi Balti
		Burusho
		Kashmiri
		Kalash
		Pathan
		Hazara
		Bal balcuh
		Sindi
		Brahui
		M.Baluch
		Mohanna
		Parsi
	Iran:	Iraninans
		Nomad Iranians
	Arabians	
E-West Mediterreans:	South Italians	
	France (part of CEPH)	
	Basques	
F-North America:	Utah (CEPH collection)	
G-East Africa:	Somalians	
H-Sub Sahara Africa:	Fulani-Sudanese	
I-North Africa:	Saharawi	
	Morrocans	
J-African-Americans		

1.2 Assay of Intestinal disacchridases

The assay of intestinal disacchridases is based on the fact that the activity of the intestinal disacchridase studied correlates directly to the amount of glucose liberated when the intestinal mucosal preparations are incubated with the respective substrate disacchrides. The assay is carried out using Tris-glucose oxidase (TGO) reagent (Dahlqvist 1964; Messer and Dahlqvist 1966). Lactase, maltase and sucrase activities were determined according to the method of Dahlqvist (Dahlqvist 1964). The assay is carried out as follow: first, the intestinal biopsies were weighted, solubilized in cold NaCI (100 ul/mg tissue) and homogenized on ice for 45 seconds and centrifuged for 15 min (4500 rpm). Second, supernatants diluted in cold NaCI, pipetted on a Deep Well plate, mixed by vortexing and incubated at 37 °C for 60 minutes. After that 300 ul of ice cold glucose oxidase (GO) reagent was added to each well and incubated again in 37 °C for 60 minutes. Third, the absorbance of the samples was measured at the wavelength of 450 nm on a Multiscan spectrophotometer and the activity of the diasacchridases calculated as units of enzyme activity (U). One unit is defined as the activity of a disacchridase needed to hydrolyze 1 umol of dissachride per minute.

Lactase deficiency was diagnosed when the activity of lactase was below 20U/g protein and lactase/sucrase ratio <0.3. Activities of maltase and sucrase were determined in order to exclude patients with secondary lactase deficiency. Low activities of all disaccharidase enzymes are seen when mucosa is injured, in cases such as celiac disease or inflammation, or if the biopsy sample is obtained from a too proximal part of the intestine (Kolho and Savilahti 2000).

1.3 Lactose Tolerance Test with ethanol (LTTE)

LTTE was performed in the families included in the adult-type hypolactasia study. Simple LTTE was conducted because it is more reliable than LTT (Sahi et al. 1973; Sahi 1974a; Sahi 1974b). After overnight fasting the patients were given 0.3 g/Kg of ethanol by mouth to inhibit galactose metabolism in the liver (Tygstrup and Lundquist 1962) and 15 minutes later 50 g lactose was given as a 12.5% solution. Capillary blood samples were taken before, 20 minutes and after 40 minutes after the lactose for determination of blood glucose and galactose concentration by the glucose oxidase and galactose oxidase methods.

The results of LTTE were expressed as a maximum rise in the blood glucose and galactose concentration in mg/100 ml above the fasting level. The criteria for hypolactasia were:
a) a maximum rise in blood glucose < 20mg/100 ml and
b) a maximum rise in blood galactose < 5 mg/100 ml.
If the subject fulfilled both criteria, a glucose-galactose tolerance test with ethanol was performed to exclude general malabsorption and the possibility of secondary hypolactasia. They were given 25 g glucose + 25 g galactose in the same amount of water as lactose in LTTE. If minimum rise was at least >20 mg/100 ml for glucose, 5 mg/100 ml for galactose, the secondary cause can be excluded. If one of the sugars shows absorption below the limit then the subject was sent for a small intestinal biopsy and histological examination to determine the disaccharidase activities.

1.4 DNA extraction

Genomic DNA was extracted from frozen peripheral blood, in accordance with standard protocols (Vandenplas et al. 1984). In addition, DNA was isolated from intestinal biopsy specimens after assaying dissachridase activities. The samples were moved into a proteinase-K buffer (0.5 % SDS, 0.1 M NaCI, 50 mM Tris pH 8.1, 20mM EDTA) and proteinase K (20 mg/ml) was added in order to dissolve the proteins. After an overnight incubation in 37°C the DNA was extracted by phenol-chloroform extraction (Sambrook 1989). Precipitation of DNA was done in - 70 °C for 15 min after addition of 2.5 vol cold ethanol and 1/10 vol 3 M NaAc. Finally the samples were centrifuged, washed with 70% ethanol, lyophilized and solubilized in 50 µl dH_2O.

2. DNA analysis

2.1 Genotyping

For genotyping in study I, a total of 10 polymorphic microsatellite markers on chromosome 2 q (Dib et al. 1996) were amplified by PCR using ^{32}P- γ ATP- labelled primers in PCR (Sambrook 1989). The order of markers was obtained mostly from the physical YAC contig map of chromosome 2 (Chumakov et al. 1995), supplemented with data from the Généthon map (Dib et al. 1996). In study II, a total of 16 microsatellite markers within the contig constructed over the LPH region were identified from the published genomic sequence of the BACs (NH034L23, NH0318L13, and RP11-329I) using the repeat Masker program were amplified by PCR using ^{32}P- γ ATP- labelled primers in PCR. PCR was performed in a total volume of 15 ul containing 12 ng template DNA, 5 pmol of the primers, 0.2 mM of each nucleotide, 20 mM TrisHCI (pH 8.8), 15mM $(NH_4)SO_4$, 1.5 mM $MgCI_2$, 0.1% Tween 20, 0.01% gelatine, and 0.25 U Taq polymerase (Dynazyme, Finnzymes). The PCR was carried on for 35 cycles, initial denaturation for 3 min then denaturation at 94°C for 30 s, annealing at various temperatures depending on the perimers, for 30 s, and extension at 72°C for 30 s; and final extension for 5 min. The amplified PRC fragments were separated on 6% polyacrylamide gels, and autoradiography was performed.

SNPs of LPH region were analyzed using either minisequencing (Syvanen et al. 1993) (SNPs 13-21), or (SNPs1-11, 23-29) using the primer extension on the chip (Pastinen et al. 2000). In addition, 3.5 kb deletion/insertion (del/Ins) polymorphism within intron one of the LPH gene was genotyped by analysis of PCR product on 1.5% agarose gel. PCR for the del/Ins was designed using 3 primers in the reaction which give to one band 900 bp for the insertion allele and 600 bp for the deleted allele.

A total of 8 SNPs and one del/ins polymorphism were genotyped in 37 populations. The DNA fragments spanning the analyzed SNPs were amplified using a biotinylated (4-10 ng), and an unbiotinylated primer (20 ng), in a 50 ul volume with genomic DNA (100ng), dNTP (200uM), 0.5 U of Taq polymerase (Dynazyme, Finnzymes) in a standard buffer. The PCR procedure was carried out as reported before **(Figure 5).**

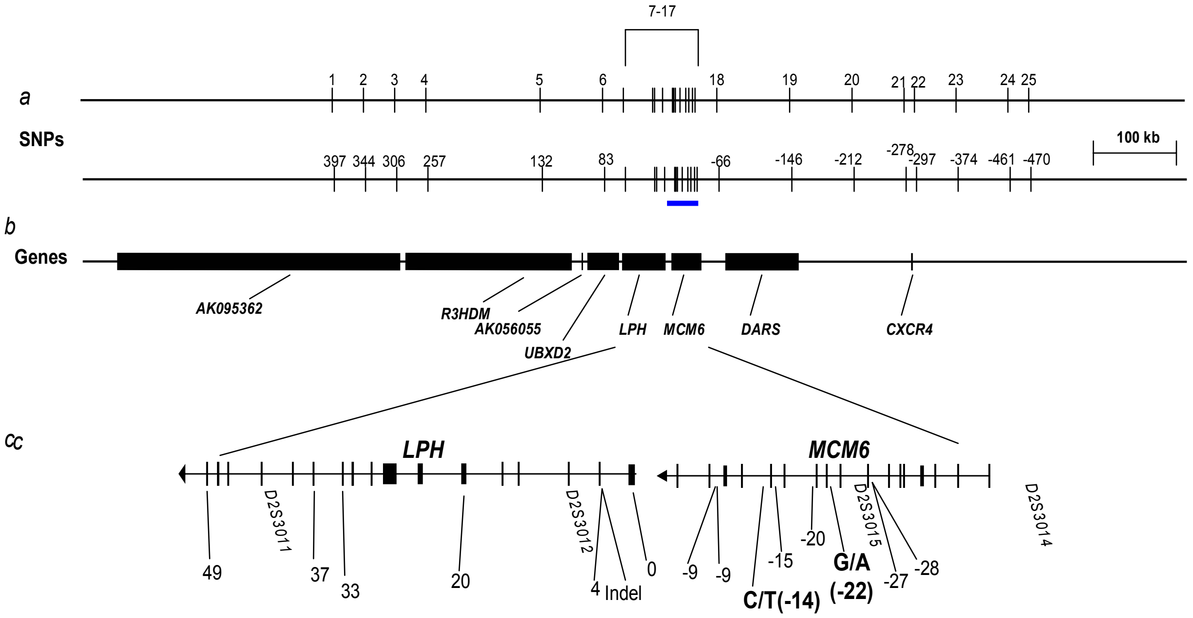


Figure 5 . (*a*) Map of the 24 SNP sites and one deletion-insertion. Number of kb from the first ATG of LPH is shown. (*b*) Genes in the region studied. (*c*) Expanded map of the LPH and MCM6 region.

2.2 Sequence and mutation analysis

The draft genomic sequence of the BACs: NH0034L23, NH0218L22 NH0318L23, and RP-329I10 that covered the critical region of adult-type hypolactasia were assembled to one contig using Sequencher 4 software (Gene Codes Corporation). Oligonucleotide primers spanning the critical region between markers D2S3013 and D2S3014 were designed and PCR amplifications were carried out in a 50 µl volume with genomic DNA (100 ng), primers (20 ng each), dNTPs (200 µM), 0.5 U of *Taq* polymerase (Dynazyme, Finnzymes) in a standard buffer. Most PCR were amplified using the following PCR cycle conditions: an initial round of denaturation at 94 ^{0}C for 3 min, then 35 cycle at 94^{0}C at 30 s, 55 ^{0}C for 30 s, and 72 ^{0}C for 1.25 min and a final extension of 72 ^{0}C for 10 min, except that in cases where the size of the PCR products were more than 1kb when we used the Dynazyme extend kit.

Purified PCR products (15-40 ng) were cycle sequenced using BigDye terminator chemistry (PE Biosystems). Data were analyzed using ABI Sequencing Analysis 3.3 (PE Biosystems) and Sequencher 4.1 (Gene Codes).

2.3 Solid-Phase Minisequencing

The DNA fragment spanning SNPs analyzed was amplified using a biotinylated and an unbiotinylated primer (study II, III, IV and V). An aliquot (10) µl of the PCR product was captured in a streptavidin coated microtitre well (Labsystems, Finland). The wells were washed, and the bound DNA was denatured as described previously (Syvanen et al. 1993). Typically, 50 µl of the minisequencing reaction mixture contained 10 pmoles of the minisequencing primers for each SNP and 0.1 µl of either ^{3}H-dNTP corresponding to the first and second allele of each SNP (Amersham, UK), and 0.05 units of DNA polymerase (Dynazyme II, Finnzymes). The microtiter plates were incubated for 20 min at 50 ºC, and the wells were washed. The detection primer was eluted, and the eluted radioactivity was measured using scintillation counter (1450 Microbeta; Perkin Elmer). Two parallel minisequencing reactions were carried out for each PCR product. The ratio (R) between the incorporated labels was calculated by the equation: R= count per minute (cpm) detecting nucleotide associated with persistent allele/ cpm detecting nucleotide associated non persistent allele.

2.4 RH mapping

A high resolution whole genome (TNG) RH panel (Research Genetics) comprising 90 human-hamster hybrid lines was analyzed together with a human lymphoblastoid cell line (RM; positive control) and non irradiated hamster recipient cells (A3; negative control). The markers were amplified, by PCR, from each RH well. PCRs were performed with 15 ng template, 60 pmol of the primers, 200 uM each dNTP, and 0.5 U Tag polymerase (Dynazyme, Finnzymes) in 50 ul buffer under the condition described above. The PCR products were visualized on 1.5% agarose gels. The screening results for the TNG panel were analyzed by use of the RH server at SHGC, and the RHMAP statistical package for RH mapping (Cox et al. 1990; Boehnke et al. 1991; Olivier et al. 2001).

2.5 Physical mapping

2.5.1 For CLD locus

At the time when we begun the project a considerable time was needed in constructing a physical map over the CLD region. The initial YAC contigs were assembled by CEPH/Généthon. YAC clones previously assigned to chromosomes 2q were ordered from the Sanger Centre (Medical Research Council, United Kingdom). The presence of the known markers of the CLD locus and three primers pairs designed from exon 1, 8, and 17 (SHGC-10723) of the LPH cDNA (Mantei et al. 1988) were tested by PCR. Similarly, a P1-derived artificial chromosome (PAC) library on 321 multiwell microtitration plates (Kindly provided by Prof. Peter de Jong, Roswell Park Cancer Institute)(Ioannou et al. 1994) was screened for the presence of the LPH gene and 5 closest markers. The PCR positive clones were picked up and cultured in Luria broth supplemented with 25 ug kanamycin/ml, and DNA was extracted from these cultures, in accordance with the standard alkaline-lysis method (Sambrook 1989). The PCR conditions were similar to those used in family studies.

2.5.2 For adult-type hypolactasia locus

The draft sequence of the BACs over the LPH region was published in 1999. These published sequences were screened with the known markers and other new markers in the region using Repeat Masker program and Sequencher 4 soft ware (GeneCodes) **(Figure 2, study II).**

3. Linkage and LD analyses

Pairwise LOD scores were calculated by use of the MLINK option of the LINKAGE program package (Lathrop and Lalouel 1984). An autosomal recessive mode of inheritance (Savilahti et al. 1983) for CLD with compete penetrance and disease allele frequency of 0.001 was assumed. LD analysis were performed by using the program HRRLAMB (Terwilliger 1995). This program applies a likelihood-ratio test for LD, calculated as parameter λ, that is independent of the number of alleles and, by extension, of the number of marker loci over a small chromosomal region. In addition, all genotyping data were analyzed using the HRRMULT program, which is designed for joint analysis of multiple loci. In this method, the recombination fraction between any given map position and each of the marker loci is fixed, and the likelihood is maximized, at that map position, over α (proportion of disease originally associated with a certain allele) and n (number of generations since introduction of the founder disease allele into the population).

For adult-type hypolactasia pairwise LOD scores were calculated using the MLINK option of the LINKAGE program package (Lathrop and Lalouel 1984). Autosomal recessive inheritance with complete penetrance, no sex difference in recombination fractions, and a disease allele frequency of 0.4 was assumed. Only individuals above 20 years of age were included in the study as the condition is manifested by that age in the Finnish population (Sahi et al. 1973; Sahi 1994a). The affection status for individuals not confirmed by LTTE was regarded as unknown. Allele frequencies and heterozygosity for the markers were estimated from family material using the Downfreq program for purposes of the parametric linkage analysis (Goring and Terwilliger 2000b). Additionally, a pseudomarker linkage and linkage disequilibrium analyses were performed, assuming autosomal recessive mode of inheritance (Goring and Terwilliger 2000c). A test of LD was performed conditional on the detected linkage treating the allele frequencies and the recombination fraction as nuisance parameters (Goring and Terwilliger 2000b; Goring and Terwilliger 2000c).

Pairwise LD between C/T$_{-13910}$ and G/A$_{-22018}$ in 1037 Samples was estimated using the D‘ statistic (Thompson et al. 1988). Haplotype frequencies were estimated by maximum likelihood using the EH program (Terwilliger 1994). D` is calculated as max (D/ D$_{max}$, D/D$_{min}$) : where the disequilibrium measure D = h$_{pq}$— p q, where h$_{pq}$ is the frequency of the haplotype with rare allele at each locus, *p* and *q* are frequencies of the rare alleles at loci 1 and 2 , and D_{max} = min *p(1-p),q(1-q)* if D>0, and D_{min} = -min *pq, (1-p)(1-q)* if D<0. The significance of deviation of D′ from 0 was determined using the statistic $D^2 \sqrt{\frac{N}{p(1-p)q(1-q)}}$ which is distributed as χ^2 with 1 df (Thompson et al. 1988) .

LD statistics were calculated in 122 Finnish trio families using Genepop program version 3.4, in which contingency tables were constructed to test for genotypic (in population samples) or gametic (for the haploid case) linkage disequilibrium for all pairs of loci in each sample. For each locus pair within each sample, the unbiased estimate of the *P*-value as well as the standard error was estimated. Then, a global test (Fisher's method) for each pair of loci was performed across samples. The results of LD analysis were then depicted using GOLD program (Abecasis and Cookson 2000). MDLBlockFinder (Koivisto et al. 2003) was used to predict the haplotype block structure underlying the region and assess the probability of a block boundary between each pair of adjacent markers.

3.1 Estimation the age of lactase persistent mutation using LD

The analyses were carried out using algorithm of Risch et al (Risch et al. 1995), without and with Luria-Delbrűck correction of the genetic clock (Labuda et al. 1996), respectively: g=log δ/log (1-Θ) and g_c=g+go, where go=-(1/d) In (Θxf_d), assuming d=0.085 (mean population growth rate) (Hastbacka et al. 1992) and f_d= 1/d . The estimate is based on the observation, in 33 lactase persistence-bearing and 52 non-persistence-bearing chromosomes, of the frequency of the allele 2 of the marker D2S3014, which showed the highest LD with the trait in the family study. The LD measure (δ) was calculated as follow, δ= (P_{per} - $P_{non\text{-}per}$)/(1- $P_{non\text{-}per}$) in which P_{per} and $P_{non\text{-}per}$ are the frequencies for the marker allele on persistence bearing and non-persistence bearing chromosomes in the Finnish families, respectively (Bengtsson and Thomson 1981)**.** The variance of (δ) was approximated by

$$\text{Var}(\delta)= \frac{1}{(1-pnon\ )^4}\sigma^2_{pnon}\left(\sigma^2_{pper}+(1-pper)^2\right)+\frac{\sigma^2_{pper}}{(1-pnon)^2}\ .$$

An approximation of the confidence interval of δ was then obtained as $\delta\pm1.96\sigma_{\delta}$, where σ^2_{δ}=Var (δ). Three conversion Factors for the genetic distance (recombination fraction,Θ) between the C/T_{-13910} variant and the marker DSS3014 were used. The first 1.2cM=1Mb inferred from the comparison of genetic and Physical maps of the interval D2S1334-D2S2196, which contains the locus. The second 1cM=0.45 Mb, was estimated from the approximate physical distance between markers D2S3010-D2S2196 in the physical map (Fig.2, Study II) to the genetic distance between them from Marshfield genetic map.
The last was the median of the first two estimations.
(http://research.marshfieldclinic.org/genetics/Physical_Maps/chromosome2.htm).

4. RNA isolation, RT-PCR, and quantitation of RNA levels

Duodenal biopsy specimens were homogenized using a sterile syringe and a 20 gauge needle. Total RNA was isolated by the guanidium thiocyanate method (RNeasy mini kit; Qiagen) followed by DNase treatment. RNA was reverse transcribed into complementary DNA (cDNA) using 20 pmol of LPH transcript specific reverse primers in exon one, exon 2, exon 6, exon 10, exon 13, and exon 17, in a 20 ul reaction mixture containing 2 mmol/l Mg CI_2, 50 mmol/l Tris-HCI (pH 8.3), 75 mmol/l KCI, 10 nmol dNTPs, 20 mmol/l DTT, 24 U RNase inhibitor (RNAguard; Amersham Pharmacia Biotech), and 200 U reverse transcriptase, Superscript II (Invitrogen, Bloomington, Minnesota, USA). RT reactions were performed at +50°C for 60 minutes and stored at - 20°C.

PCR was performed in 1 X PCR reaction in 50 ul containing 2 ul cDNA, with the previously described LPH transcript specific reverse primer and a corresponding forward primer, 30 pmol biotinylated and 60 pmol non-biotinylated primer, and 1 U of Taq polymerase.
PCR reactions were performed as described in genotype section. mRNA levels of each PCR product were quantified by the solid phase minisequencing method (Syvanen et al. 1993).
The actual relative amounts of LPH mRNA derived from the two allele of LPH have been determined from the standard curve (Table 2 & Figure 2, study III). The standard curve was constructed from analysis of RNA samples mixed with known proportions of the two alleles of the informative coding SNP (cSNP GA 593).

5. Bioinformatics, population genetics software

During the work for this thesis, flood of information was deposited in databases on daily basis and became an essential tool in molecular genetics. As an example, a considerable time has been spent in this project constructing a physical map for CLD locus, now can be done computationally in a few hours!

Information can be searched with tools maintained by different servers, the most common being the National Center for Biotechnology Information (NCBI). Sequence alignments were carried out using BLAST programs (Altschul et al. 1997) and sequence variations were viewed in SNPs database.

All statistical analyses carried in study V, except for the analysis of the LD intervals were calculated using the Arlequin software package version 2 (Schneider Stefan 2000). All haplotype estimations were made using the expectation-maximization procedure under default parameters (Excoffier and Slatkin 1995). Diversity indices were also calculated using default setting. The minimum-spanning tree among all haplotypes was calculated using Arlequin and was depicted manually for clarification by building a rooted tree in TreeView.

All statistical analysis in study IV was performed using the BMDP statistical package (BMDP Statistical Software, Los Angeles, California, USA). All variables are expressed as mean ± standard error. Frequency differences were analysed with Pearson’s chi-squared test. Differences between groups were assessed by the non-parametric Kruskal-Wallis test. A p-value less than 0.05 was considered to be statistically significant.

The detailed web addresses of the programs and data bases mentioned here are provided in the electronic database information section page 82.

Results and Discussion

1. Mapping of the CLD locus to chromosome 2q21

The LPH gene had been previously assigned to 2q21 (Kruse et al. 1988; Harvey et al. 1993) and this guided us to analyze this region as a candidate region for CLD. We analyzed the segregation of 10 polymorphic markers flanking the LPH gene in 19 CLD families. Significant evidence of linkage (LOD score >3) were observed with eight markers spanning 8 cM on chromosome 2q21 between marker D2S114 and D2S150. The highest lod score, 7.93, was obtained with marker D2S2385 **(Table 1, study I).**

Among 38 CLD (affected) chromosomes, 19 different extended haplotypes could be constructed using 7 linked markers spanning 6 cM **(Figure 3, study I).** The expected founder haplotype: ***cent-6-4-4-2-2-3-5-tel***, was present in 13 (34%) of disease chromosomes. The other haplotypes were separated from the founder haplotype by ancient recombination events. Based on the haplotype analysis the CLD locus could be restricted to 2 cM region between markers D2S314 and D2S2385, forming a core haplotype, 2-2, which was present on 82% (31/38) of affected chromosomes and on none of unaffected chromosomes. Furthermore, 92% of affected chromosomes carried a single allele 2 of the marker, D2S2385, supporting the hypothesis of one major mutation underlying CLD among Finns. This finding is in according with the experience gained from analyzed Finnish disease heritage studies (Peltonen et al. 1999; Norio 2003a; Norio 2003b; Norio 2003c).

A total of 7 markers analyzed, spanning an area >8 cM, provided evidence for LD (P<0.001) in disease alleles using HRRLAMB program, confirming the close linkage. At the time of the study the order of markers on 2q21 and location of LPH gene was not known precisely. The order of analyzed markers was defined by TNG RH mapping panel of the SHGC G3 map of chromosome 2 and by the RHMAP statistical package (Cox et al. 1990; Boehnke et al. 1991) **(Figure 4, study I)**. A physical map using YAC clones and PAC clones over the CLD region was constructed and the position of the LPH gene was confirmed by RH mapping and by construction the physical map between marker D2S114 and D2S442. According to these studies the LPH gene was positioned outside the ancient core haplotype of CLD locus, between D2S314 and D2S2385, and thus could be excluded as a causative gene for CLD. Supporting our conclusion, the mutation analysis of the cDNA coding for LPH gene and its promoter region in one of our CLD patients conducted previously by Italian group has so far not revealed any pathological sequence differences (Poggi and Sebastio 1991) and later on sequence analysis of 19 Finnish CLD patients shows that the C/T$_{-13910}$ and G/A$_{-22018}$ variants associated with adult-type hypolactasia do not correlate with the CLD phenotype providing supportive evidence of separate distinct locus underlying CLD on chromosome 2q21 (study II).

When we retrospectively; compare the data obtained with the complete genomic sequence of the CLD region available now, the order of the markers remain the same but there is a difference in the distances estimates. For example, the distance between D2S314 and D2S2384 was ~200 kb instead of ~840 kb as determined by the RH map, obviously providing a non-sufficient level of resolution **(Figure 4, study I)**.

While tracing the ancestors of the CLD patients in the Finnish population a knowledge of the birth places of grandparent revealed an enrichment in the sparsely populated eastern and northern Finland which was inhabited during the late settlement after the 16th century (Norio 2003b). The geographic distribution and wide LD interval (> 8cM) of CLD would indicate a young mutation, the history of which closely resembles that of infantile cerebellar ataxia, which has been estimated to have been introduced into the Finnish population 30-40 generations ago (Varilo et al. 1996a; Varilo 1999). In addition, the genealogical data shows that the ancestors of the three parents (8%) carrying the more rare allele 7 of D2S2385 were born in three neighbouring villages close to the border between Russia and Finland, a finding suggesting that this allele represents another minor mutation underlying CLD. Currently the work to refine the CLD locus and to isolate the gene underlying CLD is underway.

2. Identification of lactase persistence variant

2.1 Mapping & fine mapping of lactase persistence locus

After mapping the locus for CLD on 2q21, a question we asked was where is the locus for adult-type hypolactasia? Previous studies have shown that the lactase persistence/non-persistence trait is likely to be controlled by cis-acting element(s) residing within or adjacent to the LPH gene rather than by a variation in a trans-acting factor (Wang et al. 1995). In addition, further support for this hypothesis has been obtained by demonstration of a strong linkage disequilibrium (LD) across the 70 kb haplotype spanning the lactase gene (Harvey et al. 1995).

To test for linkage, we analyzed seven polymorphic markers flanking the LPH gene in nine extended Finnish families with adult-type hypolactasia (**Table 1, study II).** Significant evidence of linkage with four markers (Lod score>3), with the highest lod score (7.67) at $\theta=0$ with marker D2S2196 were observed. The linked region was defined by recombination events between markers D2S114 and D2S2385, spanning total of 5 cM. LD was monitored conditional on the detected linkage, treating the allele frequencies and the recombination fraction as nuisance parameters (Goring and Terwilliger 2000b; Goring and Terwilliger 2000c). Only marker D2S2196 showed significant evidence of LD (P=0.00010) in disease alleles, confirming the close linkage.

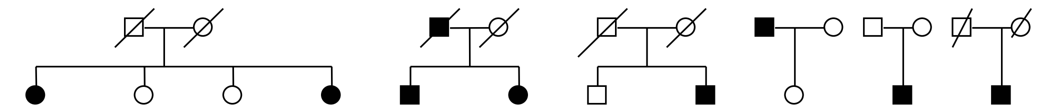
These results led us to analyze more markers for the fine mapping of the region. One Contig spanning 223 kb was built over the region by assembling four BAC clones NH0034L23, NH0218L22, NH0318L13, and RP11-329I10, and the draft sequence of these BACs was searched for new markers. A total of 9 new polymorphic markers, three within the LPH gene (D2S3011, D2S3012 and D2S3013) were found and used for fine mapping **(Figure 2, study II).** Significant evidence of LD ($P<10^{-4}$) was detected with 6 markers spanning 200 kb. The LD was strongest on the LPH gene, and 5´of LPH gene, whereas the markers 3´of LPH gene showed no evidence of LD **(Table 1, study II).**

Extended haplotypes were constructed using the markers (D2S3011-D2S3012-D2S3013-D2S3015-D2S3014-D2S3016-D2S3017) in 9 clinically phenotyped Finnish families. A total of 54 non-persistence alleles and 33 persistence alleles were identified. One major haplotype was present in 20 persistence alleles (60%) versus 3 of the non-persistence alleles (5%), whereas a more diversity of haplotypes was observed in non-persistence alleles. The remaining 13 persistence alleles (40%) differed from the ancestral haplotype in a manner; which was consistent with interruptions of the haplotype by ancient recombinations. Based on the conserved haplotype, the locus for lactase persistence could be restricted to a 47 kb interval between markers *D2S3013* and *D2S3014* **(Figure3, study II**). Furthermore, the highest LD was observed with markers D2S3012 and D2S3014 that showed the conserved core haplotype 5-2 of lactase persistent. Consistent with our data, one major lactase persistence haplotype using SNPs has been reported in other populations by others (Harvey et al. 1995; Harvey et al. 1998). However the earlier efforts to monitor that haplotype were focused solely on LPH gene and sequence in its immediate vicinity (Harvey et al. 1995).

2.2 Identification of the DNA variants associated with lactase persistence in Finnish families

We analyzed the 47-kb region in DNA of seven family members, four with lactase persistence and three with non persistence. The selection of the individuals for mutation screening was based on haplotype analysis. We chose to include individuals with homozygous lactase non persistence that had identical haplotypes to individuals with definitive heterozygous lactase persistence to ensure that we selected the right culprit **(Figure 6)**. The importance of that can be seen clearly in the case of *SNPs: G/A_{-8540} and C/G_{-8630}* , in which complete co-segregation with lactase persistence phenotype is seen except with the key sample DIV4 **(CIV3, and DIV4, table 2, figure 2; study II)**.

The region contains only one known gene, minichromosome maintenance deficient 6 (MCM6), that covers 36 kb of the critical 47 kb (Harvey et al. 1996). No sequence changes in the coding region of MCM6 were detected. For MCM6, unlike LPH, measurements of the mRNA expression in adult small intestine do not show person-to-person variation and the expression does not show restricted tissue distribution. There is no correlation in the levels of the steady state transcripts of MCM6 and LPH (Harvey et al. 1996). These result suggest that there are no functionally significant tissue specific or developmental cis-acting elements shared by the two genes (Harvey et al. 1996) . We identified a total of 52 non-coding variants in the critical MCM6 region: 43 SNPs and 9 deletion/insertion polymorphism (**Table 2, study II).** Only two of the variants (C/T$_{-13910}$, G/A$_{-22018}$) showed complete co-segregation with the lactase persistence trait in families **(Tables 2, 3; study II).** All family members with non persistence were homozygous with respect to both C$_{-13910}$ and G$_{-22018}$. The C/T$_{-13910}$ variant is located in intron 13 of MCM6 gene 13910 bp toward 5´of the initiation codon of LPH and the G/A$_{-22018}$ variant in intron 9 of MCM6 gene 22018 bp upstream of the first ATG codon of LPH.



	Family B				Family B		Family A		Family C	Family D	Family E
	IV3	IV4	IV8	IV9	IV16	IV17	IV3	IV4	IV3	IV4	III2
D2S114	9 1	8 1	9 1	9 9	2 3	3 9	6 1	9 9	9 6	4 5	3 1
P6112	9 11	9 10	9 10	9 11	12 10	10 12	7 8	5 11	9 1	9 9	9 6
D2S1334	9 11	9 10	9 10	9 11	12 10	10 12	7 8	5 11	9 1	9 9	9 6
D2S3010	9 11	9 10	9 10	9 11	12 10	10 12	7 8	5 11	9 1	9 9	9 6
D2S3011	1 6	6 6	1 6	1 6	6 6	6 6	6 6	6 5	6 6	6 6	5 5
D2S3012	3 3	4 4	3 4	3 3	4 4	4 5	4 4	5 2	4 4	2 4	2 2
D2S3013	1 4	3 2	1 2	1 4	2 4	4 5	3 2	5 7	2 3	7 3	7 7
D2S3015	4 5	5 5	4 5	4 5	5 5	5 5	5 5	5 3	5 5	3 5	2 1
D2S3014	5 5	2 4	5 4	5 5	2 2	2 5	2 2	5 5	2 2	4 2	5 5
D2S3016	4 5	5 5	4 5	4 5	5 5	5 3	5 5	3 3	5 5	3 5	3 1
D2S3017	5 3	6 6	5 6	5 3	6 6	5 6	6 6	2 1	6 5	3 6	1 5
D2S3018	8 9	3 3	9 3	9 8	4 3	3 2	3 2	4 4	3 5	4 5	5 7
D2S2196	5 9	6 1	5 1	5 9	3 6	6 9	6 5	5 9	5 4	4 6	6 5
D2S442	1 4	5 5	1 5	1 4	4 5	5 5	3 4	4 1	5 5	3 4	5 5
D2S314	4 5	6 5	4 5	4 5	4 5	5 9	5 4	4 5	5 9	5 2	5 8
D2S2385	5 5	10 5	5 5	5 5	6 6	1 5	6 5	5 4	5 8	1 5	10 10

Figure 6. The haplotypes of the seven family members used for sequencing the 47 kb region.

The restriction of the critical region to 47 kb has been criticized since the localization was claimed to rely on data from two chromosomes that differ from ancestral chromosomes at only a single marker (Grand et al. 2003; Swallow 2003; Bersaglieri et al. 2004). So, these two chromosomes could be derived from a recent mutation at one marker rather than from a recombination event. However, several lines of evidence would support our conclusion. First, the numbers reported in our paper for marker D2S3014 (2/33; 6%) are biased because these families was ascertained through lactase non-persistence probands (Enattah et al. 2002). The analysis of allele 4 of marker D2S3014 in 144 unrelated individuals with verified lactase persistence reveals the prevalence of 12%, identical with the prevalence in 118 unrelated individuals with verified lactase non persistence (unpublished data). The high frequency of allele 4 in Finnish population would support an ancient recombination event. Second, based on the LD, the highest p values for LD occurred at the markers D2S3012 (located within the intron two of the LPH gene) and marker D2S3014, the finding guiding to the 5´end of the LPH gene.

2.3 Sequencing of the regions flanking the lactase persistence/ nonpersistence locus

We extended the sequencing efforts to comprise the whole genomic region of LPH gene and we also extend the sequence effort to the marker D2S3016 **(Figure 2, study II).** The rational for this effort was to find any other sequence variations, within the LD regions, co-segregating with adult-type hypolactasia. A total of 76 SNPs, and 8 deletion insertions among the 7 individuals were identified in the 70 Kb regions covering the entire LPH gene and the 5`of marker D2S3014. None of the variation detected completely co-segregated with adult-type hypolactasia.

Within LPH region, a total of 71 variant were observed; 64 SNPs and 7 del/Ins polymorphisms, in addition to 3 microsatellite markers. We detected variations in all introns of the LPH gene except introns 8, 10, 15, and 16.

A 3640 bp deletion/insertion within intron one of LPH was monitored in the whole family material, and in sample set from different populations. Although the lactase persistence T_{-13910} allele was present almost exclusively in back-ground of the deletion allele, the deletion allele was present in all population tested: in Africa, Asia, and Europe, indicating that the deletion allele was ancient and probably occurred before modern human comes out of Africa.

On average 1 SNP/ 750 bp was thus identified within LPH gene, the coding regions of the LPH gene reveals 8 cSNPs .These variations located in exons 1, 2, 6, 10, 13, 16 and 17. 1 SNP/904 bp was identified within the 47 kb adult-type hypolactasia locus and within the region between markers D2S3014-D2S3016 there was 1 SNP/ 1154 bp **(Tables 3 & 4)**.

Overall in the 115 kb region sequenced, there is on average 1 SNP/ 845 bp, with only one Variant, C/T$_{-13910}$, co-segregating completely with the lactase persistence/ non persistence phenotype.

Table 3. The variations identified within LPH gene.

Position[a]	Variant	Lactase persistence		Lactase persistence		Lactase non-persistence		
		(Homozygous)		(Heterozygous)				
		BIV4	AIV3	BIV8	CIV3	BIV9	DIV4	EIII2[b]
+582/E1[c]	T→C	TT	TT	TC	TT	CC	ND	CC
+982	C→A	CC	CC	CA	ND	AA	CA	ND
+1046	A→G	AA	AA	AG	ND	GG	AA	ND
+2876	T→C	TT	TT	TC	TT	CC	TT	CC
+3706	Δ/I3640 bp	Δ Δ	Δ Δ	ΔI	Δ Δ	II	ΔI	II
+7936/E2	G→A	GG	GG	GG	GG	GG	N	AA
+9070	C→G	CC	CC	CG	CC	GG	GC	GG
+9303	C→T	CC	CC	CC	CC	CC	ND	TT
+9404	T→C	TT	TT	TT	TT	TT	ND	CC
+9715	Δ/I TCTC	II	II	II	II	II	ND	Δ Δ
+10204	T→A	TT	TT	TA	TT	AA	TA	AA
+10651	G→T	GG	GG	GG	GG	GG	GG	TT
+11727	T→G	TT	TT	TG	TT	GG	GT	GG
+11845	G→T	GG	GG	GG	GG	GG	GG	TT
+11857	A→G	AA	AA	AG	AA	GG	AA	AA
+11860	G→A	GG	GG	GG	GG	GG	GA	AA
+12502	G→A	GG	GG	GG	GG	GG	GA	AA
+12658	C→T	CC	CC	CC	CC	CC	CT	TT
+13362	G→A	GG	GG	GG	GG	GG	GA	AA
+14147	T→A	TT	TT	TT	TT	TT	TA	AA
+15493	T→C	TT	TT	TT	TT	TT	TC	CC
+15876	T→C	TT	TT	TT	TT	TT	TC	CC
+15991	A→G	AA	AA	AA	AA	AA	AG	GG
+16837	T→C	TT	TT	TC	TT	CC	TC	CC
+17509	C→G	CC	CC	CC	CC	CC	CC	GG

+18288	C→T	CC	CC	CT	CC	TT	ND	CC
+18399	G→A	GG	GG	GA	GG	AA	ND	AA
+18426	A→T	AA	AA	AT	AA	TT	ND	TT
+20150	G→A	GG	GG	GA	GG	AA	GA	AA
+20492	T→G	TT	TT	TG	TT	GG	ND	TT
+20646	T→C	TT	TT	TC	TT	CC	ND	TT
+22108	G→A	GG	GG	GA	GG	AA	GA	AA
+22519	G→A	GG	GG	GA	GG	AA	GG	GG
+23487/E6	A→C	AA	AA	AC	AA	CC	ND	AA
+24065	$A_8 \rightarrow A_7$	A_8 A_8	A_8 A_8	A_8 A_7	A_8 A_8	A_7 A_7	A_8 A_8	A_8 A_8
+24369	T→C	TT	TT	TC	TT	CC	TT	TT
+25253	Δ/I T	II	II	II	II	II	II	Δ Δ
+25418	A→G	AA	AA	AG	AA	GG	AA	AA
+26391	G→C	GG	GG	GC	GG	CC	GC	CC
+26694	A→T	AA	AA	AT	AA	TT	AA	AA
+26845	A→G	AA	AA	AG	AA	GG	AA	AA
+28063	T→A	TT	TT	TA	TT	AA	TT	TT
+30232	C→T	CC	CC	CC	CC	CC	CC	CT
+30463	Δ/I T	II	II	II	II	II	II	Δ Δ
+30484	A→C	AA	AA	AC	AA	CC	ND	CC
+33358	C→T	CC	CC	CT	CC	TT	CC	CC
+35438	G→C	GG	GG	GC	GG	CC	GG	GG
+36080/E10	G→C	CC	GG	CG	GG	GG	GG	GG
+36214/E10	C→T	CC	CC	CT	N	TT	ND	CC
+38684	A→G	AA	AA	AG	AA	GG	AG	AA
+40530	A→G	AA	AA	AG	N	GG	ND	AA
+41369	A→G	AA	AA	AG	AA	GG	AG	GG
+41883	G→A	GG	GG	GA	GG	AA	GG	GG
+42506	G→A	GG	GG	GG	GG	GG	AG	AA
+43029/E13	G→A	GG	GG	GA	GG	AA	ND	GG
+43798	T→C	TT	TT	TT	TT	TT	TC	CC
+43888	G→C	GG	GG	GG	GG	GG	GC	CC
+43901	G→A	GG	GG	GG	GG	GG	GA	AA
+44166	$A_{21} \rightarrow A_{20}$	$A_{21/21}$	$A_{20/20}$	$A_{21/21}$	ND	$A_{21/21}$	$A_{20/20}$	$A_{20/20}$
+45059	C→T	TT	ND	ND	TT	ND	ND	CC
+45170	A→G	AA	ND	ND	AA	ND	ND	GG
+45506	I/ΔTAT	ΔΔ	ΔΔ	ΔΔ	ΔΔ	ΔΔ	ΔΔ	II
+45507	A→T	AA	AA	AT	AA	TT	AT	TT
+45513	A→T	AA	AA	AT	AA	TT	AT	TT
+45679	G→C	GG	GG	GG	GG	GG	GC	CC
+46177	G→T	GG	GG	GG	GG	GG	GT	TT

+46186	C→T	CC	CC	CC	CC	CC	CT	TT
+46448	C→T	CC	CC	CC	CC	CC	CT	TT
+46948	A→G	AA	AA	AG	AA	GG	AG	GG
+48533	A→G	AA	AA	AG	AA	GG	AG	GG
+51334/E16	C→T	CC	TT	CT	ND	TT	ND	TT
+52577/E17	C→T	CC	CC	CT	CC	TT	ND	TT

[a] The number is from initiation translation codon (ATG) of the LPH gene (contig NT 005058), [b] the individuals sequenced from the Finnish families studied, [c] exonic SNP, ND: not determined

Table 4. The variations identified in the region between markers D2S3014 and D2S3016

Position[a]	Variant	Lactase		Lactase		Lactase		
		persistence		persistence		non-persistence		
		(Homozygous)		(Heterozygous)				
		BIV4	AIV3	BIV8	CIV3	BIV9	DIV4	EIII2[b]
-45499	A→G	AA	N	AG	AA	GG	ND	ND
-46662	A→G	AA	AA	AG	AA	GG	AA	AA
-47713	T→C	TT	TT	TC	TT	CC	C	TT
-49181	A→G	AA	ND	AG	AA	GG	ND	GG
-52181	G→A	GG	GG	GA	GG	AA	ND	GG
-54244	A_{11}→A_{10}	$A_{11/11}$	$A_{11/11}$	$A_{11/10}$	$A_{11/11}$	$A_{10/10}$	$A_{10/10}$	$A_{11/11}$
-57374	G→A	GG	GG	GG	GG	GG	GA	AA
-57775	G→A	AA	AA	AA	AA	AA	AG	GG
-57808	G→A	AA	AA	AA	AA	AA	AG	GG
- 58140	G→A	GG	GG	GA	GG	AA	GG	GG
- 58236	T→C	TT	TT	TC	TT	CC	TC	TT
- 59133	C→T	CC	CC	CC	CC	CC	ND	TT
- 59210	A→T	AA	AA	AA	AA	AA	ND	TT

[a] The number is from initiation translation codon (ATG) of the LPH gene (contig NT 005058), [b] the individuals sequenced from the Finnish families, ND: not determined

2.4 Genotype-Phenotype correlation & implication

We analyzed the two variants in Finnish DNA samples isolated from a total of 196 intestinal biopsy specimens with biochemically determined disacchridase LPH activity. All 59 samples showing primary lactase deficiency were homozygous for to the C_{-13910} allele of the C/T_{-13910}, 6 were heterozygous for to the G/A_{-22018} variant and the remaining 53 were homozygous with respect to the G_{-22018} allele. Among the 137 cases showing lactase persistence, none were homozygous with respect to alleles C_{-13910} and G_{-22018}, at C/T_{-13910} and G/A_{-22018} respectively: 74 were homozygous for alleles T_{-13910} and A_{-22018}, with 63 being heterozygous at both positions. All 53 non Finnish cases with verified disacchridase deficiency (23 from S.Korea, 22 from Italy and 8 from Germany) were homozygous CC_{-13910} with respect to C/T_{-13910}. One Italian sample was heterozygous GA for G/A_{-22018}, whereas the remaining 39 cases were homozygous GG_{-22018} at this position **(Table 3; study II)**. All 7 lactase persistence subjects from Italy were heterozygous CT and GA for C/T_{-13910} and G/A_{-22018} respectively.

Thus far, all the adult biopsy specimen results studied in study II (196 Finns, 23 South Koreans, 30 Italians, and 8 Germans) and in addition to 52 Finns studied in study III, show perfect correlation between the genotype of the C/T_{-13910} variant and phenotype of lactase persistence.

To correlate the C/T_{-13910} genotypes with disaccharidase activity in 222 Finnish adults patients includes in the studies mentioned above aged from 18 to 83 years (unpublished data), the biopsies were assayed for lactase, sucrase and maltase activity. The analysis of disacchridase activities in these samples shows a strong correlation between lactase activity and trimodal distribution of the C/T_{-13910} genotypes ($P<0.0001$). The mean level of lactase activity among CC_{-13910} genotype was 6.86±0.35 U/g protein, among CT_{-13910} genotype were 37.8±1.4 U/g protein, and among TT_{-13910} genotype were 57.6±2.4 U/g protein. In addition, our data also show significant differences in maltase activities among different C/T_{-13910} genotypes ($P=.005$). For sucrase activity, there was no significant differences among the C/T_{-13910} genotypes ($P=0.14$). However, there were no statistical differences in lactase ($P=0.84$), sucrase ($P=0.18$), and maltase activities ($P=0.24$) among different age groups. These results shows that the C/T_{-13910} genotypes correlate with a trimodal distribution of lactase activity, in addition our data show that age per se has no significant effect on the disaccharidase activity in adults aged between 18-83 years **(Tables 5 & 6)**.

Table 5. Distribution of disaccharidase activities in relation to the C/T$_{-13910}$ genotypes.

Disaccharidase[a]	Genotype				
	C/C$_{-13910}$	C/T$_{-13910}$	T/T$_{-13910}$	Total	*P*[b]
N	56	83	83	222	
Lactase	6.86±0.35(3-16)	37.8±1.4(21-99)	57.6±2.4(20-127)	37.4±1.7(3-127)	<0.0001
Sucrase	77.2±3.7(40-172)	87.2±3.8(35-240)	83.2±4.3(24-260)	83.2±2.4(24-260)	0.14
Maltase	287.7±13.2(159-599)	342.0±13.5(161-868)	305.5±14.5(140-887)	314.6±8.2(140-887)	0.005
L/S ratio[c]	0.09±0.005(0.03-0.2)	0.46±0.02(0.2-1.6)	0.73±0.02(0.25-1.35)	0.47±0.02(0.03-1.6)	<0.0001

[a] disaccharidase activities are expressed as U/g protein, mean±SE (range).

[b] The p value based on Kruskal-Wallis test. [c] lactase/sucrase ratio

Furthermore, 329 intestinal biopsy specimens from children and adolescents aged from 0.1 to 20.2 years were assayed for lactase, sucrase and maltase in other study by our team to determine the age of down-regulation of the lactase activity in children of African, Finnish and other Caucasian origins. The results shows that the C/C$_{-13910}$ genotype was associated with very low lactase activity (<10 U/g protein) in all children >12 years of age with a specificity of 100% and sensitivity of 93% of the genetic test (Rasinpera et al. 2004). These data and the unreliability of the indirect diagnostic methods of lactase non-persistence, such as the lactose tolerance test (Arola 1994), and the tedious measurements of lactase activity in jejunal biopsy sample specimens highlights the usefulness of a reliable DNA test as a first-stage screening test for adult-type hypolactasia. The functional evidence of the variant C/T$_{-13910}$ has also been recently reported by others (Olds and Sibley 2003; Troelsen et al. 2003b)

Table 6. Distribution of disaccharidase activities in different lactase variant genotypes for various age groups.

	Age groups (years)				
Disaccharidase[a]	18-39	40-59	60-83	Total	*P*[b]
N	84	94	44	222	
Lactase	35.6±2.4	39.3±3	37±3.5	37.4±1.7	0.84
Sucrase	77.3±2.8	86.6±4.5	87.4±4.5	83.2±2.4	0.18
Maltase	298±11	326±16	324±13	315±8	0.24
L/S ratio[c]	0.48±0.03	0.48±0.03	0.43±0.04	0.47±0.02	0.70

[a] disaccharidase activities are expressed as U/g protein, mean±SE (range).

[b] The p value based on Kruskal-Wallis test. [c]lactase/sucrase ratio

2.5 LD analysis of the LPH locus

We analyzed a total of 24 SNPs and one deletion/insertion polymorphism in 122 Finnish trios to analyze the intermarker LD and the haploblock boundaries at the lactase persistence locus. The analyzed region spans over 0.8 Mb, reaching from 470 kb 5` of the LPH first ATG codon, to 400 kb 3 ′ from the LPH ATG codon (**Figure 1; study V).** This set of markers should be sufficient to define the major haploblocks of the critical LPH region in various populations.

Inter-marker LD was determined for the alleles carrying lactase persistence variant, T_{-13190}, and for the alleles carrying the lactase nonpersistence variant, C_{-13910}, using Genepop program **(Figure 3, study V).** The pattern of LD on lactase persistence alleles show some level of LD (P<0.05) between all the marker pairs tested as can be observed in the plot indicating the pair-wise LD **(Figure 3a-c, study V).** However, lactase nonpersistence alleles showed less LD between most of the markers, resulting in a more rapid decay in LD intervals. These results are in good agreement with the previous data obtained with microsatellite markers that showed a significant LD up to 350 kb flanking the LPH gene (Enattah et al. 2002).

In order to further explore the genomic structure of the LPH region, we analyzed the haploblocks over the region. One haploblock extends over the complete genomic region covering over 700 kb in the lactase persistence alleles, whereas in the lactase nonpersistence alleles three blocks can be defined with sizes of 300 kb, 250 kb and >300 kb **(Figure 3 a-f, study V)**. This can also be seen when a varying number of markers is left outside the predicted haplotype blocks: the haploblock structure shows a more rapid decay in lactase nonpersistence alleles than in lactase persistence alleles **(Figure 3g-i, study V**). These results in good agreement with a recent study which shows that the haplotype containing lactase persistence alleles T_{-13910} and A_{-22018} is almost identical over 800 kb in European Americans (Bersaglieri et al. 2004).

2.6 Analysis of the DNA-Variants in different populations

The prevalence of lactase nonpersistence varies greatly in different populations, from <5% in Northern Europe to almost 100 in South East Asia (Sahi 1994a). In order to explore the prevalence of the C/T_{-13910} in different populations, we screened 938 DNA samples from anonymous Finnish blood donors collected from small parishes in Eastern and Western Finland, and in addition 92 Utah, 17 French, and 96 African-Americans samples. In Finns, the overall prevalence of lactase non-persistence genotype CC_{-13910} (170 cases) was 18.1%., with lower prevalence in the sample from western region (16.8%) than in the eastern region (18.9%, P=0.02, 1 df) **Table 7.** The prevalence of genotype CC_{-13910} in French was 41.2%, in Utah was 7.6%, and in African-American was 79.2 %. These figures were in good agreements with epidemiological data (Cuddenec et al. 1982; McLellan et al. 1984).

The test for heterogeneity of the allele frequencies was performed with maximum likelihood methods under the assumption of Hardy-Weinberg Equilibrium (HWE). No deviations from HWE were detected among allele frequencies. All pairwise comparisons: between all Finland and Utah, all Finland and France, Utah and France, are extremely significant ($p < 10^{-}3$) except for eastern vs western Finland which is marginal (p=0.02) **(Table 7).**

Table 7. Prevalence of the C/T$_{-13910}$ variant in population samples.

DNA samples (n)	Genotype			Allele frequency		% (CC) genotype	*Reported Prevalence*[a]	*P value*[b]
	CC	CT	TT	C	T			
I. Finnish population:								
1. Eastern regions (571)	108	287	176	0.440	0.560	18.9%		Eastern Vs Western Finland: 0.018
2. Western regions (367)	62	159	146	0.385	0.615	16.8%		All population
Total (938)	170	446	322	0.418	0.582	18.1%	*17%*	indentical Vs different : 0.0000001
II. CEPH parents:								Utah Vs France : 0.000003
1. Utah families (92)	7	33	52	0.255	0.745	7.6%	*5%*	All Finland Vs Utah : 0.0000005
2. French families (17)	7	9	1	0.676	0.324	41.2%	*37%*	All Finland Vs France: 0.0001
III. African Americans (AA) (96)	76	15	5	0.87	0.13	79.2%	*78 %*	All Finland Vs AA: $< 10^{-10}$

A total of 938 DNA samples of anonymous Finnish blood donors from small parishes in Eastern and Western Finland, 109 DNA samples from CEPH parents, and 96 samples from African Americans. P values for heterogeneity among allele frequencies were estimated by maximum likelihood methods, under the assumption of HWE assuming. [a] references for the reported population prevalence of adult type hypolactasia (Simoons 1978,Sahi 1994a)

The two variants C/T$_{-13910}$ and G/A$_{-22018}$ were in almost complete LD. The LD between them was calculated in the random Finnish DNA samples cohorts and reveals D′= 0.984 (χ^2 (1 df) =42.41, *P*-value=7.62x10^{-11}) **Table 8**.

Table 8. LD between C/T$_{-13910}$ and G/A$_{-22018}$ variants in random Finnish DNA-samples.

	Genotype at C/T$_{-13910}$			Total	D\`	χ^2(1 df)	*P*-value
	CC	CT	TT				
Genotype at G/A$_{-220018}$							
GG	162	2	1	165			
GA	6	440	3	449			
AA	2	4	318	324			
Total	170	446	322	938	0.984	42.41	7.62x10^{-11}

LD was calculated using D\`statistic, the *p* value is the significance of D\`from 0 as described in the Methods.

2.7 Species comparisons & Similarity Searches

The down regulation of lactase enzyme after weaning is a characteristic feature in mammals. Subsequently, sequence comparisons in different species could shed light on the functional relevance of sequence elements. We sequenced the relevant parts of intron 9 and part of intron 13 of the MCM6 gene of a Baboon (Papio hamadryas). Genotype GG and CC was present in Baboons DNA at SNPs G/A-22018 and C/T-13910. This would imply that alleles G and C, respectively, reflect the appearance of the ancient allele in which the mutations have occurred to produce the persistent status. Comparing the human sequence with other primate sequences (Orange-utans, Chimpanzee, Gorilla Gorilla, Rhesus Monkeys, and Baboons) also reveals the presence of lactase non-persistence alleles C_{-13910} and G_{-22018} at the Variants C/T_{-13910} and G/A_{-22018} respectively in all of them. This provides evidence that the ancestral state of the human lactase gene is the non persistence form. In contrast, searching mouse sequence revealed no conserved sequence between humans and mice in the regions flanking the critical variants C/T_{-13910} and G/A_{-22018}.

Sequence similarity of the region flanking the variants C/T_{-13910} and G/A_{-22018} shows that C/T_{-13910} variant is located within a short region (100 bp) that has a similarity to LINE 2 (L2) repeat element, whereas the G/A_{-22018} is located within full length Alu repeat element.
LINEs are long interspersed nuclear elements, considered to be one of the most ancient transposable elements in eukaryotic genomes. The full-length consensus sequence is 6.1 kb comprising of two open reading frames and an internal promoter within a region of 5`-UTR preceding ORF1 (Hwu et al. 1986; Smit 1996). In the human genome, three distantly related LINE families are found: LINE1, LINE2 and LINE3. Analysis of the draft genome sequence shows that LINEs comprise 20.42% of the human genome sequence, LINE2 family comprises 3.22% of the human genome sequence (Lander et al. 2001). The functional role of LINEs has been reported for L2 which has been shown to modulate the gene expression of annexin VI, potent T-cell-specific silencer (Donnelly et al. 1999). Concerning the location of the G/A_{-22018} variant the human genome contains 500,000 copies of an Alu sequence, which is a 281-bp short interspersed nucleotide repeat (Britten 1994). They have been frequently identified in the transcriptional regulatory regions of numerous genes (Maouche et al. 1994).

A search of the TRNASFAC database (http:/motif.genome.ad.jp/) for cis regulatory elements revealed high score similarity for affects stress-response element (STRE) and alcohol dehydrogenase gene regulator 1 (ADR1) binding sites at C/T$_{-13910,}$ and for Sp1 binding site at G/A$_{-22018}$. ADR1 is a transcriptional activator that regulates genes involved in carbon source utilization in Yeast (Sloan et al. 1999). ADR1 expression is repressed 3-20- fold when glucose is present in the medium (Blumberg et al. 1988; Dombek and Young 1997). The stress response element (STRE) known by the consensus core sequence AGGGG is able to mediate transcription induced by various forms of stress including carbon source starvation, heat shock and severe osmotic and oxidative stresses (Martinez-Pastor et al. 1996). Sp1 is a transcription factor whose interactions with other proteins leads to diverse effects on gene transcription, having a stimulatory effect in most cases, but also a negative effect on transcription (Shou et al. 1998). The potential functional importance of STRE, ADR1 and SP1 binding sites to the lactase persistence allele remains to be determined.

2.8 Mechanism of retaining lactase expression (lactase persistence)

We studied the effect of C/T$_{-13910}$ and G/A$_{-22018}$ on the steady state transcript level of the of LPH gene in intestinal mucosa by relative quantitation of the allele-specific transcripts by RT-PCR followed by minisequencing. The results showed that the expression level of LPH mRNA in individuals with T$_{-13910}$ and A$_{-22018}$ alleles is several times higher that found in individuals carrying C$_{-13910}$ and G$_{-22018}$ alleles **(Figures 3 and 4; study III)**. This finding suggests that these two SNPs are associated with the transcriptional regulation of the LPH gene. However, the molecular mechanism underlying the persistence of high intestinal lactase activity; still needs to be verified.

Recently reported studies have indicated that the DNA region of C/T$_{-13910}$ contains a strong enhancer, and it has been suggested that both lactase persistent T$_{-13910}$ allele and lactase non persistent C$_{-13910}$ allele have enhancer activity (Olds and Sibley 2003; Troelsen et al. 2003b). However, the T$_{-13910}$ allele enhances the LPH promoter activity several times more than the C$_{-13910}$ allele when analyzed in differentiated Caco-2 cells. In addition, a nuclear factor from an intestinal and non-intestinal extract on EMSA (gel shift assay) binds strongly to the T$_{-13910}$ allele whereas the binding to the C$_{-13910}$ allele is much weaker. Furthermore, compared with the C/T$_{-13910}$ region, the G/A$_{-13910}$ variant region result in minimal enhancements of the lactase promoter (Olds and Sibley 2003; Troelsen et al. 2003b).

Although these pieces of evidence imply the direct transcriptional regulation role for these variants, studies looking at the effect of the variants in a stably transfected cell lines or in animal models will be necessary to confirm a role for the C/T$_{-13910}$ variant region in specifying lactase persistence or non persistence.

2.9 Adult-type hypolactasia: A more complex genetic condition?

Historically, it was well known from the beginning that lactase persistence that represents the abnormal condition. Unfortunately earlier family studies referred to the adult-type hypolactasia, implying an abnormal state. This concept has it's root in the fact that these two conditions are considered as two faces of the same coin: if you are not hypolactasic then you are lactase persistent and vis versa. This is true from genetic analysis view point (family studies, linkage analysis), as you will end up with the same result if you use adult-type hypolactasia a recessive condition or lactase persistence as dominant condition. Multiple lines of evidence indicate that we have here identified the allelic variant(s) resulting in the lactase persistence phenotype. However, the situation in vivo is more complex. The expression of the LPH gene has been reported to be regulated by multiple transcription factors and their interactions (such as the caudal homologue Cdx2, HNF1α, HOX11, FREACs and GATA 4,5 and 6 factors (Troelsen et al. 1992; Troelsen et al. 1994b; Troelsen et al. 1997; Fitzgerald et al. 1998; Hollox et al. 1999; Spodsberg et al. 1999; Fang et al. 2000; Mitchelmore et al. 2000; Fang et al. 2001; Krasinski et al. 2001; van Wering et al. 2002a; van Wering et al. 2002b; Troelsen et al. 2003a) and to influence the decline of the LPH enzyme after childhood.

Lactase non persistence (adult-type hypolactasia) should perhaps not be considered a traditional recessively inherited condition, but instead it could be considered as a polygenic condition because it is a normal physiological developmental condition, involves interaction of multiple gene products (transcription factors) with 1-2 kb of the promoter of LPH to control the decline of lactase. We would like to propose that adult-type hypolactasia and lactase persistence could at the molecular level represent two different conditions affecting intestinal cells.

3. Association of lactase persistence with human diseases; type 1 and 2 diabetes as an example

In humans, intestinal lactase activity is increased in diabetes and has been shown to normalize during insulin treatment (Tandon et al. 1975). In a recent study a high frequency of lactose absorbers (lactase persistence, LP) was observed among diabetic patients in Sardinia (Meloni et al. 2001). Milk and milk products form an essential part of the Western diet. In Finland the annual consumption of milk products with an average of 220 kg milk products per capita, and thus a high frequency of lactose absorption among diabetic patients is intriguing.

To test this relationship further, we genotyped 2542 Finnish subjects for the C/T$_{-13910}$ polymorphism including 1455 type 1 diabetes (Pettersson-Fernholm et al. 2003), 615 type 2 patients, and 446 control subjects (Groop et al. 1996) (Table 1, study IV). The frequency of LP genotypes (CT and TT) was 84.7% in type 1 diabetic patients and thus did not differ from the frequency of 81.9% in the general population (Study II). Patients with the CC-genotype were slightly younger and had a shorter duration of diabetes (*P*=0.02) **(Table 2, study IV)**. The frequency of CT and TT genotypes among 465 type 2 diabetic patients from Western Finland was 91.7 % vs. 91.3% in 446 control subjects from the same region (*P*=NS). Among 120 patients with type 2 diabetes patients from Eastern Finland, the prevalence of the persistence CT and TT- genotypes was 82.5 % vs. 81.1% in the general population of Eastern Finland (study II) **(Table 2, study IV**). The LP genotypes were associated with waist to hip ratio (WHR) in type 2 diabetic females from Western Finland (P=0.03) and in type 2 diabetic males from Eastern Finland (*P*=0.03) **(Table 2, study IV**). Thus the analysis of 2516 subjects, could not demonstrate any association between LP genotypes and type 1 or type 2 diabetes. This suggests that lactase persistence is not a risk factor for diabetes in Finland, and the finding contradicts the results obtained in the Sardinians (Meloni et al. 2001). It also highlights the methodology and importance of subject selection in a case-control study of lactase persistence to avoid false positive associations when the prevalence of the condition varies considerably among and between populations (Sahi 1994).

4. Tracing the history of lactase persistence

4.1 Prevalence & geographic distribution of lactase persistence variant *in global populations*

The population frequencies of lactase persistence allele were examined in 37 population samples from different geographic areas **(Figure 3, study V).** Again we use lactase persistence in the prevalence figures instead of hypolactasia. The data are summarized in **Table 1, study V**. The frequency of the lactase persistence (C/T_{-13910}) variant in various populations systematically reflected the reported prevalences of lactase persistence, determined by disaccharidase activities in intestinal biopsies, and/or lactose tolerance tests in these populations (Sahi et al. 1973; Simoons 1978; Sahi 1994a; Hollox et al. 2001).

As expected, the lactase persistent allele was almost non existent among East Asians. In the Finno-Ugric populations, the prevalence of lactase persistence (measured by C/T_{-13910} and T/T_{-13910} genotypes) gradually increases from <10% in the areas east of Ural (Ob-Ugric speakers) to 30-60% in the populations west of Ural (Komi, Udmorts, Mokshas, Erzas, Saami) reaching the peak prevalence of 84% in Finland. In West Asia we analyzed 12 different ethnic groups from Pakistan and found the prevalence distribution to show a gradual increase of lactase persistence allele from north-east (4 out of 6 groups shows < 10% frequency) to south-west (the frequency range in other 6 groups from 28% to 65%). The exceptional prevalence in Kashmiri and Pathan regions could reflect a very young history of these two ethnic groups, in north from Pakistan. Among Iranians, nomadic Iranians and Arabic samples from Iraq, Syria, Lebanon, the region where the agriculture and domestication of animals is considered to begin, a relatively low frequency of lactase persistent allele (< 20%) was observed **(Table1, study V)**. Among Europeans populations, the prevalence in southern part of Italy is as low as 11%, in line with epidemiological data. In France, the frequency is as high as 59%. In Basques, an isolated population from the Paleolithic era of Europe, the prevalence of lactase persistence is very high, up to 92% like the Utah population in Northern America, population with northern European ancestry. In Africa, the prevalence in East (Somalia) is low (<10%), whereas in Northern Africa moderate prevalence of lactase persistence allele is observed in Moroccans 49%, and nomadic population (Saharawi) 63%. In Subsharan Africa, a high frequency of lactase persistence was found among the Sudanese-Fulanese tribe living 11 km east of capital city of Sudan (Khartoum) 70%. This result was in contrast to previous report in Sudanese samples which show the lack of lactase persistence allele. The panel of mixed African-American DNA samples showed a prevalence of 12% **(Table 1, study V).**

4.2 Identification of the likely place of origin based on the haplotype analysis

To gain a better understanding of the allelic diversity underlying lactase persistence/non-persistence, we selected a subset of SNPs to analyze the detailed haplotypes in 37 populations. A total of 8 SNPs and one deletion/insertion polymorphism were monitored, located in 30 kb area flanking the lactase variants **(Figure 5)**. The observed haplotypes were ranked based on their prevalence in the combined sample. A total of 83 different haplotypes were identified in the complete study sample. The prevalence distribution was determined using all 9 SNPs for common haplotypes 1-5 whereas for more rare haplotypes 6-13 the figures represent combined pooled haplotypes based on the 2 SNPs, and deletion/insertion polymorphism to facilitate presentation (**Table 1, Table 2 and Table 3; study V).** The diversity of the LPH haplotypes (which reflecting the heterozygosity) in most populations varied between 60 (±0.03) to 89 (±0.03), the Utah sample showing the lowest value of 0.44 (±0.04).

Four major lactase non persistence haplotypes with a high frequency (>5%) (haplotypes 1-4) were observed in almost all populations tested **(Table 2 and Figure 3; study V**). Haplotype 1-3 which account for majority of lactase non persistence alleles are highly divergent from the major haplotype of lactase persistence, haplotype 5 **(Table 2 and Figure 3; study V)**, from haplotype 1 they differ at every SNP site tested. . In contrast to this, the haplotype 4 differs from haplotype 5 only at the positions of the two critical variants (C/T$_{-13910}$ and G/A$_{-22018}$). For lactase persistence haplotypes, one major haplotype distinctly predominates (haplotype 5) in almost all population tested, with only a few notably exceptions. In Komi and Mokshas, the lactase persistence haplotype 10, in which lactase persistence mutation is atypically found on background of the insertion allele, was observed at high frequency (10% and 11%).

The information provides a picture of the relatively limited haplotype diversity among lactase non persistence alleles, somewhat in contrast to our previous findings using microsatellites (Enattah et al. 2002). For example, in South Korea, Han Chinese, Somi balti and Kalash showing a fixation for lactase non-persistence alleles, the haplotypes 1-4 explain 76- 97% of haplotype diversity in these populations. Since haplotypes 2, and 3 differ only in one SNP, this would imply only two major lactase non persistent alleles, a highly restricted number of ancestral alleles **(Table 2 and Figure 3; study V)**. This is not too surprising since SNP profiles reflect the haploblocks of old alleles whereas the multiallelic markers expose more recent mutational events of the markers.

In our search for the population which could be represent the place of origin of lactase persistent variant and thus provide hints about the evolutionary history of this mutation, we aimed to find populations with a high frequencies and maximal diversity of lactase persistent specific allelic haplotypes. Trans-Ural populations were found to fulfill these criteria **(Table 2 and Figure 3; Study V).** Haplotype 4 represents a lactase non persistent haplotype differing from lactase persistent haplotypes at only two positions, the variants most tightly associated with lactase persistence. This haplotype could be considered as pre-lactase persistence and represent allelic background on which lactase persistence mutation(s) occurred. However, we cannot exclude the possibility that this haplotype could have arisen from haplotype 5 as a result of a gene conversion event.

We monitored the frequency pattern of the haplotype 4 in our global populations to assess how and when the lactase persistence alleles were introduced. The highest prevalence of this haplotype was found in Trans- Ural populations, in the eastern part of Urals, among, among Ob-Ugric speakers, reaching the prevalence of 33%. This population has less than 5% frequency of lactase persistence allele. In the Western part of Ural, the populations located in the region between the Ural mountain and Volga river, the frequency of haplotype 4 is as high as 35% among Komi, Udmurtains, Lapp Saami. The high prevalence of haplotype 4 extends to Han Chinese (36%), to populations totally lacking the lactase persistence allele. The high prevalence of this allelic background could also exist in South Korea, in which the pooled haplotype 7 (1xx1xx1xx) (**Table 2; study V)** in this population representing only one ancestral haplotype (122122122) (22%) carried the same variation of haplotype 4 but lacked the deletion polymorphism (data not shown).

The pooled lactase persistence haplotypes 9 and 10 (which could be considered as earliest haplotypes where lactase persistence variants occurred on an insertion polymorphism background (ancestral lactase non persistence) allele) showed the highest frequencies (5-15%) in the populations of Western slope of Ural mountains like Komi, Udmurtians, Mokshas and Erzayas populations. For example the pooled haplotype 10 in Mokshas represents only one haplotype (111211211) with 11% frequency. This haplotype carries lactase persistence variants on the haplotype background characteristic of lactase nonpersistence haplotype 1 (111111111). This result could imply that the route to lactase persistence could have occurred through two lineages in human history: one lineage is represented on the background of haplotypes 1 and 2 resulting in the less frequent lactase persistence haplotypes 9 and 10. This lineage would represent a very early event **(Figure 4; study V**). The other lineage originates from the background of lactase non persistence haplotype 4, resulting in the most common lactase persistence haplotype 5. This lactase persistence mutation obviously occurred more recent in history **(Figure 4 and Table 2; study V**).

To investigate the phylogenetic relationship between haplotypes, we constructed the Minimum Spanning Tree (MST) of the haplotypes from the matrix of pairwise distances calculated between all pairs of haplotypes using Arelquin program (**Figure 4; study V)** and the tree was depicted using TreeView program. The major haplotypes carrying lactase non persistent allele are shown in yellow (haplotypes1-3) and red (haplotype 4), the major haplotypes carrying both lactase persistent variants are shown in green (haplotype 5) **(Figure 4 and Table 2; study V).** The other haplotypes were pooled based on 2 SNPs (lactase variants: C/T$_{-13910}$ and G/A$_{-22018}$) and deletion/insertion polymorphism of intron one of the LPH gene (Table 2; study V). The color codes for pooled lactase non-persistence and persistence haplotypes are shown in **Figure 4 and Table 2 of study V.**

There are two common lactase non-persistent haplotypes 1, 2 differing by only one mutational step and both these haplotypes are found in the sequence of all primates (Chimpanzee, Orangeutans, Gorilla Gorilla, and Rhesus Monkey). Based on the primate haplotype the tree can be rooted for lactase non persistence haplotypes 1 as ancestral haplotype. Lactase persistent haplotype 5 is highly divergent from haplotype 1, they differ for every SNP (n=9) tested. The lactase non-persistence haplotypes 1, 2 are highly divergent from lactase non-persistent haplotype 4 and the frequency of intervening haplotypes between them are low, being most probably lost due to the drift. Although the most common lactase persistence haplotype seen in all population tested is haplotype 5, we detected traces of other lineage of lactase persistence haplotype (haplotypes 10a-10b-10c-6a). The option that haplotype 5 originates from this relatively rare lineage still remains especially if haplotype 4 has risen via gene conversion from haplotype 5. In both cases the trans Uralic populations have the highest prevalence of haplotype 4 as well as the lineage10a-10b-10c-6a (Figures 3 and 4; study V) which makes this geographical region the most likely focal origin of the most common lactase persistence allele, representing haplotype 5. Finally, the star pattern surrounding haplotypes 1, 2 and 4 is consistent with a history of population expansion.

We used linkage disequilibrium to estimate the number of generations to the most recent common ancestor (MRCA) of lactase persistence trait in Finnish population by analyzing the frequency of the allele 2 of the marker D2S3014, which showing highest LD with the trait in our previous family study (Bersaglieri et al. 2004). Assuming a 20-years generation time, this estimate would indicate that the MRCA dates back between 3200 (95 %CI; 1040-8000) to 4400 (95% CI, 2260-9240) years ago without and with adopting Luria-Delbrűck genetic clock respectively **(Table 9).** This date is in good agreement with the history of nomadic tribes who came to Europe at that time from the east **(Figure 3; study V**) (Cavalli-Sforza 1994). This estimates are in good agreement with recent report in which the estimated date using Scandinavian families between 1625-3188 years ago and using CEPH between 2188-20650 years (Bersaglieri et al. 2004).

Table 9. Estimation of the most common recent ancestor of the C/T$_{-13910}$ variant of the Finnish population.

Marker	*D2S3014*		
Allele	*2*		
Distance Kb	31		
Conversion factor	1.2cM=1Mb[a]	1cM=0.42Mb[b]	1cM=0.625Mb[c]
Θ	0.00037	0.00074	0.00049
P_{per} (N)	0.939 (33)		
$P_{non\text{-}per}(N)$	0.192 (52)		
LD (δ)[d]	0.924		
Estimated age[e]			
g (95%CI)	213 (69-529)	107 (35-264)	160 (52-400)
	6390 (2070-15870)	3210 (1050-7920)	4800 (1560-12000)
g_c (95%CI)	277 (160-593)	163 (91-320)	220 (113-461)
	8310 (4800-17790)	4890 (2730-9600)	6600 (3390-13830)

[a] Based on the comparison of genetic and Physical maps of the interval D2S1334-D2S2196, which contains the variant from Marshfield Map (Medical Research foundation). [b] Inferred from approximate physical distance between markers D2S3010-D2S2196 (Fig. 2) to genetic distance from Marshfield genetic map [c] The median Θ obtained from the previous conversions [d] linkage disequilibrium index calculated as described in the method [e] g and g_c are estimated ages (generations) obtained using of Risch algorithm, without and with Luria-Delbrűck correction of the genetic clock respectively as described in the methods. We used a generation time =30 years.

4.3 The proposed history of lactase persistence

The presence of highly divergent haplotypes like the major lactase persistence haplotype (haplotype 5) and the lactase non persistent haplotype 4 has been recently shown to be very common in human genome, and it is referred as yin yang haplotype (Zhang et al. 2003). In the analysis of 175 random genomic regions in humans, the proportion of the genome spanned by yin yang haplotypes is suggested to be 75-85%. Interestingly, this could be explained by strictly neutral evolution in a well-mixed populations by coalescence simulations (Zhang et al. 2003).

The unsolved issue of selection against lactase nonpersistence allele could result in a biased interpretation of population history. Despite this problem, a plausible hypothesis for the origin of lactase persistence, based on the frequency and diversity of the haplotype the age of lactase persistence allele could be estimated based on LD and incorporating the genetic, archaeological and historical evidences (Simoons 1978; Guglielmino et al. 1990; Sokal et al. 1992; Barbujani et al. 1995; Weng and Sokal 1995; Holden and Mace 1997; Dudd and Evershed 1998; Semino et al. 2000; Derbeneva et al. 2002).

We propose that the genetic differentiation for lactase persistence has occurred within the pastoral nomads in Central Asia adjacent to the eastern slope of the Ural 4800-6600 years ago (supported by high frequency of pre lactase persistence haplotype 4 in this region) **(Figure 3; study V**). These nomads moved west towards the western slope of Ural and in this location the mutation for lactase persistence allele has most likely arisen within pastoral nomads in the region between the Volga and the Urals, north of the Caucasus and the Black Sea (this is supported by high frequency of haplotype 4 and haplotype diversity of lactase persistence alleles in this region). This scenario is in good agreement with the hypothesis of Gimbutas who suggested that there was a westward expansion from the Kurgan area above the Caucasus, of pastoral nomads speaking Indo-European languages between 2500 and 1500 B.C. **(Figure 3; study V**) (Cavalli-Sforza 1994).

CONCLUDING REMARKS

The first publication of this thesis presents the localization and fine mapping of the congenital lactase deficiency (CLD) locus on chromosome 2q21, using linkage, linkage disequilibrium and ancestral haplotype analyses, between markers D2S314-D2S2385, about 1 Mb 5′ of LPH gene. The result was a clear indication that two separate genetic loci underlie congenital lactase deficiency and lactase persistence residing on chromosome 2q21. This result was also confirmed by mutational analysis of the cDNA of LPH gene in one CLD patient by an Italian group, which reveal no sequence changes in CLD, and no correlation between CLD and the variants C/T$_{-13910}$ and G/A$_{-22018}$ associated with lactase persistence. The work to identify the CLD culprit is underway and will be reported elsewhere.

The second publication of the thesis presents the localization and identification of the mutations underlies one of the most common conditions in human lactase persistence/ non persistence trait. Initially we localized the locus flanking the LPH gene between markers D2S114-D2S2385 in nine extended Finnish families. Based on linkage disequilibrium and haplotype analysis we restricted the region to 47 kb interval between markers D2S3014-D2S3012. Sequence analysis revealed two variants, C/T$_{-13910}$ and G/A$_{-22018}$, correlating with lactase persistence/non persistence trait in both families and case-control study materials.
Analysis of both SNPs in different populations gives supportive evidence that SNP C/T$_{-13910}$ is most probably the causing variant of lactase persistence trait, in which one major haplotype carrying the persistent T$_{-13910}$ is present in all populations studied (Hollox et al. 2001; Enattah et al. 2002). The analysis of disacchridase activities in these samples shows a strong statistically significant correlation between lactase activity and trimodal distribution of the C/T$_{-13910}$ genotypes ($P<0.0001$). In addition our data show that age per se has no significant effect on the disaccharidase activity in adults, aged between 18-83 years. This result enabled us to test the potential association between lactase persistence with diabetes and other clinical conditions. Finally, we analyzed the diversity of critical 30 kb region in 36 populations. Based on haplotype analysis, genetic variations, and LD intervals we propose that the region adjacent to the Eastern and Western slopes of Ural is the most likely region of origin for the lactase persistence mutation(s). The major lactase persistent haplotype most probably originated in nomadic population tribe between the western slope of Urals and Volga River some (4800-6600 years ago) (Cavalli-Sforza 1994) and the mutation then spread with

migration of tribes westward towards Europe in the direction from southeast to northwest and in the southeast ward towards Western Asia and Middle East. The presence of one mutation in many populations underlies the lactase persistence/non persistence trait enabled us to develop a DNA based diagnostic test for this condition. While the work of this thesis was proceeding, the functional convincing evidence for SNP C/T$_{-13910}$ emerged in the literature (Olds and Sibley 2003; Troelsen et al. 2003b). Analysis of the critical DNA region in various populations has shed light on the possible place of origin of the mutation and paved the way to study the effect of different evolutionary forces shaping the diversity in this region such as the role of selection, genetic drift and population structure.

Electronic database Information

BLAST service at NCBI, http://www.ncbi.nlm.nih.gov/BLAST/

Cold Spring Harbor Laboratory, http://www.cshl.edu

dbSNP home page, http://www.genome.gov/SNP/

Finnish disease Heritage database, http://www.findis.org

GeneBank at NCBI, http:// www.ncbi.nlm.nih.gov/Genebank/

Genemap 99 at NCBI, http://www.ncbi.nlm.nih.gov/genemap99/

Genome Data Base (GDB), http://www.gdb.org/

Human Genome organization (HUGO), http://www.gene.ucl.ac.uk/hugo/

Marshfield genetic map, http://research.marshfieldclinic.org/genetics/Physical_Maps/

National Center for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.giv/

National Human Genome Research Institute (NHGRI), http://www.genome.gov/

Online Mendelian Inheritance in Man (OMIM), http://www.genome.gov/Omim/

The Baylor College of Medicine, http://searchlauncher.bcm.tmc.edu/

The HapMap project, http://www.hapmap.org

TRNASFAC database, http:/motif.genome.ad.jp/

White head Institute, http://www-genome.wi.mit.edu/

Acknowledgments

First of all, praise and thanks be to God, who provided me with the power and means to accomplish this work. This study was carried out between years 1997-2004 in the Department of Molecular Medicine at the National Public Health Institute in Helsinki.
I wish to thank the former head of the institute, Professor Jussi Huttunen, and the new director Professor Pekka Puska, for providing me with excellent research facilities.

I wish to express my deepest gratitude to my first supervisor, my mentor Academy Professor Leena Peltonen-Palotie, Head of the Molecular Medicine Department at National Health Institute, Helsinki. Her encouraging attitude, her continued support and broad understanding of genetics and biology, her constructive and timely discussions, her exceptional ability to take into account both scientific issues and people when doing research and her creation of a free, yet stimulating academic atmosphere, which without I would never have succeeded in accomplishing this work. She is an outstanding scientist and it has been one of luckiest events in my life that I have the opportunity to work under your supervision. I also wish to thank my second supervisor Docent Irmä Jarvelä for her support and valuable advice throughout the period of my study. Her patience with me was unbelievable especially when things went wrong! I was never forget the first time I met you in the Lab and the Linkage Analysis Book you gave to me at that time when I knew nothing about genetics! She has been an admirable example of a hard working researcher, with her contagious enthusiasm, positive attitude and efficiency.

I am greatly indebted to Professor Jaakko Ignatius and Docent Tarja Ruuska for carefully reviewing the manuscript of my thesis and for their valuable comments. M.Sc. Donald Smart is thanked for reviewing the language of my thesis.

I have been very lucky to work with several very talented senior scientists. The talent Clinicians: Docent Timo Sahi, and Professor Erkki Savilahti, This project and its achievements would not have been possible without the contribution of each of these individuals. Professor Joseph Terwilliger, your great skills in Statistical genetics are beyond dispute. I am greatly indebted to Professor Antti Sajantila, Docent Per-Henrik Groop, Dr Tiinamaija Tuomi, Dr Carol Forsblom, Docent Markus Perols, Professor Raivo Sulkava, Professor Matti Valimaki and Professor Kimmo Kontula, for their collaboration in many aspects of this project. The seniors in our Lab Docent Ismo Ulmanen and current head of the Lab Docent Anu Jalanko are thanked for kind attitude toward all the projects in our Lab. Sari Kiviko, Sari mustala, Tuija Svahnbäck and Helen Knuuttila, and Minna Partanen are thanked for their secretarial help and great kindness. Teemu Perheentupa, Jari Raikko, for not only their help with computer problems, which makes the work environment in our Lab totally different, but also their unbelievable kindness. I have had a pleasure to share the room with nice people from MLO. Mari Auranen, during the years she spent in our Lab, for the nice conversations. Tero Ylisaukko-oja, thanks for your friendship, scientific sharing, kindness and help when i need. The lactase team: Mikko Kuokkanen and Heli Rasinperä, you add to the project the taste which makes lactase papers digestible! Elli Kempas, Anne Vikman, Minna Levander are thanked for excellent laboratory assistance. Special thanks go to Teppo Varilo , who was the first one offer to help me in the Lab, on my the first day! Other present and past members of our Lab are thanked for support and friendly atmosphere. Joni Saarela your kindness, ready to help and calmness is really great, Ville Holmberg your open mind make the nice conversations with you something different. Jyrki Kuokkanen is thanked for your cheerful and kindly spirit, Lasse Lonqvist your kindness, and being ready to help is unbelievable. Pekka Ellonen, The master of the sequencing in the Lab, your help and kindness is really appreciated.

Special thanks go to the Libyan Friends in Finland (Helsinki, Turku, and Tampere), Canada, Germany and Libya for continues care and support. Especially, to Dr. Ahmed Ali (Biomedicum, Helsinki) for his humour and cheerful times we spent eating Kebab!

My warmest gratitude go to my dear family. I am grateful to my parents, sisters, brothers, especially Dr. Siraj (MD, Ph.D), Salah and Zakaria for their help and continous support during all these years in Helsinki, Finland. The kids of my brother Siraj: Abdu and libia; whose love of chocolate made it easier to convince them to make their lives here in Helsinki, and whose presence here for my first few years made my life much easier. Special thanks to my sister in law Dr Amna, for her supporting during all the time she spent in Helsinki. I would like to give my best thanks to my wife Noor-Elhuda, who comes as a gift to me from sky in the most critical time of my life, for her encouragement, loving support and understanding. Finally, the last of my supporters, is my new baby, (Hedaya) who has made my outlook on life totally different.

I am indebted to the patients and the families for participating in this study.
This study was financially supported by the Sigrid Juselius Foundation, Helsinki University Hospital, Academy of Finland, Emil Aaltonen Foundation Tampere and University of Helsinki.

Helsinki, January 2005

References

(1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium. Science 282:2012-8

(2003) The International HapMap Project. Nature 426:789-96

Abecasis GR, Cookson WO (2000) GOLD--graphical overview of linkage disequilibrium. Bioinformatics 16:182-3

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, et al. (2000) The genome sequence of Drosophila melanogaster. Science 287:2185-95

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-402

Antequera F, Bird A (1993) Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci U S A 90:11995-9

Aoki K (1986) A stochastic model of gene-culture coevolution suggested by the "culture historical hypothesis" for the evolution of adult lactose absorption in humans. Proc Natl Acad Sci U S A 83:2929-33

Aoki K (2001) Theoretical and empirical aspects of gene-culture coevolution. Theor Popul Biol 59:253-61

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3:299-309

Arola H (1994) Diagnosis of hypolactasia and lactose malabsorption. Scand J Gastroenterol Suppl 202:26-35

Arola H, Tamm A (1994) Metabolism of lactose in the human body. Scand J Gastroenterol Suppl 202:21-5

Arribas JC, Herrero AG, Martin-Lomas M, Canada FJ, He S, Withers SG (2000) Differential mechanism-based labeling and unequivocal activity assignment of the two active sites of intestinal lactase/phlorizin hydrolase. Eur J Biochem 267:6996-7005

Asp NG, Dahlqvist A (1974) Intestinal beta-galactosidases in adult low lactase activity and in congenital lactase deficiency. Enzyme 18:84-102

Asp NG, Dahlqvist A, Kuitunen P, Launiala K, Visakorpi JK (1973) Complete deficiency of brush-border lactase in congenital lactose malabsorption. Lancet 2:329-30

Astbury WT (1951) X-Ray studies of Nucleic Acids. Symp. Soc. Exp. Biol. I. Nucleic Acids 1:66-76

Auricchio S, Rubino A, Landolt M, Semenza G, Prader A (1963) Isolated Intestinal Lactase Deficiency in the Adult. Lancet 13:324-6

Avery OTM, C.; McCarty, M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcus types. I. Exp. Med 79:137-158

Barbujani G, Sokal RR, Oden NL (1995) Indo-European origins: a computer-simulation test of five hypotheses. Am J Phys Anthropol 96:109-32

Bengtsson B, Steen B, Dahlqvist A, Jagerstad M (1984) Does lactose intake induce cataract in man? Lancet 1:1293-4

Bengtsson BO, Thomson G (1981) Measuring the strength of associations between HLA antigens and diseases. Tissue Antigens 18:356-63

Berg NO, Dahlqvist A, Lindberg T, Studnitz W (1969) Severe familial lactose intolerance--a gastrogen disorder? Acta Paediatr Scand 58:525-7

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN (2004) Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74:1111-20

Birge SJ, Jr., Keutmann HT, Cuatrecasas P, Whedon GD (1967) Osteoporosis, intestinal lactase deficiency and low dietary calcium intake. N Engl J Med 276:445-8

Blumberg H, Hartshorne TA, Young ET (1988) Regulation of expression and activity of the yeast transcription factor ADR1. Mol Cell Biol 8:1868-76

Boehnke M, Lange K, Cox DR (1991) Statistical methods for multipoint radiation hybrid mapping. Am J Hum Genet 49:1174-88

Bolin TD, Davis AE (1970) Primary lactase deficiency: genetic or acquired? Am J Dig Dis 15:679-92

Bolin TD, McKern A, Davis AE (1971) The effect of diet on lactase activity in the rat. Gastroenterology 60:432-7

Bolin TD, Pirola RC, Davis AE (1969) Adaptation of intestinal lactase in the rat. Gastroenterology 57:406-9

Boll W, Wagner P, Mantei N (1991) Structure of the chromosomal gene and cDNAs coding for lactase-phlorizin hydrolase in humans with adult-type hypolactasia or persistence of lactase. Am J Hum Genet 48:889-902

Borgstrom B, Dahlqvist A, Lundh G, Sjovall J (1957) Studies of intestinal digestion and absorption in the human. J Clin Invest 36:1521-36

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314-31

Britten RJ (1994) Evidence that most human Alu sequences were inserted in a process that ceased about 30 million years ago. Proc Natl Acad Sci U S A 91:6148-50

Britton JA, Westhoff C, Howe GR, Gammon MD (2000) Lactose and benign ovarian tumours in a case-control study. Br J Cancer 83:1552-5

Brunser O, Castillo C, Araya M (1976) Fine structure of the small intestinal mucosa in infantile marasmic malnutrition. Gastroenterology 70:495-507

Buller HA, Kothe MJ, Goldman DA, Grubman SA, Sasak WV, Matsudaira PT, Montgomery RK, Grand RJ (1990) Coordinate expression of lactase-phlorizin hydrolase mRNA and enzyme levels in rat intestine during development. J Biol Chem 265:6978-83

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268:78-94

Cardon LR, Abecasis GR (2003) Using haplotype blocks to map human complex trait loci. Trends Genet 19:135-40

Carrera E, Nesheim MC, Crompton DW (1984) Lactose maldigestion in Ascaris-infected preschool children. Am J Clin Nutr 39:255-64

Carroccio A, Montalto G, Cavera G, Notarbatolo A (1998) Lactose intolerance and self-reported milk intolerance: relationship with lactose maldigestion and nutrient intake. Lactase Deficiency Study Group. J Am Coll Nutr 17:631-6

Cavalli-Sforza LL (1973) Analytic review: some current problems of human population genetics. Am J Hum Genet 25:82-104

Cavalli-Sforza LL, Menozzi, P., Piazza, A. (1994) The History and Geography of Human Genes. Princeton University Press, Princeton, New Jersey

Chargaff E (1951) Structure and function of nucleic acids as cell constituents. Fed Proc 10:654-9

Chargaff EV, E; Doniger, R.; Green, C; Misani, F (1949) The composition ofthe Deoxypentose Nucleic Acids ofthe Thymus and Spleen. J. Biol. Chem 177:405-416

Chumakov IM, Rigault P, Le Gall I, Bellanne-Chantelot C, Billault A, Guillou S, Soularue P, Guasconi G, Poullier E, Gros I, et al. (1995) A YAC contig map of the human genome. Nature 377:175-297

Church DM, Stotler CJ, Rutter JL, Murrell JR, Trofatter JA, Buckler AJ (1994) Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. Nat Genet 6:98-105

Claudio JO, Marineau C, Rouleau GA (1994) The mouse homologue of the neurofibromatosis type 2 gene is highly conserved. Hum Mol Genet 3:185-90

Cohen D, Chumakov I, Weissenbach J (1993) A first-generation physical map of the human genome. Nature 366:698-701

Cohen SN, Chang AC, Boyer HW, Helling RB (1973) Construction of biologically functional bacterial plasmids in vitro. Proc Natl Acad Sci U S A 70:3240-4

Cold Spring Harbor Publications CSH, N.Y. (1966) The genetic code. Cold Spring Harbor symposia on Quantitative Biology. Cold Spring Harbor Publications, Cold Spring Harbor, N.Y. 31

Collins FS, Green ED, Guttmacher AE, Guyer MS (2003) A vision for the future of genomics research. Nature 422:835-47

Colombo V, Lorenz-Meyer H, Semenza G (1973) Small intestinal phlorizin hydrolase: the "beta-glycosidase complex". Biochim Biophys Acta 327:412-24

Cook GC, al-Torki MT (1975) High intestinal lactase concentrations in adult Arbs in Saudi Arabia. Br Med J 3:135-6

Corazza GR, Benati G, Di Sario A, Tarozzi C, Strocchi A, Passeri M, Gasbarrini G (1995) Lactose intolerance and bone mass in postmenopausal Italian women. Br J Nutr 73:479-87

Cox DR, Burmeister M, Price ER, Kim S, Myers RM (1990) Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 250:245-50

Cramer DW (1989) Lactase persistence and milk consumption as determinants of ovarian cancer risk. Am J Epidemiol 130:904-10

Cuddenec Y, Delbruck H, Flatz G (1982) Distribution of the adult lactase phenotypes--lactose absorber and malabsorber--in a group of 131 army recruits. Gastroenterol Clin Biol 6:776-9

Dahlqvist A (1964) Method for Assay of Intestinal Disaccharidases. Anal Biochem 57:18-25

Dahlqvist A, Borgstrom B (1961) Digestion and absorption of disaccharides in man. Biochem J 81:411-8

Dahlqvist A, Hammond JB, Crane RK, Dunphy JV, Littman A (1963) Intestinal Lactase Deficiency and Lactose Intolerance in Adults. Preliminary Report. Gastroenterology 45:488-91

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. Nat Genet 29:229-32

Danielsen EM, Sjostrom H, Noren O (1981) Biosynthesis of intestinal microvillar proteins. Putative precursor forms of microvillus aminopeptidase and sucrase--isomaltase isolated from Ca2+-precipitated enterocyte membranes. FEBS Lett 127:129-32

Danielsen EM, Skovbjerg H, Noren O, Sjostrom H (1984) Biosynthesis of intestinal microvillar proteins. Intracellular processing of lactase-phlorizin hydrolase. Biochem Biophys Res Commun 122:82-90

de Vrese M, Stegelmann A, Richter B, Fenselau S, Laue C, Schrezenmeir J (2001) Probiotics--compensation for lactase insufficiency. Am J Clin Nutr 73:421S-429S

Deloukas P, Bentley D (2004) The HapMap project and its application to genetic studies of drug response. Pharmacogenomics J 4:88-90

Deloukas P, Schuler GD, Gyapay G, Beasley EM, Soderlund C, Rodriguez-Tome P, Hui L, et al. (1998) A physical map of 30,000 human genes. Science 282:744-6

Derbeneva OA, Starikovskaya EB, Wallace DC, Sukernik RI (2002) Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. Am J Hum Genet 70:1009-14

Devlin B, Risch N (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 29:311-22

Di Stefano M, Veneto G, Malservisi S, Strocchi A, Corazza GR (2001) Lactose malabsorption and intolerance in the elderly. Scand J Gastroenterol 36:1274-8

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152-4

Dombek KM, Young ET (1997) Cyclic AMP-dependent protein kinase inhibits ADH2 expression in part by decreasing expression of the transcription factor gene ADR1. Mol Cell Biol 17:1450-8

Donnelly SR, Hawkins TE, Moss SE (1999) A conserved nuclear element with a role in mammalian gene regulation. Hum Mol Genet 8:1723-8

Dudd SN, Evershed RP (1998) Direct demonstration of milk as an element of archaeological economies. Science 282:1478-81

Duluc I, Jost B, Freund JN (1993) Multiple levels of control of the stage- and region-specific expression of rat intestinal lactase. J Cell Biol 123:1577-86

Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I (2002) Identification of a variant associated with adult-type hypolactasia. Nat Genet 30:233-7

Escher JC, de Koning ND, van Engen CG, Arora S, Buller HA, Montgomery RK, Grand RJ (1992) Molecular basis of lactase levels in adult humans. J Clin Invest 89:480-3

Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12:921-7

Fajardo O, Naim HY, Lacey SW (1994) The polymorphic expression of lactase in adults is regulated at the messenger RNA level. Gastroenterology 106:1233-41

Fang R, Olds LC, Santiago NA, Sibley E (2001) GATA family transcription factors activate lactase gene promoter in intestinal Caco-2 cells. Am J Physiol Gastrointest Liver Physiol 280:G58-67

Fang R, Santiago NA, Olds LC, Sibley E (2000) The homeodomain protein Cdx2 regulates lactase gene promoter activity during enterocyte differentiation. Gastroenterology 118:115-27

Ferguson A, Maxwell JD (1967) Genetic aetiology of lactose intolerance. Lancet 2:188-90

Finkenstedt G, Skrabal F, Gasser RW, Braunsteiner H (1986) Lactose absorption, milk consumption, and fasting blood glucose concentrations in women with idiopathic osteoporosis. Br Med J (Clin Res Ed) 292:161-2

Fitzgerald K, Bazar L, Avigan MI (1998) GATA-6 stimulates a cell line-specific activation element in the human lactase promoter. Am J Physiol 274:G314-24

Flatz G (1984) Gene-dosage effect on intestinal lactase activity demonstrated in vivo. Am J Hum Genet 36:306-10

Flatz G (1987) Genetics of lactose digestion in humans. Adv Hum Genet 16:1-77

Flatz G, Rotthauwe HW (1973) Lactose nutrition and natural selection. Lancet 2:76-7

Flatz G, Rotthauwe HW (1977) The human lactase polymorphism: physiology and genetics of lactose absorption and malabsorption. Prog Med Genet 2:205-49

Franklin SE, Gosling RG (1953) Molecular configuration in sodium thymonucleate. Nature 171:740-1

Freund JN, Duluc I, Raul F (1991) Lactase expression is controlled differently in the jejunum and ileum during development in rats. Gastroenterology 100:388-94

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225-9

Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, et al. (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493-521

Gilat T, Benaroya Y, Gelman-Malachi E, Adam A (1973) Genetics of primary adult lactase deficiency. Gastroenterology 64:562-8

Gilat T, Russo S, Gelman-Malachi E, Aldor TA (1972) Lactase in man: a nonadaptable enzyme. Gastroenterology 62:1125-7

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) Life with 6000 genes. Science 274:546, 563-7

Goodman MT, Wu AH, Tung KH, McDuffie K, Kolonel LN, Nomura AM, Terada K, Wilkens LR, Murphy S, Hankin JH (2002) Association of dairy products, lactose, and calcium with the risk of ovarian cancer. Am J Epidemiol 156:148-57

Goring HH, Terwilliger JD (2000a) Linkage analysis in the presence of errors I: complex-valued recombination fractions and complex phenotypes. Am J Hum Genet 66:1095-106

Goring HH, Terwilliger JD (2000b) Linkage analysis in the presence of errors III: marker loci and their map as nuisance parameters. Am J Hum Genet 66:1298-309

Goring HH, Terwilliger JD (2000c) Linkage analysis in the presence of errors IV: joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. Am J Hum Genet 66:1310-27

Grand RJ, Montgomery RK, Chitkara DK, Hirschhorn JN (2003) Changing genes; losing lactase. Gut 52:617-9

Griffith F (1928) The significance of Pneumococcal types. J.Hyg 27:113-159

Groop L, Forsblom C, Lehtovirta M, Tuomi T, Karanko S, Nissen M, Ehrnstrom BO, Forsen B, Isomaa B, Snickars B, Taskinen MR (1996) Metabolic consequences of a family history of NIDDM (the Botnia study): evidence for sex-specific parental effects. Diabetes 45:1585-93

Guglielmino CR, Piazza A, Menozzi P, Cavalli-Sforza LL (1990) Uralic genes in Europe. Am J Phys Anthropol 83:57-68

Guigo R, Knudsen S, Drake N, Smith T (1992) Prediction of gene structure. J Mol Biol 226:141-57

Guo SW (1997) Linkage disequilibrium measures for fine-scale mapping: a comparison. Hum Hered 47:301-14

Harvey CB, Fox MF, Jeggo PA, Mantei N, Povey S, Swallow DM (1993) Regional localization of the lactase-phlorizin hydrolase gene, LCT, to chromosome 2q21. Ann Hum Genet 57 (Pt 3):179-85

Harvey CB, Hollox EJ, Poulter M, Wang Y, Rossi M, Auricchio S, Iqbal TH, Cooper BT, Barton R, Sarner M, Korpela R, Swallow DM (1998) Lactase haplotype frequencies in Caucasians: association with the lactase persistence/non-persistence polymorphism. Ann Hum Genet 62 (Pt 3):215-23

Harvey CB, Pratt WS, Islam I, Whitehouse DB, Swallow DM (1995) DNA polymorphisms in the lactase gene. Linkage disequilibrium across the 70-kb region. Eur J Hum Genet 3:27-41

Harvey CB, Wang Y, Darmoul D, Phillips A, Mantei N, Swallow DM (1996) Characterisation of a human homologue of a yeast cell division cycle gene, MCM6, located adjacent to the 5' end of the lactase gene on chromosome 2q21. FEBS Lett 398:135-40

Hastbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nat Genet 2:204-11

Hastbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, Hamilton BA, Kusumi K, Trivedi B, Weaver A, et al. (1994) The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. Cell 78:1073-87

Hauri HP, Sterchi EE, Bienz D, Fransen JA, Marxer A (1985) Expression and intracellular transport of microvillus membrane hydrolases in human intestinal epithelial cells. J Cell Biol 101:838-51

Heilskov NS (1951) Studies on animal lactase. I. Lactase activity determination. Acta Physiol Scand 22:267-76

Herrinton LJ, Weiss NS, Beresford SA, Stanford JL, Wolfla DM, Feng Z, Scott CR (1995) Lactose and galactose intake and metabolism in relation to the risk of epithelial ovarian cancer. Am J Epidemiol 141:407-16

Ho MW, Povey S, Swallow D (1982) Lactase polymorphism in adult British natives: estimating allele frequencies by enzyme assays in autopsy samples. Am J Hum Genet 34:650-7

Holden C, Mace R (1997) Phylogenetic analysis of the evolution of lactose digestion in adults. Hum Biol 69:605-28

Hollox EJ, Poulter M, Wang Y, Krause A, Swallow DM (1999) Common polymorphism in a highly variable region upstream of the human lactase gene affects DNA-protein interactions. Eur J Hum Genet 7:791-800

Hollox EJ, Poulter M, Zvarik M, Ferak V, Krause A, Jenkins T, Saha N, Kozlov AI, Swallow DM (2001) Lactase haplotype diversity in the Old World. Am J Hum Genet 68:160-172

Holzel A (1967) Sugar malabsorption due to deficiencies of disaccharidase activities and of monosaccharide transport. Arch Dis Child 42:341-52

Holzel A, Mereu T, Thomson ML (1962) Severe lactose intolerance in infancy. Lancet 2:1346-8

Holzel A, Schwarz V, Sutcliffe KW (1959) Defective lactose absorption causing malnutrition in infancy. Lancet 1:1126-8

Honkanen R, Kroger H, Alhava E, Turpeinen P, Tuppurainen M, Saarikoski S (1997) Lactose intolerance associated with fractures of weight-bearing bones in Finnish women aged 38-57 years. Bone 21:473-7

Honkanen R, Pulkkinen P, Jarvinen R, Kroger H, Lindstedt K, Tuppurainen M, Uusitupa M (1996) Does lactose intolerance predispose to low bone density? A population-based study of perimenopausal Finnish women. Bone 19:23-8

Horowitz M, Wishart J, Mundy L, Nordin BE (1987) Lactose and calcium absorption in postmenopausal osteoporosis. Arch Intern Med 147:534-6

Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu SH, et al. (1995) An STS-based map of the human genome. Science 270:1945-54

Hwu HR, Roberts JW, Davidson EH, Britten RJ (1986) Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. Proc Natl Acad Sci U S A 83:3875-9

Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nat Genet 6:84-9

Jackson DA, Symons RH, Berg P (1972) Biochemical method for inserting new genetic information into DNA of simian virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. Proc Natl Acad Sci U S A 69:2904-2909

Jacob R, Weiner JR, Stadge S, Naim HY (2000) Additional N-glycosylation and its impact on the folding of intestinal lactase-phlorizin hydrolase. J Biol Chem 275:10630-7

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA (2001) Haplotype tagging for the identification of common disease genes. Nat Genet 29:233-7

Johnson JD, Kretchmer N, Simoons FJ (1974) Lactose malabsorption: its biology and history. Adv Pediatr 21:197-237

Keller P WH, Semenza G, Mantei N. (1993) Structure of lactase-phlorizin hydrolase
and it gene. In: Auricchio S, Semenza G (ed.). Common food intolerances 2:
Milk in human nutrition and adult-type hypolactasia. 3 . Basel,
Switzerland: Karger:76-84

Keller P, Zwicker E, Mantei N, Semenza G (1992) The levels of lactase and of sucrase-isomaltase along the rabbit small intestine are regulated both at the mRNA level and post-translationally. FEBS Lett 313:265-9

Keusch GT, Troncale FJ, Miller LH, Promadhat V, Anderson PR (1969) Acquired lactose malabsorption in Thai children. Pediatrics 43:540-5

Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, Linnoila M, Goldman D, Long JC (1998) Dual origins of Finns revealed by Y chromosome haplotype variation. Am J Hum Genet 62:1171-9

Koivisto M, Perola M, Varilo T, Hennah W, Ekelund J, Lukk M, Peltonen L, Ukkonen E, Mannila H (2003) An MDL method for finding haplotype blocks and for estimating the strength of haplotype block boundaries. Pac Symp Biocomput:502-13

Kolho KL, Savilahti E (2000) Ethnic differences in intestinal disaccharidase values in children in Finland. J Pediatr Gastroenterol Nutr 30:283-7

Kornberg A (1960) Biologic synthesis of deoxyribonucleic acid. Science 131:1503-8

Kosnai I, Kuitunen P, Savilahti E, Rapola J, Kohegyi J (1980) Cell kinetics in the jejunal crypt epithelium in malabsorption syndrome with cow's milk protein intolerance and in coeliac disease of childhood. Gut 21:1041-6

Krasinski SD, Estrada G, Yeh KY, Yeh M, Traber PG, Rings EH, Buller HA, Verhave M, Montgomery RK, Grand RJ (1994) Transcriptional regulation of intestinal hydrolase biosynthesis during postnatal development in rats. Am J Physiol 267:G584-94

Krasinski SD, Upchurch BH, Irons SJ, June RM, Mishra K, Grand RJ, Verhave M (1997) Rat lactase-phlorizin hydrolase/human growth hormone transgene is expressed on small intestinal villi in transgenic mice. Gastroenterology 113:844-55

Krasinski SD, Van Wering HM, Tannemaat MR, Grand RJ (2001) Differential activation of intestinal gene promoters: functional interactions between GATA-5 and HNF-1 alpha. Am J Physiol Gastrointest Liver Physiol 281:G69-84

Kretchmer N (1971) Lactose and lactase--a historical perspective. Gastroenterology 61:805-13

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347-63

Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 57:439-54

Kruse TA, Bolund L, Grzeschik KH, Ropers HH, Sjostrom H, Noren O, Mantei N, Semenza G (1988) The human lactase-phlorizin hydrolase gene is located on chromosome 2. FEBS Lett 240:123-6

Laan M, Paabo S (1997) Demographic history and linkage disequilibrium in human populations. Nat Genet 17:435-8

Labuda M, Labuda D, Korab-Laskowska M, Cole DE, Zietkiewicz E, Weissenbach J, Popowska E, Pronicka E, Root AW, Glorieux FH (1996) Linkage disequilibrium analysis in young populations: pseudo-vitamin D-deficiency rickets and the founder effect in French Canadians. Am J Hum Genet 59:633-43

Lacey SW, Naim HY, Magness RR, Gething MJ, Sambrook JF (1994) Expression of lactase-phlorizin hydrolase in sheep is regulated at the RNA level. Biochem J 302 (Pt 3):929-35

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409:860-921

Lathrop GM, Lalouel JM (1984) Easy calculations of lod scores and genetic risks on small computers. Am J Hum Genet 36:460-5

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci U S A 81:3443-6

Launiala K, Kuitunen P, Visakorpi JK (1966) Disaccharidases and histology of duodenal mucosa in congenital lactose malabsorption. Acta Paediatr Scand 55:257-63

Lebenthal E, Sunshine P, Kretchmer N (1973) Effect of prolonged nursing on the activity of intestinal lactase in rats. Gastroenterology 64:1136-41

Lee SY, Wang Z, Lin CK, Contag CH, Olds LC, Cooper AD, Sibley E (2002) Regulation of intestine-specific spatiotemporal expression by the rat lactase promoter. J Biol Chem 277:13099-105

Leichter J (1973) Effect of dietary lactose on intestinal lactase activity in young rats. J Nutr 103:392-6

Lember M, Tamm A (1988) Lactose absorption and milk drinking habits in Estonians with myocardial infarction. Br Med J (Clin Res Ed) 296:95-6

Lisker R, Cervantes G, Perez-Briceno R, Alva G (1988) Lack of relationship between lactose absorption and senile cataracts. Ann Ophthalmol 20:436-8

Lloyd M, Mevissen G, Fischer M, Olsen W, Goodspeed D, Genini M, Boll W, Semenza G, Mantei N (1992) Regulation of intestinal lactase in adult hypolactasia. J Clin Invest 89:524-9

Lobban PE, Kaiser AD (1973) Enzymatic end-to end joining of DNA molecules. J Mol Biol 78:453-71

Lorenz-Meyer H, Blum AL, Haemmerli HP, Semenza G (1972) A second enzyme defect in acquired lactase deficiency: lack of small-intestinal phlorizin-hydrolase. Eur J Clin Invest 2:326-31

Lovett M (1994) Fishing for complements: finding genes by direct selection. Trends Genet 10:352-7

Macdonald I (1989) Galactose and ovarian cancer. Lancet 2:452

Mahmood A, Pathak RM, Agarwal N (1978) Effect of chronic alloxan diabetes and insulin administration on intestinal brush border enzymes. Experientia 34:741-2

Mainguet P, Faille I, Destrebecq L, Devogelaer JP, Nagant de Deuxchaisnes C (1991) Lactose intolerance, calcium intake, and osteopenia. Lancet 338:1156-7

Maiuri L, Rossi M, Raia V, Garipoli V, Hughes LA, Swallow D, Noren O, Sjostrom H, Auricchio S (1994) Mosaic regulation of lactase in human adult-type hypolactasia. Gastroenterology 107:54-60

Mantei N, Villa M, Enzler T, Wacker H, Boll W, James P, Hunziker W, Semenza G (1988) Complete primary structure of human and rabbit lactase-phlorizin hydrolase: implications for biosynthesis, membrane anchoring and evolution of the enzyme. Embo J 7:2705-13

Maouche L, Cartron JP, Chretien S (1994) Different domains regulate the human erythropoietin receptor gene transcription. Nucleic Acids Res 22:338-46

Martinez-Pastor MT, Marchler G, Schuller C, Marchler-Bauer A, Ruis H, Estruch F (1996) The Saccharomyces cerevisiae zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). Embo J 15:2227-35

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci U S A 74:560-4

McCracken RD (1970) Adult lactose tolerance. Jama 213:2257-60

McLellan T, Jorde LB, Skolnick MH (1984) Genetic distances between the Utah Mormons and related populations. Am J Hum Genet 36:836-57

McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, et al. (2001) A physical map of the human genome. Nature 409:934-41

Meloni GF, Colombo C, La Vecchia C, Pacifico A, Tomasi P, Ogana A, Marinaro AM, Meloni T (2001) High prevalence of lactose absorbers in Northern Sardinian patients with type 1 and type 2 diabetes mellitus. Am J Clin Nutr 73:582-5

Meloni GF, Colombo C, La Vecchia C, Ruggiu G, Mannazzu MC, Ambrosini G, Cherchi PL (1999) Lactose absorption in patients with ovarian cancer. Am J Epidemiol 150:183-6

Mendel G (1866) Versuche Über Pflanzen-Hybriden. Verhandlungen des narurforschenden Vereines in BrÜnn:3-47

Mendel LBM, P.H. (1907) Chemical studies on growth.I.The inversting enzymes of the alimentary tract, escpecially in the embryo. Amer J Physiol 20:81-96

Messer M, Dahlqvist A (1966) A one-step ultramicro method for the assay of intestinal disaccharidases. Anal Biochem 14:376-92

Metneki J, Czeizel A, Flatz SD, Flatz G (1984) A study of lactose absorption capacity in twins. Hum Genet 67:296-300

Mettlin CJ, Piver MS (1990) A case-control study of milk-drinking and ovarian cancer risk. Am J Epidemiol 132:871-6

Metz G, Jenkins DJ, Peters TJ, Newman A, Blendis LM (1975) Breath hydrogen as a diagnostic method for hypolactasia. Lancet 1:1155-7

Mitchelmore C, Troelsen JT, Spodsberg N, Sjostrom H, Noren O (2000) Interaction between the homeodomain proteins Cdx2 and HNF1alpha mediates expression of the lactase-phlorizin hydrolase gene. Biochem J 346 Pt 2:529-35

Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277-318

Murakami I, Ikeda T (1998) Effects of diabetes and hyperglycemia on disaccharidase activities in the rat. Scand J Gastroenterol 33:1069-73

Naim HY (2001) Molecular and cellular aspects and regulation of intestinal lactase-phlorizin hydrolase. Histol Histopathol 16:553-61

Naim HY, Lacey SW, Sambrook JF, Gething MJ (1991) Expression of a full-length cDNA coding for human intestinal lactase-phlorizin hydrolase reveals an uncleaved, enzymatically active, and transport-competent protein. J Biol Chem 266:12313-20

Naim HY, Lentze MJ (1992) Impact of O-glycosylation on the function of human intestinal lactase-phlorizin hydrolase. Characterization of glycoforms varying in enzyme activity and localization of O-glycoside addition. J Biol Chem 267:25494-504

Naim HY, Naim H (1996) Dimerization of lactase-phlorizin hydrolase occurs in the endoplasmic reticulum, involves the putative membrane spanning domain and is required for an efficient transport of the enzyme to the cell surface. Eur J Cell Biol 70:198-208

Naim HY, Sterchi EE, Lentze MJ (1987) Biosynthesis and maturation of lactase-phlorizin hydrolase in the human small intestinal epithelial cells. Biochem J 241:427-34

Nemeth K, Plumb GW, Berrin JG, Juge N, Jacob R, Naim HY, Williamson G, Swallow DM, Kroon PA (2003) Deglycosylation by small intestinal epithelial cell beta-glucosidases is a critical step in the absorption and metabolism of dietary flavonoid glycosides in humans. Eur J Nutr 42:29-42

Newcomer AD, Hodgson SF, McGill DB, Thomas PJ (1978) Lactase deficiency: prevalence in osteoporosis. Ann Intern Med 89:218-20

Newcomer AD, McGill DB, Thomas PJ, Hofman AF (1975) Prospective comparison of indirect methods for detecting lactase deficiency. N Engl J Med 293:1232-6

Nirenberg M (2004) Historical review: Deciphering the genetic code--a personal account. Trends Biochem Sci 29:46-54

Nirenberg MW, Matthaei JH (1961) The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci U S A 47:1588-602

Norio R (2003a) Finnish Disease Heritage I: characteristics, causes, background. Hum Genet 112:441-56

Norio R (2003b) Finnish Disease Heritage II: population prehistory and genetic roots of Finns. Hum Genet 112:457-69

Norio R (2003c) The Finnish Disease Heritage III: the individual diseases. Hum Genet 112:470-526

Norio R, Nevanlinna HR, Perheentupa J (1973) Hereditary diseases in Finland; rare flora in rare soul. Ann Clin Res 5:109-41

Olby RC (1966) The growth of ideas about inheritance and variation that eventually led to the Mendelian solution. Origin of Mendelism.London: Constable and Complany Ltd.

Olds LC, Sibley E (2003) Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a cis regulatory element. Hum Mol Genet 12:2333-40

Olivier M, Aggarwal A, Allen J, Almendras AA, Bajorek ES, Beasley EM, Brady SD, et al. (2001) A high-resolution radiation hybrid map of the human genome draft sequence. Science 291:1298-302

Ott J (1999) Analysis of Human Genetic Linkage,3rd. Johns Hopkins University Press, Baltimore

Pastinen T, Perola M, Ignatius J, Sabatti C, Tainola P, Levander M, Syvanen AC, Peltonen L (2001) Dissecting a population genome for targeted screening of disease mutations. Hum Mol Genet 10:2961-72

Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, Syvanen AC (2000) A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. Genome Res 10:1031-42

Peltonen L, Jalanko A, Varilo T (1999) Molecular genetics of the Finnish disease heritage. Hum Mol Genet 8:1913-23

Peltonen L, McKusick VA (2001) Genomics and medicine. Dissecting human disease in the postgenomic era. Science 291:1224-9

Peltonen L, Palotie A, Lange K (2000) Use of population isolates for mapping complex traits. Nat Rev Genet 1:182-90

Pennisi E (1997) Laboratory workhorse decoded. Science 277:1432-4

Pettersson-Fernholm K, Forsblom C, Hudson BI, Perola M, Grant PJ, Groop PH (2003) The functional -374 T/A RAGE gene polymorphism is associated with proteinuria and cardiovascular disease in type 1 diabetic patients. Diabetes 52:891-4

Poggi V, Sebastio G (1991) Molecular analysis of the lactase gene in the congenital lactase deficiency. Am J Hum Genet Suppl 49:105

Rasinpera H, Savilahti E, Enattah NS, Kuokkanen M, Totterman N, Lindahl H, Jarvela I, Kolho K-L (2004) Genetic test, which can be used to diagnose adult type hypolactasia in children. Gut

Rasinpera H SE, Enattah NS,Kuokkanen M, Totterman N, Lindahl H, Jarvela I, Kolho K-L (2004) Genetic test, which can be used to diagnose adult-type hypolactasia in children. Gut 53::1571-1576

Rinaldi E, Albini L, Costagliola C, De Rosa G, Auricchio G, De Vizia B, Auricchio S (1984) High frequency of lactose absorbers among adults with idiopathic senile and presenile cataract in a population with a high prevalence of primary adult lactose malabsorption. Lancet 1:355-7

Risch HA, Jain M, Marrett LD, Howe GR (1994) Dietary lactose intake, lactose intolerance, and the risk of epithelial ovarian cancer in southern Ontario (Canada). Cancer Causes Control 5:540-8

Risch N, de Leon D, Ozelius L, Kramer P, Almasy L, Singer B, Fahn S, Breakefield X, Bressman S (1995) Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. Nat Genet 9:152-9

Rossi M, Maiuri L, Fusco MI, Salvati VM, Fuccio A, Auricchio S, Mantei N, Zecca L, Gloor SM, Semenza G (1997) Lactase persistence versus decline in human adults: multifactorial events are involved in down-regulation after weaning. Gastroenterology 112:1506-14

Roth J (1987) Subcellular organization of glycosylation in mammalian cells. Biochim Biophys Acta 906:405-36

Saarela T, Simila S, Koivisto M (1995) Hypercalcemia and nephrocalcinosis in patients with congenital lactase deficiency. J Pediatr 127:920-3

Sahi T (1974a) The inheritance of selective adult-type lactose malabsorption. Scand J Gastroenterol Suppl 30:1-73

Sahi T (1974b) Lactose malabsorption in Finnish-speaking and Swedish-speaking populations in Finland. Scand J Gastroenterol 9:303-8

Sahi T (1978) Dietary lactose and the aetiology of human small-intestinal hypolactasia. Gut 19:1074-86

Sahi T (1994a) Genetics and epidemiology of adult-type hypolactasia. Scand J Gastroenterol Suppl 202:7-20

Sahi T (1994b) Hypolactasia and lactase persistence. Historical review and the terminology. Scand J Gastroenterol Suppl 202:1-6

Sahi T, Isokoski M, Jussila J, Launiala K (1972) Lactose malabsorption in Finnish children of school age. Acta Paediatr Scand 61:11-6

Sahi T, Isokoski M, Jussila J, Launiala K, Pyorala K (1973) Recessive inheritance of adult-type lactose malabsorption. Lancet 2:823-6

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-91

Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C, Paabo S (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc Natl Acad Sci U S A 93:12035-9

Saltzman JR, Russell RM, Golner B, Barakat S, Dallal GE, Goldin BR (1999) A randomized trial of Lactobacillus acidophilus BG2FO4 to treat lactose intolerance. Am J Clin Nutr 69:140-6

Sambrook JF, E. F.; Maniatis, T. (1989) Molecular cloning: A laboratory manual.Second edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989

Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94:441-8

Savilahti E, Launiala K, Kuitunen P (1983) Congenital lactase deficiency. A clinical study on 16 patients. Arch Dis Child 58:246-52

Schedl HP, Al-Jurf AS, Wilson HD (1983) Elevated intestinal disaccharidase activity in the streptozotocin-diabetic rat is independent of enteral feeding. Diabetes 32:265-70

Schlegel-Haueter S, Hore P, Kerry KR, Semenza G (1972) The preparation of lactase and glucoamylase of rat small intestine. Biochim Biophys Acta 258:506-19

Schneider Stefan RD, Excoffier Laurent (2000) A software for population genetic data analysis.Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Sebastio G, Villa M, Sartorio R, Guzzetta V, Poggi V, Auricchio S, Boll W, Mantei N, Semenza G (1989) Control of lactase in human adult-type hypolactasia and in weaning rabbits and rats. Am J Hum Genet 45:489-97

Segall JJ (1980) Hypothesis is lactose a dietary risk factor for ischaemic heart disease? Int J Epidemiol 9:271-6

Segall JJ (1994) Dietary lactose as a possible risk factor for ischaemic heart disease: review of epidemiology. Int J Cardiol 46:197-207

Segall JJ (2003) Digestive and nutritional factors may explain lower prevalence of coronary disease in indigenous peoples. Bmj 327:449-50

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. Science 290:1155-9

Shou Y, Baron S, Poncz M (1998) An Sp1-binding silencer element is a critical negative regulator of the megakaryocyte-specific alphaIIb gene. J Biol Chem 273:5716-26

Simoons FJ (1969) Primary adult lactose intolerance and the milking habit: a problem in biological and cultural interrelations. I. Review of the medical research. Am J Dig Dis 14:819-36

Simoons FJ (1970) Primary adult lactose intolerance and the milking habit: a problem in biologic and cultural interrelations. II. A culture historical hypothesis. Am J Dig Dis 15:695-710

Simoons FJ (1978) The geographic hypothesis and lactose malabsorption. A weighing of the evidence. Am J Dig Dis 23:963-80

Simoons FJ (1980) Age of onset of lactose malabsorption. Pediatrics 66:646-8

Simoons FJ (1982) A geographic approach to senile cataracts: possible links with milk consumption, lactase activity, and galactose metabolism. Dig Dis Sci 27:257-64

Simoons FJ (2001) Persistence of lactase activity among Northern Europeans: a weighing of evidence for the calcium absorption hypothesis. Ecology of Food and Nutrition 40:397-469

Skovbjerg H, Danielsen EM, Noren O, Sjostrom H (1984) Evidence for biosynthesis of lactase-phlorizin hydrolase as a single-chain high-molecular weight precursor. Biochim Biophys Acta 798:247-51

Skovbjerg H, Noren O, Sjostrom H, Danielsen EM, Enevoldsen BS (1982) Further characterization of intestinal lactase/phlorizin hydrolase. Biochim Biophys Acta 707:89-97

Skovbjerg H, Sjostrom H, Noren O (1981) Purification and characterisation of amphiphilic lactase/phlorizin hydrolase from human small intestine. Eur J Biochem 114:653-61

Slatkin M (1994) Linkage disequilibrium in growing and stable populations. Genetics 137:331-6

Slemenda CW, Christian JC, Hui S, Fitzgerald J, Johnston CC, Jr. (1991) No evidence for an effect of lactase deficiency on bone mass in pre- or postmenopausal women. J Bone Miner Res 6:1367-71

Sloan JS, Dombek KM, Young ET (1999) Post-translational regulation of Adr1 activity is mediated by its DNA binding domain. J Biol Chem 274:37575-82

Smit AF (1996) The origin of interspersed repeats in the human genome. Curr Opin Genet Dev 6:743-8

Smith HO, Wilcox KW (1970) A restriction enzyme from Hemophilus influenzae. I. Purification and general properties. J . Mol. Biol 51:379-391

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE (1986) Fluorescence detection in automated DNA sequence analysis. Nature 321:674-9

Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 58:1323-37

Sokal RR, Oden NL, Thomson BA (1992) Origins of the Indo-Europeans: genetic evidence. Proc Natl Acad Sci U S A 89:7669-73

Spinelli D, Vota MG, Formenti F, Accetta S, Careddu P, Roggero P, Imbriano A, Volpe C (1987) Idiopathic presenile and senile cataract formation and changes in lactase activity. Fortschr Ophthalmol 84:666-8

Spodsberg N, Troelsen JT, Carlsson P, Enerback S, Sjostrom H, Noren O (1999) Transcriptional regulation of pig lactase-phlorizin hydrolase: involvement of HNF-1 and FREACs. Gastroenterology 116:842-54

Sriratanaban A, Symynkywicz LA, Thayer WR, Jr. (1971) Effect of physiologic concentration of lactose on prevention of postweaning decline of intestinal lactase. Am J Dig Dis 16:839-44

Sterchi EE, Mills PR, Fransen JA, Hauri HP, Lentze MJ, Naim HY, Ginsel L, Bond J (1990) Biogenesis of intestinal lactase-phlorizin hydrolase in adults with lactose intolerance. Evidence for reduced biosynthesis and slowed-down maturation in enterocytes. J Clin Invest 86:1329-37

Strachan TaR, A. (1999) Human Molecular genetics. 2ed.Bios Scientific Publishers

Sutton WS (1903) The chromosome in Heredity. Biol. Bull 4:231-251

Swallow DM (2003) Genetics of lactase persistence and lactose intolerance. Annu Rev Genet 37:197-219

Syvanen AC, Sajantila A, Lukka M (1993) Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing. Am J Hum Genet 52:46-59

Tandon RK, Srivastava LM, Pandey SC (1975) Increased disaccharidase activity in human diabetics. Am J Clin Nutr 28:621-5

Terwilliger JD (1995) A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. Am J Hum Genet 56:777-87

Terwilliger JD, Goring HH (2000) Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Hum Biol 72:63-132

Terwilliger JD, Zollner S, Laan M, Paabo S (1998) Mapping genes through the use of linkage disequilibrium generated by genetic drift: 'drift mapping' in small populations with no demographic expansion. Hum Hered 48:138-54

Terwilliger JDO, J (1994) Hand book of human genetic analysis. Johns Hopkins University Press, Baltimore (1994)

Thompson EA, Deeb S, Walker D, Motulsky AG (1988) The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. Am J Hum Genet 42:113-24

Torp N, Rossi M, Troelsen JT, Olsen J, Danielsen EM (1993) Lactase-phlorizin hydrolase and aminopeptidase N are differentially regulated in the small intestine of the pig. Biochem J 295 (Pt 1):177-82

Troelsen JT, Mehlum A, Olsen J, Spodsberg N, Hansen GH, Prydz H, Noren O, Sjostrom H (1994a) 1 kb of the lactase-phlorizin hydrolase promoter directs post-weaning decline and small intestinal-specific expression in transgenic mice. FEBS Lett 342:291-6

Troelsen JT, Mitchelmore C, Olsen J (2003a) An enhancer activates the pig lactase phlorizin hydrolase promoter in intestinal cells. Gene 305:101-11

Troelsen JT, Mitchelmore C, Spodsberg N, Jensen AM, Noren O, Sjostrom H (1997) Regulation of lactase-phlorizin hydrolase gene expression by the caudal-related homoeodomain protein Cdx-2. Biochem J 322 (Pt 3):833-8

Troelsen JT, Olsen J, Mitchelmore C, Hansen GH, Sjostrom H, Noren O (1994b) Two intestinal specific nuclear factors binding to the lactase-phlorizin hydrolase and sucrase-isomaltase promoters are functionally related oligomeric molecules. FEBS Lett 342:297-301

Troelsen JT, Olsen J, Moller J, Sjostrom H (2003b) An upstream polymorphism associated with lactase persistence has increased enhancer activity. Gastroenterology 125:1686-94

Troelsen JT, Olsen J, Noren O, Sjostrom H (1992) A novel intestinal trans-factor (NF-LPH1) interacts with the lactase-phlorizin hydrolase promoter and co-varies with the enzymatic activity. J Biol Chem 267:20407-11

Tygstrup N, Lundquist F (1962) The effect of ethanol on galactose elimination in man. J Lab Clin Med 59:102-9

Ulshen MH, Rollo JL (1980) Pathogenesis of escherichia coli gastroenteritis in man--another mechanism. N Engl J Med 302:99-101

Van Den Oord EJ, Neale BM (2003) Will haplotype maps be useful for finding genes? Mol Psychiatry

van Wering HM, Huibregtse IL, van der Zwan SM, de Bie MS, Dowling LN, Boudreau F, Rings EH, Grand RJ, Krasinski SD (2002a) Physical interaction between GATA-5 and hepatocyte nuclear factor-1alpha results in synergistic activation of the human lactase-phlorizin hydrolase promoter. J Biol Chem 277:27659-67

van Wering HM, Moyer L, Grand RJ, Krasinski SD (2002b) Novel interaction at the Cdx-2 binding sites of the lactase-phlorizin hydrolase promoter. Biochem Biophys Res Commun 299:587-93

Vandenplas S, Wiid I, Grobler-Rabie A, Brebner K, Ricketts M, Wallis G, Bester A, Boyd C, Mathew C (1984) Blot hybridisation analysis of genomic DNA. J Med Genet 21:164-72

Wang Y, Harvey CB, Hollox EJ, Phillips AD, Poulter M, Clay P, Walker-Smith JA, Swallow DM (1998) The genetically programmed down-regulation of lactase in children. Gastroenterology 114:1230-6

Wang Y, Harvey CB, Pratt WS, Sams VR, Sarner M, Rossi M, Auricchio S, Swallow DM (1995) The lactase persistence/non-persistence polymorphism is controlled by a cis-acting element. Hum Mol Genet 4:657-62

Varilo T (1999) The age of the mutations in the Finnish disease heritage; genealogical and linakge disequilibrium study. PhD thesis, National Public Health Institute, Helsinki, Finland

Varilo T, Nikali K, Suomalainen A, Lonnqvist T, Peltonen L (1996a) Tracing an ancestral mutation: genealogical and haplotype analysis of the infantile onset spinocerebellar ataxia locus. Genome Res 6:870-5

Varilo T, Savukoski M, Norio R, Santavuori P, Peltonen L, Jarvela I (1996b) The age of human mutation: genealogical and linkage disequilibrium analysis of the CLN5 mutation in the Finnish population. Am J Hum Genet 58:506-12

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520-62

Watson JD, Crick FH (1953a) Genetical implications of the structure of deoxyribonucleic acid. Nature 171:964-7

Watson JD, Crick FH (1953b) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171:737-8

Webb PM, Bain CJ, Purdie DM, Harvey PW, Green A (1998) Milk consumption, galactose metabolism and ovarian cancer (Australia). Cancer Causes Control 9:637-44

Weissenbach J (1993) A second generation linkage map of the human genome based on highly informative microsatellite loci. Gene 135:275-8

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M (1992) A second-generation linkage map of the human genome. Nature 359:794-801

Welsh JD (1970) Isolated lactase deficiency in humans: report on 100 patients. Medicine (Baltimore) 49:257-77

Welsh JD, Poley JR, Bhatia M, Stevenson DE (1978) Intestinal disaccharidase activities in relation to age, race, and mucosal damage. Gastroenterology 75:847-55

Weng Z, Sokal RR (1995) Origins of Indo-Europeans and the spread of agriculture in Europe: comparison of lexicostatistical and genetic evidence. Hum Biol 67:577-94

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, et al. (2001) The sequence of the human genome. Science 291:1304-51

Venter JC, Adams MD, Sutton GG, Kerlavage AR, Smith HO, Hunkapiller M (1998) Shotgun sequencing of the human genome. Science 280:1540-2

Wheadon M, Goulding A, Barbezat GO, Campbell AJ (1991) Lactose malabsorption and calcium intake as risk factors for osteoporosis in elderly New Zealand women. N Z Med J 104:417-9

Villako K, Maaroos H (1994) Clinical picture of hypolactasia and lactose intolerance. Scand J Gastroenterol Suppl 202:36-54

Witte J, Lloyd M, Lorenzsonn V, Korsmo H, Olsen W (1990) The biosynthetic basis of adult lactase deficiency. J Clin Invest 86:1338-42

von Heijne G (1986) A new method for predicting signal sequence cleavage sites. Nucleic Acids Res 14:4683-90

Wuthrich M, Grunberg J, Hahn D, Jacob R, Radebach I, Naim HY, Sterchi EE (1996) Proteolytic processing of human lactase-phlorizin hydrolase is a two-step event: identification of the cleavage sites. Arch Biochem Biophys 336:27-34

Xu Y, Mural R, Shah M, Uberbacher E (1994) Recognizing exons in genomic sequence using GRAIL II. Genet Eng (N Y) 16:241-53

Zhang J, Rowe WL, Clark AG, Buetow KH (2003) Genomewide Distribution of High-Frequency, Completely Mismatching SNP Haplotype Pairs Observed To Be Common across Human Populations. Am J Hum Genet 73:1073-81

Zhang MQ (1997) Identification of protein coding regions in the human genome by quadratic discriminant analysis. Proc Natl Acad Sci U S A 94:565-8

PREVIOUSLY PUBLISHED IN THIS SERIES BY THE DEPARTMENT OF MOLECULAR MEDICINE

1. **Irma Järvelä**, Molecular distinction of neuronal ceroid-lipofuscinoses: Assignment of separate gene loci forinfantile and juvenile forms. NPHI A2/1991
2. **Päivi Helminen**, Hypervariable regions of human genome applied to paternity testing and detection of malignant cell clones. NPHI A1/1992
3. **Elina Ikonen**, Molecular genetics of aspartylglucosaminuria. NPHI A2/1992
4. **Antti Sajantila**, DNA analysis in forensic medicine: Application of the polymerase chain reaction (PCR) to the identification of individuals. NPHI A3/1992
5. **Katariina Kainulainen**, Molecular genetics of Marfan syndrome. NPHI A4/1992
6. Raili Kauppinen, Prognosis of acute porphyrias and molecular genetics of acute intermittent porphyria in Finland. NPHI A8/1992
7. **Miikka Vikkula**, The human type II collagen gene and cartilage diseases. NPHI A1/1993
8. **Anu Suomalainen**, Mutations of mitochondrial DNA in human disease. NPHI A4/1993
9. **Pentti Tienari**, Genetic susceptibility in multiple sclerosis. NPHI A10/1993
10. **Nina Enomaa**, Aspartylglucosaminuria: Molecular pathogenesis and in vitro correction of the enzyme defect. KTL A9/1994
11. **Tiina Paunio**, Molecular pathogenesis of familial amyloidosis, Finnish type. NPHI A5/1995
12. **Jouni Vesa**, The molecular defect in infantile neuronal ceroid lipofuscinosis. KTL A12/1995
13. **Elina Hellsten**, Positional cloning of the infantile neuronal ceroid lipofuscinosis gene. KTL A16/1995
14. **Pekka Nokelainen**, Genetic analyses in myotonic dystrophy and tibial muscular dystrophy in Finland. NPHI A5/1996
15. **Ritva Tikkanen**, Human lysosomal aspartylglucosaminidase: Structure, function and intracellular targeting. KTL A4/1996
16. **Aija Riikonen-Kyttälä**, Intracellular maturation of aspartylglucosaminidase. KTL A7/1996
17. **Leena Karttunen**, Molecular pathogenesis of Marfan syndrome. KTL A9/1996
18. **Johanna Aaltonen**, Molecular genetics of APECED (Autoimmune PolyEndocrinopathy-Candidiasis-Ectodermal Dystrophy). KTL A3/1998
19. **Terhi Rantamäki-Häkkinen**, Fibrillin defects in Marfan syndrome: Impact on DNA diagnosis and molecular pathogenesis. KTL A5/1998
20. **Minna Peltola**, Aspartylglucosaminuria (AGU): Lysosomal targeting of AGA, the cellular consequences of mutations and an attempt at gene therapy in the AGU mouse. NPHI A12/1998
21. **Annukka Uusitalo**, Aspartylglucosaminuria: Disease pathogenesis, developmental expression and regulation of the aspartylglucosaminidase gene. NPHI A15/1998
22. **Satu Kuokkanen**, Search for gene loci predisposing to multiple sclerosis in the Finnish population. NPHI A13/1998
23. **Kaisu Nikali**, Molecular genetics of infantile onset spinocerebellar ataxia. NPHI A14/1998
24. **Lasse Lönnqvist**, Molecular pathology of type-1 fibrillinopathies. KTL A16/1998
25. **Kai Tenhunen**, Mouse aspartylglucosaminidase gene and mouse model for aspartylglucosaminuria. KTL A17/1998
26. **Petra Pekkarinen**, Genetic mapping of the loci for a monogenic and multifactorial neuropsychiatric disorder: PLO-SL and familial bipolar disorder. KTL A19/1998
27. **Iiris Hovatta**, Molecular genetics of familial schizophrenia and PLO-SL. KTL A20/1998
28. **Paulina Paavola**, Molecular genetics of Meckel syndrome. KTL A21/1998
29. **Tuomas Klockars**, Positional cloning of the CLN5 gene. KTL A22/1998
30. **Päivi Pajukanta**, The search for familial combined hyperlipidemia susceptibility genes. KTL A26/1998
31. **Markus Perola**, Molecular genetics of hypertension and related traits. KTL A8/1999
32. **Teppo Varilo**, The age of mutations in the Finnish disease heritage; a genealogical and linkage disequilibrium study. KTL A21/1999
33. **Petra Björses**, Autoimmune polyendocrinopathy – Candidiasis – Ectodermal Dystrophy (APECED): From locus to defective protein. KTL A24/1999
34. **Minna Savukoski**, Molecular genetics of the late infantile neuronal ceroid lipofuscinosis

(LINCL): One gene (CLN5) and two gene loci (CLN2 and CLN6). KTL A25/1999

35. **Jyrki Kaukonen**, Autosomal dominant progressive external ophthalmoplegia (adPEO): A tale of two genomes. KTL A4/2000
36. **Tomi Pastinen**, Scoring human genomic SNPs and mutations: Multiplexed primer extension with manifolds and microarrays as solid-support. KTL A5/2000
37. **Hannele Kangas**, Familial amyloidosis of the Finnish type (FAF) – consequences of amyloidosis-associated mutation for gelsolin processing and function. KTL A9/2000
38. **Miina Öhman**, The search for genes predisposing to obesity. KTL A3/2001
39. **Jesper Ekelund**, Molecular genetics of schizophrenia and comorbid and related traits. KTL A17/2001
40. **Tarja Salonen**, Molecular and cellular biology of infantile neuronal ceroid lipofuscinosis (INCL). KTL77A16/2001
41. **Sonja Jaari**, Proteins involved in high density lipoprotein metabolism: A special reference to apolipoprotein AI, hepatic lipase and phospholipid transfer protein. KTL A1/2002
42. **Mari Auranen**, Molecular genetics of autism spectrum disorders in the Finnish population. KTL A23/2002
43. **Ilona Visapää**, Molecular genetics of the GRACILE syndrome. KTL A28/2002
44. **Saara Laitinen**, Family of human oxysterol binding protein homologues: ORP2 is a new regulator of cellular lipid metabolism. KTL A30/2002
45. **Juha Isosomppi**, Molecular and cell biology of infantile (CLN1) and variant late infantile (CLN5) neuronal ceroid lipofuscinoses. KTL A3/2003
46. **Maria Halonen**, Monogenic model for autoimmune diseases: Molecular basis of autoimmune polyendocrinopathy - candidiasis - ectodermal dystrophy (APECED). KTL A4/2003
47. **Juha Paloneva**, Two genes behind PLOSL: Molecular and pathological characteristics of the disease. KTL A8/2003
48. **Titta S. Blom**, Characterisation of cellular defects in Niemann-Pick type C disease. KTL A11/2003
49. **Nina Aula**, Molecular pathogenesis of Salla disease. KTL A19/2003
50. **Henna Haravuori**, Molecular genetics of tibial muscular dystrophy (TMD) and a novel distal myopathy. KTL A24/2003
51. **Riikka Nissinen**, Immunological features of chronic active rheumatoid arthritis. KTL A22/2003
52. **Jani Saarela**, Characterization of aspartylglucosaminidase activation and aspartylglucosaminuria mutations. KTL A1/2004
53. **Ville Holmberg**, CLN5 - from mutation to defective protein and clinical phenotype. KTL A2/2004
54. **Heidi Lilja**, Searching for genes predisposing to common dyslipidemias. KTL A16/2004