Annu. Rev. Genet. 2003. 37:197–219
doi: 10.1146/annurev.genet.37.110801.143820

First published online as a Review in Advance on July 11, 2003

GENETICS OF LACTASE PERSISTENCE AND LACTOSE INTOLERANCE

Dallas M. Swallow
Galton Laboratory, Department of Biology, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, England; email: dswallow@hgmp.mrc.ac.uk

Key Words milk, polymorphism, population distribution, developmental decline, functional DNA element

■ **Abstract** The enzyme lactase that is located in the villus enterocytes of the small intestine is responsible for digestion of lactose in milk. Lactase activity is high and vital during infancy, but in most mammals, including most humans, lactase activity declines after the weaning phase. In other healthy humans, lactase activity persists at a high level throughout adult life, enabling them to digest lactose as adults. This dominantly inherited genetic trait is known as lactase persistence. The distribution of these different lactase phenotypes in human populations is highly variable and is controlled by a polymorphic element *cis*-acting to the lactase gene. A putative causal nucleotide change has been identified and occurs on the background of a very extended haplotype that is frequent in Northern Europeans, where lactase persistence is frequent. This single nucleotide polymorphism is located 14 kb upstream from the start of transcription of lactase in an intron of the adjacent gene *MCM6*. This change does not, however, explain all the variation in lactase expression.

CONTENTS

INTRODUCTION TO LACTASE AND ITS VARIABLE DEVELOPMENTAL DECLINE 198
PROPERTIES OF THE LACTASE ENZYME 198
THE IDENTIFICATION OF DISTINCT LACTASE PHENOTYPES, LACTASE PERSISTENCE, AND LACTASE NONPERSISTENCE 199
EVIDENCE FOR A GENETIC CAUSE 200
POPULATION DISTRIBUTION OF THE LACTASE PERSISTENCE POLYMORPHISM 201
MOLECULAR BASIS OF THE LACTASE PERSISTENCE POLYMORPHISM 202
LCT PROMOTER STUDIES 204
SEARCH FOR NUCLEOTIDE CHANGES WITHIN LCT, HAPLOTYPES AND EVIDENCE OF A *CIS*-ACTING MECHANISM 205
A PUTATIVE CAUSAL POLYMORPHISM 207
EVIDENCE OF OTHER FACTORS THAT MAY INFLUENCE LACTASE ACTIVITY 212

CONGENITAL ALACTASIA 212
NUTRITIONAL CONSEQUENCES OF THE LACTASE PERSISTENCE POLYMORPHISM 213
THE ROLE OF SELECTION 213
EVOLUTIONARY PHYLOGENY OF THE *LCT* GENE SEGMENT 214

INTRODUCTION TO LACTASE AND ITS VARIABLE DEVELOPMENTAL DECLINE

The main carbohydrate in milk is the disaccharide lactose, which requires hydrolysis in the intestinal tract in order that the component monosaccharides galactose and glucose can be taken up into the enterocytes and used as a source of energy. This hydrolysis is catalyzed by the enzyme lactase, a β-galactosidase located in the brush border membrane of small-intestinal enterocytes. Lactase activity is high and vital during infancy, when milk is the main source of nutrition. In most mammals, however, lactase activity declines after the weaning phase, and this is also the case in the majority of humans throughout the world who are described as lactase nonpersistent. In other healthy humans, lactase activity persists at a high level throughout adult life, and this trait is known as lactase persistence. People who are lactase persistent can usually hydrolyze large amounts of lactose and can thus consume large quantities of fresh milk without complication. People with lactase nonpersistence (also referred to as adult-type hypolactasia or lactase restriction) have a much lower lactose digestion capacity (LDC)[1] than those with lactase persistence, and thus often, but not always, show symptoms of lactose intolerance after consumption of fresh milk.

The distribution of these different lactase phenotypes in human populations is highly variable, an observation that has long been a source of interest in relation to evolutionary genetics. This review describes the current information on the molecular and population genetics of this fascinating polymorphism.

PROPERTIES OF THE LACTASE ENZYME

Lactase or lactase-phlorizin hydrolase (E.C. 3.2.1.108), as it is often called, occurs on the apical surface of brush border enterocytes where it is anchored into the membrane by its C-terminal end, with the bulk of the molecule projecting into the lumen of the gut. It is a large glycoprotein, with two active sites, that can catalyze the hydrolysis of a variety of β glucosides, including phlorizin, flavonoid glucosides

[1]Abbreviations: SNP, single nucleotide polymorphism; LPH, lactase phlorizin hydrolase; LCT, lactase gene; LDC, lactose digestion capacity. For other gene symbols see: http://www.gene.ucl.ac.uk/nomenclature/ and for transcription factor acronyms (HNF1, Cdx, FREAC, GATA, see the associated references).

(57), and pyridoxine-5′-β-D glucoside (44), and β galactosides in addition to lactose. Lactase is encoded by a single gene (*LCT*) of approximately 50 kb (7) located on chromosome 2 (23). The gene has 17 exons (7) and encodes an mRNA transcript of 6274 nt (Genbank X07994) and a preproprotein of 1927 amino-acid residues (47). This is composed of a putative signal peptide of 19 amino-acid residues, a large pro-portion of 849 amino acids and a mature protein that contains two catalytic sites, and at the C-terminal end, a membrane-spanning domain and short cytoplasmic domain. *LCT* shows a fourfold internal homology, which suggests that it arose by two duplication events. Pro-lactase is proteolytically processed to a smaller protein (36, 49, 90), and two of the four homologous regions occur in the cleaved pro-portion of the molecule, which does not have a catalytic function, but probably has a chaperone function, in that it seems to play a role in transporting the molecule to the cell surface (33, 34, 53, 54, 59, 60). There is one active site in each of the domains of the mature protein. 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 (3, 91).

There is evidence that lactase is O-glycosylated through serines and threonines as well as N-glycosylated (through asparagine), and this glycosylation probably affects enzymatic activity as well as folding and intracellular transport (55).

Lactase expression is restricted to the enterocytes or absorptive cells of the small intestine; it is expressed at highest level in the mid-jejunum. This pattern of expression closely parallels that of another digestive hydrolase, sucrase-isomaltase (58). However, lactase is expressed only at low levels throughout fetal life, whereas sucrase is expressed at high levels in the small intestine of early fetuses, and also transiently in fetal colon (84).

THE IDENTIFICATION OF DISTINCT LACTASE PHENOTYPES, LACTASE PERSISTENCE, AND LACTASE NONPERSISTENCE

Although it was known since the end of the nineteenth century that lactase activity was lower in baby mammals than in adults (63), it was long believed that all adult humans had high levels of lactase, probably because most early research was conducted in countries where lactose tolerance was the most frequent phenotype. The first examples of lactase nonpersistence were therefore considered as an abnormal trait (4) and described as lactase deficiency, but it was soon recognized that this supposed abnormality was the most frequent trait worldwide (see 76 for review).

The adult lactase phenotypes can be determined directly by assay of the lactase from a small-intestinal biopsy or indirectly by lactose-tolerance tests (76),

although the discriminatory power of the tests is variable. Early population and family studies measured increase in blood glucose after giving a lactose load of 50 gm. People with high lactase activity have a significant rise in blood glucose concentration within 15 to 45 min following lactose administration, whereas those with low levels do not. Another method involves determination of urinary galactose after inclusion of ethanol with the lactose load (2, 32). In lactase-nonpersistent people, undigested lactose reaches the colon, where it is fermented, leading to the production of fatty acids and gases, including hydrogen, which is excreted in the breath. This property can be exploited in a more convenient and fairly reliable lactose-tolerance test, which involves testing breath hydrogen after a lactose load. Some investigators include all three measures and have obtained data on the correspondence of results and thus likely error rates (61).

Direct determination of lactase activities in intestinal samples is better than lactose-tolerance tests but is not a practical proposition for family studies. Also, even this procedure has some limitations. It is important to exclude secondary loss of lactase, due to epithelial damage. In addition to careful histology, determination of ratios of disaccharidase activities is more powerful than simple lactase activities in identifying adults with primary low lactase activity (18, 27, 28), because secondary loss is usually characterized by parallel loss of other brush border functions.

In lactase-nonpersistent adults, the reduction of lactase activity and onset of lactose intolerance usually begins between 2 and 3 years and is complete by the age of 5 to 10 years (71, 77, 85, 89). However, exceptions to this timing have been reported; for example, it occurs during adolescence in Finland (67).

EVIDENCE FOR A GENETIC CAUSE

Family studies, using lactose-tolerance tests, showed that these interindividual differences in lactase activity were due to a genetic polymorphism. The most convincing evidence came from large families from Finland (66). The studies suggested that adults with low LDC (who are lactase nonpersistent) are homozygous for an autosomal recessive allele that causes the post-weaning decline of lactase activity, whereas people who are lactase persistent (high LDC) are either heterozygous or homozygous for a dominant allele *LCT*P* that allows lactase to persist. Swallow & Hollox (76) summarize the published family studies from different parts of the world. Only eight cases have been reported in which individuals with high LDC were found in the progeny of parents who both had low LDC (20, 42). In none of these studies was nonpaternity excluded, so this fit with a genetic model is surprisingly good in view of difficulties in making a certain diagnosis!

Further evidence of the genetic nature of this variation came from a study of lactose digestion in twins. Concordance of the lactose-tolerance phenotypes was complete in monozygous twins, and the distribution of dizygous twin pairs corresponded with Hardy-Weinberg predictions (50). Studies of lactase enzyme activities in intestinal samples from adults were also in agreement: Using disaccharidase

activity ratios, a trimodal distribution of sucrase:lactase or maltase:lactase activity ratios was observed (18, 28). This activity distribution was compatible with the existence of three genotypes: homozygous persistent, homozygous nonpersistent, and heterozygotes. The observation of intermediate activity in the putative heterozygotes also implied that *cis*-acting differences (i.e., within or neighboring the lactase gene) are responsible for the polymorphism. The distribution of lactase activities shown as a sucrase/lactase ratio in children and adults is shown in Reference 76. Young babies show a unimodal distribution that shifts progressively to a trimodal distribution with age.

Lactase persistence behaves as a dominant trait because half levels of lactase activity are sufficient to show significant digestion of lactose. However, as the levels of intestinal hydrolases are not present in vast excess over requirement (87), under conditions of stress or mild pathology, heterozygotes may be more prone than lactase-persistent homozygotes to become lactose intolerant. In contrast, lactase-nonpersistent individuals do not necessarily experience symptoms of lactose intolerance, particularly if they consume only modest amounts of milk or other lactose-containing foods, e.g., the amount added to tea. Many genetic and nutritional factors may influence lactose digestion in a single individual (19) so that self-assessment of lactose-tolerance status is not very accurate. People seem to adjust their dietary intake of milk and milk products consciously or unconsciously to their individual lactose-tolerance threshold, and some adaptation can occur via changes in colonic flora. However, there is no evidence for adaptive alteration in lactase expression (37).

POPULATION DISTRIBUTION OF THE LACTASE PERSISTENCE POLYMORPHISM

The frequency of lactase persistence varies dramatically in different populations (76 and references therein). Lactase persistence is most prevalent in Northwestern Europe, with the highest frequency in Swedes and Danes, and there is a decline in frequency as one moves south and west (Figure 1*a*,*b*). A similar cline is seen from north to south of India (76). In general, the frequency is low in the rest of the world including Asiatic populations. Lactase persistence is, however, also frequent in milk-dependent nomads of the Afro-Arabian desert zone. In these regions, the pastoralists tend to have higher frequencies of lactase persistence than the neighboring nonpastoralists in the same countries (Figure 1*c*). However, some groups in which lactase nonpersistence is the most frequent phenotype do have milk as part of their diet, for example, the Mongols, the Herero, the Nuer (76), and the Dinka (Figure 1*c*). Several groups such as the Dinka may use cows as a status symbol, rather than as a significant source of nutrition. Milk is also often fermented to products such as yoghurt and cheese, which have a much reduced lactose content. Nevertheless, this correlation of culture and genetic trait led to the idea that the lactase persistence allele reached high frequency as a result of natural

selection in pastoralist populations that developed dairying during the Neolithic period (see section on The Role of Selection).

MOLECULAR BASIS OF THE LACTASE PERSISTENCE POLYMORPHISM

Although reports on lactase mRNA levels were conflicting in the earlier studies (13, 52, 69), it is now generally agreed that nonpersistent individuals have lower levels of lactase mRNA (14, 64, 84). The difference in mRNA level is more marked in some studies than others, with relatively higher levels of lactase mRNA found in lactase-nonpersistent individuals from Naples (64, 86). Genetic, or just conceivably environmental/dietary, differences may be responsible for this discrepancy. However, the discrimination is most likely clearer in duodenal samples, which were used in most studies, than in the jejunal samples used in Naples, even though this result would not have been predicted by the early studies of Newcomer & McGill (58), in which it was shown that the relative lactase and sucrase activities remained similar throughout the length of the intestine.

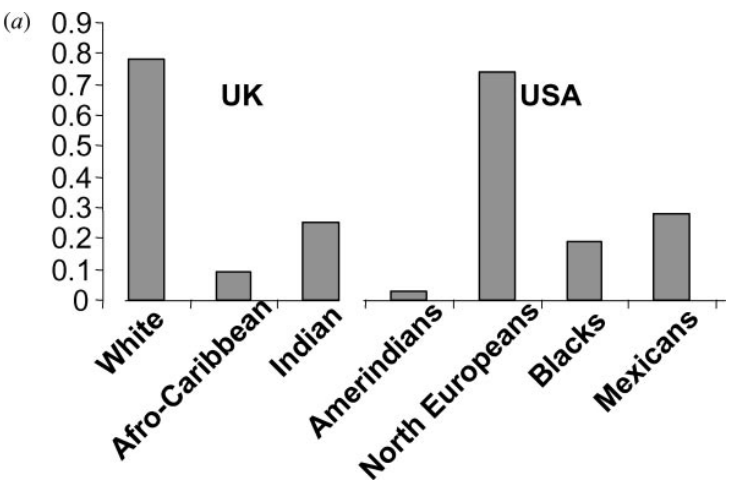

Figure 1 Frequency distribution of the lactase persistence allele *LCT*P* in different human populations. Data are taken from (76) and references therein, and *LCT*P* frequency is shown on the vertical axis. Representative population groups of 50 or more individuals are included. *a*. Populations of different ancestry in the United States and the United Kingdom. *b*. European populations shown in order of *LCT*P* frequency, illustrating the North/South and West/East clines. *c*. Comparison of frequencies in different Arab and African groups showing milk-drinking pastoralists and non-milk-drinkers from neighboring communities. Country of origin is shown below the population groups.

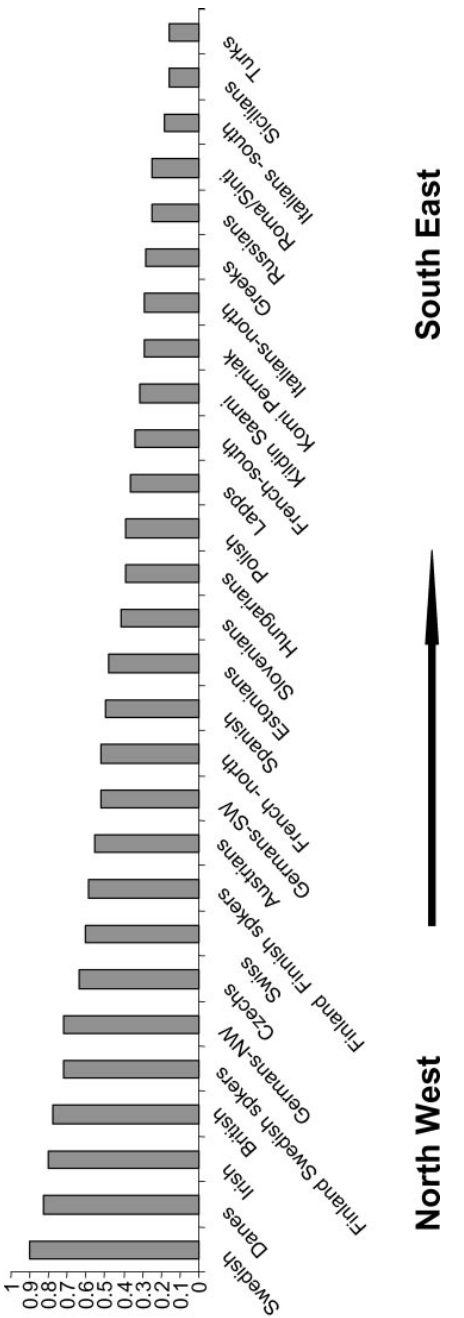

Figure 1 (*Continued*).

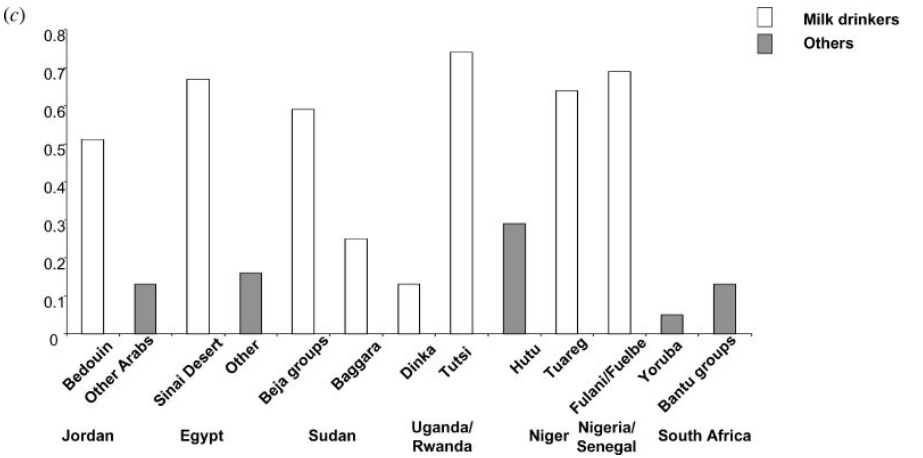

Figure 1 (*Continued*).

LCT PROMOTER STUDIES

Pig and rat lactase promoter functions have been studied in transgenic mice (38, 41, 78) and in both cases, 1–2 kb upstream from the start of transcription was sufficient for the normal tissue-specific expression and developmental down-regulation.

A number of *cis* elements with a putative role in *LCT* transcription have been identified in the sequences immediately upstream of the TATA box of the pig promoter and in other species. Experiments by a number of groups, which involve transfection of reporter constructs, yeast one-hybrid cloning, gel-shift assays together with the use of specific antibodies, introduction of specific mutations, and cotransfection of transcription factors, show evidence of interaction with the caudal homologue Cdx2, HNF1, HOXC11, FREACs, and GATA factors (15, 17, 30, 39, 51, 74, 79–82).

The human, rat, and pig promoters show stretches of homology, though the human promoter is disrupted by a pair of tail-to-tail Alu elements (Figure 2). Upstream of the Alu elements clear evidence of sequence homology is found with the pig sequence, but not the rat sequence. A diagram of the human promoter region is shown in Figure 2, and some putative (by homology) and experimentally determined transcriptional factor binding sites (using the naming given in the pig sequence) and areas that show sequence conservation across species are indicated, together with the positions of the upstream *LCT* SNPs.

Most of the well-characterized elements are located in regions where no single nucleotide polymorphisms are found. However, putative elements were found in a highly polymorphic region upstream of the Alu repeats that contained several SNPs (30). This region was studied in detail because, despite its high variability,

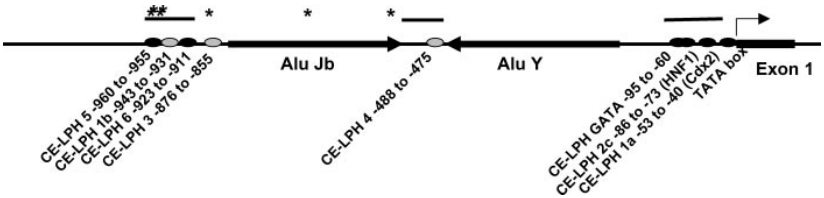

Figure 2 Diagram of the immediate upstream region of the *LCT* gene showing the positions of the reported SNPs, the Alu elements, and the *cis*-acting elements that have been identified. The SNPs (25, 30) are shown as asterisks; the likely functional transcription factor binding sites as black ovals (see text for references) and also putative sites identified by sequence homology with the known pig sequences and or transcription factor binding sites, as gray ovals. The elements are numbered according to the pig elements [see publications of Troelsen and colleagues (74)], with the two extra sites shown in the highly polymorphic region given the next numbers in the series. There are two GATA elements in the immediate promoter that surround the HNF1 site (39) and these *trans*-acting proteins are thought to act synergistically (81), as do HNF1α and Cdx-2 (51). An enhancer sequence that contains at least three *cis*-elements is located further upstream in the pig promoter (-814 to -894) but it is not yet clear whether this exists in humans (79).

it is highly conserved between human and pig, and contains sequence elements similar to Cdx sites. Evidence of protein binding was obtained, and for one of these elements located at approximately –950 to –965, the T allele of a CT SNP at –958 disrupts binding of a protein (8, 30). Transfection studies with both variants confirmed the effect in Caco-2 cells (8). However, this site cannot be causal of the phenotypic polymorphism because it does not show appropriate high association with persistence/nonpersistence.

SEARCH FOR NUCLEOTIDE CHANGES WITHIN LCT, HAPLOTYPES AND EVIDENCE OF A *CIS*-ACTING MECHANISM

Although sequencing of the region immediately upstream of *LCT*, and also of the exons, failed to identify any nucleotide difference that could be causal of the phenotypic polymorphism, a large number of single nucleotide polymorphisms (SNPs) were found (7, 25, 30, 43) (Figures 2, 3), which were useful as genetic markers. Eleven of the SNPs (shown as circles in Figure 3) were also used to define haplotypes across *LCT*. Only four common 60-kb 11 site haplotypes (**A**, **B**, **C**, and **U**, see Figure 3) were found worldwide, and of these, only three (**A**, **B**, **C**) exist in Europe (31).

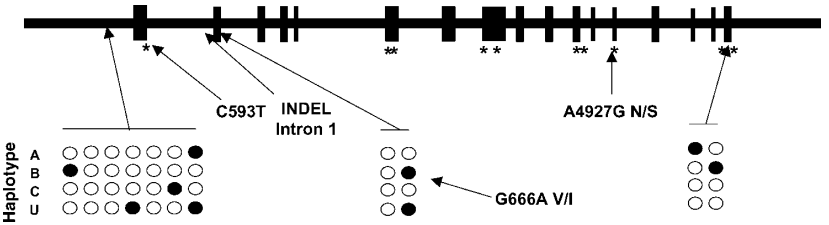

Figure 3 Diagram of the *LCT* gene showing the positions of the reported exonic and intronic SNPs described in (7, 25, 62). The 11 sites shown as circles are those used to construct the original haplotypes. The common haplotypes are also shown with the ancestral allele as white and the derived alleles as black. The ancestral allele was in most cases assumed by consideration of the allele in the chimpanzee at that position. The additional asterisked sites are as reported in (7), and those marked with arrows have been studied by us and are mentioned in the text. The two confirmed SNPs that lead to an amino-acid change (G666A V/I, and A4927G N/S) are indicated. The details on the precise position of the SNPs can be found on my home page: http://www.gene.ucl.ac.uk/mucin/.

Studies on mRNA levels that use the exonic polymorphisms to distinguish the two transcripts demonstrated clearly that in most lactase-persistent individuals, who were also heterozygous for the SNPs, only one of the two *LCT* homologues is expressed at high levels, the other accounting for 10% or less of the total lactase message (86). These individuals were interpreted as lactase-persistent heterozygotes who carry one persistent and one nonpersistent allele. Lactase-nonpersistent individuals who were heterozygous for exonic SNPs were shown to express equal but very low amounts of the two alleles (86). In contrast, young babies who are heterozygous for the same markers showed high equal expression of two transcripts (85). Evidence of progressive but variable down-regulation of expression of certain lactase transcripts was observed in children aged 20 months to 12 years (85) (see Figure 6, below). Informative fetuses showed low expression of both transcripts (85). These observations confirmed the notion that the sequence difference between individuals responsible for the differential down-regulation must be *cis*-acting, i.e., reside within or near *LCT*.

Association studies and analysis of mRNA transcript expression show highly significant association of high lactase expression with a single haplotype—the **A** haplotype, which is the most frequent in Europeans (24) and accounts for 87% of Northern European chromosomes.

Independent evidence that the lactase-persistence polymorphism was controlled by a sequence difference *cis*-acting to the gene came from linkage studies in the Finnish families (12), in which no recombination was detected between lactase persistence and microsatellites in a 3-Mb region of chromosome 2q21–22.

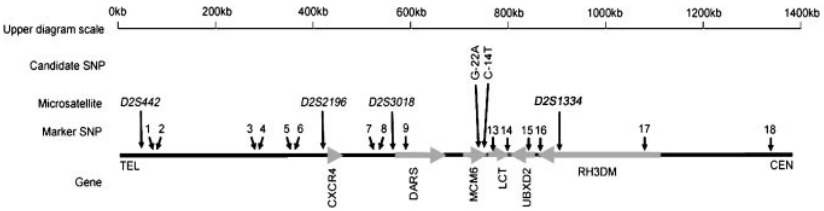

Figure 4 Region of linkage disequilibrium surrounding the lactase gene (*LCT*) showing the positions of markers used for the extended haplotypes in relation to physical distance and the known genes in the area. Statistically significant LD between sites 1 and 18 was found in Northern Europeans. Further details can be found in (62). Explanations of the gene symbols can be found at http://www.gene.ucl.ac.uk/nomenclature/. The position of the CT −14 kb and GA –22 kb SNPs described by (12) are also shown. The details on the precise position of the SNPs 1 to 18 can be found on my home page: http://www.gene.ucl.ac.uk/mucin/.

A PUTATIVE CAUSAL POLYMORPHISM

Haplotype and association studies using polymorphic markers at greater distances from the gene show that *LCT* resides within a very large region of Linkage Disequilibrium (LD). In our own studies using the CEPH (Centre d'Etude du Polymorphisme Humain) families, significant LD was observed across 1 Mb (62) (Figure 4), and the **A**, **B**, and **C** haplotypes extend over hundreds of kilobases.

Analysis, by another group (12), of microsatellite polymorphisms in the Finnish population showed a region of 200 kb of linkage disequilibrium. Within this region, these investigators identified a 47-kb region (located directly upstream of intron 1) that was identical with respect to these markers in all the lactase-persistent individuals whom they tested. It was assumed that this region was identical by descent and represented the largest region not disrupted by ancestral recombinations. They sequenced the 47-kb region in its entirety in a set of individuals of known lactase-persistence phenotype and genotype and identified a very large number of SNPs. They found a CT SNP at –14 kb in complete association with lactase persistence in the Finns from the original extended families (66) (n = 145, phenotyped by lactose tolerance) as well as unrelated individuals (n = 196, phenotyped by lactase assay) (12); the T allele was present in all persistent individuals and absent in nonpersistent individuals. A second SNP (G-22 kbA) was concordant with phenotype in all but a few rare individuals. Both SNPs were located within introns of the adjacent gene, *MCM6* (26 and references therein), a gene involved in the cell cycle and intriguingly, but probably coincidentally, first identified as being differentially expressed in intestinal crypts rather than villi. These findings together with the fact that the T allele at –14 kb is present at appropriate frequency in other populations (Table 1) led the authors to suggest that this

TABLE 1 Frequency of the CT –14-kb genotypes and T allele in comparison with reported frequencies of *LCT*P* in similar populations[a]

Population	Source of CT −14 kb data	N	TT	CT	CC	−14 kb T	*LCT*P* In similar population[a]	N
UK N. European	Own data[b]	37	20	15	2	0.74	0.78	163
USA N. Europeans (Utah CEPHs)	(12)	92	52	33	7	0.74	0.74	188
France (North) CEPHs	Own data[b]	24	5	12	7	0.49	0.52	62
Finns	(12)	938	322	446	170	0.58	0.59	449
S. European (mainly Italy south)	Own data[b]	40	0	11	29	0.14	0.18	197
Indians (North)	Own data[b]	12	2	4	6	0.33	0.48	264
USA African Americans	(12)	96	5	15	76	0.112	0.19	390

[a]From Table in Reference 76.

[b]From Reference 62 and G. Polanco & D.M. Swallow, unpublished.

nucleotide change is causal of persistence. Indeed, two laboratories in Finland are offering a diagnostic DNA test (Hyks-Laboratoridiagnostiikka, Medix Laboratoriot) as an alternative to lactose-tolerance testing. The nucleotide change is located within a potential AP-2 site, but it is not known whether this is functionally relevant.

Results from my own laboratory involving analysis of CEPH families show that both –22 kbA and –14 kbT occur uniquely on the background of a very extended **A** haplotype (62). Interestingly, another recently studied polymorphism, which involves a large (3.5 kb) insertion/deletion in intron 1 (62) (Figure 3), also subdivides the A haplotype, with the L allele being ancestral to S. This is better associated with phenotype than other **A** haplotype markers but less well associated than CT –14 kb or GA –22 kb (62). Eighteen of 22 of the T-carrying chromosomes tested show a common extended **A** haplotype (with intron 1 S) of more than 1 Mb in length (from points 3 to 18, on Figure 4). The shortest region carrying the T allele (points 8 to 18 on Figure 4) was more than 800 kb. In contrast, one of six –14 kb C carrying **A** haplotype chromosomes showed the alleles of the extended haplotype over the full 1 Mb region, and two of these showed the ancestral long allele in intron 1 (62). The T allele was also always found in association with **A** haplotype in individuals of Northern European, Southern European, Finnish, and Indian ancestry (Table 2). A similar difference in haplotype length was also observed for the deduced –14 kb T/**A** haplotypes and –24 kb C/**A** in the Finnish cohort we tested (62).

TABLE 2 Genotype for the CT −14-kb SNP compared with haplotype, described as **A** or not **A (X)**[a]

Genotype/ Haplotype*	French CEPH parents or grandparents			Finns			N. Europeans			S. Europeans			Indians		
	TT	**CT**	**CC**	**TT**	**CT**	**CC**	**TT**	**CT**	**CC**	**TT**	**CT**	**CC**	**TT**	**CT**	**CC**
AA	5	2	0	5	5	2	19	4	0	0	2	3	1	0	1
AX	**0**	10	5	**0**	7	15	**0**	10	1	**0**	12	11	**0**	8	4
XX	**0**	**0**	2	**0**	**0**	11	**0**	**0**	0	**0**	**0**	15	**0**	**0**	3

[a]Haplotype of all the French CEPH individuals and half the Indian individuals was deduced from family structure, whereas most of the others were assumed from the genotype data and the known haplotype frequencies in these populations. Note that the T allele is never found in association with non-**A** haplotype–relevant zero results shown in bold. In all 10 doubly heterozygous CEPHs the T allele was co-inherited with the **A** haplotype chromosome. Also note that the Finnish population was selected to consist of more nonpersistent than persistent individuals and is therefore not included in Table 1. Data from Reference 62 and G. Polanco & D.M. Swallow, unpublished.

Our own studies (62) also show a high level of association of this SNP with phenotype in Finns and in other groups we have tested (Table 3). This marker is clearly a better candidate than any others tested, including the GA SNP at −22 kb. However, one exception was found in the Finns and two clear exceptions were found (lactase-persistent individuals without a T allele) in the Italian cohort

TABLE 3 CT −14-kb genotypes in persistent and nonpersistent individuals, of Northern European ancestry (UK) Finland, Southern Europe (resident in Italy and UK) and India (resident in UK)[a]

Origin	Status	TT	CT	CC
Northern European	Persistent	20	15	0
	Nonpersistent	0	0	2
Finnish	Persistent	5	11	**1**
	Nonpersistent	0	0	24
Southern European	Persistent	0	11	**2**
	Nonpersistent	0	0	27
	Borderline	0	1	0
Indian	Persistent	2	4	0
	Nonpersistent	0	0	6

[a]Most of the Finnish and Indian samples were phenotyped by lactose-tolerance testing, whereas the others were done by enzyme assay. The discrepant results (i.e., lactase persistence in individuals not carrying a T allele) are shown in bold. The discrepancy observed in the Finnish cohort is discussed in detail in Reference 62 and probably reflects the inherent inaccuracies of the testing methods for lactose tolerance. The discrepancies observed in the Italian samples are further examples of non-**A** haplotype chromosomes associated with high lactase activity (24, 62). Data from Reference 62 and G. Polanco & D.M. Swallow, unpublished.

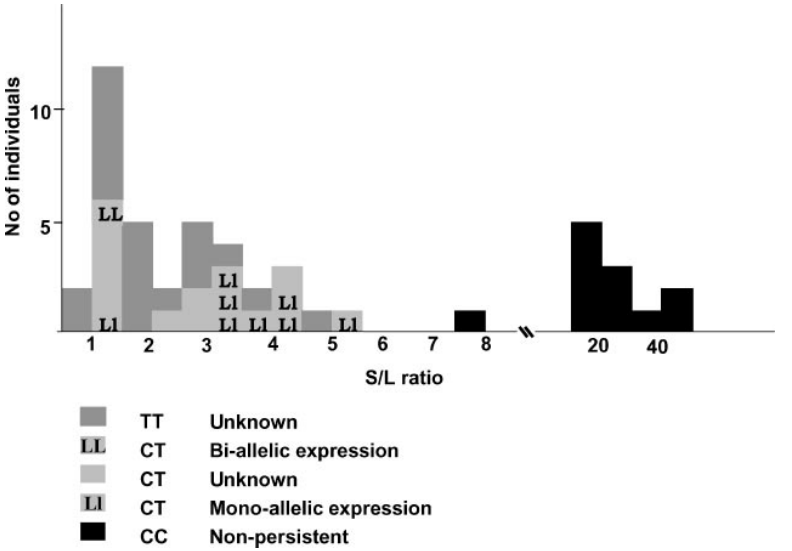

Figure 5 Histogram showing distribution of lactase activities, shown as sucrase/lactase ratios, in a cohort of individuals recruited in London, showing the trimodal distribution reported previously (27, 28, 62, 86). Individuals with low ratio are persistent whereas those with high ratio are nonpersistent. The cross slashes indicate the original cut point for determining lactase-persistence status. The one individual just to the left of the cut point was reported as "uncertain." The TT –14-kb genotype individuals are shown in dark gray shading, the CT –14-kb genotype in light gray shading, and CC individuals as black. Those individuals who show monoallelic expression and are thus heterozygotes for lactase persistence are indicated as Ll. The one individual showing high expression of both transcripts (see Figure 6) and thus a presumed homozygote is shown as LL.

(24) we tested. Although these exceptions may result from the inaccuracies of the phenotypic testing for lactose tolerance or lactase persistence, in our London samples the correlation with lactase activity was surprisingly not as good as might have been expected for a causal change. Figure 5 shows the distribution of sucrase/lactase ratios. In particular, those individuals already known from transcript expression studies to be heterozygous for lactase persistence showed the expected lower lactase activities than the whole group of persistent individuals, but the other lactase-persistent CT carriers did not: The known functional heterozygotes showed significantly lower lactase activities ($p = 0.03$) than the CT heterozygotes who were homozygous **A** haplotype and thus noninformative for transcript expression studies (62). This distinction was surprising and suggested a lack of complete correlation of expression level and CT status.

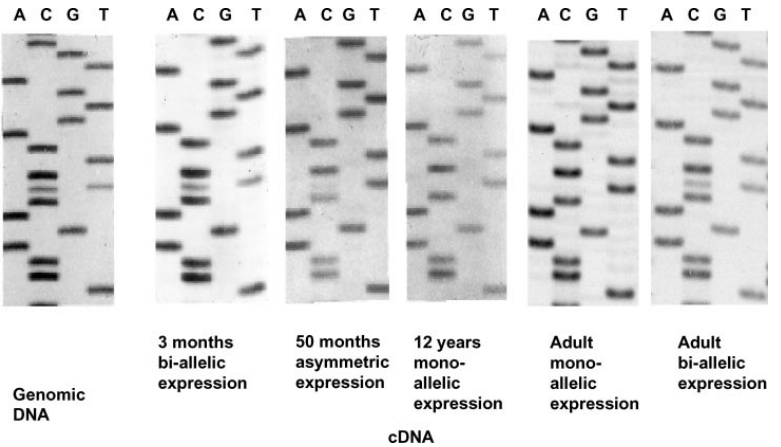

Figure 6 Sequencing runs from a representative genomic DNA sample of a heterozygote for C/T at nt 593 in exon 1 and from four cDNA samples from heterozygotes of different age and showing different levels of expression of the C allele (which is associated with **B** and **C** haplotypes) (85).

Furthermore, in one very well studied individual, in whom there was unambiguous high expression of lactase, we could show high (and equal) expression of both mRNA transcripts (as occurs in young babies), one of which was from a non-T-carrying chromosome, with extended **B** haplotype (Figure 6).

In contrast, in a recent Finnish study where relative transcript expression was also measured in CT heterozygotes using exonic SNPs, there was no difference in lactase activity between the informative and noninformative CT heterozygotes (40). Although there was overlap, the CT heterozygotes all fell nicely into an intermediate activity group.

Two explanations are possible for these discrepant observations: that the nucleotide is itself not causal but simply a highly associated marker, which postdates the causal mutation, and is somewhat more highly associated in Finns than in other Europeans; or that there is genetic heterogeneity. If the CT-14 kb is merely a highly associated marker, then the region that has been completely sequenced does not cover the entire candidate region, which seems to be much longer than 47 kb. Inspection of the microsatellite data shows that at the *LCT* side of the region a single marker disrupts the haplotype identity of the persistent chromosomes, although two more distant ones do not. This, and perhaps other of the microsatellites, are possibly highly mutable and may have diverged since the mutation leading to lactase persistence.

It remains to be seen from functional studies whether the CT at −14 kb is located in a genuinely functional element, whether it is responsible for persistence in other population groups such as African pastoralists, and whether there are any

other causal mutations. A separate change could be responsible for those cases where high expression is not associated with **A** haplotype. Such a change could in fact be *trans*-acting and might even not be under simple genetic control at all.

EVIDENCE OF OTHER FACTORS THAT MAY INFLUENCE LACTASE ACTIVITY

In measuring lactase in relation to lactase mRNA (but not in relation to sucrase), Rossi and colleagues (64) found several nonpersistent individuals with high lactase enzymatic activity in relation to their mRNA level. In some cases, evidence has been presented for abnormal or slower processing (75, 88) and abnormal electrophoretic mobility (27; C.B. Harvey & D.M. Swallow, unpublished data). These differences could perhaps reflect additional genetic variation. There are two exonic SNPs that alter amino acid sequence, one in the mature protein and one in the pro-protein (Figure 3). There have been no reported studies to determine whether either of these affects enzymatic activity, processing, or stability. There are also known interpersonal differences in terminal glycosylation of the carbohydrate side chains of lactase due to the ABO, Lewis and secretor polymorphisms (21), and these differences might possibly influence protein stability, since other glycosylation differences have been shown to do so (55).

Lactase expression is also undoubtedly affected by factors such as inflammation and is curiously more resistant to recovery from secondary loss in children than sucrase-isomaltase (83). There is a hint in the literature that *LCT* expression may also be affected by specific epigenetic effects. This suggestion came from the observation of a mosaic pattern of expression in jejunal samples from lactase-nonpersistent Neapolitans. Two distinct populations of cells were observed, some expressing and some not expressing lactase (45, 46), which suggests that groups of cells may have escaped the transcriptional control or silencing. Curiously, the spatial relationship of these cells was not consistent with the known clonal origin of intestinal villus cells, which would have been expected to appear as ribbons emerging from a single crypt stem cell (22).

CONGENITAL ALACTASIA

Although secondary loss of lactase in children is a frequent problem resulting from viral infection and allergy (83), true congenital deficiency of lactase or alactasia is very rare indeed, a cluster of cases has been reported in Finland (68), where this is one of the "Finnish recessive disorders." Very little is known about the molecular basis of this condition. No coding mutations have yet been reported, but the locus responsible for Finnish congenital alactasia has been genetically mapped 5′ of *LCT* (35), and these authors suggest that it is at a considerable distance from the gene (2 Mb). Elucidating the molecular changes that lead to this condition should throw further light on the regulation of lactase.

NUTRITIONAL CONSEQUENCES OF THE LACTASE PERSISTENCE POLYMORPHISM

The nutritional consequences of this polymorphism have long been a matter of controversy. Fresh milk supplies an excellent source of dietary calcium, which is generally perceived to be beneficial, as well as protective against osteoporosis (11), although some authors suggest that it confers increased risk of heart disease (70). Likewise, the higher consumption of galactose has been implicated in risk of cataract (6). The potential risks or benefits of digestion of plant glycosides have not yet been considered.

THE ROLE OF SELECTION

Most authors assume that selection has played a major role in determining the very variable frequencies of lactase persistence in different human populations. The possible selective factors suggested were nutritional benefit of milk, source of water in desert zones (9, 10), and in Northwestern Europe where there is low solar irradiation, improved calcium absorption (5, 19), on the assumption that rickets and osteomalacia were strong selective forces.

Several authors (16) have attempted to make mathematical models, often assuming coevolution of the cultural trait of milk drinking, to explain how modern gene frequencies of lactase persistence were reached within 6000–9000 years, the period since humans first domesticated livestock and practiced milk-based pastoralism. All models make the assumption that the ancestral state was non-persistence (the normal mammalian state), and this is strongly supported by the haplotype phylogenies (see below). Most of the models require high selection coefficients, typically 5%, and a reasonable starting gene frequency to reach a frequency of 0.70 in 6000–9000 years. However, genetic drift followed by later selection is perhaps a more probable explanation (56) of present-day frequencies, and the relevant mutation may have originated before the geographical expansion of modern humans. Little is known about the lactose tolerance status of higher primates, and it is an appealing idea that lactase persistence may have evolved as part of the changes that enabled higher primates to delay weaning and protect their young for longer, and space their births. However, study of the phylogeny of the *LCT* haplotypes (see section below) suggests that this mutation arose in humans. One group put forward the hypothesis that selection against lactase persistence and late weaning, by malaria, might account for the near-absence of this allele in much of Africa (1), but this hypothesis is not supported by other data.

Holden & Mace (29), in investigating the nature of the selective force while controlling for population (genetic) relatedness, suggested that lactase persistence is an adaptation to pastoralism (pastoralism developing before lactase persistence). However, no evidence was obtained for the added selective effect of highly arid environments or high latitudes (global solar irradiation).

EVOLUTIONARY PHYLOGENY OF THE *LCT* GENE SEGMENT

Studies on SNPs in the immediate vicinity of *LCT* show much more haplotype diversity in Africans than elsewhere in the world, where only four main 11 site haplotypes are common, **A**, **B**, **C**, and **U** (**U** not occurring in Europeans) (31). The four common haplotypes and their simple derivatives (haplotypes **A**, **B**, **C**, **U**, **O**, **D**, **N**, and **F**) comprise 89% of the total haplotype diversity in non-Africans, but only 51% in sub-Saharan Africans. Both the San and Bantu have high frequency of haplotypes (30% and 28%, respectively), which are the intermediate steps between the four common haplotypes, but these are almost absent in non-African populations ($\leq$3%). Haplotype networks were constructed to show the relationships of the common haplotypes to each other and to the root haplotype (Figure 7). The resultant networks emphasize that the **A**, **B**, and **U** haplotypes are not directly connected to the root and must have been generated by intermediate haplotypes that are now rare or absent from the non-African populations. This distribution suggested a population bottleneck or genetic drift, and loss of haplotype diversity outside of Africa. This bottleneck is likely to have occurred before modern humans were geographically spread over the Old World but recently enough for linkage disequilibrium to be observed across the region in modern non-African populations.

The **A** haplotype varies dramatically in frequency across Europe, with particularly high frequency in Northern Europe, mirroring the frequency of lactase

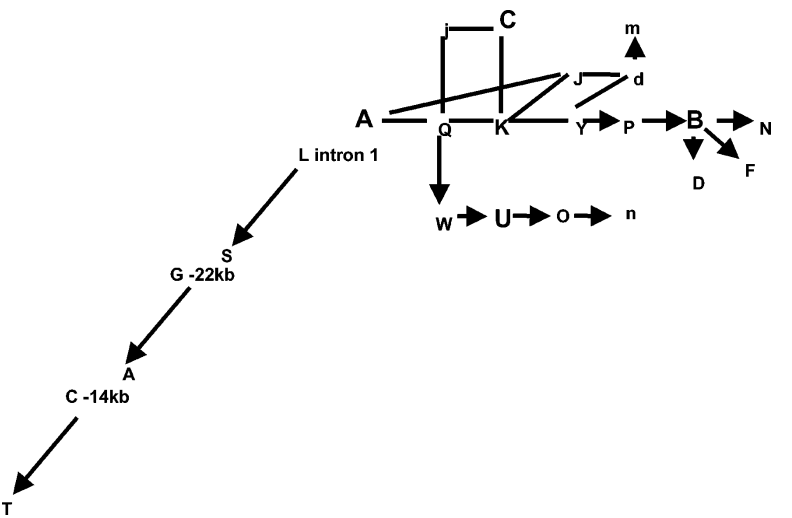

Figure 7 Haplotype network adapted from (31) showing the distant relationship of the **A**, **B**, **C**, and **U** haplotypes, and added onto the network, the succession of subsequent mutations on **A** haplotype chromosomes including the −14-kb C to T change.

persistence. This suggested a selective sweep, with selection acting on the causative locus within the same region of linkage disequilibrium. The T allele at −14 kb which is carried on an extended **A** haplotype may be that locus. The existence of this very extended **A** haplotype is also suggestive of a recent mutation and recent directional selection (48, 72, 73).

ACKNOWLEDGMENTS

I thank Charlotte Mulcare and Jesper Troelsen for their helpful comments on this manuscript, Mauro Rossi and Tariq Iqbal for supplying DNA samples from people typed for lactase persistence status, and Ed Hollox, Mark Poulter, and many other present and previous members of my group for their various contributions.

The *Annual Review of Genetics* is online at http://genet.annualreviews.org

LITERATURE CITED

1. Anderson B, Vullo C. 1994. Did malaria select for primary adult lactase deficiency? *Gut* 35:1487–89
2. Arola H, Koivula T, Jokela H, Jauhiainen M, Keyrilainen O, et al. 1988. Comparison of indirect diagnostic methods for hypolactasia. *Scand. J. Gastroenterol.* 23:351–57
3. Arribas JC, Herrero AG, Martin-Lomas M, Canada FJ, He S, et al. 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
4. Auricchio S, Rubino A, Semenza G, Landolt M, Prader A. 1963. Isolated intestinal lactase deficiency in the adult. *Lancet* 2:324–26
5. Birlouez-Aragon I. 1988. Effect of lactose hydrolysis on calcium absorption during duodenal milk perfusion. *Reprod. Nutr. Dev.* 28:1465–72
6. Birlouez-Aragon I, Ravelontseheno L, Villate-Cathelineau B, Cathelineau G, Abitbol G. 1993. Disturbed galactose metabolism in elderly and diabetic humans is associated with cataract formation. *J. Nutr.* 123:1370–76
7. 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
8. Chitkara DK, Krasinski SD, Grand RJ, Montgomery RK. 2001. Regulation of human lactase phlorizin hydrolase (LPH) gene by proteins binding to sites 5′ to the Alu sequence. *Gastroenterology* 120:A304
9. Cook GC. 1978. Did persistence of intestinal lactase into adult life originate in the Arabian peninsula? *Man* 13:418–27
10. Cook GC, Al-Torki MT. 1975. High intestinal lactase concentrations in adult Arabs in Saudi Arabia. *Br. Med. J.* 3:135–36
11. Di Stefano M, Veneto G, Malservisi S, Cecchetti L, Minguzzi L, et al. 2002. Lactose malabsorption and intolerance and peak bone mass. *Gastroenterology* 122:1793–99
12. Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, et al. 2002. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* 30:233–37
13. Escher JC, De Koning N, Van Engen C, Arora S, Büller HA, et al. 1992. Molecular basis of lactase levels in adult humans. *J. Clin. Invest.* 89:480–83
14. 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

15. 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
16. Feldman MW, Cavalli-Svorza LL. 1989. On the theory of evolution under genetic and cultural transmission with application to the lactose absorption problem. In *Mathematical Evolutionary Theory*, ed. MW Feldman, pp. 145. Princeton: Princeton Univ. Press
17. 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
18. Flatz G. 1984. Gene dosage effect on intestinal lactase activity demonstrated in vivo. *Am. J. Hum. Genet.* 36:306–10
19. Flatz G, Rotthauwe HW. 1971. Evidence against nutritional adaption to tolerance to lactase. *Humangenetik* 13:118–25
20. Gilat T, Benaroya Y, Gelman-Malachi E, Adam A. 1973. Genetics of primary adult lactase deficiency. *Gastroenterology* 64: 562–68
21. Green FR, Greenwell P, Dickson L, Griffiths B, Noades J, et al. 1988. Expression of the ABH, Lewis, related antigens on the glycoproteins of the human jejunal brush border. *Subcell. Biochem.* 12:119–53
22. Hamano Y, Kodama H, Fujikawa Y, Tanaka Y, Nishimura K, et al. 1988. Use of immunocytochemical analysis of a duodenal biopsy specimen to identify a carrier of ornithine transcarbamylase deficiency. *New Engl. J. Med.* 318:1521–23
23. Harvey CB, Fox MF, Jeggo PA, Mantei N, Povey S, et al. 1993. Regional localization of the lactase-phlorizin hydrolase gene, LCT, to chromosome 2q21. *Ann. Hum. Genet.* 57:179–85
24. Harvey CB, Hollox EJ, Poulter M, Wang Y, Rossi M, et al. 1998. Lactase haplotype frequencies in Caucasians: association with the lactase persistence/non-persistence polymorphism. *Ann. Hum. Genet.* 62: 215–23
25. 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
26. Harvey CB, Wang Y, Darmoul D, Phillips A, Mantei N, et al. 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
27. Harvey CB, Wang Y, Hughes LA, Swallow DM, Thurrell WP, et al. 1995. Studies on the expression of intestinal lactase in different individuals. *Gut* 36:28–33
28. 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.* 34:650–57
29. Holden C, Mace R. 1997. Phylogenetic analysis of the evolution of lactose digestion in adults. *Hum. Biol.* 69:605–28
30. 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
31. Hollox EJ, Poulter M, Zvarik M, Ferak V, Krause A, et al. 2001. Lactase haplotype diversity in the Old World. *Am. J. Hum. Genet.* 68:160–72
32. Isokoski M, Jussila J, Sarna S. 1972. A simple screening method for lactose malabsorption. *Gastroenterology* 62:28–32
33. Jacob R, Bulleid NJ, Naim HY. 1995. Folding of human intestinal lactase-phlorizin hydrolase. *J. Biol. Chem.* 270:18678–84
34. Jacob R, Peters K, Naim HY. 2002. The prosequence of human lactase-phlorizin hydrolase modulates the folding of the mature enzyme. *J. Biol. Chem.* 277:8217–25
35. Jarvela I, Enattah NS, Kokkonen J, Varilo T, Savilahti E, et al. 1998. Assignment of the locus for congenital lactase deficiency to 2q21, in the vicinity of but separate from

the lactase-phlorizin hydrolase gene. *Am. J. Hum. Genet.* 63:1078–85
36. Keller P, Zecca L, Boukamel R, Zwicker E, Gloor S, et al. 1995. Furin, PC1/3, and/or PC6A process rabbit, but not human, prolactase-phlorizin hydrolase to the 180-kDa intermediate. *J. Biol. Chem.* 270:25722–28
37. Keusch GT, Troncale FJ, Thavaramara B, Prinyanont P, Anderson PR, et al. 1969. Lactase deficiency in Thailand: effect of prolonged lactose feeding. *Am. J. Clin. Nutr.* 22:638–41
38. Krasinski SD, Upchurch BH, Irons SJ, June RM, Mishra K, et al. 1997. Rat lactase-phlorizin hydrolase/human growth hormone transgene is expressed on small intestinal villi in transgenic mice. *Gastroenterology* 113:844–55
39. 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
40. Kuokkanen M, Enattah NS, Oksanen A, Savilahti E, Orpana A, et al. 2003. Transcriptional regulation of the lactase-phlorizin hydrolase gene by polymorphisms associated with adult-type hypolactasia. *Gut* 52:647–52
41. Lee SY, Wang Z, Lin CK, Contag CH, Olds LC, et al. 2002. Regulation of intestine-specific spatiotemporal expression by the rat lactase promoter. *J. Biol. Chem.* 277: 13099–105
42. Lisker R, Gonzalez B, Daltabuit M. 1975. Recessive inheritance of the adult type of intestinal lactase deficiency. *Am. J. Hum. Genet.* 27:662–64
43. Lloyd M, Mevissen G, Fischer M, Olsen W, Goodspeed D, et al. 1992. Regulation of intestinal lactase in adult hypolactasia. *J. Clin. Invest.* 89:524–29
44. Mackey AD, Henderson GN, Gregory JF 3rd. 2002. Enzymatic hydrolysis of pyridoxine-5′-beta-D-glucoside is catalyzed by intestinal lactase-phlorizin hydrolase. *J. Biol. Chem.* 277:26858–64
45. Maiuri L, Raia V, Potter J, Swallow D, Ho MW, et al. 1991. Mosaic pattern of lactase expression by villous enterocytes in human adult-type hypolactasia. *Gastroenterology* 100:359–69
46. Maiuri L, Rossi M, Raia V, Garipoli V, Hughes LA, et al. 1994. Mosaic regulation of lactase in human adult-type hypolactasia. *Gastroenterology* 107:54–60
47. Mantei N, Villa M, Enzler T, Wacker H, Boll W, et al. 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
48. McPeek MS, Strahs A. 1999. Assessment of linkage disequilibrium by the decay of haplotype sharing, with application to fine-scale genetic mapping. *Am. J. Hum. Genet.* 65:858–75
49. Mesonero JE, Gloor SM, Semenza G. 1998. Processing of human intestinal prolactase to an intermediate form by furin or by a furin-like proprotein convertase. *J. Biol. Chem.* 273:29430–36
50. Metneki J, Czeizel A, Flatz S, Flatz G. 1984. A study of lactose absorption capacity in twins. *Hum. Genet.* 67:296–300
51. 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(Part 2):529–35
52. Montgomery RK, Büller HA, Rings EHHM, Grand RJ. 1991. Lactose intolerance and the genetic regulation of intestinal lactase-phlorizin hydrolase. *FASEB J.* 5:2824–32
53. Naim HY. 1995. The pro-region of human intestinal lactase-phlorizin hydrolase is enzymatically inactive towards lactose. *Biol. Chem. Hoppe-Seyler* 376:255–58
54. Naim HY, Jacob R, Naim H, Sambrook JF, Gething MJ. 1994. The pro region of human intestinal lactase-phlorizin hydrolase. *J. Biol. Chem.* 269:26933–43

55. 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
56. Nei M, Saitou N. 1986. Genetic relationship of human populations and ethnic differences in relation to drugs and food. In *Ethnic Differences in Reactions to Drugs and Other Xenobiotics*, ed. W Kalow, HW Goedde, DP Agarwal, pp. 21–37. New York: Alan R Liss
57. Nemeth K, Plumb GW, Berrin JG, Juge N, Jacob R, et al. 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
58. Newcomer A, McGill D. 1966. Distribution of disaccharidase activity in the small bowel of normal and lactase-deficient subjects. *Gastroenterology* 51:481–88
59. Oberholzer T, Mantei N, Semenza G. 1993. The pro sequence of lactase-phlorizin hydrolase is required for the enzyme to reach the plasma membrane. An intramolecular chaperone? *FEBS Lett.* 333:127–31
60. Ouwendijk J, Peters WJ, van de Vorstenbosch RA, Ginsel LA, Naim HY, et al. 1998. Routing and processing of lactase-phlorizin hydrolase in transfected Caco-2 cells. *J. Biol. Chem.* 273:6650–55
61. Peuhkuri K, Poussa T, Korpela R. 1998. Comparison of a portable breath hydrogen analyser (Micro H2) with a Quintron MicroLyzer in measuring lactose maldigestion, and the evaluation of a Micro H2 for diagnosing hypolactasia. *Scand. J. Clin. Lab. Invest.* 58:217–24
62. Poulter M, Hollox E, Harvey CB, Mulcare CM, Peuhkuri K, et al. 2003. The causal element for the lactase persistence/non-persistence polymorphism is located in a 1Mb region of linkage disequilibrium in Europeans. *Ann. Hum. Genet.* 67:298–311
63. Röhmann F, Nagano J. 1903. Über die Resorption und die fermentative Spaltung der Disaccharide im Dünndarm des ausgewachsenen Hundes. *Arch. Ges. Physiol.* 95:533–605
64. Rossi M, Maiuri L, Fusco MI, Salvati VM, Fuccio A, et al. 1997. Lactase persistence versus decline in human adults: Multifactorial events are involved in down-regulation after weaning. *Gastroenterology* 112:1506–14
65. Deleted in proof
66. Sahi T. 1974. The inheritance of selective adult-type lactose malabsorption. *Scand. J. Gastroenterol.* 9:1–73
67. Sahi T, Launiala K, Laitinen H. 1983. Hypolactasia in a fixed cohort of young Finnish adults, a follow-up study. *Scand. J. Gastroenterol.* 18:865–70
68. Savilahti E, Launiala K, Kuitunen P. 1983. Congenital lactase deficiency. A clinical study on 16 patients. *Arch. Dis. Child.* 58:246–52
69. Sebastio G, Villa M, Sartorio R, Guzzetta V, Poggi V, et al. 1989. Control of lactase in human adult-type hypolactasia and in weaning rabbits and rats. *Am. J. Hum. Genet.* 45:489–97
70. Seely S. 2000. Possible connection between milk and coronary heart disease: the calcium hypothesis. *Med. Hypotheses* 54:701–3
71. Simoons F. 1980. Age of onset of lactose malabsorption. *Pediatrics* 66:646–48
72. Slatkin M. 2000. Balancing selection at closely linked, overdominant loci in a finite population. *Genetics* 154:1367–78
73. Slatkin M, Bertorelle G. 2001. The use of intraallelic variability for testing neutrality and estimating population growth rate. *Genetics* 158:865–74
74. Spodsberg N, Troelsen JT, Carlsson P, Enerback S, Sjostrom H, et al. 1999. Transcriptional regulation of pig lactase-phlorizin hydrolase: involvement of HNF-1 and FREACs. *Gastroenterology* 116:842–54
75. Sterchi E, Mills P, Fransen J, Hauri H,

Lentze M, et al. 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

76. Swallow DM, Hollox EJ. 2000. The genetic polymorphism of intestinal lactase activity in adult humans. In *The Metabolic and Molecular Basis of Inherited Disease*, ed. CR Scriver, AL Beaudet, WS Sly, D Valle, pp. 1651–62. New York: McGraw-Hill
77. Thomas S, Walker-Smith JA, Senewiratne B, Hjelm M. 1990. Age dependency of the lactase persistence and lactase restriction phenotypes among children in Sri Lanka and Britain. *J. Trop. Pediatr.* 36:80–85
78. Troelsen JT, Mehlum A, Olsen J, Spodsberg N, Hansen GH, et al. 1994. 1 kb of the lactase-phlorizin hydrolase promoter directs post-weaning decline and small intestinal-specific expression in transgenic mice. *FEBS Lett.* 342:291–96
79. Troelsen JT, Mitchelmore C, Olsen J. 2003. An enhancer activates the pig lactase phlorizin hydrolase promoter in intestinal cells. *Gene* 305:101–11
80. Troelsen JT, Olsen J, Mitchelmore C, Hansen GH, Sjostrom H, et al. 1994. Two intestinal specific nuclear factors binding to the lactase-phlorizin hydrolase and sucrase-somaltase promoters are functionally related oligomeric molecules. *FEBS Lett.* 342:297–301
81. van Wering HM, Huibregtse IL, van der Zwan SM, de Bie MS, Dowling LN, et al. 2002. 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
82. van Wering HM, Moyer L, Grand RJ, Krasinski SD. 2002. Novel interaction at the Cdx-2 binding sites of the lactase-phlorizin hydrolase promoter. *Biochem. Biophys. Res. Commun.* 299:587–93
83. Walker-Smith J. 1997. Diarrheal disease. *Nestle Nutr. Workshop Ser.* 38
84. Wang Y, Harvey C, Rousset M, Swallow DM. 1994. Expression of human intestinal mRNA transcripts during development: analysis by a semiquantitative RNA polymerase chain reaction method. *Pediatr. Res.* 36:514–21
85. Wang Y, Harvey CB, Hollox EJ, Phillips AD, Poulter M, et al. 1998. The genetically programmed down-regulation of lactase in children. *Gastroenterology* 114:1230–36
86. Wang Y, Harvey CB, Pratt WS, Sams VR, Sarner M, et al. 1995. The lactase persistence/non-persistence polymorphism is controlled by a *cis*-acting element. *Hum. Mol. Genet.* 4:657–62
87. Weiss SL, Lee EA, Diamond J. 1998. Evolutionary matches of enzyme and transporter capacities to dietary substrate loads in the intestinal brush border. *Proc. Natl. Acad. Sci. USA* 95:2117–21
88. Witte J, Lloyd M, Lorenzsonn V, Korsmo H, Olsen W. 1990. The biosynthetic basis of adult lactase deficiency. *J. Clin. Invest.* 86:1338–42
89. Wittenberg D, Moosa A. 1990. Lactose maldigestion—age specific prevalence in black and Indian children. *S. Afr. Med. J.* 78:470–72
90. Zecca L, Mesonero JE, Gloor SM, Semenza G. 1999. Species differences in the sites of cleavage of pro-lactase to lactase supports lack of selective pressure. *Biochim. Biophys. Acta* 1435:51–60
91. Zecca L, Mesonero JE, Stutz A, Poiree JC, Giudicelli J, et al. 1998. Intestinal lactase-phlorizin hydrolase (LPH): the two catalytic sites; the role of the pancreas in pro-LPH maturation. *FEBS Lett.* 435:225–28

Annual Review of Genetics
Volume 37, 2003

CONTENTS

IRA HERSKOWITZ (1946–2003), *The Editors* 1

TRANSPOSON-BASED STRATEGIES FOR MICROBIAL FUNCTIONAL GENOMICS AND PROTEOMICS, *Finbarr Hayes* 3

ERROR-PRONE DNA POLYMERASES: WHEN MAKING A MISTAKE IS THE ONLY WAY TO GET AHEAD, *Alison J. Rattray and Jeffrey N. Strathern* 31

GENETICS OF HAIR AND SKIN COLOR, *Jonathan L. Rees* 67

THIOL-BASED REGULATORY SWITCHES, *Mark S.B. Paget and Mark J. Buttner* 91

PSEUDOGENES: ARE THEY "JUNK" OR FUNCTIONAL DNA? *Evgeniy S. Balakirev and Francisco J. Ayala* 123

UNUSUAL LIFE STYLE OF GIANT CHLORELLA VIRUSES, *James L. Van Etten* 153

GENETICS OF LACTASE PERSISTENCE AND LACTOSE INTOLERANCE, *Dallas M. Swallow* 197

CELL POLARITY AND THE CYTOSKELETON IN THE *CAENORHABDITIS ELEGANS* ZYGOTE, *Stephan Q. Schneider and Bruce Bowerman* 221

THE SPINDLE ASSEMBLY AND SPINDLE POSITION CHECKPOINTS, *Daniel J. Lew and Daniel J. Burke* 251

LATERAL GENE TRANSFER AND THE ORIGINS OF PROKARYOTIC GROUPS, *Yan Boucher, Christophe J. Douady, R. Thane Papke, David A. Walsh, Mary Ellen R. Boudreau, Camilla L. Nesbø, Rebecca J. Case, and W. Ford Doolittle* 283

GENETICS OF AGING IN THE FRUIT FLY, *DROSOPHILA MELANOGASTER*, *Stephen L. Helfand and Blanka Rogina* 329

NATURAL SELECTION AND THE EVOLUTION OF GENOME IMPRINTING, *Elena de la Casa-Esperón and Carmen Sapienza* 349

THE NEED FOR WINTER IN THE SWITCH TO FLOWERING, *Ian R. Henderson, Chikako Shindo, and Caroline Dean* 371

TRANSMISSION RATIO DISTORTION IN MICE, *Mary F. Lyon* 393

STRUCTURE, DIVERSITY, AND EVOLUTION OF PROTEIN TOXINS FROM SPORE-FORMING ENTOMOPATHOGENIC BACTERIA, *Ruud A. de Maagd, Alejandra Bravo, Colin Berry, Neil Crickmore, and H. Ernst Schnepf* 409

YEAST VACUOLE INHERITANCE AND DYNAMICS, *Lois S. Weisman* 435

HEPARAN SULFATE CORE PROTEINS IN CELL-CELL SIGNALING, *Kenneth L. Kramer and H. Joseph Yost* 461

RETROTRANPOSONS PROVIDE AN EVOLUTIONARILY ROBUST NON-TELOMERASE MECHANISM TO MAINTAIN TELOMERES, *Mary-Lou Pardue and P.G. DeBaryshe* 485

A CYANOBACTERIAL CIRCADIAN TIMING MECHANISM, *S.S. Golden, J.L. Ditty, and S.B. Williams* 513

REGULATION OF CELL CYCLES IN DROSOPHILA DEVELOPMENT: INTRINSIC AND EXTRINSIC CUES, *Laura A. Lee and Terry L. Orr-Weaver* 545

RECOGNITION AND RESPONSE IN THE PLANT IMMUNE SYSTEM, *Zachary Nimchuk, Thomas Eulgem, Ben F. Holt III, and Jeffrey L. Dangl* 579

RECA-DEPENDENT RECOVERY OF ARRESTED DNA REPLICATION FORKS, *Justin Courcelle and Phillip C. Hanawalt* 611

INDEXES

Subject Index 647

Cumulative Index of Contributing Authors, Volumes 33–37 679

Cumulative Index of Chapter Titles, Volumes 33–37 682

ERRATA

An online log of corrections to *Annual Review of Genetics* chapters may be found at http://genet.annualreviews.org/errata.shtml