

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

**A DIVERSIDADE DO GENE *LCT* E A PERSISTÊNCIA DA
LACTASE NA POPULAÇÃO BRASILEIRA.**

Deise Cristine Friedrich

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutor em Ciências (Genética e Biologia Molecular).

Orientadora: Dra Mara Helena Hutz

Porto Alegre, junho de 2013

INSTITUIÇÕES E FONTES FINANCIADORAS

O presente trabalho foi desenvolvido no Laboratório de Genética Humana Molecular, Departamento de Genética da Universidade Federal do Rio Grande do Sul e no Laboratório de Gastroenterologia Pediátrica da Stanford University, com financiamento do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq e Programa Ciência sem Fronteiras.

AGRADECIMENTOS

Esta tese significa uma meta alcançada, o título de doutorado. Gostaria de aqui agradecer a todos que fizeram parte desta conquista, principalmente a minha orientadora, Dra Mara Hutz, que me deu a oportunidade de realizar um sonho, deu apoio, muitos ensinamentos que levarei comigo e tenho certeza que serão muito importantes e úteis sempre. Para ela, meu eterno muito obrigada!

Gostaria de agradecer a todos os colegas do Laboratório de Genética Humana Molecular pelo companheirismo, pelos cafés com risadas pós almoço e pelas indispensáveis ajudas que me deram. Estes 4 anos foram magníficos e muito agradáveis ao lado de vocês!

À Profa Sídia Callegari Jacques meu muito obrigada por toda a paciência e ajuda com as análises estatísticas.

Ao Dr Eric Sibley, agradeço pela oportunidade e por ter me recebido na Stanford University.

À Gianina Teribele Venturin, minha eterna roommate, gostaria de agradecer pela amizade que fez com que o período do doutorado sanduíche fosse muito divertido e inesquecível.

Um muito obrigada também para meus amigos, que entenderam as ausências, aos professores do Departamento de Genética e ao Elmo Cardoso, indispensável na nossa luta para que tudo dê certo.

Finalmente, um obrigada de todo o coração à minha família e ao Eduardo Rosa Franco, meu amor, por sempre terem me apoiado, mesmo nas decisões mais dolorosas como ficar meses longe de casa, e por terem acreditado no meu sonho que agora está se realizando. Devo tudo que sou a vocês! Obrigada!

SUMÁRIO

Lista de abreviaturas	6
Resumo	7
Abstract	9
Capítulo 1: Introdução	11
1.1 O leite	12
1.2 A hipolactasia primária adulta e a intolerância à lactose	13
1.3 Padrões de expressão da enzima e a persistência da lactase	14
1.4 A enzima lactase e o gene <i>LCT</i>	18
1.5 Frequência dos polimorfismos presentes na região <i>enhancer</i>	21
1.6 Estudos funcionais	27
1.7 A diversidade do gene <i>LCT</i> e evolução da persistência da lactase	29
1.8 Saúde e considerações médicas	35
1.9 Justificativa e Objetivo	40
Capítulo 2: Stability or variation? Patterns of lactase gene and its enhancer region distributions in Brazilian Amerindians.	42
Capítulo 3: Several different lactase persistence associated alleles and high diversity of the lactase gene in the admixed Brazilian population.	51
Capítulo 4: The lactase persistence genotype is a protective factor for the metabolic syndrome.	65
Capítulo 5: Discussão	77
5.1 Estudos populacionais e evolutivos	78
5.2 Considerações sobre a saúde	79
5.3 O uso de -13910C>T como diagnóstico molecular	81
5.4 A importância dos estudos funcionais	81
5.5 Causalidade da persistência da lactase	82
5.6 Considerações finais	83
Referências	84
Anexos	
Anexo 1: A análise funcional da variante -13937G>A em células Caco-2	97

Referências do anexo 1	105
Anexo 2: Tabela de frequências dos alelos derivados dos polimorfismos do gene <i>LCT</i> nas populações brasileiras estudadas	106

LISTA DE ABREVIATURAS

IMC – índice de massa corporal

Kb – quilo bases ou mil pares de bases

KDa – quilo Daltons

LCT- gene da lactase

Mb – mega bases

NPL – não persistência da lactase

OR – *odds ratio*, razão de chances

pb – pares de base

PCR-RFLP – Reação de polimerização em cadeia e análise por tamanho de fragmento de restrição (*Polymerase Chain Reaction – Restriction Fragment Length Polymorphism*)

PL – persistência da lactase

SM – síndrome metabólica

SNP – polimorfismo de base única

RESUMO

A hipolactasia do tipo adulto é o fenótipo determinado pela diminuição da expressão da lactase após o período de lactação. Ela ocorre em um grande número de adultos em todo o mundo. A lactase é produzida pelos enterócitos e sua função principal é hidrolisar a lactose, que é o carboidrato do leite. Os indivíduos intolerantes à lactose irão apresentar sintomas como inchaço, flatulência, náusea e diarreia causados pela fermentação da lactose.

A persistência da lactase (PL) é o fenótipo no qual a expressão da lactase se mantém elevada durante toda a vida. Na Europa, a PL foi relacionada a um polimorfismo de base única (SNP) localizado a aproximadamente 14 Kb do sítio de início da transcrição do *LCT* (gene da lactase), dentro de um íntron do gene *MCM6*, sendo este SNP uma troca de C para T na posição -13910 (rs4988235). Na África e Oriente Médio os seguintes SNPs foram relacionados a PL: -13907C>G (rs41525747), -13915T>G (rs41380347), -14010G>C (rs145946881).

O gene *LCT* também possui SNPs na região codificadora e na região promotora que não estão envolvidos com a PL. Estes SNPs apresentam alto desequilíbrio de ligação formando haplótipos, sendo que os haplótipos A, B, C e U são os mais frequentes na maioria das populações.

No Brasil, dados sobre os alelos relacionados com a persistência da lactase são escassos. Além disso, dados populacionais relacionados à diversidade do gene *LCT* ainda não foram descritos para nossas populações. Portanto, o objetivo deste trabalho foi estudar a diversidade do gene *LCT*, da região codificadora do gene, da região promotora proximal e da região *enhancer* população brasileira.

Um total de 1297 indivíduos foram analisados. As populações estudadas foram nativos brasileiros (Kaingang N=72, Xavante N=101, Guarani-Kaiowá N=84 e Guarani-Ñandeva N=59), eurodescendentes de Porto Alegre (Rio Grande do Sul, N=337), afrodescendentes de Porto Alegre (N=182), miscigenados de Belém (Pará, N=200) e de Recife (Pernambuco, N=262).

Doze SNPs foram analisados, 10 nas regiões codificadora e promotora do *LCT* e 2 na região *enhancer*. As metodologias utilizadas na genotipagem destes SNPs foram PCR-RFLP, discriminação alélica pelo sistema TaqMan e

sequenciamento. O sequenciamento também foi utilizado na busca de novos alelos da região *enhancer*. Com relação à população nativa, o único alelo de PL encontrado foi o -13910*T, variando de 0,5% em Xavante a 7,6% nos Guarani-Ñandeva. O gene *LCT* foi altamente polimórfico apresentando 15 haplótipos com distribuição heterogênea nas populações nativas.

Na população brasileira em geral, a frequência do alelo -13910*T foi maior (0,295) nos eurodescendentes de Porto Alegre e menor (0,175) na população de Belém. Nos grupos de afrodescendentes de Porto Alegre, Belém e Recife, 4 outras variantes, previamente descritas, da região *enhancer* foram encontradas: -13779G>C, -13937G>A, -14010G>C, -14011C>T. Vinte e seis haplótipos previamente descritos foram identificados.

O estudo de associação da presença do alelo -13910*T com a presença da síndrome metabólica nos eurodescendentes de Porto Alegre demonstrou que os indivíduos persistentes apresentam menor risco do que os não persistente de ter síndrome metabólica (OR=0,47; p=0,023).

Na tentativa de auxiliar no entendimento das causas da PL foi realizado um estudo funcional da variante -13937G>A. Os resultados demonstraram que o alelo derivado não direciona maior expressão do gene repórter em células em cultura.

Considerando os dados obtidos no presente trabalho e os disponíveis na literatura, ressaltamos a importância dos estudos que buscam compreender a PL pela busca de novos alelos, por estudos de correlação fenótipo-genótipo e também pelos estudos funcionais para a caracterização das variantes encontradas em relação ao fenótipo da lactase.

ABSTRACT

Adult-type hypolactasia is the phenotype determined by the decreased lactase expression after weaning. It occurs in a high number of adults in the world. Lactase is produced by the enterocytes and its major function is to hydrolyze lactose, the milk carbohydrate. The lactose intolerant individuals will have symptoms like bloating, flatulence, nausea and diarrhea caused by the lactose fermentation.

Lactase persistence (LP) is the high lactase expression during adulthood. In Europe, the LP was related to a single nucleotide polymorphism (SNP) located approximately 14 Kb from the *LCT* (lactase gene) transcription initiation site, within a *MCM6* gene intron, and this SNP is a C to T mutation in the -13910 position (rs4988235). In Africa and Middle East, the following SNPs were related to LP: -13907C>G (rs41525747), -13915T>G (rs41380347), -14010G>C (rs145946881).

LCT gene also has SNPs in the coding and promoter region that are not involved in the LP. These SNPs have high linkage disequilibrium forming haplotypes, with the A, B, C and U being the most frequent haplotypes in the majority of the populations.

In Brazil, data about the LP related alleles are rare. Moreover, population data related to *LCT* gene diversity was not described for our population. Hence, the aim of this work was to study the *LCT* gene diversity in the coding region, in the proximal promoter region, and in the enhancer region in the Brazilian population.

In total, 1297 individuals were investigated. The populations studied were Brazilian natives (Kaingang N=72, Xavante N=101, Guarani-Kaiowá N=84 and Guarani-Ñandeva N=59), Eurodescendants from Porto Alegre (Rio Grande do Sul state N=337), Afrodescendants from Porto Alegre (N=182), admixed individuals from Belém (Pará state, N=200) and from Recife (Pernambuco state, N=262).

We analyzed 12 SNPs, 10 in the coding and promoter region of the *LCT* gene and 2 in the enhancer region. The genotyping methodologies applied were PCR-RFLP, allelic discrimination by TaqMan system and sequencing. Sequencing was also employed for new alleles identification in the enhancer region. In relation to the native population, the only LP allele found was -13910*T, and the frequency

ranged from 0.5% in Xavante to 7.6% in Guarani-Ñandeva. The *LCT* gene was highly polymorphic showing 15 haplotypes with heterogeneous distribution in the native populations.

In the general population, the frequency of the -13910*T was higher (0.295) in Eurodescendants from Porto Alegre and lower (0.175) in the Belém population. In the groups of Afrodescendants from Porto Alegre, Belém and Recife, 4 other previously described variants in the enhancer region were found: -13779G>C, -13937G>A, -14010G>C, -14011C>T. Twenty-six haplotypes previously described were found in the Brazilian population.

The association study of the -13910*T allele and of the presence of the metabolic syndrome in the Eurodescendants from Porto Alegre showed that the persistent individuals have lower risk than the non-persistent of developing metabolic syndrome (OR=0.47, p=0.023).

In an attempt to disclose LP causality, a functional study of the -13937G>A variant was performed. The results showed that the derived allele does not drive a higher expression of the reporter gene in cells in culture.

Considering the results of this study and the data available in the literature, we emphasize the importance of the studies that try to determine the LP looking for new alleles, phenotype-genotype studies, and functional studies to characterize the variants found related to the lactase phenotype.

Capítulo 1: INTRODUÇÃO

1.1 O leite

O leite é a primeira e principal fonte de nutrição de recém-nascidos. Os nutrientes básicos presentes no leite materno são: água, proteínas, hidratos de carbono, lipídeos, minerais e vitaminas. Os hidratos de carbono constituem a fonte de energia mais comum, de fácil digestão e absorção desde o início da vida. A lactose é um dos principais componentes do leite e seu carboidrato predominante (Battochio *et al*, 2003). De todos os mamíferos, o leite humano contém a maior concentração de lactose, 7% (Montgomery *et al*, 1991).

A lactose (Figura 1) é um dissacarídeo que ocorre naturalmente apenas no leite (Nelson e Cox, 2002). Este dissacarídeo é sintetizado pela enzima lactose sintetase exclusivamente nas glândulas mamárias dos mamíferos placentários (com exceção do leão marinho) no final da gestação e durante a lactação (Montgomery *et al*, 1991).

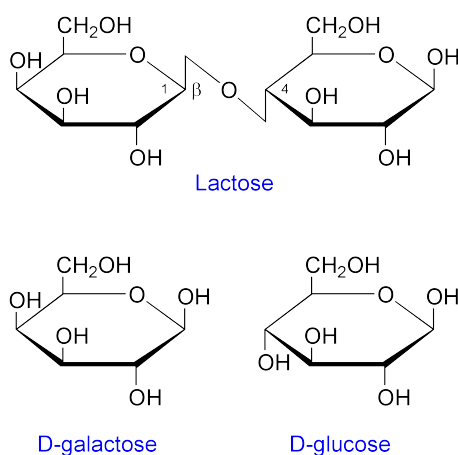


Figura 1: Estrutura do dissacarídeo lactose e dos monossacarídeos que o compõem.

Fonte: http://upload.wikimedia.org/wikipedia/commons/6/6b/Lactose_etc.png

A lactose é hidrolisada nos seus monômeros galactose e glicose (Figura 1) e a enzima responsável por esta quebra é a lactase que fica ancorada nos enterócitos absorptivos, as células epiteliais do intestino delgado. Os monossacarídeos liberados na hidrólise são absorvidos por estas células passando para a corrente sanguínea e sendo transportados até os diferentes tecidos onde serão metabolizados na via da glicólise. A função final destes monossacarídeos será a geração de energia para as células de diferentes partes do organismo (Nelson e Cox, 2002).

Como a lactose sozinha não pode ser transportada através da membrana celular, a lactase é essencial para a nutrição de mamíferos recém-nascidos, para os quais a única fonte de nutrição é o leite (Ingram e Swallow, 2009).

1.2 A hipolactasia primária adulta e a intolerância à lactose

A hipolactasia primária adulta (MIM 223100) ou deficiência da lactase é o declínio geneticamente programado da expressão da enzima lactase que é necessária para a hidrólise da lactose presente na dieta (Boll *et al*, 1991; Montgomery *et al*, 1991). A consequente má digestão da lactose pode, mas não necessariamente irá, causar sintomas (He *et al*, 2008).

O nome hipolactasia primária adulta foi dado para contrastar com a ausência de lactase ao nascimento causada por mutações na região codificadora do gene, que é a alactasia hereditária, e também da hipolactasia secundária causada por doenças gastrointestinais que danificam a parede intestinal.

A intolerância à lactose se refere aos sintomas gastrointestinais associados com a incompleta digestão da lactose (He *et al*, 2008). Aproximadamente 65% da população mundial apresenta essa característica, que não é considerada uma doença (Montgomery *et al*, 1991; Brüssow, 2013), apesar de gerar problemas para adultos quando da ingestão de leite e derivados.

Na falta da enzima lactase, a lactose não pode ser hidrolisada e absorvida no intestino delgado e chega ao cólon do intestino onde é fermentada, por bactérias, produzindo ácidos graxos e gases (incluindo o hidrogênio) (Nelson e Cox, 2002; Olds e Sibley, 2003; Krawczyk *et al*, 2008). A lactose não digerida tem dois efeitos fisiológicos. Primeiro, um gradiente osmótico é gerado na parede do intestino, o que resulta em um influxo de água, causando a diarreia. Segundo, os produtos da fermentação causam desconforto, dor, inchaço e flatulência (Ingram *et al*, 2009a).

Os sintomas da intolerância à lactose e a intensidade destes variam de pessoa para pessoa, podendo ser influenciado por outros fatores além da lactose mal digerida. Sendo assim, algumas pessoas podem tolerar pequenas quantidades de lactose e outras podem tolerar grandes quantidades. Essas diferenças individuais podem estar relacionadas a variação na composição da flora intestinal, na

percepção individual, presença de desordens gastrointestinais, no trânsito gastrointestinal, na quantidade de lactose ingerida e na microbiota intestinal (He *et al*, 2008; Ingram e Swallow, 2009).

O diagnóstico da hipolactasia primária adulta pode ser feito de forma direta e indireta. A quantificação direta da lactase via biópsia do intestino delgado é o método mais acurado, porém invasivo. Diversos métodos indiretos foram desenvolvidos com o propósito de diagnóstico, todos utilizando a digestão da lactose para informar sobre a expressão da enzima. Para estes métodos indiretos, a prática geral é administrar uma carga de lactose, após jejum de 12 horas. Após a carga de lactose, dois exames diferentes podem ser realizados, o teste de glicose no sangue ou o teste do hidrogênio expirado. Dentre estes dois métodos, o teste do hidrogênio expirado é mais acurado e o menos invasivo (Krawczyk *et al*, 2008; Ingram *et al*, 2009a; Ingram e Swallow, 2009).

Para o teste de glicose, uma medida basal da glicose sanguínea é realizada antes da ingestão da lactose e, após a ingestão, em intervalos de 30 minutos durante 2 horas. O aumento da glicose indica a digestão da lactose e o não aumento é indicativo da não digestão, ou seja, de deficiência da lactase (Ingram e Swallow, 2009).

O teste do hidrogênio expirado é um procedimento que mede a quantidade de hidrogênio exalado que é produzido em decorrência da fermentação da lactose, quando esta não foi degradada pela lactase (Krawczyk *et al*, 2008). O aumento na concentração de hidrogênio no ar expirado é indicativo de hipolactasia (Pretto *et al*, 2002).

Nos dois testes, pontos de corte arbitrários devem ser estabelecidos para distinguir os dois fenótipos e os dois métodos informam a habilidade da pessoa em digerir a lactose e não o nível de expressão da lactase. Dessa forma, o fenótipo da expressão é inferido a partir destes testes permitindo erros em ambas as direções, o indivíduo pode expressar a enzima e ser classificado como não digestor de lactose ou não expressar a enzima e ser classificado como digestor (Ingram e Swallow, 2009).

1.3 Padrões de expressão da enzima e a persistência da lactase

A lactase apresenta regulação posicional exibindo um padrão de expressão altamente controlado do eixo proximal ao distal do intestino em animais, como pode ser observado na Figura 2, e em humanos, com altos níveis de expressão na porção mediana e reduzidos níveis no duodeno e íleo (Montgomery *et al*, 2007).

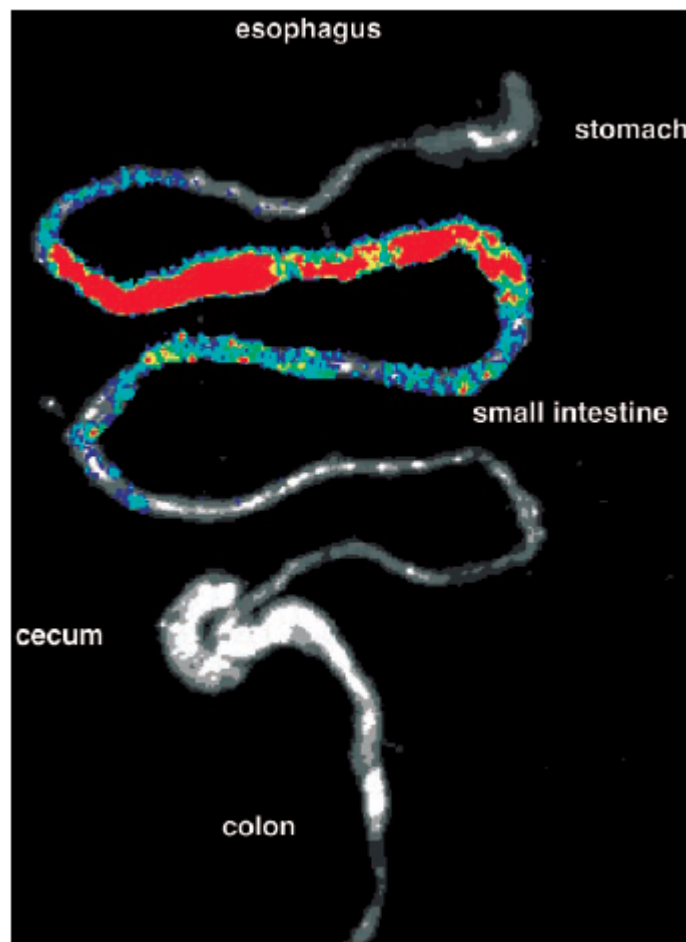


Figura 2: Imagem do trato gastrointestinal de camundongos transgênicos que expressam luciferase. A imagem mostra o trato gastrointestinal de um camundongo adulto transgênico que foi transfectado com um plasmídeo apresentando a região promotora do gene da lactase de ratos controlando o gene repórter da luciferase. A emissão da luz ocorre pela degradação do substrato luciferina quando da presença da enzima luciferase. A cor azul ilustra a menor intensidade e a cor vermelha os sinais mais intensos de emissão. Extraído de Lee *et al*, 2002.

A regulação mais marcante na expressão da lactase é a regulação temporal. Nos mamíferos não humanos, os níveis de RNA mensageiro e de proteína são baixos até pouco antes do nascimento, quando estes aumentam drasticamente. Os níveis se mantêm elevados até o desmame, quando declinam para menos de 10% dos níveis dos neonatos e a atividade reduzida é mantida ao longo da vida adulta. O declínio é regulado pela redução da taxa de transcrição, a consequente diminuição da síntese proteica e da atividade enzimática no jejuno e a virtual extinção da expressão no duodeno e no íleo distal (Montgomery *et al*, 2007).

O padrão temporal de expressão da lactase é diferente nos humanos. Antes das 24 semanas de gestação, a atividade da lactase é baixa. Após esse período, ela começa a aumentar e durante o terceiro trimestre, a atividade aumenta marcadamente até os níveis encontrados nos neonatos e que são similares às taxas de bebês de 2 a 11 meses. Assim como nos outros animais mamíferos, na maioria das populações humanas, a atividade da lactase diminui durante a infância (em média aos 5 anos), resultando em baixos níveis dessa idade em diante (Montgomery *et al*, 2007).

Como já mencionado anteriormente, o fenótipo mais frequente é a intolerância à lactose. Porém, em algumas populações especialmente no norte da Europa, os adultos mantêm a capacidade de digerir a lactose ao longo da vida e este fenótipo é chamado de tolerância à lactose. Biologicamente, essas pessoas mantêm a expressão da enzima lactase alta durante toda a vida, a chamada persistência da lactase (PL) (Montgomery *et al*, 2007; Ingram *et al*, 2009a). Como pode ser visto na Figura 3, a frequência da PL é maior do que 90% em suecos e dinamarqueses, diminuindo progressivamente para o sul da Europa chegando à frequência aproximada de 50% em espanhóis e franceses. Em algumas populações do norte da África, da Ásia e da Arábia também há o predomínio de PL (Boll *et al*, 1991; Harvey *et al*, 1998; Tishkoff *et al*, 2007; Ingram e Swallow, 2009).

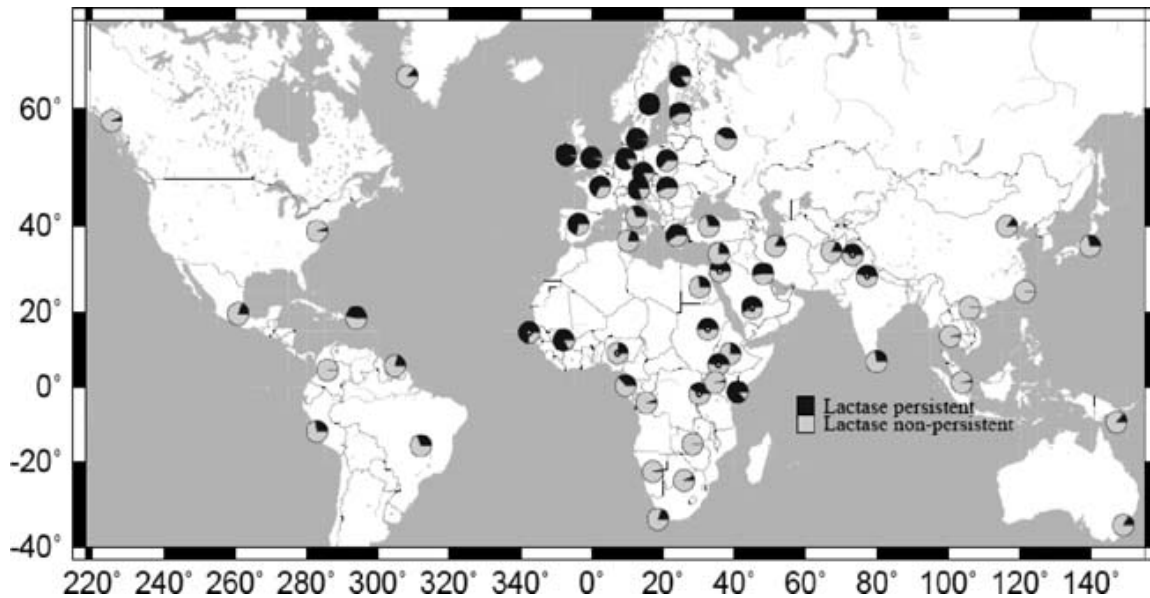


Figura 3: Distribuição mundial da persistência da lactase. Extraído de Ingram e Swallow, 2009.

A não persistência da lactase (NPL) é o declínio natural do nível de expressão da enzima lactase durante a infância. A NPL é predominante em populações nativas da Austrália e Oceania, leste e sudeste da Ásia e África tropical (aproximadamente 1% de indivíduos persistentes em chineses e 5% a 20% em populações do oeste da África, onde a subsistência é baseada na agricultura) (Boll *et al*, 1991; Harvey *et al*, 1998; Tishkoff *et al*, 2007).

Inicialmente, acreditava-se que a contínua ingestão de leite durante a vida adulta manteria a expressão da lactase alta. Esta hipótese foi aceita até os anos 70. Mais tarde, quando estudos mostraram que nem a ingestão prolongada de lactose e nem mesmo a sua exclusão da dieta influenciam na atividade da enzima, esta teoria foi rejeitada. Paralelamente, surgiu a hipótese de que a deficiência da lactase seria de origem genética e resultado da homozigose de um gene autossômico e a PL seria o fenótipo dominante (Boll *et al*, 1991; Montgomery *et al*, 1991). Os adultos heterozigotos persistentes apresentam níveis intermediários de atividade enzimática que são suficientes para hidrolisar a lactose da dieta e a carga de lactose do teste de tolerância (Wang *et al*, 1998).

1.4 A enzima lactase e o gene *LCT*

A lactase (*lactase-phlorizin hydrolase* – LPH, EC 3.2.1.23-62) é constituída por 1927 resíduos de aminoácidos (Mantei *et al*, 1988). Ela é uma glicoproteína transmembrânica de 160 KDa com a porção C-terminal intracelular e a N-terminal encontrada na superfície luminal da bicamada lipídica dos enterócitos absortivos. A lactase é uma enzima multifuncional com especificidade por diversos substratos e tem a capacidade de hidrolisar, além da lactose, a lactosilceramida, a celobiose, a celotriose e a florizina (Montgomery *et al*, 2007).

O gene que codifica a enzima lactase está localizado no braço longo do cromossomo 2 (2q21) (Kruse *et al*, 1988). O *LCT* (MIM 603202), possui aproximadamente 56 Kb de extensão e 17 éxons. Os estudos com este gene mostraram que ele não sofre *splicing* alternativo nem edição de RNA (Boll *et al*, 1991) e que a enzima presente no adulto é idêntica à da criança (Potter *et al*, 1985).

Com relação à região promotora do *LCT*, vários fatores de transcrição importantes para a expressão da lactase foram identificados. O fator Cdx-2 está implicado na regulação de muitos genes expressos no intestino e possui vários sítios de ligação *upstream* ao *LCT*. Este fator aumenta a expressão de *LCT in vitro*. O fator-1 α nuclear de hepatócito (HNF-1 α) também modula a expressão da lactase e atua sinergisticamente com o Cdx-2. Os fatores de transcrição GATA-4, -5 e -6 ligam-se ao sítio GATA e interagem com o Cdx-2 e HNF-1 α ativando a transcrição da lactase (Troelsen *et al*, 1992; Troelsen *et al*, 1997; Fitzgerald *et al*, 1998; Fang *et al*, 2000; Fang *et al*, 2001; Krasinski *et al*, 2001; van Wering *et al*, 2002). Os sítios de ligação destes fatores estão agrupados nos 100 pb anteriores ao TATA *box*. A região promotora e os respectivos sítios de ligação dos fatores de transcrição estão representados na Figura 4.

```

          GATA          HNF-1α          GATA
- 104 TGTGATCATA GATAA CCCAG TAAATATTA AGTCTTAATT ATCACTTAGT
          Cdx-2          TATA
- 54  ATTTTACAAC CTCAGTTGCA GTTATAAAGT AAGGGTTCCA CATACTCCT
- 4  AACAGTTCCT AGAAAATGGA

```

Figura 4: Região promotora do gene *LCT* humano. Os sítios de ligação e os respectivos fatores de transcrição estão destacados. O início do éxon 1 está em negrito.

Embora o HNF-1 α seja necessário para a expressão do *LCT*, o padrão espacial específico da expressão é definido pela correção de GATA-4 (Bosse *et al*, 2006). A abundância de RNA mensageiro de GATA-4 é maior nos segmentos intestinais proximais e medianos e reduzido nos segmentos distais, coincidente com a redução do RNA mensageiro do *LCT* nesta região (van Wering *et al*, 2004). O controle espacial também é gerado pelos fatores de transcrição PDX-1 e CDP, que reprimem a transcrição nas porções proximais e distais do intestino delgado (Wang *et al*, 2004; Fang *et al*, 2006).

Nos 1000 primeiros pares de base da região 5'-flanqueadora do *LCT* de humanos encontram-se duas sequências *Alu* de aproximadamente 300 pb cada. Porém não se sabe se esses segmentos repetitivos afetam a expressão da lactase (Grand *et al*, 2003; Montgomery *et al*, 2007).

Mesmo que os sítios e os correspondentes fatores de transcrição do promotor proximal sejam importantes para a regulação basal da expressão da lactase, eles não são suficientes para a correta expressão temporal do gene e, portanto, esta região não está envolvida na PL. De fato, o sequenciamento do *LCT* e da região promotora imediata em europeus não mostrou nenhuma mudança de nucleotídeo que fosse absolutamente associada com a PL (Ingram e Swallow, 2009).

Em 1995 foi sugerido que fatores em *cis* ao gene *LCT* determinam a alta ou a baixa expressão do RNA mensageiro do gene da lactase (Wang *et al*, 1995). Mais tarde, foi demonstrado que variantes na sequência de íntrons do gene *MCM6* (*Minichromosome Maintenance 6*, MIM 601806) apresentam alto desequilíbrio de ligação com o gene *LCT* (Hollox *et al*, 2001).

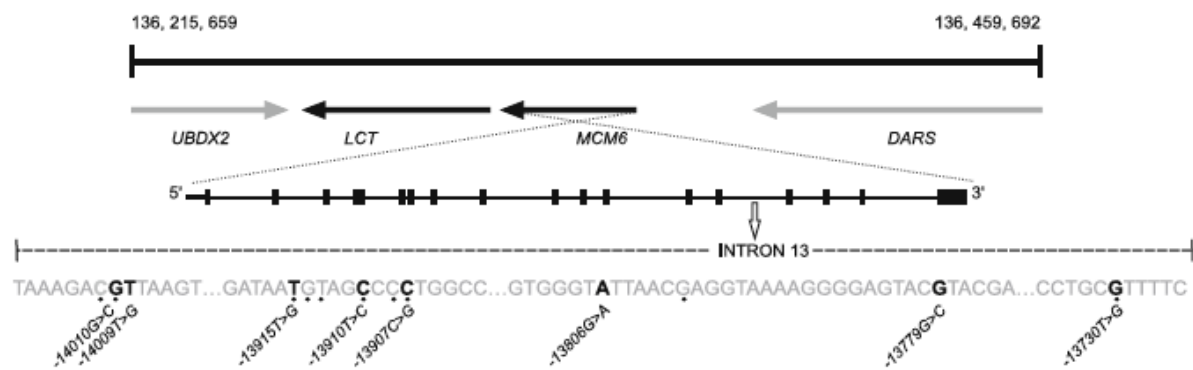


Figura 6: Diagrama esquemático do íntron 13 do gene *MCM6* onde os nucleotídeos em negrito são as posições dos SNPs encontrados no estudo citado e os pontos são as variantes já descritas anteriormente na literatura. Extraído de Ingram *et al*, 2009b.

Outros SNPs também foram descritos nas populações africanas e do Oriente Médio: -13779G>C, -13806A>G, -13906T>C, -13914G>A -14009T>G, -14091C>T, -14107G>A, -14176A>C (Tag *et al*, 2007; Ingram *et al*, 2009b; Torniainen *et al*, 2009). Na população tibetana, China, foram descritos -13838G>A, -13906T>A e -13908C>T (Peng *et al*, 2012).

1.5 Frequência dos polimorfismos presentes na região *enhancer*

A frequência do alelo -13910*T varia de 5 a 81,5% na Europa, como pode ser visto na Tabela 1. A frequência é maior no norte e oeste e decresce em direção ao sul e leste, chegando a números tão baixos quanto 5% na população do sul da Itália (Enattah *et al*, 2007a). Nenhum outro alelo relacionado a PL foi descrito na Europa.

Em algumas regiões africanas, a frequência da PL é bastante alta, porém o alelo -13910*T é muito raro ou ausente na maior parte do continente, com exceção da população Fulani, onde a frequência desse alelo varia de 37% a 48% (Tabela 1). A população Hausa de Camarões também é exceção pois a frequência desse alelo é compatível com a de PL. Por outro lado, algumas populações apresentam frequências bastante altas de outros SNPs, como por exemplo o alelo -14010*C que foi observado em 75% dos cromossomos no Quênia ou o alelo -13907*G em 25% no Sudão (Mulcare *et al*, 2004; Enattah *et al*, 2007a; Imtiaz *et al*, 2007; Ingram *et al*, 2007; Lokki *et al*, 2011).

No continente asiático, o alelo -13910*T é ausente em países como o Japão e a Coreia do Sul, porém pode atingir frequências elevadas na Rússia (Tabela 1) (Bersaglieri *et al*, 2004; Enattah *et al*, 2007a).

No Brasil, a frequência do alelo T na população do sudeste do país é de 24,7% em brancos, 23% em pardos e 12% em negros (Mattar *et al*, 2009).

Tabela 1: Frequência, em porcentagem, dos alelos relacionados a persistência da lactase em diferentes populações.

População	-13910T*	-22018A	-13915G*	-14010C*	-13779C	-13907G	-13913C	-13937A	-14011T	Referência
África										
África do Sul										
Miscigenados	21,8		0	8,1		0		0		Torniainen et al, 2009
Bantus	0	0								Bersaglieri et al, 2004
Xhosa	0		0	12,8		0		1,4		Torniainen et al, 2009
Argélia	21,7	21,7								Bersaglieri et al, 2004
Angola	0		0	1-6		0	0		0	Coelho et al, 2009
Camarões	38,9		0-6,3			0	0-2,4			Mulcare et al, 2004; Ingram et al, 2007
Etiópia	0		5,6-15			5,3-20	0-5,6			Mulcare et al, 2004; Ingram et al, 2007
Fulani	37-48	55	0,6							Enattah et al, 2007a; Ingram et al, 2007; Lokki et al, 2011
Marrocos	18-21	16	8,3			0	0			Enattah et al, 2007a; Enattah et al, 2008
Moçambique	0		0	0		0	0		0	Coelho et al, 2009
Nigéria	0	0								Mulcare et al, 2004; Bersaglieri et al, 2004
Quênia	0	0	0-50	0-75		0-12,5	0		0	Bersaglieri et al, 2004; Tishkoff et al, 2007
Senegal	0-2,6									Mulcare et al, 2004
Sudão	0-0,6	0	0-24,4	0		0-25	0-0,6		0	Mulcare et al, 2004; Ingram et al, 2007;

									Tishkoff et al, 2007
Tanzânia	0	0-1,1	0-3,8	0-62,5		0	0	0	Tishkoff et al, 2007
Uganda	0								Mulcare et al, 2004
Europa									
Espanha	38,6								Agueda et al, 2010
Finlândia	56-81,5		0			0			Bersaglieri et al, 2004; Enattah et al, 2007 b; Enattah et al, 2008
França	34-66,7	37-66,7							Bersaglieri et al, 2004; Enattah et al, 2007a
Grécia	9	12							Anagnostou et al, 2009
Inglaterra	74,7								Smith et al, 2009
Itália	5-35,7	6-35,7							Bersaglieri et al, 2004; Enattah et al, 2007a; Anagnostou et al, 2009
Portugal	34,9								Manco et al, 2013
Oriente Médio									
Arábia Saudita	0-13	10	10,5-76,9	0		0-0,8	0-1,1		Enattah et al, 2007a; Imtiaz et al, 2007; Ingram et al, 2007; Enattah et al, 2008
Iran	5-10	5-7	0			0	0		Enattah et al, 2007a; Enattah et al, 2008
Israel	2,6-3,9	2,1-4,1	0-13,2			0	0		Bersaglieri et al, 2004, Ingram et al, 2007
Jordânia	5,4-6,5		5,4-39,1			0	0-0,9		Enattah et al, 2008; Ingram et al, 2007

Paquistão	0-41	0-43							Bersaglieri et al, 2004; Enattah et al, 2007a
Américas									
Afrodescendentes	9-14	9-13,3							Bersaglieri et al, 2004; Enattah et al, 2007a
Brasil									
Caritiana	0	0							Bersaglieri et al, 2004
Suruí	0	0							Bersaglieri et al, 2004
Branços	24,7								Mattar et al, 2009
Pardos	23								Mattar et al, 2009
Negros	12								Mattar et al, 2009
Descendentes de Japoneses	0								Mattar et al, 2009
Colômbia	0-10	0							Bersaglieri et al, 2004; Torres et al, 2011
Chile									
Hispânicos	22,3								Morales et al, 2011
Ameríndios	58								Morales et al, 2011
Eurodescendentes	77,2	77,1							Bersaglieri et al, 2004
México	0-2	0-2							Bersaglieri et al, 2004
Ásia									
Camboja	0	0							Bersaglieri et al, 2004
China	0-10	0-10							Bersaglieri et al, 2004; Enattah et al, 2007a
Coréia do Sul	0	0							Enattah et al, 2007a
Índia	10,3-22,5	24,4	0	0	2,4	0		0,4	Romero et al, 2012,

Japão	0	0	Kuchay et al, 2013 Bersaglieri et al, 2004
Rússia	7-62	7-62	Bersaglieri et al, 2004; Enattah et al, 2007a; Khabarova et al, 2009
Sibéria	6	6	Bersaglieri et al, 2004

*Comprovada relação a PL por estudos funcionais

1.6 Estudos funcionais

Para testar a relação dos SNPs descritos com a persistência da lactase diversos grupos realizaram estudos funcionais utilizando a região de interesse em conjunto com a região promotora da lactase de camundongos ou humanos no controle da expressão do gene repórter da luciferase. Na maioria dos estudos, as construções foram transfectadas em células Caco-2. Caco-2 é uma linhagem celular humana derivada de adenocarcinoma que imita o fenótipo de enterócitos com relação a expressão de diversas hidrolases digestivas, incluindo a lactase (Hauri *et al*, 1994).

Dois estudos funcionais com o polimorfismo -13910C>T demonstraram que o alelo T direciona uma expressão maior do gene repórter do que o alelo C. O alelo A do polimorfismo -22018G>A gerou um incremento mínimo na expressão ou atuou como repressor (Olds e Sibley, 2003; Troelsen *et al*, 2003). A associação dos dois polimorfismos no mesmo plasmídeo resultou em um incremento um pouco maior que o proporcionado somente pelo alelo T, aproximadamente 1,2 vezes (Olds e Sibley, 2003). Porém, como -13910*T e -22018*A apresentam alto desequilíbrio de ligação, não se pode descartar a hipótese de que ambos sejam necessários *in vivo* para a PL ocorrer (Kuokkanen *et al*, 2003; Enattah *et al*, 2008).

A partir de biópsias, foi demonstrado que a quantidade de enzima produzida é maior nos indivíduos homozigotos -13910TT, intermediária nos heterozigotos e menor nos homozigotos CC (Kuokkanen *et al*, 2003; Enattah *et al*, 2007b).

Os resultados mostram que a região -13910 atua como um *enhancer* da transcrição do gene da lactase. A explicação para a persistência da lactase seria que o T tem afinidade maior pelos fatores de transcrição que o alelo C. A disponibilidade de alguns dos fatores de transcrição necessários pode ser diminuída durante o desenvolvimento porque os genes envolvidos na digestão de uma dieta rica em amido são regulados positivamente após o desmame (Troelsen *et al*, 2003).

A posição -13910 faz parte do sitio de ligação do fator de transcrição Oct-1 (*Octamer-binding Transcription Factor 1*) (Figura 7). A proteína Oct-1 tem efeito sinérgico com a proteína HNF-1 α (*Hepatocyte Nuclear Factor 1 alfa*) que se liga

ao sítio HNF-1 no promotor de *LCT* (Lewinsky *et al*, 2005; Enattah *et al*, 2008). A proteína Oct-1 recruta fatores modificadores da cromatina que são capazes tanto de aumentar quanto silenciar o gene, dependendo do tipo celular e da arquitetura do promotor. É possível que Oct-1 ligando-se ao alelo -13910*T, *in vivo*, induza mudanças na cromatina próxima ao gene *LCT* que estão envolvidas no fenótipo da PL (Lewinsky *et al*, 2005).

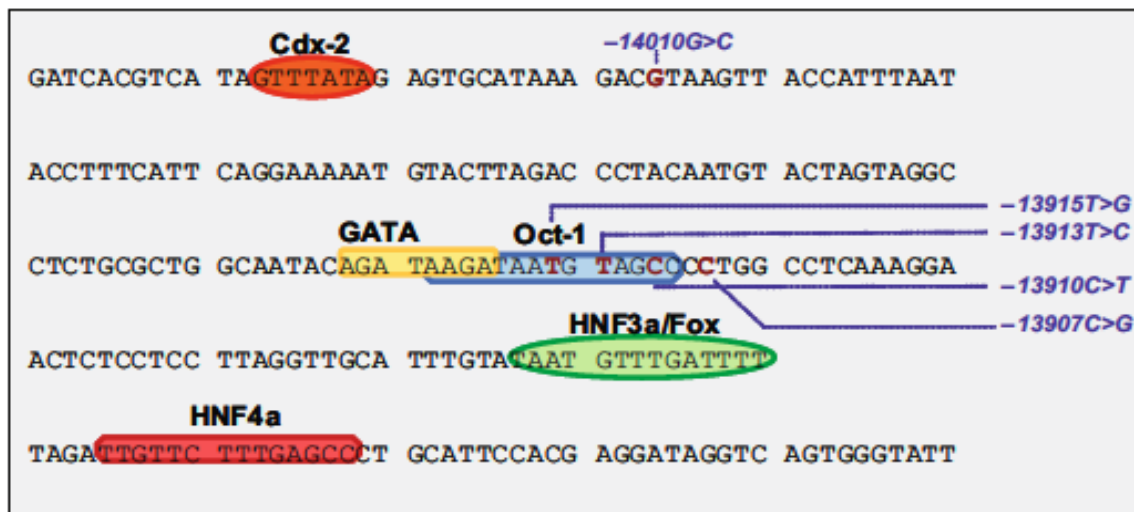


Figura 7: Sítios de ligação e os respectivos fatores de transcrição presentes na região a - 14 Kb do gene *LCT*. Os polimorfismos relacionados à persistência a lactase estão destacados. Extraído de Ingram e Swallow, 2009.

A região -14 Kb possui ainda os sítios de ligação para outros fatores de transcrição que são expressos no intestino (GATA-6, Fox, Cdx-2 e HNF-4 α), como pode ser visto na Figura 7. Esses 4 sítios seriam necessários, conjuntamente, para a ativação do *LCT* (Lewinsky *et al*, 2005).

Os alelos -14010*C, -13915*G e -13907*G também funcionaram como *enhancers* mais potentes do que os alelos ancestrais em cultura de células (Tishkoff *et al*, 2007). Foram identificados, ao redor da posição -14010, sítios de ligação para os fatores de transcrição Oct-1 e HNF-1 α , sendo que a proteína Oct-1 tem maior afinidade pelo alelo C que pelo G (Jensen *et al*, 2011). A proteína Oct-1 também interage com o DNA na posição -13915 e é capaz de aumentar a atividade promotora da construção promotor da lactase-gene repórter que possui o alelo -13915*G em cultura de células (Olds *et al*, 2011).

1.7 A diversidade do gene *LCT* e evolução da persistência da lactase

O gene *LCT* é altamente variável apresentando diversos haplótipos pela combinação de sítios de polimorfismos de base única (SNPs) na região promotora proximal e intragênicos (Figura 8). A nomenclatura do haplótipo é baseada na composição alélica seguindo o esquema de presença e ausência mostrado na Figura 9.

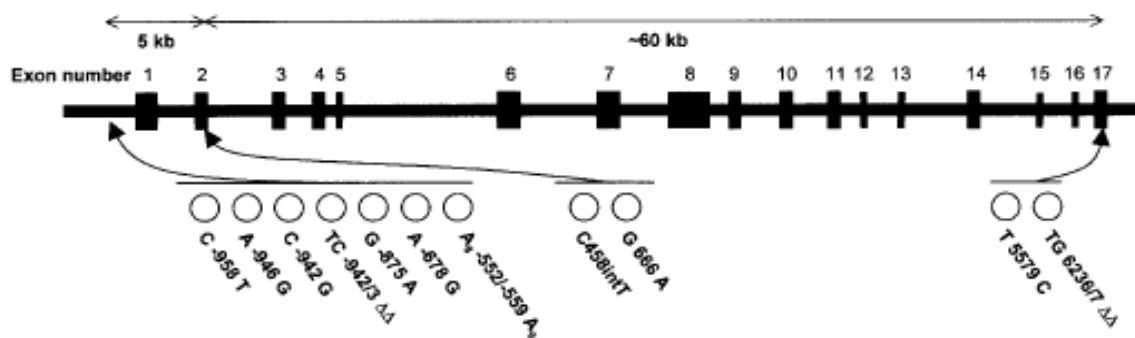


Figura 8: Diagrama representando o gene da lactase e as posições dos polimorfismos que formam os haplótipos. Extraído de Hollox *et al*, 2001.

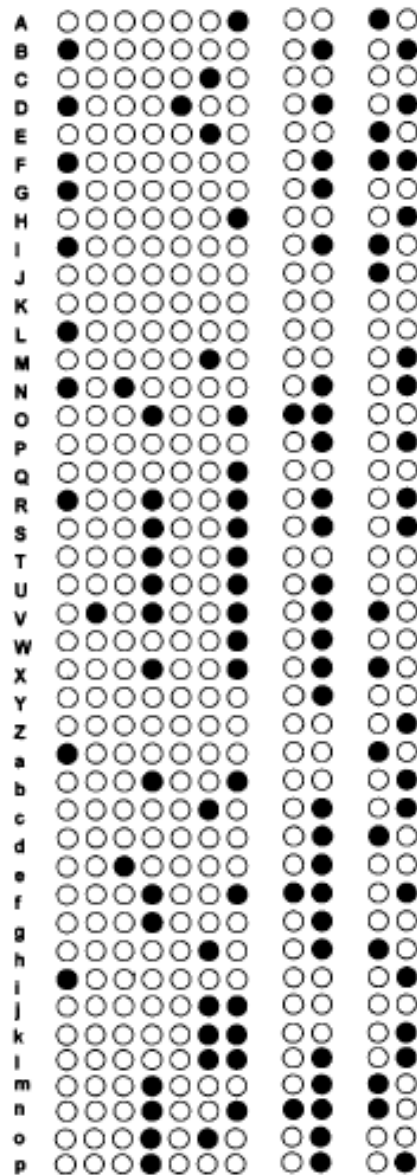


Figura 9. Lista dos haplótipos deduzidos do *LCT* baseados na presença do alelo ancestral ou derivado. O círculo vazio indica a presença do alelo ancestral para o SNP naquela posição, o círculo preto indica o alelo derivado. A sequência dos SNPs da esquerda para a direita é: -958C>T, -946A<G, -942C>G, -942/3TC<ΔΔ, -875G<A, -678A<G, -552/-559A₈<A₉, 458C<intT, 666G<A, 5579T<C e 6236/7TG<ΔΔ. Modificado de Hollox *et al*, 2001.

Os haplótipos mais frequentes nas populações europeias, indianas, asiáticas e africanas são o A (chegando a 86% no norte da Europa e apresentando gradiente da frequência em direção ao sul e nos africanos varia de 10 a 30%); o B (presente em todas as populações com exceção dos Bantus da África do Sul); o C (presente

em todas as populações, porém raro no norte da Europa) e o U (com frequência variada). Todos os outros haplótipos descritos na literatura possuem frequências menores que 5% nas populações estudadas. Entre os haplótipos citados, na população europeia o A está fortemente associado a PL (Harvey *et al*, 1998; Hollox *et al*, 1999).

Nas populações do norte da Europa, a heterozigose de haplótipos do *LCT* é bastante baixa (0,26), refletindo a alta prevalência do haplótipo A. Já nas populações africanas estudadas, a taxa de heterozigose varia de 0,87 a 0,91, com muitos haplótipos presentes em baixa frequência. Nas populações do sul da Europa, nos indianos e nos asiáticos a heterozigosidade varia de 0,65 a 0,78 (Hollox *et al*, 2001).

Todos os SNPs apresentam desequilíbrio de ligação, aos pares, superior a 0,8 (Hollox *et al*, 2001). O forte desequilíbrio de ligação que se estende por mais de 1 Mb evidencia a ausência de recombinação, o que gera indício de seleção positiva recente em favor do haplótipo ligado a PL (Poulter *et al*, 2003).

Em todas as populações nas quais a PL é frequente, existe o histórico de domesticação de animais, com início há aproximadamente 9.000 anos, período no qual o leite tornou-se um alimento disponível em grande quantidade (Dudd e Evershed, 1998).

A idade estimada do alelo -13910*T varia entre 5.000 e 12.300 anos (Coelho *et al*, 2005; Enattah *et al*, 2007a; Ingram *et al*, 2009a), sendo que este alelo estava ausente ou, então, presente em baixa frequência nos europeus do neolítico (a aproximadamente 12.000 a 10.000 anos atrás) (Burger *et al*, 2007; Malmström *et al*, 2010). Porém, haveria também um alelo -13910*T presente em um haplótipo menos frequente e com surgimento mais recente, há aproximadamente 1.400 a 3.000 anos, demonstrando que este alelo foi independentemente introduzido mais de uma vez na população mundial (Enattah *et al*, 2007a).

A idade estimada para o alelo -13915*G, presente na população árabe, é de 4.000 anos (Enattah *et al*, 2008) e para o alelo -14010*C esta situa-se entre 3.000 e 7.000 anos (Tishkoff *et al*, 2007).

Análises de coalescência determinaram que os alelos -13910*T e -13907*G originaram-se do mesmo haplótipo ancestral e tornaram-se frequentes juntamente com o evento de domesticação do gado que inseriu o leite de vaca na dieta dessas populações e que -13915*G originou-se independentemente coevoluindo com a domesticação do camelo (Enattah *et al*, 2008).

A identificação de vários alelos associados a PL sugere que esta característica aumentou em frequência e foi selecionada positivamente de forma independente em diferentes populações humanas. É provável que a capacidade de digerir leite tenha sido extremamente vantajosa, ao menos para algumas populações, há poucos mil anos (Ingram *et al*, 2009a).

As assinaturas de seleção positiva existentes no loco da PL são: um excesso de variantes raras, uma grande diferença na frequência alélica entre as populações (indicando efeito diferencial da seleção que aumentou dramaticamente a frequência dos alelos em algumas, mas não em todas as populações) e a existência de um haplótipo comum que se mantém intacto em longas distâncias (indicando que um certo alelo aumentou rapidamente em frequência antes que a recombinação pudesse desmembrar o haplótipo onde se encontra esse alelo) (Bersaglieri *et al*, 2004).

A redução na variabilidade de microssatélites nos haplótipos que contêm os alelos -13910*T e -22018*A indica uma aparente ausência de recombinação dentro de 61,4 Kb (Coelho *et al*, 2005). Na verdade, os dados com SNPs mostram um longo haplótipo homozigoto, com mais de 1 Mb, onde reside o alelo -13910*T, diferentemente dos outros haplótipos que possuem alta taxa de recombinação (Poulter *et al*, 2003; Bersaglieri *et al*, 2004). O coeficiente de seleção em favor de -13910*T calculado para a população europeia varia de 0,014 a 0,15 (Bersaglieri *et al*, 2004).

Para o alelo -14010*C também houve intensa seleção positiva. Este alelo está inserido num haplótipo de mais de 2 Mb, apresentando coeficiente de seleção de 0,04 a 0,097 (Tishkoff *et al*, 2007).

A forte pressão seletiva fez com que os alelos selecionados aumentassem rapidamente em frequência. A frequência do haplótipo no qual o alelo ocorre aumentou de forma correspondente e, sendo recente, não houve tempo suficiente

para a recombinação quebrar o haplótipo e, dessa forma, todos os SNPs que o formam aumentaram em frequência, em um evento chamado de carona genética, evidenciado pelos altos valores de desequilíbrio de ligação (Bersaglieri *et al*, 2004). Recentemente, as pesquisas focando na diversidade de haplótipos observada no *LCT* em várias populações sugerem que a frequência altamente divergente observada no norte da Europa comparada a outras populações seria melhor explicada pela combinação da deriva genética com a seleção positiva para a PL (Ingram e Swallow, 2009).

Mas qual a vantagem evolutiva que conferiu a persistência da lactase e que poderia então explicar essa seleção positiva? Diferentes autores sugeriram diferentes hipóteses, como as que seguem.

Uma hipótese diz que o leite era uma fonte de reserva alimentar em períodos de fome ou anterior à próxima colheita, período no qual as reservas de comida estavam baixas (Brüssow, 2013). De fato, as populações com alta frequência de indivíduos intolerantes desenvolveram adaptações para a utilização do leite, que envolvem a fermentação para produção de queijo e iogurte. Estes produtos apresentam menor quantidade de lactose e, portanto, os sintomas são menores ou ausentes (Kolars *et al*, 1984). Desse modo, teria havido uma coevolução entre este gene e hábitos culturais, com favorecimento dos indivíduos que podiam tomar leite (Feldman e Cavalli-Sforza, 1989).

A hipótese de coevolução gene-cultura é a teoria que mais se adapta a provável história da PL na África onde as altas frequências desse fenótipo são preferencialmente encontradas em populações pastoralistas (Gerbault *et al*, 2009). Mas, essa hipótese também pode ser aplicada na Europa onde ela é apoiada pela coincidência geográfica entre a alta diversidade dos genes de proteínas do leite bovino (essa alta diversidade implica em maior tamanho populacional no passado), locais de pecuária no período neolítico e a presente tolerância a lactose (Beja-Pereira *et al*, 2005).

Outra teoria seria a hipótese do clima árido. Nesta teoria especula-se que em climas desérticos onde a água e a comida são escassos, grupos nômades poderiam sobreviver utilizando o leite como comida e fonte de água limpa e descontaminada. Os benefícios aos persistentes se tornavam ainda mais

pronunciados durante eventos de doenças diarreicas, quando os intolerantes não seriam capazes de utilizar o leite como fonte de água sem exacerbar sua condição. Obviamente essa hipótese não pode ser aplicada a população do norte da Europa (Ingram e Swallow, 2009).

A explicação mais adequada à população europeia seria a hipótese da absorção de cálcio porque os baixos níveis de luminosidade que ocorrem no hemisfério norte estão associados com o risco elevado de desenvolver raquitismo e osteomalacia pela falta de vitamina D. O cálcio ajuda a prevenir o raquitismo impedindo a quebra da vitamina D pelo fígado e é um mineral essencial a saúde do osso. Por outro lado, os intolerantes poderiam obter cálcio do iogurte e do queijo, laticínios estes que contém quantidade reduzida de lactose. Entretanto, as proteínas do leite juntamente com a lactose facilitam a absorção do cálcio pelo organismo e, assim, eles não obteriam a mesma vantagem ingerindo os derivados do leite (Ingram e Swallow, 2009).

Na Europa, a história da população teve um papel importante na difusão da PL pelo continente. Além disso, a pressão de seleção sobre a PL foi muito alta no noroeste do continente, ao contrario da parte sudeste onde a deriva genética sozinha pode explicar as frequências observadas. Este aumento da pressão da seleção com a latitude é compatível com a hipótese de assimilação do cálcio enquanto que a hipótese de coevolução gene-cultura não pode ser descartada (Gerbault *et al*, 2009).

Um trabalho com simulações de modelos de seleção levando em conta processos demográficos concluiu que na Europa, o alelo -13910*T primeiro sofreu seleção entre os fazendeiros que produziam laticínios há 7.500 anos. Nas simulações, a seleção natural favorecendo a PL, pela necessidade de obtenção de vitamina D da dieta, não foi maior no norte do continente. Essas simulações sugerem que uma seleção homogênea no tempo e espaço em combinação com processos demográficos são suficientes para explicar a correlação PL/latitude na Europa (Itan *et al*, 2009).

1.8 Saúde e considerações médicas

A intolerância a lactose não causa apenas os sintomas clássicos citados anteriormente, mas também náusea e vômito e muitos pacientes apresentam constipação pela reduzida motilidade intestinal ao invés de diarreia (Tabela 2). Além disso, alguns autores relatam sintomas sistêmicos causados pela intolerância a lactose, como os descritos na Tabela 2 (Matthews e Campbell, 2000a; Matthews e Campbell, 2000b; Treudler *et al*, 2002; Matthews *et al*, 2005). Porém, temos que levar em consideração que esses sintomas sistêmicos possam estar ocorrendo de forma coincidente com a intolerância a lactose.

Tabela 2: Sintomas apresentados por pacientes intolerantes à lactose*.

	Frequência em %
<u>Gastrointestinais</u>	
Dor abdominal	100
Distensão abdominal	100
Borborigmo	100
Flatulência	100
Diarreia	70
Constipação	30
Náusea	78
Vômito	78
<u>Sistêmicos</u>	
Dor de cabeça e tontura	86
Perda de concentração e pobre memória de curto prazo	82
Cansaço severo prolongado	63
Dor muscular	71
Dor nas articulações	71
Alergia (eczema, prurido, rinite, sinusite, asma)	40
Arritmia cardíaca	24
Úlcera bucal	30
Frequência aumentada de micção	< 20

Uma das confirmações do diagnóstico de intolerância a lactose é a retirada da lactose da dieta, porém nem sempre ocorre a melhora nos sintomas, muitas vezes levando a descartar esse diagnóstico. Nos EUA, esse problema tem sido chamado de lactose “escondida” na comida. A lactose pode ser usada como agente escurecedor de pães, em misturas de pães e bolos e adicionado a carnes processadas como salsichas e bife de hambúrguer. Além disso, pode ser injetada em carnes de frango e adicionada em refrigerantes e cerveja *lager*. Os produtos para emagrecimento podem conter tanta lactose quanto o leite (Matthews *et al*, 2005).

A intolerância à lactose está sendo utilizada para explicar até mesmo problemas dermatológicos. Um estudo mostrou que 64% de pacientes intolerantes que apresentavam prurido crônico de origem desconhecida tiveram melhora nos sintomas após uma dieta livre de lactose. Dessa forma, a intolerância à lactose pode ser um fator causal independente na elucidação do prurido crônico (Grundmann *et al*, 2011).

Diversos estudos exploraram o possível benefício da ingestão do leite para a saúde. Por exemplo, o leite possui peptídeos que promovem um efeito hipotensor que auxilia pessoas com hipertensão. Esse efeito hipotensor ocorre porque as proteínas do leite são fontes de peptídeos inibidores da enzima conversora de angiotensina I. Essa enzima hidrolisa angiotensina I em angiotensina II, sendo que a angiotensina II é um potente vasoconstritor. A inibição desse sistema leva a diminuição da pressão arterial (FitzGerald *et al*, 2004).

O consumo de leite também teria efeito sobre o coração, observa-se uma tendência em direção a diminuição do risco de infarto e doença cardíaca isquêmica quanto maior a quantidade consumida. Um estudo caso-controle sueco encontrou uma associação inversa significativa entre o consumo de leite e risco de primeiro infarto de miocárdio, um outro estudo norueguês observou uma tendência não significativa (Pfeuffer e Schrezenmeir, 2006). Porém, a ausência de evidência de associação do consumo de leite com risco de doença coronariana e ataque cardíaco também já foi relatada (Dalmeijer *et al*, 2012).

O cálcio presente no leite também confere vários benefícios, ele inibe a absorção de gorduras, a lipogênese, acelera a lipólise, aumenta a termogênese e suprime o acúmulo de gordura e ganho de peso. Estudos mostram que dietas com baixa quantidade de cálcio impedem a perda de gordura. Além disso, o cálcio do leite e derivados exibe um efeito maior na atenuação do ganho de gordura e ainda acelera a perda. Esse efeito aumentado provavelmente se deve a compostos bioativos adicionais do leite que atuam sinergisticamente com o cálcio na atenuação da adiposidade. Assim, leite e derivados são importantes produtos na prevenção e tratamento da obesidade (Zemel, 2003; Mirmiran *et al*, 2005).

Ainda em relação aos benefícios que o consumo do leite pode trazer à saúde, foi demonstrada a associação inversa com a Síndrome Metabólica (SM) e seus componentes individuais, com marcada redução na prevalência destes (Pereira *et al*, 2002; Azadbakht *et al*, 2005; Elwood *et al*, 2007). A SM é um agrupamento de fatores metabólicos de risco, incluindo obesidade central, hipertensão, dislipidemia, resistência à insulina e hiperglicemia, sendo preditiva de risco aumentado de desenvolver diabetes tipo 2 e doença cardiovascular (Després e Pérusse, 2008). A associação inversa com consumo de leite também foi demonstrada com relação ao stress oxidativo e inflamatório da SM (Stancliffe *et al*, 2011), com o IMC e porcentagem de gordura corporal em meninas (Mirmiran *et al*, 2005; Abreu *et al*, 2012).

Por outro lado, alguns autores relatam a ausência de associação do consumo de laticínios com mudanças no peso, distribuição de gordura, variáveis metabólicas e com a incidência da SM, excluindo a hipótese do fator protetivo dos laticínios (Rajpathak *et al*, 2006; Snijder *et al*, 2008).

Contudo, associações positivas do consumo de leite também já foram relatadas com relação à concentração de triglicérides e IMC (Lawlor *et al*, 2005). Além disso, o consumo maior de leite foi significativamente associado com maior altura em meninas (Berkey *et al*, 2009), dado este que poderia ser explicado pela grande quantidade de hormônios presente no leite bovino e que pode influenciar a produção de hormônios endógenos (Hartmann *et al*, 1998). Foi também demonstrado que o leite apresenta a capacidade de aumentar o nível plasmático de IGF-I (*Insulin-like growth factor-I*) (Giovannucci *et al*, 2003; Hoppe *et al*, 2004).

A PL está ligada ao maior consumo de leite e derivados (Almon *et al*, 2010; Corella *et al*, 2011), tendo relação direta com o genótipo, ou seja, indivíduos -13910TT consomem maior quantidade de leite, os heterozigotos quantidade intermediária e os -13910CC consomem menos (Travis *et al*, 2013). Por outro lado, a NPL tende a influenciar a quantidade de leite que pode ser tolerada e não a incapacidade de tomar leite (Smith *et al*, 2009).

Observando a relação do consumo do leite com o genótipo de PL (presença do alelo -13910*T) e também a influência do consumo de leite para a saúde, alguns estudos tentam relacionar a presença desse alelo com algumas condições ou características. De fato, um estudo associou a atividade aumentada da lactase com a diabetes, em pacientes que consumiam leite. Isso se deve ao fato de esses pacientes estarem mais expostos a glicose e a galactose do que se a lactose não fosse metabolizada, o que acontece nos intolerantes (Meloni *et al*, 2001).

Associações positivas da PL, presença do alelo -13910*T, foram feitas com a prevalência de SM (OR maior para os persistentes), maior quantidade de gordura corporal em crianças e adolescentes (Almon *et al*, 2010), maior estatura (Campbel *et al*, 2005; Agueda *et al*, 2010), maior IMC, maior peso e circunferência da cintura (Agueda *et al*, 2010; Kettunen *et al*, 2010; Corella *et al*, 2011; Malek *et al*, 2013). Porém, os resultados são contraditórios porque alguns autores não encontraram as mesmas associações, como nos exemplos a seguir. Smith e colaboradores, 2009, relatam que não há evidências de que o SNP -13910C>T esteja associado a Síndrome Metabólica (SM) e nem aos componentes metabólicos individuais dessa síndrome. Também não está associado a velocidade de crescimento e a altura final, tanto em meninos quanto em meninas (Lehtimaki *et al*, 2006), nem ao peso e ao IMC (Gugatschka *et al*, 2005).

O declínio na atividade da lactase parece ser um fator de risco para o desenvolvimento de osteoporose porque as pessoas intolerantes evitam o consumo de produtos a base de leite ou, talvez, pela interferência da lactose não digerida com a absorção de cálcio (Lee e Krasinski, 1998). Essa associação está evidenciada em um estudo onde homens jovens com o genótipo -13910CC são mais suscetíveis a perda óssea. Entretanto, os próprios autores discutem que o consumo de cálcio prediz mudanças maiores na massa óssea que o genótipo de

lactase (Laaksonen *et al*, 2009). Em outro estudo esse genótipo foi considerado fator de risco para fraturas em idosos (Enattah *et al*, 2005). Por outro lado, a ausência de associação foi demonstrada na população masculina finlandesa e espanhola, onde o genótipo CC não afeta a aquisição de massa óssea e não representa fator de risco a fraturas (Enattah *et al*, 2004; Agueda *et al*, 2010).

Outra relação que tem sido estudada é com a doença inflamatória do intestino. 70% dos pacientes com doença inflamatória são sensíveis a lactose (teste do hidrogênio expirado positivo e presença de sintomas), porém somente 7% dessa população branca do norte da Europa possui o genótipo -13910CC. Neste caso, a genotipagem é recomendada como triagem para excluir os indivíduos geneticamente NPL do teste de tolerância a lactose, visto que os sintomas apresentados por estes pacientes perduram por dias (Eadala *et al*, 2011).

Também foi demonstrada uma tendência a menor taxa de infecção por malária de persistentes, porém os dados não alcançaram significância estatística. A baixa suscetibilidade a malária causada pelo *Plasmodium falciparum* e a alta taxa de persistência da lactase na população africana Fulani levou a essa teoria de que o leite confere proteção contra o desenvolvimento da malária. A primeira hipótese seria de o leite promover melhora do estado nutricional de forma geral sendo rico em energia, proteínas e ácidos graxos. A segunda, seria de que diversos componentes imunomoduladores presentes no leite auxiliariam na proteção. A terceira, seria de que uma dieta dominada por laticínios leva a uma deficiência relativa de ácido para-aminobenzóico (PABA), protegendo contra a malária porque o PABA é necessário ao parasito. Adicionalmente, é possível que outros fatores desconhecidos no leite promovam a proteção contra a malária e outras infecções (Lokki *et al*, 2011).

Ainda, há vários estudos relacionando a PL com hábitos alimentares e risco aumentado de tipos de cânceres. Por exemplo, foram observadas menores taxas de câncer de ovário em populações com maiores índices de hipolactasia. O *odds ratio* do desenvolvimento de câncer de ovário em tolerantes foi de 2,51. A explicação seria a de que a galactose gerada pela hidrólise da lactose é citotóxica aos oócitos (Meloni *et al*, 1999; Kuokkanen *et al*, 2005). Para os homens, o leite forneceria proteção ao câncer de próstata. A explicação está no fato do leite

possuir diversos ácidos graxos poli-insaturados que exibem efeito inibitório de crescimento celular. A esfingomielina, um desses componentes, pode diminuir o crescimento de células tumorais através de seus metabolitos ativos, a esfingosina e a ceramida. Dessa forma, a gordura do leite poderia proteger contra o câncer. De fato, a presença de câncer de próstata foi associada ao consumo de leite com baixo teor de gordura. Já com relação ao polimorfismo de PL, não houve associação deste com risco de câncer de próstata em finlandeses e suecos (Torniainen *et al*, 2007).

Com relação ao câncer colorretal, o genótipo -13910CC está significativamente associado com risco aumentado em finlandeses, mas não em britânicos e espanhóis. O resultado encontrado na população finlandesa provavelmente reflete o efeito protetor do leite em indivíduos tolerantes. Porém a associação pode ser indireta, pois já foi demonstrado que a microflora intestinal de intolerantes é diferente da de tolerantes e que o câncer colorretal é influenciado pela microflora do paciente (Rasinperä *et al*, 2005).

1.9 Justificativa e Objetivo

No Brasil, dados sobre os alelos relacionados com a persistência da lactase são escassos conforme descrito na Introdução desse trabalho. Além disso, os estudos existentes restringiram-se ao polimorfismo -13910. Adicionalmente, dados populacionais relacionados à diversidade do gene *LCT* ainda não foram descritos para nossas populações.

O conhecimento do espectro de mutações que determinam a PL, é fundamental para o diagnóstico molecular da intolerância à lactose. O diagnóstico genético, quando disponível, acaba sendo preferido pelos pacientes intolerantes por ser pouco invasivo e não gerar o desconforto e sintomas associados ao teste que se baseia no consumo de lactose e acompanhamento por dosagem da glicose no sangue ou do hidrogênio expirado.

Portanto, o objetivo deste trabalho foi estudar a diversidade do gene *LCT*, tanto de sua região codificadora, quanto da região promotora proximal e da região

enhancer, assim como identificar os polimorfismos que pudessem estar relacionados com a persistência da lactase na população brasileira.

Capítulo 2: Artigo intitulado “Stability or variation? Patterns of lactase gene and its enhancer region distributions in Brazilian Amerindians” publicado na American Journal of Physical Anthropology, 2012.

Stability or Variation? Patterns of Lactase Gene and Its Enhancer Region Distributions in Brazilian Amerindians

Deise C. Friedrich,¹ Sidia M. Callegari-Jacques,^{1,2} M. Luiza Petzl-Erler,³ Luiza Tsuneto,³ Francisco M. Salzano,¹ and Mara H. Hutz^{1*}

¹Departamento de Genética, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970, Porto Alegre, RS, Brazil

²Departamento de Estatística, Universidade Federal do Rio Grande do Sul, Campus do Vale, 91509-900 Porto Alegre, RS, Brazil

³Departamento de Genética, Universidade Federal do Paraná, Caixa Postal 19071, 81531-990 Curitiba, PR, Brazil

KEY WORDS Amerindian populations; selection; genetic structure; *LCT* gene

ABSTRACT Lactase persistence (LP) is the phenotypic trait in which lactase secretion is maintained during adulthood. LP is due to mutations in the *LCT* enhancer region, located 14-kb upstream of the gene. In Europeans, the -13910^*T allele is associated with LP. In Africans this allele is rare while other mutations in this same region were related to LP. The *LCT* is highly polymorphic in human populations, but so far Brazilian Amerindians had not been investigated for these polymorphisms or for the presence of LP mutations. We describe the genetic diversity of the *LCT* region and the presence of LP enhancer mutations in four native Brazilian populations (Guarani-Kaiowá, Guarani-Nandeva, Kaingang, and Xavante). Twelve polymorphisms were genotyped by PCR-

based methods. The -13910^*T allele varied from 0.5% in the Xavante to 7.6% in the Guarani-Nandeva. These frequencies probably derive from European sources and they correlate with non-native admixture proportions previously estimated for these groups. But since admixture is virtually absent in the Xavante, we suggest that the presence of the LP allele could have been determined by a de novo mutation. No other mutations in the -14 kb enhancer region were found. The *LCT* was highly polymorphic in the present sample showing 15 haplotypes with a heterogeneous distribution among the four Amerindian populations. This diversity could be due to drift, as indicated by the neutrality test performed. *Am J Phys Anthropol* 147:427–432, 2012. © 2012 Wiley Periodicals, Inc.

Brazil was colonized by people of Asian origin, who arrived in the American continent through the Bering Strait about 15–18 thousand years ago (Salzano, 2011). These peoples spread all over North, Central, and South America and they are now known as Amerindians. At the time of the Conquest, the Spanish and Portuguese explorers found an aboriginal population estimated between 7 and 39 million individuals. Some societies depended heavily on agriculture, others practiced a mix of farming, fishing, hunting, and gathering whereas others were hunters and gatherers only (Salzano and Callegari-Jacques, 1988). Cattle were not present at that time; however it was introduced in the Americas by Columbus in his second trip in 1493, and in Brazil the first herds were brought by the Portuguese in 1534 (Primo, 1992).

Polymorphisms across the 70-kb lactase gene (*LCT*) have been previously described (Boll et al., 1991; Harvey et al., 1995). The worldwide distribution of their derived haplotypes showed that four haplotypes (A, B, C, and U) were observed at frequencies higher than 0.05. Haplotype A is the most common in all populations except in Sub-Saharan Africans; it has its highest frequency in northern Europeans. The B haplotype is present in all populations except Bantu-speaking South Africans. Haplotype C is also present in all populations but it is more common in southern Europe and India whereas, the U haplotype is rare in Indo-European populations (Harvey et al., 1998; Hollox et al., 2001).

Most human adults downregulate the production of intestinal lactase after weaning. Lactase is necessary for

the digestion of lactose, the main carbohydrate in milk, and without it, milk consumption can lead to bloating, flatulence, cramps, and nausea (Swallow, 2003). Persistence of small-intestinal lactase production into adult life in humans is caused by genetic differences *cis*-acting near the lactase gene (Wang et al., 1995), which enables some alleles to escape the developmental downregulation characteristic of the ancestral state. The first allele to be identified, -13910^*T (rs4988235) was discovered in a north-European sample and is in high linkage disequilibrium with -22018^*A allele (Enattah et al., 2002). The -13910^*T is present on the background of a single very extended Haplotype A (Hollox et al., 2001; Poulter et al., 2003; Bersaglieri et al., 2004; Coelho et al., 2005). The

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil.

*Correspondence to: Mara H. Hutz, Departamento de Genética, UFRGS, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. E-mail: mara.hutz@ufrgs.br

Received 10 August 2011; accepted 7 December 2011

DOI 10.1002/ajpa.22010

Published online 23 January 2012 in Wiley Online Library (wileyonlinelibrary.com).

TABLE 1. Characteristics of the populations investigated

Characteristics	Populations			
	Kaingang	Guarani-Kaiowá	Guarani-Ñandeva	Xavante
Localities	Nonoai	Amambai; Limão Verde	Amambai; Porto Lindo	Pimentel Barbosa
Geographic coordinates	52°W, 27°S	55°W, 23°S; 54°W, 23°S	55°W, 23°S; 54°W, 23°S	51°W, 13°S
Region (Brazil)	South	Central	Central	Central
Linguistic group ^a	Jê	Tupi	Tupi	Jê
Nonnative influence	Extensive	Extensive	Extensive	Minimal
Number of individuals studied	72	84	59	101

^a According to Greenberg, 1987.

–13910*T allele is rare in most African populations except in the Fulani (Lokki et al., 2011). Other alleles with similar effect [–13915T>G (rs41380347), –13907C>G (rs41525747), –13913T>C (rs41456145), 14010G>C (rs145946881)], were identified in African and Middle-Eastern populations (Enattah et al., 2002, 2008; Ingram et al., 2007; Tishkoff et al., 2007; Imtiaz et al., 2009; Tornaiainen et al., 2009). Alleles –13915*G and –14010*C are also present on single extended haplotypes that are distinct from the A haplotype and from each other (Tishkoff et al., 2007; Enattah et al., 2008). These polymorphisms are all located within *MCM6* (the upstream adjacent gene) intron, in a sequence that acts as an enhancer for lactase expression in vitro (Troelsen et al., 2003). The –13910*T allele acts as a stronger enhancer in vitro than the –13910*C allele (Troelsen et al., 2003). Recently, higher activity of the –13915*G and the –14010*C alleles were demonstrated in vitro (Jensen et al., 2011; Olds et al., 2011), whereas the functional role of other variants is suggested by their position in the enhancer region. In this study, we will use LP to refer to the lactase persistence, and *LCT* to the lactase gene.

The LP phenotype 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. A similar cline is seen from north to south of India. The frequency is low or absent in Asian and African populations except in pastoralist populations from Africa and the Middle-East (Tishkoff et al., 2007; Ingram et al., 2009).

A few old studies were performed with physiological tests to access the LP phenotype in native Americans, its frequency varies with admixture rates; in nonadmixed populations this trait was absent (for a review see Salzano and Callegari-Jacques, 1988).

Despite many studies conducted with Amerindians, only one recent investigation reported the frequency of the –13910*T allele in Chilean Indians (Morales et al., 2011). The objectives of our study were: a) since previous studies suggested the absence of lactase persistence in adulthood, and consequently of variation in the *LCT* enhancer region, would this uniformity also be found in its 5' flanking and coding regions? (b) if variability would be found among the latter, how would it compare with those observed in other ethnic groups?

SUBJECTS AND METHODS

Subjects

A total of 316 individuals from four Brazilian Amerindian populations were investigated. Table 1 provides information about these populations. They live in southern and central Brazil, speak languages of two distinct

groups, and have a varied degree of non-native influences. The Indian leaders and the study participants were adequately informed about the aims of this study and gave their approval. Our research program was approved by the Brazilian Ethics National Committee (Resolution 123/98). The LP status of the individuals sampled was not evaluated.

Genotyping

The –13910C>T (rs4988235) polymorphism was genotyped by PCR-RFLP as previously described (Bulhões et al., 2007). In an attempt to identify other polymorphisms in the –14-kb region, a fragment of 427 bp was amplified with the MCM6i13 and LAC-CL2 primers (Ingram et al., 2007). PCR products were purified with *Exonuclease I* and Shrimp Alkaline Phosphatase enzymes and then sequenced at MACROGEN (Seoul, Republic of Korea) using the MCM6i13 primer.

A total of 12 polymorphisms were genotyped at the *LCT* locus and in the enhancer region. A list of them is shown in Table 2. SNPs 666G>A (rs3754689), 5579T>C (rs2278544), 6236TG>ΔΔ (rs10552864), –22018G>A (rs182549) were genotyped by allelic discrimination using TaqMan assays in a real time PCR equipment (StepOne Plus, Applied Biosystems). The others were genotyped by sequencing a 597-bp fragment in the 5' flanking region of the *LCT*. Primers used to generate this fragment were forward 5'-ACC CAG GCA TTC TGA CTC CTGT-3' and reverse 5'-GCA GAG GAA CCC GGA AAA TGT CT-3'. PCR products were purified with *Exonuclease I* and Shrimp Alkaline Phosphatase enzymes and then were sequenced at MACROGEN (Seoul, Republic of Korea). Chromatograms were examined using Codon-Code Aligner Software v.3.5.7.

Statistical analyses

Allele frequencies were obtained directly by gene counting. Hardy-Weinberg equilibrium was tested for each locus within each population using Markov chain as implemented in Arlequin v.3.0 program (Excoffier et al., 2005). Haplotypes were inferred using a Bayesian algorithm implemented in Phase v.2.1 software (Stephens et al., 2001; Stephens and Donnelly, 2003).

Exact tests of population differentiation based on genotype frequencies (Raymond and Rousset, 1995; Goudet et al., 1996), and Wright's F_{ST} (Weir and Cockerham, 1984; Excoffier et al., 1992; Weir, 1996) were performed using Arlequin v.3.0, and Bonferroni's correction was applied when necessary. Heterozygosity (Nei, 1987) was calculated with the DISPAN software (Ota, 1993). Arlequin was also employed to perform the Ewens-Watterson neutrality test (infinite allele model) (Ewens, 1972;

TABLE 2. Polymorphisms investigated and their method of identification

Polymorphism position ^a	rs number	Position in chromosome 2	Method of identification
-22018G>A	rs182549	136450486	TaqMan
-13910C>T	rs4988235	136442378	PCR-RFLP
-958C>T	rs56064699	136595706	Sequencing
-946A>G	NA ^b	136595694	Sequencing
-942/3TC> $\Delta\Delta$	rs148142676	136595690	Sequencing
-942C>G	NA ^b	136595690	Sequencing
-875G>A	rs78205226	136595623	Sequencing
-678A>G	rs56211644	136595426	Sequencing
-552/9A ₈ >A ₉	rs113260867	136595299;136595300	Sequencing
666G>A	rs3754689	136590746	TaqMan
5579T>C	rs2278544	136546110	TaqMan
6236/7TG> $\Delta\Delta$	rs10552864	136545452	TaqMan

^a In base pairs, upstream and within the *LCT* gene.

^b NA = not available in the dbSNP.

Watterson, 1975; Stewart, 1977). Multidimensional Scaling analysis was performed with SPSS v.18.

RESULTS

All SNPs in the *LCT* and its 5' flanking region were in Hardy-Weinberg equilibrium in all populations. Polymorphisms in the 5' flanking region and within the *LCT* were used to derive *LCT* haplotypes as previously described by Hollox et al. (2001). The two polymorphisms investigated in the enhancer region were employed to infer LP genotypes.

The *LCT* was highly polymorphic showing 15 haplotypes with a highly heterogeneous distribution in Amerindians ($P < 0.00001$, Table 3). All four worldwide most common haplotypes (A, B, C, and U) were present in our sample. The most frequent haplotype in Amerindians is C (41%) while in other ethnic groups this figure is about 20% in Asians and Africans, and only 9% in Europeans (Supporting Information Table 1). Another striking feature in the Amerindian haplotype distribution is the high frequency of Haplotype E, a probable recombinant between A and C, (average 10%; Table 3, Supporting Information Table 1). Comparisons among the four native Brazilian populations showed that Haplotype A is more frequent in Guarani than in Je-speaking groups, where the C haplotype is the most frequent (Table 3). In one Xavante chromosome the inferred haplotype is unknown; it is probably a new, not previously described recombinant (Table 3). Average heterozygosity varied from 55% in the Xavante to 73% in the Guarani-Kaiowá (Table 3). The Ewens-Watterson neutrality test showed that the probability of observing random samples with F values identical or smaller than the original sample could be accepted suggesting that these haplotypes are not under selective pressure (Table 4), accepting the neutrality hypothesis. So the present-day haplotype distribution in Amerindians probably resulted from genetic drift. Multidimensional scaling using F_{ST} distances calculated using *LCT* haplotype frequencies is shown in Figure 1. The Xavante was the most isolated population whereas the Guarani-Kaiowá was closer to Asian and Southern European populations.

Frequencies of SNPs -22018G>A and -13910C>T in the enhancer region are shown in Table 3. These SNPs were in linkage disequilibrium ($P < 0.0001$) in all populations investigated forming an AT haplotype. The AT haplotype, which is associated with lactase persistence in the literature was detected together with an A haplotype in both Guarani populations and in the Kaingang.

In the Xavante this haplotype was observed associated with a C haplotype (Table 3). The second chromosome of this individual was inferred as B. Sequencing of the -14-kb enhancer region in these populations did not reveal other LP-related alleles in these populations. The enhancer AT haplotype varied from 0.5% in the Xavante to 6.8% in the Guarani-Nandeva (Table 3). This haplotype differs among Amerindian populations ($P < 0.00001$). The rare combination of the allele -13910*T allele and -22018*G was only present in one Guarani-Nandeva chromosome (0.8%), whereas the rare AC haplotype was present in the Kaingang only (0.7%). The -13910*T/-22018*G combination was observed in a B/C *LCT* haplotype heterozygous individual. Overall the -13910*T allele varied from 0.5% in the Xavante to 7.6% in the Guarani-Nandeva (Table 5).

DISCUSSION

The present investigation reports the first study of *LCT* polymorphism diversity in Brazilian natives. The *LCT* was highly polymorphic in this ethnic group showing 15 haplotypes with a heterogeneous distribution among the four populations investigated. The only mutation associated to LP observed was the -13910*T.

South American Indians inhabit regions with different environmental conditions, which include the Andean very high altitudes, the largest tropical rainforest of the world, and large stretches of desert land. These conditions lead to distinct use, by the Amerindians, of the flora and fauna available to them. Thus, the Xavante until recently were hunter and gatherers, while Kaingang and Guarani were agriculturalists and fishermen (Salzano and Callegari-Jacques, 1988).

The number of years of contact of these Amerindians with the colonizers varied from about 196 years for the Guarani, 180 years for the Kaingang, and only 30 years for the Xavante. This contact led to acculturation and also admixture with their non-Indian neighbors. The percentage of admixture ranges from 3 to 19% in the Guarani to 6% in the Kaingang (Callegari-Jacques and Salzano 1999). The AT LP haplotype, which is associated with lactase persistence was detected on an A haplotype background in both the Guarani and Kaingang. The LP allele frequencies and admixture estimates correlate perfectly, suggesting that this haplotype is probably of European descent.

Enattah et al. (2007) showed that the -13910*T variant is found on two different, highly divergent A subhaplotype backgrounds in the global populations. One is

TABLE 3. Combined enhancer region haplotype frequency and LCT haplotype frequencies (\pm standard error) in four Amerindian populations

Enhancer region	LCT haplotypes ^a															
	A	B	C	D	E	J	K	P	Q	S	T	U	X	c	j	q ^b
Guarani-Kaiowá (N ^c = 84)																
GC	0.440 \pm 0.038	0.113 \pm 0.024	0.161 \pm 0.028	0.155 \pm 0.028	0.006							0.119 \pm 0.025				
AT	0.006															
Guarani-Ñandeva (N = 59)																
GC	0.322 \pm 0.043	0.051 \pm 0.020	0.340 \pm 0.044	0.008	0.127 \pm 0.030					0.008		0.060 \pm 0.022		0.008		
GT			0.008													
AT	0.068 \pm 0.023															
Kaingang (N = 72)																
GC	0.257 \pm 0.036	0.097 \pm 0.025	0.423 \pm 0.041	0.090 \pm 0.024					0.007	0.021 \pm 0.012		0.028 \pm 0.014	0.021 \pm 0.012			
AC	0.007															
AT	0.049 \pm 0.018															
Xavante (N = 101)																
GC	0.094 \pm 0.020	0.099 \pm 0.021	0.653 \pm 0.033	0.045 \pm 0.014					0.005	0.005		0.005	0.069 \pm 0.018		0.010 \pm 0.007	0.005
AT			0.005													0.005
Amerindian (N = 316)																
	0.292 \pm 0.018	0.093 \pm 0.011	0.412 \pm 0.020	0.002	0.101 \pm 0.012	0.002	0.002	0.002	0.002	0.006 \pm 0.003	0.002	0.072 \pm 0.010	0.003 \pm 0.002	0.005 \pm 0.003	0.002	0.002

If only one chromosome was identified, the standard error was not estimated. Heterozygosity over the whole haplotype (Nei, 1987): Guarani-Kaiowá = 0.728, Guarani-Ñandeva = 0.711, Kaingang = 0.708, Xavante = 0.549; Exact test of sample differentiation (Raymond and Rosset, 1995; Goudet et al., 1996) for LP haplotype frequencies: $P < 0.00001$; for LCT haplotype frequencies: $P < 0.00001$.

^a Haplotype designation as given in Hollox et al., 2001.

^b New haplotype identified with the following arrangement: -958*C; -946*A; -942/3*TC; -942*C; -875*G; -678*G; -552/9*A₈; 666*A; 5579*T; 6236/7*TG.

^c N = number of samples (2n).

TABLE 4. Ewens-Watterson neutrality test of the summarized LCT haplotype frequencies (F) in which P-value is generated by simulation

	Guarani-Kaiowá	Guarani-Ñandeva	Kaingang	Xavante
Observed F value	0.131	0.128	0.139	0.268
Expected F value	0.147	0.104	0.120	0.124
Watterson F ^a P-value	0.450	0.850	0.763	0.764

^a Ewens, 1972; Watterson, 1975; Stewart, 1977.

the most common LP haplotype, present in all populations, whereas the other, which originated from the same ancestral allelic haplotype, is found in geographically restricted populations living west of the Urals and north of the Caucasus. In one Guarani-Ñandeva and in one Xavante chromosome the European lactase persistence allele -13910*T was observed on a C haplotype background. Two explanations are possible for these observations: a) Recombination or b) a de novo mutation. The occurrence of the coincident mutation at -22018*A as well as the existence of one GT haplotype in Guarani supports the notion that there has been some recombination. Although Harvey et al. (1998) reported rare cases of this phenotype on a C haplotype background in Italians, we cannot exclude that the -13910*T allele in Amerindians would be de novo mutations, mainly in the Xavante where admixture with non-Indians is virtually absent. In addition, this individual were also carrier of a Haplotype B, in which no LP -13910*T allele was observed (Harvey et al., 1998). These data lend strong support to the concept of convergent and still-ongoing LP adaptation in response to adult milk consumption in different human populations (Enattah et al., 2007; Torriainen et al., 2009).

The -22018G>A and -13910C>T polymorphisms are in high linkage disequilibrium, and they were originated according to a GC \rightarrow AC \rightarrow AT phylogenetic sequence. The intermediate AC haplotype is rare (Coelho et al., 2005), and was seen only in the Kaingang with a frequency of 0.7%. The AC haplotype was also reported in Portugal (0.01%) and São Tomé (0.02%) populations. On the other hand, the GT haplotype could only be possible due to a recombination event. But there seems to be an apparent lack of recombination within a region of 61.4 kb surrounding these polymorphisms (Coelho et al., 2005). The -22018 G>A polymorphism emerged before the -13910 C>T polymorphism in human populations (Coelho et al., 2005). If the -22018*A allele was present in the original population that colonized the American continent only one mutational event would be necessary to create the AT haplotype in Amerindians. The occurrence of a GT haplotype also suggest that the -13910*T allele in Amerindians is a de novo mutation.

The haplotype distribution of the LCT showed great heterogeneity, but only five haplotypes (A, B, C, E, and U) are present with a frequency of ≥ 0.05 . These haplotypes were probably present in the original population that colonized the American continent. Overall these haplotype frequencies were different from those reported in other major ethnic groups (Supporting Information Table 1). For instance, when they were compared to those of Asians, with whom they share more recent common ancestors, the C haplotype presented higher frequencies and A and B lower prevalence. In a two-dimensional plot (Fig. 1) the Xavante are the most different, and the Guarani-Kaiowá the most similar to Asian and

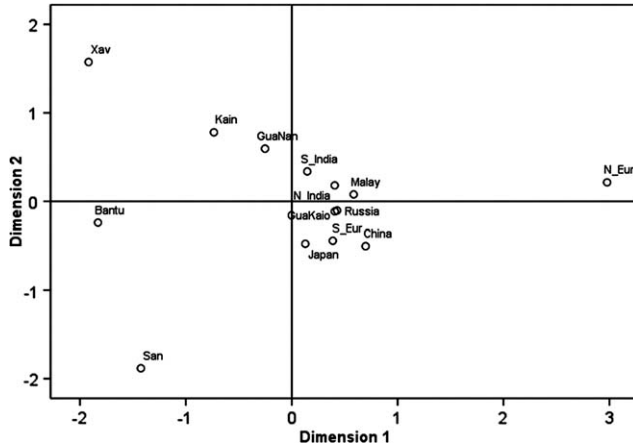


Fig. 1. Multidimensional scaling showing the relationships among Amerindians and other populations described by Hollox et al., 2001 in a two-dimension plot based on F_{ST} distances calculated using haplotype frequencies. Xav = Xavante, Kain = Kaingang, GuaKaio = Guarani-Kaiowá, GuaNan = Guarani-Nandeva, S_India = Southern India, N_India = Northern India, N_Eur = Northern Europe, S_Eur = Southern Europe. Bartu and San are African populations.

TABLE 5. Heterozygotes and homozygotes proportions of the polymorphism -13910C>T and allele frequency

Population	N	-13910 C>T genotype; number of individuals			Allele frequency (%)	
		CC	CT	TT	C	T
Guarani-Kaiowá	84	83	1		99.4	0.6
Guarani-Nandeva	59	50	9		92.4	7.6
Kaingang	72	66	5	1	95.1	4.9
Xavante	101	100	1		99.5	0.5

Southern Europeans. This similarity is not unexpected due to the admixture level observed in the Guarani. Northern Europe was also isolated but this difference in frequencies probably occurred by LP allele positive selection that increased the A haplotype frequency by genetic hitchhiking.

In relation to diversity, only the Xavante presented decreased haplotypic heterozygosity (0.55), which reflects the high frequency of the C haplotype in this population, whereas the other investigated groups showed high values (0.71–0.73). The mean heterozygosity of these groups were previously estimated with STR markers, and presented similar results for Kaingang (0.77 ± 0.02), Guarani (0.71 ± 0.02), and Xavante (0.69 ± 0.03), (Kohlrausch et al., 2005).

Patterns of haplotype diversity can be attributed to relatively recent events, probably generated by LP selection, and older events affecting a large subset of populations were probably generated by drift (Hollox et al., 2001). Because in the original population that migrated to America the LP allele was probably absent, the haplotype diversity now found may be due to drift, as indicated by the neutrality test performed. This result was expected since pastoralism was not common in historical times.

CONCLUSION

It is therefore clear that the absence of original variation in the enhancer region of the *LCT* gene is not reflected in the variability found in its 5' flanking and

coding regions; and that the variation found in the latter is distinctive. Different evolutionary factors can be at play even in relatively close DNA portions.

ACKNOWLEDGMENTS

Thanks are due to the Fundação Nacional do Índio for authorizing this study and for logistic assistance. The Indian leaders and the subjects of investigation were adequately informed about the aims of the study and gave their approval, which is gratefully acknowledged. The research program was approved by the Brazilian Ethics National Committee (Resolution 123/98).

LITERATURE CITED

Bersaglieri T, Sabeti P, Patterson N, Vanderploeg T, Schaffner S, Drake J, Rhodes M, Reich D, Hirschhorn J. 2004. Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* 74:1111–1120.

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.

Bulhões A, Goldani H, Oliveira F, Matte U, Mazzuca R, Silveira T. 2007. Correlation between lactose absorption and the C/T-13910 and G/A-22018 mutations of the lactase-phlorizin hydrolase (*LCT*) gene in adult-type hypolactasia. *Braz J Med Biol Res* 40:1441–1446.

Callegari-Jacques SM, Salzano FM. 1999. Brazilian Indian/non-Indian interactions and their effects. *Cienc Cult* 51:166–174.

Coelho M, Luiselli D, Bertorelle G, Lopes A, Seixas S, Destro-Bisol G, Rocha J. 2005. Microsatellite variation and evolution of human lactase persistence. *Hum Genet* 117:329–339.

Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Järvelä I. 2002. Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 30:233–237.

Enattah NS, Trudeau A, Pimenoff V, Maiuri L, Auricchio S, Greco L, Rossi M, Lentze M, Seo JK, Rahgozar S, Khalil I, Alifrangis M, Natah S, Groop L, Shaat N, Kozlov A, Vershchubskaya G, Comas D, Bulayeva K, Mehdi SQ, Terwilliger JD, Sahi T, Savilahti E, Perola M, Sajantila A, Järvelä I, Peltonen L. 2007. Evidence of still-ongoing convergence evolution of the lactase persistence T-13910 alleles in humans. *Am J Hum Genet* 81:615–625.

Enattah NS, Jensen TG, Nielsen M, Lewinski R, Kuokkanen M, Rasinpera H, El-Shanti H, Seo JK, Alifrangis M, Khalil IF, Natah A, Ali A, Natah S, Comas D, Mehdi SQ, Groop L, Vestergaard EM, Imtiaz F, Rashed MS, Meyer B, Troelsen J, Peltonen L. 2008. Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am J Hum Genet* 82:57–72.

Ewens WJ. 1972. The sampling theory of selectively neutral alleles. *Theor Popul Biol* 3:87–112.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50.

Goudet J, Raymond M, de Meeüs T, Rousset F. 1996. Testing differentiation in diploid populations. *Genetics* 144:1933–1940.

Greenberg J. 1987. *Language in the Americas*. Stanford, CA: Stanford University Press.

Harvey CB, Pratt W, 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, Hollox EJ, Poulter M, Wang Y, Rossi M, Auricchio S, Iqbal TH, Cooper BT, Barton R, Sarnar M, Korpela R,

- Swallow DM. 1998. Lactase haplotype frequencies in Caucasians: association with the lactase persistence/nonpersistence polymorphism. *Ann Hum Genet* 62:215–223.
- Hollox EJ, Poulter M, Zvarik M, Ferak V, Krause A, Jenkins T, Saha N, Kozlov A, Swallow DM. 2001. Lactase haplotype diversity in the Old World. *Am J Hum Genet* 68:160–172.
- Imtiaz F, Savilahti E, Sarnesto A, Trabzuni D, Al-Kahtani K, Kagevi I, Rashed MS, Meyer BF, Järvelä I. 2007. The T/G-13915 variant upstream of the lactase gene (*LCT*) is the founder allele of lactase persistence in an urban Saudi population. *J Med Genet* 44:e89.
- Ingram CJE, Elamin MF, Mulcare CA, Weale ME, Tarekegn A, Raga TO, Bekele E, Elamin FM, Thomas MG, Bradman N, Swallow DM. 2007. A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Hum Genet* 120:779–788.
- Ingram CJE, Mulcare CA, Itan Y, Thomas MG, Swallow DM. 2009. Lactose digestion and the evolutionary genetics of lactase persistence. *Hum Genet* 124:579–591.
- Jensen TGK, Liebert A, Lewinsky R, Swallow DM, Olsen J, Troelsen JT. 2011. The -14010°C variant associated with lactase persistence is located between an Oct-1 and HNF1 α binding site and increases lactase promoter activity. *Hum Genet* 130:483–493.
- Kohlrausch FB, Callegari-Jacques SM, Tsuneto LT, Petzl-Erler ML, Hill K, Hurtado AM, Salzano FM, Hutz MH. 2005. Geography influences microsatellite polymorphism diversity in Amerindians. *Am J Phys Anthropol* 126:463–470.
- Lokki AI, Järvelä I, Israelsson E, Maiga B, Troye-Blomberg M, Dolo A, Doumbo OK, Meri S, Holmberg V. 2011. Lactase persistence genotypes and malaria susceptibility in Fulani of Mali. *Malar J* 10:9.
- Morales E, Azocar L, Maul X, Perez C, Chianale J, Miquel JF. 2011. The European lactase persistence genotype determines the lactase persistence state and correlates with gastrointestinal symptoms in the Hispanic and Amerindian Chilean population: a case-control and population-based study. *BMJ Open* 1:e000125.
- Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.
- Olds LC, Ahn JK, Sibley E. 2011. -13915*G DNA polymorphism associated with lactase persistence in Africa interacts with Oct-1. *Hum Genet* 129:111–113.
- Ota T. 1993. *DISPAN: genetic distance and phylogenetic analysis*. University Park, PA: Institute of Molecular Evolutionary Genetics, Pennsylvania State University.
- Poulter M, Hollox EJ, Harvey CB, Mulcare CA, Peuhkuri K, Kajander K, Sarner M, Korpela R, Swallow DM. 2003. The causal element for the lactase persistence/nonpersistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. *Ann Hum Genet* 67:298–311.
- Primo AT. 1992. El ganado bovino iberico en las Americas: 500 años después. *Arch Zootec* 41:421–431.
- Raymond M, Rousset F. 1995. An exact test for population differentiation. *Evolution* 49:1280–1283.
- Salzano FM. 2011. The prehistoric colonization of the Americas: evidence and models. *Evol Educ Outreach* 4:199–204.
- Salzano FM, Callegari-Jacques S. 1988. *South American Indians: a case study in evolution*. Oxford: Clarendon Press.
- Stephens M, Donnelly P. 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73:1162–1169.
- Stephens M, Smith N, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989.
- Stewart F. 1977. Computer algorithm for obtaining a random set of allele frequencies for a locus in an equilibrium population. *Genetics* 86:482–483.
- Swallow DM. 2003. Genetics of lactase persistence and lactose intolerance. *Annu Rev Genet* 37:197–219.
- Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, Ibrahim M, Omar SA, Lema G, Nyambo TB, Ghorri J, Bumpstead S, Pritchard JK, Wray GA, Deloukas P. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* 39:31–40.
- Torniainen S, Parker MI, Holmberg V, Lahtela E, Dandara C, Jarvela I. 2009. Screening of variants for lactase persistence/nonpersistence in populations from South Africa and Ghana. *BMC Genet* 10:31.
- Troelsen JT, Olsen J, Møller J, Sjöström H. 2003. An upstream polymorphism associated with lactase persistence has increased enhancer activity. *Gastroenterology* 125:1686–1694.
- Wang Y, Harvey CB, Pratt WS, Sams VR, Sarner M, Rossi M, Auricchio S, Swallow DM. 1995. The lactase persistence/nonpersistence polymorphism is controlled by a cis-acting element. *Hum Mol Genet* 4:657–662.
- Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276.
- Weir B. 1996. *Genetic data analysis II: methods for discrete population genetic data*. Sunderland, MA: Sinauer.
- Weir B, Cockerham C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.

SUPPLEMENTARY TABLE 1. LCT gene haplotype range frequency (%) and weighted average (%) in Amerindians and other populations.

Haplotypes	Amerindians ^a		Africans ^b		Europeans ^b		Asians ^b	
	Range	Average	Range	Average	Range	Average	Range	Average
A	9.4 - 44.6	29.2	6 - 10	8.8	36 - 86	53.6	37 - 49	45.1
B	5.1 - 11.3	9.3	0 - 3	0.9	6 - 32	23.6	10 - 22	15.2
C	16.1 - 65.4	41.2	3 - 31	23.0	3 - 15	9.5	9 - 31	19.6
D	0 - 0.8	0.2			1 - 3	2.0		
E	5 - 15.5	10.1			2 - 6	3.5	0 - 2	1.5
G					0 - 4	1.8	0 - 6	2.8
H			0 - 4	1.2	0 - 1	0.2	0 - 1	0.4
I					0 - 2	0.7	0 - 2	0.7
J	0 - 0.6	0.2			0 - 3	1.1	0 - 1	0.6
K	0 - 0.5	0.2	0 - 8	5.6	0 - 3	1.0	0 - 1	0.1
L					0 - 1	0.3		
M			0 - 5	3.5	0 - 1	0.7	0 - 1	0.7
N							0 - 2	0.5
O			6 - 20	10.0				
P	0 - 0.5	0.2	4 - 10	5.8			0 - 1	0.4
Q	0 - 0.7	0.2	0 - 4	2.8	0 - 3	1.0	1 - 2	1.6
S	0.85 - 2.1	0.6	0 - 2	1.4			0 - 4	1.2
T	0 - 0.5	0.2	1 - 3	1.6				
U	2.8 - 11.9	7.2	7 - 17	10.0	0 - 1	0.2	0 - 24	8.5
V			1 - 3	1.6				
W							0 - 2	0.3
X	0 - 2.1	0.3	0 - 7	5.0				
Y			0 - 13	3.8				

a					0 - 1	0.4	0 - 1	0.3
c	0 - 1	0.5						
d			0 - 4	1.2				
e							0 - 1	0.2
f			0 - 1	0.7				
g			0 - 1	0.7				
j	0 - 0.5	0.2	0 - 5	3.6				
k			0 - 1	0.7	0 - 1	0.4		
l			0 - 6	4.2				
m			0 - 3	0.9				
n			0 - 10	3.0				
p							0 - 1	0.3
q	0 - 0.5	0.2						

^aFrequencies and averages from this study.

^bFrequencies and nomenclature as given in Hollox et al., 2001.

Capítulo 3: Artigo intitulado “Several different lactase persistence associated alleles and high diversity of the lactase gene in the admixed Brazilian population” publicado na PLoS ONE, 2012.

Several Different Lactase Persistence Associated Alleles and High Diversity of the Lactase Gene in the Admixed Brazilian Population

Deise C. Friedrich¹, Sidney E. B. Santos², Ândrea K. C. Ribeiro-dos-Santos², Mara H. Hutz^{1*}

¹ Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, ² Laboratório de Genética Humana e Médica, Universidade Federal do Pará, Belém, Brazil

Abstract

Adult-type hypolactasia is a common phenotype caused by the lactase enzyme deficiency. The $-13910\text{ C}>\text{T}$ polymorphism, located 14 Kb upstream of the lactase gene (*LCT*) in the *MCM6* gene was associated with lactase persistence (LP) in Europeans. This polymorphism is rare in Africa but several other variants associated with lactase persistence were observed in Africans. The aims of this study were to identify polymorphisms in the *MCM6* region associated with the lactase persistence phenotype and to determine the distribution of *LCT* gene haplotypes in 981 individuals from North, Northeast and South Brazil. These polymorphisms were genotyped by PCR based methods and sequencing. The -13779^*C , -13910^*T , -13937^*A , -14010^*C , -14011^*T LP alleles previously described in the *MCM6* gene region that acts as an enhancer for the *LCT* gene were identified in Brazilians. The most common LP allele was -13910^*T . Its frequency was highly correlated with European ancestry in the Brazilian populations investigated. The -13910^*T was higher (0.295) in southern Brazilians of European ancestry and lower (0.175) in the Northern admixed population. *LCT* haplotypes were derived from the 10 *LCT* SNPs genotyped. Overall twenty six haplotypes previously described were identified in the four Brazilian populations studied. The Multidimensional Scaling analysis showed that Belém, in the north, was closer to Amerindians. Northeastern and southern Afro-descendants were more related with Bantu-speaking South Africans whereas the Southern population with European ancestry grouped with Southern and Northern Europeans. This study shows a high variability considering the number of *LCT* haplotypes observed. Due to the highly admixed nature of the Brazilian populations, the diagnosis of hypolactasia in Brazil, based only in the investigation of the -13910^*T allele is an oversimplification.

Citation: Friedrich DC, Santos SEB, Ribeiro-dos-Santos ÁKC, Hutz MH (2012) Several Different Lactase Persistence Associated Alleles and High Diversity of the Lactase Gene in the Admixed Brazilian Population. PLoS ONE 7(9): e46520. doi:10.1371/journal.pone.0046520

Editor: Dana C. Crawford, Vanderbilt University, United States of America

Received: May 9, 2012; **Accepted:** September 1, 2012; **Published:** September 28, 2012

Copyright: © 2012 Friedrich et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Financial support was provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (Brazil; grant # 470937/2010-6). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mara.hutz@ufrgs.br

Introduction

Adult-type hypolactasia or lactose intolerance (OMIM #223100) is a worldwide common phenotype determined by lactase deficiency, it is due to lactase activity decline after weaning. Lactase or lactase-phlorizin hydrolase enzyme (EC 3.2.1.23-62) is encoded by the *LCT* gene and it is located in the brush border membrane of small-intestinal enterocytes. The lactase enzyme activity is to hydrolyze lactose, the main carbohydrate in milk [1]. Most intolerant subjects present symptoms like bloating, flatulence, nausea, and diarrhea after consumption of fresh milk [2–4]. Moreover, adult-onset lactase decline appears to be a risk factor for osteoporosis due to avoidance of dairy products or undigested lactose interference with calcium absorption [5].

The regulation of *LCT* expression in humans has been studied extensively. No causative differences in the *LCT* gene sequence have been found within the gene. However, a T/C polymorphism at position -13910 and an A/G polymorphism at position -22018 from the start codon of the *LCT* gene have been identified. Although these nucleotide variants are located in introns 9 and 13 of the neighboring *MCM6* gene, the -13910^*C allele associates 100% and the -22018^*G allele associates

approximately 97% with the lactase nonpersistent phenotype [6–8]. The region surrounding the -13910 position has been described to function as an enhancer stimulating the *LCT* promoter activity. The derived allele -13910T increases promoter activity [8–11].

The *LCT* gene (OMIM #603202) was mapped on 2q21 [12]. Several single nucleotide polymorphisms (SNPs) were described across the lactase gene, and these polymorphic sites were used to derive *LCT* haplotypes [13,14]. The two SNPs associated with lactase persistence (LP) phenotype are linked to an A-haplotype background in European populations [6]. Mulcare et al. [15] showed that the -13910^*T allele cannot be causal of lactase persistence in most Africans, although it could possibly explain lactase persistence in some Cameroonians. In that study, it was suggested that the presence of the -13910^*T allele in Cameroon is due to introgression from outside sub-Saharan Africa.

In nomadic pastoralist and non-pastoralists groups from East and South Africa and Middle East populations, other polymorphisms at the same enhancer region or on its vicinity were also related to the LP phenotype. For example, $-13907\text{C}>\text{G}$ (rs41525747) and $-13915\text{T}>\text{G}$ (rs41380347) were both identified

in Ethiopia, Kenya, Saudi Arabia, Sudan, and Tanzania populations, whereas the $-13915T>G$ was also found in Ethiopian Somali, Morocco, and Jordan [16–19], whereas the $-14010G>C$ (rs145946881) polymorphism was described in Kenya, Tanzania and Xhosa-speaking South Africans [17,20]. Functional studies demonstrated the role of the $-13910*T$, $-13907*G$, $-13915*G$, and $-14010*C$ alleles in the maintenance of the enzyme expression during adulthood [9–11].

The Brazilian populations were formed by successive migratory waves. Amerindian people occupied the Brazilian territory when the Portuguese arrived in 1500 and colonized the country. Then between the 16th and 19th centuries, West and Southwest Africans were brought to Brazil as slaves. In addition to the Portuguese, other migratory waves occurred in the 19th and 20th centuries, mainly from Italy, Germany and Spain [21]. All of these migratory events contributed to the formation of a multi-ethnic and highly admixed population. This heterogeneity was documented in several genetic studies that used uniparental or autosomal markers to demonstrate a typical, although non-uniform, tri-ethnic (European, African and Amerindian) pattern for the Brazilian population. This admixture process occurred in different ways in the various geographic regions of the country. In Northeastern Brazil, the African contribution is high; in the North, the contribution of Native Americans is pronounced; and in the South, there are reduced Amerindian and African influences when compared with the other geographic regions [22,23].

The aims of this study were (1) to determine the prevalence of LP related alleles; and (2) to describe the distributions patterns of the *LCT* haplotypes in the Brazilian population.

Results

Identification of SNPs in the *LCT* Enhancer Region

The overall $-13910*T$ allele frequency varied from 17.5% in the Northern admixed population to 29.5% in Southern Brazilians of European ancestry (Table 1). The $-13910*T$ allele frequency is higher in the Southern Euro-descendants ($p = 1.7 \times 10^{-5}$) than in the other Brazilian populations investigated. As LP is a dominant trait [24], the LP predicted phenotype frequency based on $-13910C>T$ genotypes was inferred (Table 1). The majority of the population from Belém, Recife, and the Afro-descendants from Porto Alegre are lactose intolerant (CC genotype with almost 70% frequency). In the Euro-descendants individuals from Porto Alegre, the lactase persistence frequency is higher than 50%.

Comparisons of the $-13910*T$ allele frequencies of the four Brazilian populations with frequencies available in dsSNP-NCBI database showed that overall these frequencies differ from those described, although some similarities could also be observed. Subjects with European ancestry from Porto Alegre are similar to those described at the global frequency from the NIH Polymorphism Discovery Resource (PDR90) only. The Afro-descendants from Porto Alegre, Recife, and Belém populations did not differ from African Americans from Southwest USA (HapMap-ASW). The Recife population also did not differ from the frequency described at the 1000Genomes project. These results are shown in the Table S1.

The $-13910C>T$ and $-22018G>A$ polymorphisms are in high linkage disequilibrium (data not shown). The combined frequencies of these two alleles are presented in Table 2.

Table 3 shows other variants found by sequencing of the *LCT* enhancer region in the Brazilian population. Four different polymorphisms were observed: $-13937*A$ (rs4988234), $-14010*C$ (rs145946881), $-14011*T$ (rs4988233) and $-13779*C$ (not included in the dsSNP). Their regional distribu-

tion is also shown in Table 3. A total of 9 heterozygous individuals for each of these alleles were detected. The Northeastern population was the more variable with 3 different alleles besides the $-13910*T$ whereas Euro-descendants subjects from the South presented only the $-13910*T$ allele.

LCT Haplotypes

A total of 26 haplotypes were observed in the Brazilian population. The most variable population was Recife in the northeast that presented 21 haplotypes. The most frequent haplotype in all populations was the A. The haplotypes observed and their frequencies are shown in Table 4. The F_{ST} values were calculated for haplotype frequencies of the four Brazilian and their parental populations (Amerindians, Bantu-speaking population from Africa, and Southern and Northern Europeans). The degree of differentiation among populations is not high. The highest F_{ST} value is between the Euro-descendants from Porto Alegre and the Bantu-speaking population from Africa (0.202, $p < 0.0001$, Table 5).

The nonmetric Multidimensional Scaling (MDS) analysis (Figure 1) showed the relationships among Brazilians with their parental populations based on D_A genetic distance. Belém, in the north, was closer to Amerindians. Recife and southern Afro-descendants were more related with Bantu-speaking South Africans whereas the Porto Alegre population with European ancestry grouped with Southern and Northern Europeans. The stress of this model is 0.047.

The $-13910*T$ allele was observed in other haplotypes than the A in the Brazilian population. The association of the $-13910*T$ and *LCT* haplotypes in the four Brazilian populations are shown in the Tables S2, S3, S4, and S5.

Discussion

The $-13910*T$ allele is present in the three Brazilian regions studied. The highest frequency of this allele was found in Euro-descendants southern subjects. The Afro-descendants from Porto Alegre also have a high frequency of this allele (18.4%) probably due to the high proportion of European contribution (43.1%) [23]. The $-13910*T$ LP European allele was also present in Northern and Northeastern Brazil, both populations with high contributions of European ancestry to their gene pool (69.7% and 60.6%, respectively) [23].

The $-13910C>T$ and $-22018G>A$ polymorphisms are in high linkage disequilibrium in the four Brazilian populations studied, as in Northern Europeans and in the African Fulbe population [6,25]. The $G \rightarrow A$ mutation might have occurred only shortly before the $C \rightarrow T$ mutation and it has been suggested that there was not enough time for recombination to break the TA haplotype [25]. Nevertheless we found the $-13910*T$ allele in combination with the $-22018*G$ allele in southern Brazilians with European ancestry and in the admixed population from the northeast (Tables S2, and S5). This combination would be possible under three situations: if a *de novo* mutation generating a $-13910*T$ allele occurred in a chromosome that carried a $-22018*G$ allele; or if a recombination event occurred; this second situation being more plausible in a highly admixed population as Brazilians. Moreover the fact that the TG combination is present in four different haplotypes (Tables S2, and S5) reinforces the recombination hypothesis. Another possibility for this finding is the contribution of Amerindian genomic ancestry. The TG haplotype was also observed in this ethnic group [26].

The CA combination was found on an A-haplotype background (Tables S2, S3, S4, S5). This CA haplotype is considered rare but

Table 1. Observed allele and genotype frequencies of the $-13910\text{ C}>\text{T}$ polymorphism in four Brazilian populations.

Population	$-13910*\text{T}$ allele	$-13910\text{ C}>\text{T}$ genotypes			LP predicted phenotype frequency
		TT	CT	CC	
		N %	N %	N %	
Porto Alegre ^a					
European ancestry (n = 337)	0.295	26 7.7	147 43.6	164 48.6	0.513
African ancestry (n = 182)	0.184	9 4.9	49 26.9	124 68.1	0.320
Belém ^b (n = 200)	0.175	8 4.0	54 27.0	138 69.0	0.310
Recife ^c (n = 262)	0.204	11 4.2	85 32.4	166 63.4	0.366

n = number of individuals; the genotypes are at Hardy-Weinberg equilibrium.

^aSouthern Brazil.

^bNorthern Brazil.

^cNortheastern Brazil.

Frequencies in bold are lower than the frequency observed in individuals with European ancestry from Porto Alegre. Heterogeneity Chi-square test: $p = 1.7 \times 10^{-5}$. Results from Chi-square test partition according to Everitt [50].

$p = 2 \times 10^{-4}$ for the comparison between individuals from European and African ancestries from Porto Alegre.

$p = 2.2 \times 10^{-5}$ for the comparison between population from Belém and from Porto Alegre with European ancestry.

$p = 0.001$ for the comparison between population from Recife and from Porto Alegre with European ancestry.

doi:10.1371/journal.pone.0046520.t001

it was also reported in the population from London [27], Portugal, São Tomé Island [25], and Kaingang [26]. It is interesting that the São Tomé Island has a colonization history similar to Brazil, where Portuguese settlers imported slaves from the Gulf of Guinea and from Congo and Angola region [28]. Maybe the CA combination is a genomic vestige of the Portuguese settlers that colonized both São Tomé Island and Brazil.

The $-13910*\text{T}$ allele has been first reported to occur exclusively on a *LCT* A haplotype background [27]. More recently, it has been shown that the $-13910*\text{T}$ occurs on two divergent A subhaplotypes suggesting more than one origin for the lactase persistence allele in Europeans [29]. Ingram et al [19] described this allele on an F haplotype in this ethnic group. In Brazil the $-13910*\text{T}$ allele was found on A, B, J, and K haplotypes backgrounds (Tables S2, S3, S4, S5). Probably recombination was the source of this diversity in this tri-ethnic population.

This study reports the presence of other substitutions in the *LCT* enhancer region than the $-13910\text{C}>\text{T}$ in the Brazilian population. These variants had been previously described in Africans. The $-14010*\text{C}$ allele observed in one individual with African ancestry from Porto Alegre is common in East and South Africa. This allele is considered a LP allele because it has been demonstrated *in vitro* that it increases gene transcription [30]. It occurs in 32% and 39% of Kenyans and Tanzanians respectively [18]. It has also been reported in the Somali [19]. In Black Xhosa-

speaking South Africans the $-14010*\text{C}$ allele was observed in 13.3% of the individuals investigated [20]. This allele was also observed in low frequencies (1–6%) in Angola, Southwest Africa [31]. Two explanations for these observations in different parts of Africa were hypothesized: 1) a direct migratory link between East and Southwest Africa; 2) a first contact between East African and South pastoralists, followed by $-14010*\text{C}$ allele transfer to Southwest pastoralists [31].

In Brazil, the $-14010*\text{C}$ allele occurred in a heterozygous individual for A and P haplotypes. This allele was first described on an F haplotype [17] and after on a B-haplotype background [19]. In this last report the $-958\text{C}>\text{T}$ (rs56064699) polymorphism that discriminates between B and P haplotypes was not tested; therefore the B haplotype of Ingram et al. [19] might be the same P haplotype observed herein.

Less is known about the other variants detected. The $-13779*\text{C}$ allele was described in a lactose non-digester individual (frequency 1/107) from a Somali cohort [19]. But it was common in an Indian herder sample (0.024) [32]. Now this allele was detected in two admixed subjects from Recife. The $-13937*\text{A}$ allele was observed in one individual of African ancestry from Porto Alegre and in one individual from Recife. This allele was described at low frequencies (0.014) in the Black Xhosa-speaking people from South Africa, a population that has the habit of consuming fermented milk [20]. Functional studies about the role

Table 2. Frequencies of the combinations found between the $-13910\text{ C}>\text{T}$ and $-22018\text{ G}>\text{A}$ alleles in the Brazilian population.

Alleles	Porto Alegre		Belém	Recife
	European ancestry (n = 337)	African ancestry (n = 182)	(n = 200)	(n = 262)
CG	0.686	0.811	0.822	0.784
TA	0.279	0.184	0.175	0.204
CA	0.022	0.005	0.003	0.008
TG	0.013	–	–	0.004

doi:10.1371/journal.pone.0046520.t002

Table 3. Number of -13779*C, -13937*A, -14010*C, and -14011*T alleles according to the population.

Allele	Porto Alegre African ancestry		Recife
		Belém	
-13779*C			2 ^d
-13937*A	1 ^a		1 ^e
-14010*C	1 ^b		
-14011*T		1 ^c	3 ^f

^aHeterozygous A and U *LCT* haplotypes.

^bHeterozygous A and P *LCT* haplotypes.

^cHomozygous A *LCT* haplotype.

^dHeterozygous C and U *LCT* haplotypes; heterozygous C and B *LCT* haplotypes.

^eHeterozygous A and g *LCT* haplotypes.

^fHomozygous A *LCT* haplotype; heterozygous A and E *LCT* haplotypes;

heterozygous A and S *LCT* haplotypes.

doi:10.1371/journal.pone.0046520.t003

of these two variants in the transcription of the *LCT* have not been performed yet.

The -14011C>T was described in the Estonian and Indian population [32,33] and its global frequency is 0.006. The functional role of this variant in the *LCT* transcription is unknown. But its location is a good predictor of functionality since it is close to the -14010G>C that interacts with transcription factors that increase lactase promoter activity [30]. This variant was observed in four admixed Brazilian subjects, three from the Northeast and one from the North.

This diversity at the *LCT* enhancer region is not unexpected if we consider the Brazilian roots. The slaves brought to Brazil were mainly from West and Southwest African areas and highly admixed with the European colonizers [23,34].

Although the high admixture rate in Brazil determine a high number of *LCT* haplotypes, the A haplotype was the most frequent. The A haplotype prevalence is explained by the large contribution of European ancestry to the Brazilian population gene pool: 69.7% in the North, 60.6% in the Northeast, and 94% in the South [23,34].

The MDS analysis shows the close relation of parental populations to Brazilians. Southern Europeans are closer to Porto Alegre Euro-descendants. The Porto Alegre Afro-descendants and Recife population are more related to the Bantu-speaking South

Table 4. *LCT* haplotypes frequencies \pm standard error in the Brazilian population.

<i>LCT</i> haplotype ^a	Porto Alegre		Belém	Recife
	European ancestry (n ^b =321)	African ancestry (n=182)	(n=200)	(n=258)
A	0.460 \pm 0.019	0.343 \pm 0.025	0.380 \pm 0.024	0.378 \pm 0.021
B	0.194 \pm 0.015	0.127 \pm 0.017	0.147 \pm 0.018	0.168 \pm 0.016
C	0.176 \pm 0.015	0.225 \pm 0.022	0.242 \pm 0.021	0.205 \pm 0.018
D	0.028 \pm 0.006	0.011 \pm 0.005	0.025 \pm 0.008	0.008 \pm 0.004
E	0.060 \pm 0.009	0.036 \pm 0.009	0.055 \pm 0.011	0.029 \pm 0.007
G	0.001	0.003	0.005 \pm 0.003	0.002
H			0.005 \pm 0.003	0.002
I	0.005 \pm 0.003	0.008 \pm 0.005	0.003	0.010 \pm 0.004
J	0.003 \pm 0.002	0.003		0.012 \pm 0.005
K	0.005 \pm 0.003	0.036 \pm 0.009	0.012 \pm 0.005	0.025 \pm 0.007
M	0.001	0.003	0.003	0.002
P	0.014 \pm 0.005	0.038 \pm 0.01	0.030 \pm 0.008	0.023 \pm 0.006
Q	0.005 \pm 0.003	0.008 \pm 0.005	0.003	0.012 \pm 0.005
S	0.006 \pm 0.003	0.044 \pm 0.01	0.017 \pm 0.006	0.015 \pm 0.005
T			0.003	0.002
U	0.030 \pm 0.007	0.066 \pm 0.013	0.045 \pm 0.010	0.062 \pm 0.010
W	0.001			
X	0.008 \pm 0.003	0.030 \pm 0.009	0.025 \pm 0.008	0.027 \pm 0.007
C	0.001	0.008 \pm 0.005		0.002
D	0.001			
G				0.010 \pm 0.004
H		0.003		
M				0.004 \pm 0.003
O		0.005 \pm 0.004		0.002
P	0.001			
Q		0.003		

If only one chromosome was identified, the standard error was not estimated.

^aNomenclature according to Hollox et al. [14].

^bn = number of individuals.

doi:10.1371/journal.pone.0046520.t004

Table 5. F_{ST} values of haplotype frequencies among Brazilians and with the parental populations (Amerindians, Bantu-speaking population from Africa, Southern and Northern Europeans).

	Porto Alegre		Belém	Recife
	African ancestry	European ancestry		
Porto Alegre European ancestry	0.01981			
Belém	0.00202	0.00796		
Recife	0.00055	0.00861	0.00047	
Brazilian Amerindian ^a	0.03714	0.06281	0.02874	0.04412
North Europe ^b	0.19868	0.13396	0.18029	0.17467
South Europe ^b	0.02886	0.01916	0.02607	0.01738
Bantu ^b	0.10271	0.20200	0.14165	0.13205

F_{ST} values in bold are statistically significant ($p < 0.0001$).

^aHaplotype frequencies from Friedrich et al. [26].

^bHaplotype frequencies from Hollox et al. [14].

doi:10.1371/journal.pone.0046520.t005

Africans, whereas Brazilian Amerindians have a closer relation with the Belém population.

Two previous studies validated the screening of the $-13910C>T$ polymorphism for hypolactasia molecular diagnosis in Brazil [35,36]. A third study concluded that the $-22018G>A$ polymorphism is a better predictor of lactase persistence in Japanese-Brazilians than the $-13910C>T$ [37]. In our study we

demonstrated that a more comprehensive screening would be needed since we found four variants in the enhancer region besides the $-13910C>T$. If only the $-13910C>T$ polymorphism would be tested in Recife, for example, 6 individuals would be considered lactose intolerant and they are carrier of *LCT* enhancer region variants that could be causal of the lactase persistence phenotype. In heterogeneous populations like Brazilians a single test for the

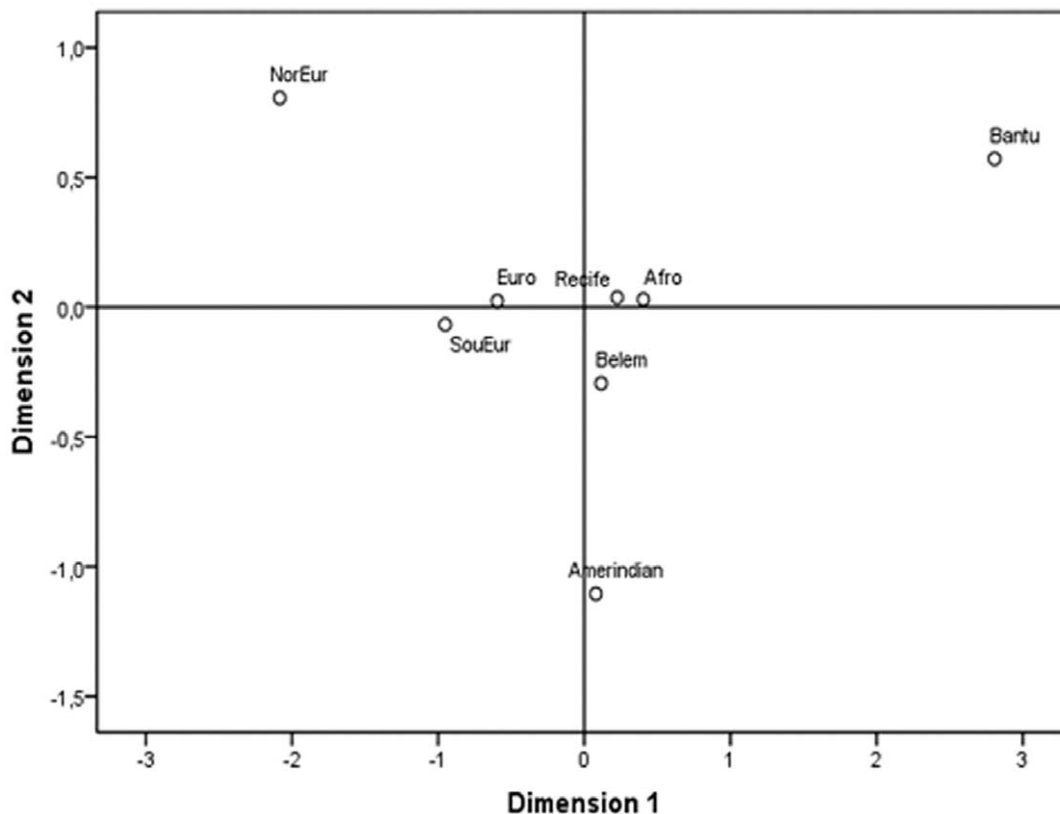


Figure 1. MDS of several populations *LCT* haplotypes. Nonmetric Multidimensional Scaling analysis of *LCT* haplotypes based on D_A distance showing the relationships among the four Brazilian populations with their parental groups: Brazilian Amerindians [26]; Southern Europeans, Northern Europeans, and Bantu-speaking South Africans [14]. NorEur = Northern European, SouEur = Southern European, Euro = Porto Alegre Euro-descendant, Afro = Porto Alegre Afro-descendant.
doi:10.1371/journal.pone.0046520.g001



Figure 2. Geographic location of the 3 Brazilian cities where the samples were collected.
doi:10.1371/journal.pone.0046520.g002

–13910C>T polymorphism is an underestimate of the LP phenotype.

Materials and Methods

Ethics Statement

All enrolled subjects provided their written informed consent to participate. The study protocol was approved by the ethics committees of the Federal University of Rio Grande do Sul, Federal University of Pará and of the Instituto Materno Infantil de Pernambuco, the three institutions that participated in blood sample collection.

Subjects

The study cohort consisted of 981 individuals recruited from the North, Northeast, and South regions of Brazil (Figure 2). All individuals from the Southern sample were selected at random at the Clinical Analysis Laboratory of the Pharmacy School of the Federal University of Rio Grande do Sul, Porto Alegre, among those who came from several city health centers for free routine blood determinations. This sample included 337 individuals of European ancestry and 182 African Brazilians. European and African ancestry were ascertained by visual inspection of skin color and morphological characteristics. These samples have been fully described in previous publications [38–40]. The Northeastern sample consisted of 262 healthy adolescents ascertained at Instituto Materno-Infantil de Pernambuco at Recife, the capital of the Brazilian Northeastern state of Pernambuco. The

characteristics of this sample were previously described [41]. The Northern sample consisted of 200 individuals from the city of Belém in the Brazilian Amazon region ascertained at the Federal University of Pará. We did not stratify by ethnicity the Northeastern and Northern populations because skin color is not very indicative of genomic ancestry in these populations [23,34]. The LP status of the individuals sampled was not evaluated.

Identification of SNPs in the LCT Enhancer Region

The –13910C>T (rs4988235) polymorphism was genotyped by PCR-RFLP as previously described [35]. In an attempt to identify other polymorphisms in the –14 kb region, a fragment of 427 bp was amplified with the MCM6i13 and LAC-CL2 primers [16] in all non-carriers of the –13910*T allele. PCR products were purified with *Exonuclease I* and Shrimp Alkaline Phosphatase enzymes and then sequenced at MACROGEN (Seoul, Republic of Korea) using the MCM6i13 primer. The –22018G>A (rs182549) polymorphism was genotyped by allelic discrimination using TaqMan assays in a real time PCR equipment (StepOne Plus, Applied Biosystems).

LCT Haplotypes

From the eleven polymorphisms described across the 70 Kb *LCT* gene [42], ten were genotyped in the present study to infer *LCT* haplotypes. Seven out of 10 polymorphisms investigated mapped between –1099 and –502 at the 5' flanking region and were identified by sequencing a fragment of 597 pb at MACROGEN as previously described [26]. The other three polymorphisms

666G>A (rs3754689), 5579T>C (rs2278544), and 6236TG>ΔΔ (rs10552864) reside within the *LCT* gene, and were genotyped by TaqMan assays. The haplotypes were designated according to the nomenclature previously described [14].

Statistical Analysis

Chromatograms were examined using CodonCode Aligner Software v.3.5.7. Allele frequencies were obtained directly by gene counting. Hardy-Weinberg equilibrium was tested by Chi-Square with the WINPEPI software [43]. This software was also used for allele frequency comparisons by heterogeneity Chi-Square test. *LCT* haplotypes were inferred using a Bayesian algorithm implemented in Phase v.2.1 software [44,45]. Wright's F_{ST} [46,47] was calculated using Arlequin v.3.0 [48]. D_A distance [49] matrix for the *LCT* haplotypes frequencies was generated with POPTREE software version 1. This matrix was used in the nonmetric Multidimensional Scaling analysis performed with SPSS v.18.

Supporting Information

Table S1 P-values from the Chi-square test of the -13910*T allele frequencies among the studied populations from Brazil and the data available in dsSNP-NCBI. (DOC)

Table S2 Distribution of the -13910 C>T and -22018 G>A haplotype on the *LCT* haplotypes in the Euro-

descendants from Porto Alegre city, Rio Grande do Sul State, Brazil.

(DOC)

Table S3 Distribution of the -13910 C>T and -22018 G>A haplotype on the *LCT* haplotypes in the Afro-descendants from Porto Alegre city, Rio Grande do Sul State, Brazil.

(DOC)

Table S4 Distribution of the -13910 C>T and -22018 G>A haplotype on the *LCT* haplotypes in the Belém city population, Pará State, Brazil.

(DOC)

Table S5 Distribution of the -13910 C>T and -22018 G>A haplotype on the *LCT* haplotypes in the Recife city population, Pernambuco State, Brazil.

(DOC)

Acknowledgments

The authors thank Everaldo de França for help in sample collection in Recife, and Ursula Silveira Matte for a -13910*T control sample.

Author Contributions

Conceived and designed the experiments: DCF MHH. Performed the experiments: DCF. Analyzed the data: DCF MHH. Contributed reagents/materials/analysis tools: AKRS SEBS MHH. Wrote the paper: DCF AKRS SEBS MHH.

References

- 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: 2705–2713.
- 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.
- Montgomery R, Büller H, Rings E, Grand R (1991) Lactose intolerance and the genetic regulation of intestinal lactase-phlorizin hydrolase. *FASEB J* 5: 2824–2832.
- Swallow DM (2003) Genetics of lactase persistence and lactose intolerance. *Annu Rev Genet* 37: 197–219.
- Lee MF, Krasinski SD (1998) Human adult-onset lactase decline: an update. *Nutr Rev* 56: 1–8.
- 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: 233–237.
- Kuokkanen M, Enattah N, 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: 647–652.
- Troelsen JT, Olsen J, Møller J, Sjöström H (2003) An upstream polymorphism associated with lactase persistence has increased enhancer activity. *Gastroenterology* 125: 1686–1694.
- Olds L, Sibley E (2003) Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a cis regulatory element. *Hum Mol Genet* 12: 2333–2340.
- Lewinsky R, Jensen T, Møller J, Stensballe A, Olsen J, et al. (2005) T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. *Hum Mol Genet* 14: 3945–3953.
- Enattah NS, Kuokkanen M, Forsblom C, Natak S, Oksanen A, et al. (2007) Correlation of intestinal disaccharidase activities with the C/T-13910 variant and age. *World J Gastroenterol* 13: 3508–3512.
- Kruse T, Bolund L, Grzeschik K, Ropers H, Sjöström H, et al. (1988) The human lactase-phlorizin hydrolase gene is located on chromosome 2. *FEBS Lett* 240: 123–126.
- Harvey C, Hollox E, 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: 215–223.
- Hollox E, Poulter M, Zvarik M, Ferak V, Krause A, et al. (2001) Lactase haplotype diversity in the Old World. *Am J Hum Genet* 68: 160–172.
- Mulcare C, Weale M, Jones A, Connell B, Zeitlyn D, et al. (2004) The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (*LCT*) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. *Am J Hum Genet* 74: 1102–1110.
- Ingram C, Elamin M, Mulcare C, Weale M, Tarekegn A, et al. (2007) A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Hum Genet* 120: 779–788.
- Tishkoff S, Reed F, Ranciaro A, Voight B, Babbitt C, et al. (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* 39: 31–40.
- Enattah NS, Jensen TG, Nielsen M, Lewinski R, Kuokkanen M, et al. (2008) Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am J Hum Genet* 82: 57–72.
- Ingram C, Raga T, Tarekegn A, Browning S, Elamin M, et al. (2009) Multiple rare variants as a cause of a common phenotype: several different lactase persistence associated alleles in a single ethnic group. *J Mol Evol* 69: 579–588.
- Torniaainen S, Parker MI, Holmberg V, Lahtela E, Dandara C, et al. (2009) Screening of variants for lactase persistence/non-persistence in populations from South Africa and Ghana. *BMC Genet* 10: 31.
- Callegari-Jacques SM, Grattapaglia D, Salzano FM, Salamoni SP, Crossetti SG, et al. (2003) Historical genetics: spatiotemporal analysis of the formation of the Brazilian population. *Am J Hum Biol* 15: 824–834.
- Salzano FM, Bortolini MC (2002) The evolution and genetics of Latin American Population. Cambridge, UK: Cambridge University Press.
- Pena SD, Di Pietro G, Fuchshuber-Moraes M, Genro JP, Hutz MH, et al. (2011) The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One* 6: e17063.
- Wang Y, Harvey CB, Hollox EJ, Phillips AD, Poulter M, et al. (1998) The genetically programmed down-regulation of lactase in children. *Gastroenterology* 114: 1230–1236.
- Coelho M, Luiselli D, Bertorelle G, Lopes A, Seixas S, et al. (2005) Microsatellite variation and evolution of human lactase persistence. *Hum Genet* 117: 329–339.
- Friedrich DC, Callegari-Jacques SM, Petzl-Erler ML, Tsuneto L, Salzano FM, et al. (2012) Stability or variation? Patterns of lactase gene and its enhancer region distributions in Brazilian Amerindians. *Am J Phys Anthropol* 147: 427–432.
- Poulter M, Hollox E, Harvey C, Mulcare C, Peuhkuri K, et al. (2003) The causal element for the lactase persistence/non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. *Ann Hum Genet* 67: 298–311.
- Tomás G, Seco L, Seixas S, Faustino P, Lavinha J, et al. (2002) The peopling of São Tomé (Gulf of Guinea): origins of slave settlers and admixture with the Portuguese. *Hum Biol* 74: 397–411.
- Enattah NS, Trudeau A, Pimenoff V, Maiuri L, Auricchio S, et al. (2007) Evidence of still-ongoing convergence evolution of the lactase persistence T-13910 alleles in humans. *Am J Hum Genet* 81: 615–625.

30. Jensen TG, Liebert A, Lewinsky R, Swallow DM, Olsen J, et al. (2011) The -14010°C variant associated with lactase persistence is located between an Oct-1 and HNF1 α binding site and increases lactase promoter activity. *Hum Genet* 130: 483–493.
31. Coelho M, Sequeira F, Luiselli D, Belez S, Rocha J (2009) On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evol Biol* 9: 80.
32. Gallego Romero I, Basu Mallick C, Liebert A, Crivellaro F, Chaubey G, et al. (2012) Herders of Indian and European cattle share their predominant allele for lactase persistence. *Mol Biol Evol* 29: 249–260.
33. Lember M, Torniaainen S, Kull M, Kallikorm R, Saadla P, et al. (2006) Lactase non-persistence and milk consumption in Estonia. *World J Gastroenterol* 12: 7329–7331.
34. Santos NP, Ribeiro-Rodrigues EM, Ribeiro-Dos-Santos AK, Pereira R, Gusmão L, et al. (2010) Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Hum Mutat* 31: 184–190.
35. Bulhões A, Goldani H, Oliveira F, Matte U, Mazzuca R, et al. (2007) Correlation between lactose absorption and the C/T-13910 and G/A-22018 mutations of the lactase-phlorizin hydrolase (LCT) gene in adult-type hypolactasia. *Braz J Med Biol Res* 40: 1441–1446.
36. Mattar R, Monteiro MoS, Villares CA, dos Santos AF, Carrilho FJ (2008) Single nucleotide polymorphism C/T(-13910), located upstream of the lactase gene, associated with adult-type hypolactasia: validation for clinical practice. *Clin Biochem* 41: 628–630.
37. Mattar R, Monteiro MoS, Silva JM, Carrilho FJ (2010) LCT-22018G>A single nucleotide polymorphism is a better predictor of adult-type hypolactasia/lactase persistence in Japanese-Brazilians than LCT-13910C>T. *Clinics (Sao Paulo)* 65: 1399.
38. Mattevi VS, Zembruški VM, Hutz MH (2002) Association analysis of genes involved in the leptin-signaling pathway with obesity in Brazil. *Int J Obes Relat Metab Disord* 26: 1179–1185.
39. Fiegenbaum M, de Andrade FM, Hutz MH (2007) Association between plasma lipid parameters and APOC3 genotypes in Brazilian subjects: effect of gender, smoking and APOE genotypes. *Clin Chim Acta* 380: 175–181.
40. De Andrade FM, Maluf SW, Schuch JB, Voigt F, Barros AC, et al. (2011) The influence of the S19W SNP of the APOA5 gene on triglyceride levels in southern Brazil: interactions with the APOE gene, sex and menopause status. *Nutr Metab Cardiovasc Dis* 21: 584–590.
41. De França E, Alves JG, Hutz MH (2004) Apolipoprotein E polymorphism and its association with serum lipid levels in Brazilian children. *Hum Biol* 76: 267–275.
42. 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.
43. Abramson JH (2004) WINPEPI (PEPI-for-Windows): computer programs for epidemiologists. *Epidemiol Perspect Innov* 1: 6.
44. Stephens M, Smith N, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68: 978–989.
45. Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73: 1162–1169.
46. Wright S (1969) *Evolution and the Genetics of Populations*. Chicago: University of Chicago Press.
47. Wright S (1978) *Evolution and the Genetics of Populations*. Chicago: University of Chicago Press.
48. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1: 47–50.
49. Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J Mol Evol* 19: 153–170.
50. Everitt BS (1992) *The analysis of contingency tables*. London: Chapman & Hall. 164 p.

Table S1. P-values from the Chi-square test of the -13910*T allele frequencies among the studied populations from Brazil and the data available online in dsSNP-NCBI.

	Porto Alegre		Belém	Recife
	European	African		
	ancestry	ancestry		
HapMap-CEU ^a	1x10 ⁻⁸	9x10 ⁻⁹	9.3x10 ⁻⁹	9.7x10 ⁻⁹
HapMap-ASW ^b	2.5x10 ⁻⁵	0.277	0.399	0.079
1000Genomes ^c	0.001	0.038	0.01	0.157
PDR90 ^d	0.712	0.015	5.6x10 ⁻³	0.048

^a Utah residents with Northern and Western European ancestry from the CEPH collection; -13910*T allele frequency: 0.73; number of samples: 180

^b Individuals with African ancestry in Southwest USA; -13910*T allele frequency: 0.143; number of samples: 90

^c 1000Genome phase 1 genotype data, released in the May 2011 dataset; -13910*T allele frequency: 0.233; number of samples: 1094

^d The NIH Polymorphism Discovery Resource; -13910*T allele frequency: 0.276; number of samples: 90

Table S2. Distribution of the -13910 C>T and -22018 G>A haplotype on the *LCT* haplotypes in the Euro-descendants from Porto Alegre city, Rio Grande do Sul State, Brazil.

<i>LCT</i> haplotype	CG	TA	CA	TG
A	0.157	0.276	0.022	0.005
B	0.187			0.007
C	0.176			
D	0.028			
E	0.060			
G	0.001			
I	0.005			
J		0.003		
K	0.003			0.001
M	0.001			
P	0.014			
Q	0.005			
S	0.007			
U	0.030			
W	0.001			
X	0.008			
c	0.001			
d	0.001			
p	0.001			

Total number of individuals = 321

Table S3. Distribution of the -13910 C>T and -22018 G>A haplotype on the *LCT* haplotypes in the Afro-descendants from Porto Alegre city, Rio Grande do Sul State, Brazil.

<i>LCT</i> haplotype	CG	TA	CA
A	0.157	0.181	0.005
B	0.127		
C	0.225		
D	0.011		
E	0.036		
G	0.003		
I	0.008		
J		0.003	
K	0.036		
M	0.003		
P	0.038		
Q	0.008		
S	0.044		
U	0.066		
X	0.030		
c	0.008		
h	0.003		
o	0.005		
q	0.003		

Total number of individuals = 182

Table S4. Distribution of the -13910 C>T and -22018 G>A haplotype on the *LCT* haplotypes in the Belém city population, Pará State, Brazil.

<i>LCT</i> haplotype	CG	TA	CA
A	0.202	0.175	0.003
B	0.147		
C	0.242		
D	0.025		
E	0.055		
G	0.005		
H	0.005		
I	0.003		
K	0.012		
M	0.003		
P	0.030		
Q	0.003		
S	0.017		
T	0.003		
U	0.045		
X	0.025		

Total number of individuals = 200

Table S5. Distribution of the -13910 C>T and -22018 G>A haplotype on the *LCT* haplotypes in the Recife city population, Pernambuco State, Brazil.

<i>LCT</i> haplotype	CG	TA	CA	TG
A	0.171	0.198	0.008	0.002
B	0.167			
C	0.205			
D	0.008			
E	0.029			
G	0.002			
H	0.002			
I	0.010			
J	0.004	0.006		0.002
K	0.025			
M	0.002			
P	0.023			
Q	0.012			
S	0.015			
T	0.002			
U	0.062			
X	0.027			
c	0.002			
g	0.010			
m	0.004			
o	0.002			

Total number of individuals = 258

Capítulo 4: Manuscrito intitulado “The lactase persistence genotype is a protective factor for the metabolic syndrome” em preparação.

The lactase persistence genotype is a protective factor for the metabolic syndrome

Deise C. Friedrich¹, Sidia M. Callegari-Jacques^{1,2}, and Mara H. Hutz¹.

¹ Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

² Departamento de Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Corresponding author:

Mara H. Hutz

Departamento de Genética, UFRGS

Caixa Postal 15053

91501-970, Porto Alegre, RS, Brazil

Tel: 55 51 3308 6720

Fax: 55 51 3308 7311

E-mail: mara.hutz@ufrgs.br

Abbreviated title: Metabolic syndrome and lactase persistence

Key terms: Metabolic syndrome, lactase persistence, hypolactasia, Brazil

Word count: 1139

Disclosure of Potential Conflict of Interest: The authors have no conflict of interest to declare

1 **Abstract**

2

3 1. Context. Metabolic syndrome (MetS) is defined as a pattern of metabolic disturbances, including
4 central obesity, insulin resistance and hyperglycemia, dyslipidemia, and hypertension. Milk has
5 been promoted as a healthy beverage that can improve the management of MetS. People presenting
6 the lactase persistence phenotype do not have symptoms when consume dairy products.

7 2. Objective. The aim of the study was to determine if there is an association between genetically
8 determined lactase persistence phenotype and MetS.

9 3. Design. This study is cross-sectional.

10 4. Setting. The population was randomly enrolled and it has European ancestry.

11 5. Participants. Participants were selected at random at the Clinical Analysis Laboratory of the
12 Pharmacy School of the Federal University of Rio Grande do Sul, Porto Alegre, Brazil, among
13 those who came from several city health centers for free routine blood determinations.

14 6. Main Outcome Measure. The hypothesis being tested was formulated after data collection.

15 7. Results. The lactase persistence genotype is protective to MetS, OR=0.47 (95% CI 0.245-0.899,
16 p=0.023).

17 8. Conclusions. Tolerant people have a lower odds ratio of presenting MetS than the intolerant
18 probably due to their possibility of drinking milk and then to be benefited by this beverage.

19

1 **Introduction**

2

3 The metabolic syndrome (MetS) has been defined as a cluster of metabolic risk factors, including
4 central obesity, hypertension, dyslipidemia, insulin resistance and hyperglycemia. It is predictive of
5 an increased risk of type-2 diabetes and cardiovascular disease (CVD) (1, 2).

6 The management of the MetS can have benefits with therapeutic changes in lifestyle as increasing
7 physical activity and healthy diet. Milk is considered a healthy beverage, so it can have beneficial
8 effects on MetS. In observational studies, dairy consumption was inversely associated with the
9 occurrence of one or more individual components of the MetS, i.e., a protective effect of milk
10 against the metabolic syndrome (3, 4), and also against high blood pressure (5). A possible
11 explanation for those results was that dairy intake could be associated with other healthy dietary
12 habits or lifestyles that promote health (4). Only one study reported that women who never drank
13 milk had lower insulin resistance scores, triglyceride levels and body mass index (BMI), higher
14 High-density Lipoprotein (HDL) cholesterol, and suffered less frequently from diabetes than those
15 who drank milk (4).

16 Milk has large amounts of lactose. The lactase enzyme that is secreted by small-intestinal
17 enterocytes hydrolyzes it. The gene that encodes lactase is the *LCT* (OMIM #603202) mapped in
18 2q21 (6). A C to T transition SNP (rs4988235) located -14 Kb upstream the *LCT* is responsible for
19 the lactase persistence (LP) phenotype in the European population. The CC genotype determines
20 the lactase non-persistence (LNP) also known as adult-type hypolactasia or lactose intolerance (7).
21 Tolerant individuals can consume milk and dairy products without symptoms. Other
22 polymorphisms in the same enhancer region or in its vicinity were also related to the LP phenotype
23 in populations from East Africa and Middle East (8-10).

24 We have recently published a study about the lactase gene in the Brazilian population. The
25 estimated LP phenotype (-13910 CT and TT genotypes) in individuals with European ancestry
26 from the southernmost Brazilian state was 51.4%. No other variant in the -14 Kb enhancer region
27 was identified by sequencing (11).

1 Considering milk benefits over metabolism we hypothesized if lactase persistent genotypes carriers
2 are protected against MetS.

3

4 **Materials and Methods**

5

6 The sample

7 From the 337 individuals enrolled in the previous published study (11), 3 were excluded because
8 there were missing data. The -13910 C>T genotypes frequencies are as follow: 48.8% CC, 43.4%
9 CT, and 7.8% TT. Descriptive characteristics and metabolic variables from the sample were
10 collected and they have been fully described in previous publications (12-14).

11

12 Definition of metabolic syndrome

13 We used the National Cholesterol Education Program - Adult Treatment Panel III (NCEP-ATP III)
14 to define the MetS (2). Several large-scale studies have confirmed the ability of the NCEP-ATP III
15 to predict both CVD and type-2 diabetes.

16

17 Data analysis

18 Differences between means of continuous variables were analyzed by Student *t* test and Mann-
19 Whitney U test was applied when there was a deviation from normal distribution. Chi-square
20 analysis was performed for categorical variables. A binary logistic regression was applied to
21 analyze the prevalence of MetS in the LP and LNP groups. Gender, age, and physical activity
22 (dichotomous variable, sedentary or not) were included in the model as confounders. The statistical
23 analyses were performed using the SPSS v. 18 package. Two sided P-values <0.05 were considered
24 for statistical significance.

25

26

27

1 **Results and Discussion**

2

3 The major finding reported herein is the lower risk of developing the metabolic syndrome (MetS)
4 for the LP (CT and TT genotypes) individuals. The prevalence of MetS was 19.6% in the LNP
5 group and 11.1% in the LP group. The odds ratio of the LP group for developing MetS was 0.47
6 (95% CI 0.245-0.899, $p=0.023$). This result is shown in Table 1, as well as the results for the
7 singular components of MetS. In the binary logistic regression model, adjustments for gender, age
8 and physical activity (categorical dichotomous variable) were applied and the age had a minimum
9 but significant effect ($OR=1.04$, $p<0.001$).

10 The causal relation between LP and MetS can be the possibility of the tolerant individuals drink
11 milk with no symptoms and then have benefits from this beverage. It has already been reported that
12 this group consumes significantly more milk than intolerant individuals but this higher
13 consumption is not linked to higher measures of body fat mass in children and adolescents (15).

14 There are many evidences that some milk components would have a direct benefit to insulin
15 sensitivity, weight, blood pressure, and cholesterol levels (4). Calcium would be one of it and many
16 possible mechanisms concerning its activity were suggested: a) it could play a critical role in the
17 regulation of energy metabolism; b) the promotion of increased faecal fat excretion by formation of
18 insoluble calcium-fatty acid soaps or by binding of bile acids that impair the formation of micelles
19 (4); c) it promotes blood pressure reduction (16); and d) it decreases the oxidative stress and the
20 expression of inflammatory cytokines present in the MetS (17). Milk protein-derived peptides are
21 benefic dairy component as well. They show hypotensive effects and they also improve the
22 bioavailability of other minerals and trace elements like magnesium, manganese, zinc, selenium
23 and iron (5).

24 But we need to consider that the inverse relation between dairy consumption and MetS may be
25 attributed to the healthy lifestyle associated with higher intakes of dairy because people who
26 consumed higher amounts of dairy also consumed higher amounts of fiber, fruit, vegetables, and
27 whole grains (18).

1 One limitation of our study is that we do not have information about the dietary pattern of the
2 participants, so we could not test if the LP group has higher milk intake than the LNP group. We
3 also do not have the blood pressure measure that can be included in the characterization of the
4 MetS. Maybe the MetS prevalence would be greater in both groups if the blood pressure had been
5 taken into account.

6 Contrary to our result, there are only two studies, one showing that milk avoidance is associated
7 with a reduced risk of insulin resistance and metabolic syndrome (19), and another showing that LP
8 subjects had significant higher frequency of MetS, and higher OR (OR=1.56), when compared to
9 LNP subjects (15).

10 Therapeutic lifestyle changes including healthy eating habits and weight loss significantly improve
11 most metabolic abnormalities of the MetS. This confirms that nutrition as an environmental factor
12 plays a key role in the development of the MetS (20). Considering the results of milk intake, one
13 change suggested is the milk consumption, especially the low-fat milk, as recommended by the
14 DASH (Dietary Approaches to Stop Hypertension) diet. However, the benefits versus harm of milk
15 intake for the intolerant people must be analyzed because for them the very uncomfortable
16 symptoms are prejudicial to their health, mainly the diarrhea.

17 In conclusion, creating subgroups of individuals with MetS regarding the lactase persistence
18 phenotype can be a valuable tool for the management of this syndrome during the nutrition
19 counseling.

20

21 **Acknowledgments**

22

23 None to declare.

24

25

26

27

1 **References**

2

3 1. **Després JP, Pérusse L** 2008 Genetic and nutritional determinants of the metabolic
4 syndrome: introduction. *J Nutrigenet Nutrigenomics* 1:97-99

5 2. **Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in**
6 **Adults** 2001 Executive Summary of The Third Report of The National Cholesterol
7 Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of
8 High Blood Cholesterol In Adults (Adult Treatment Panel III). *JAMA* 285:2486-2497

9 3. **Elwood PC, Pickering JE, Fehily AM** 2007 Milk and dairy consumption, diabetes and
10 the metabolic syndrome: the Caerphilly prospective study. *J Epidemiol Community Health*
11 61:695-698

12 4. **Pfeuffer M, Schrezenmeir J** 2007 Milk and the metabolic syndrome. *Obes Rev* 8:109-118

13 5. **FitzGerald RJ, Murray BA, Walsh DJ** 2004 Hypotensive peptides from milk proteins. *J*
14 *Nutr* 134:980S-988S

15 6. **Kruse T, Bolund L, Grzeschik K, Ropers H, Sjöström H, Norén O, Mantei N,**
16 **Semenza G** 1988 The human lactase-phlorizin hydrolase gene is located on chromosome
17 2. *FEBS Lett* 240:123-126

18 7. **Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Järvelä I** 2002
19 Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 30:233-237

20 8. **Ingram C, Elamin M, Mulcare C, Weale M, Tarekegn A, Raga T, Bekele E, Elamin F,**
21 **Thomas M, Bradman N, Swallow D** 2007 A novel polymorphism associated with lactose
22 tolerance in Africa: multiple causes for lactase persistence? *Hum Genet* 120:779-788

23 9. **Tishkoff S, Reed F, Ranciaro A, Voight B, Babbitt C, Silverman J, Powell K,**
24 **Mortensen H, Hirbo J, Osman M, Ibrahim M, Omar S, Lema G, Nyambo T, Ghorri J,**
25 **Bumpstead S, Pritchard J, Wray G, Deloukas P** 2007 Convergent adaptation of human
26 lactase persistence in Africa and Europe. *Nat Genet* 39:31-40

- 1 10. **Torniainen S, Parker MI, Holmberg V, Lahtela E, Dandara C, Jarvela I** 2009
2 Screening of variants for lactase persistence/non-persistence in populations from South
3 Africa and Ghana. *BMC Genet* 10:31
- 4 11. **Friedrich DC, Santos SE, Ribeiro-Dos-Santos AK, Hutz MH** 2012 Several different
5 lactase persistence associated alleles and high diversity of the lactase gene in the admixed
6 brazilian population. *PLoS One* 7:e46520
- 7 12. **Mattevi VS, Zembrzusi VM, Hutz MH** 2002 Association analysis of genes involved in
8 the leptin-signaling pathway with obesity in Brazil. *Int J Obes Relat Metab Disord*
9 26:1179-1185
- 10 13. **Fiegenbaum M, de Andrade FM, Hutz MH** 2007 Association between plasma lipid
11 parameters and APOC3 genotypes in Brazilian subjects: effect of gender, smoking and
12 APOE genotypes. *Clin Chim Acta* 380:175-181
- 13 14. **De Andrade FM, Maluf SW, Schuch JB, Voigt F, Barros AC, Lucatelli JF, Hutz MH**
14 2011 The influence of the S19W SNP of the APOA5 gene on triglyceride levels in
15 southern Brazil: interactions with the APOE gene, sex and menopause status. *Nutr Metab*
16 *Cardiovasc Dis* 21:584-590
- 17 15. **Almon R, Alvarez-Leon EE, Engfeldt P, Serra-Majem L, Magnuson A, Nilsson TK**
18 2010 Associations between lactase persistence and the metabolic syndrome in a cross-
19 sectional study in the Canary Islands. *Eur J Nutr* 49:141-146
- 20 16. **Hajjar IM, Grim CE, Kotchen TA** 2003 Dietary calcium lowers the age-related rise in
21 blood pressure in the United States: the NHANES III survey. *J Clin Hypertens*
22 (Greenwich) 5:122-126
- 23 17. **Zemel MB, Sun X** 2008 Dietary calcium and dairy products modulate oxidative and
24 inflammatory stress in mice and humans. *J Nutr* 138:1047-1052
- 25 18. **Azadbakht L, Mirmiran P, Esmailzadeh A, Azizi F** 2005 Dairy consumption is
26 inversely associated with the prevalence of the metabolic syndrome in Tehranian adults.
27 *Am J Clin Nutr* 82:523-530

- 1 19. **Lawlor DA, Ebrahim S, Timpson N, Davey Smith G** 2005 Avoiding milk is associated
2 with a reduced risk of insulin resistance and the metabolic syndrome: findings from the
3 British Women's Heart and Health Study. *Diabet Med* 22:808-811
- 4 20. **Levesque J, Lamarche B** 2008 The metabolic syndrome: definitions, prevalence and
5 management. *J Nutrigenet Nutrigenomics* 1:100-108

TABLE

Table 1. Descriptive characteristics (mean \pm standard deviation), metabolic variables (mean \pm standard deviation), and prevalence of metabolic syndrome grouped according to genotype.

Variables	Genotypes		P-value
	CC	CT + TT ^a	
Total number	163	171	
Age (years)	40.6 \pm 14.6	42.3 \pm 15.6	0.452 ^b
Gender (% of male)	47.9	46.8	0.913 ^b
Waist circumference (cm)	92.1 \pm 13.6	89.9 \pm 11.9	0.127 ^c
Cholesterol total	199.1 \pm 45.2	198.2 \pm 42.9	0.902 ^b
HDL- Cholesterol (mg/dL)	44.1 \pm 10.9	44.1 \pm 12.4	0.798 ^b
LDL- Cholesterol (mg/dL)	125 \pm 37.9	127.3 \pm 38	0.711 ^b
Triglycerides (mg/dL)	138.3 \pm 105.1	128.9 \pm 83.2	0.417 ^b
Glycaemia (mg/dL)	96.5 \pm 28	92.4 \pm 24.7	0.081 ^b
Physical activity (% of sedentary individuals)	66	61.2	0.364 ^d
Metabolic syndrome (%)	19.6	11.1	0.023 ^e

^a CT and TT genotypes were grouped because the presence of the -13910*T allele denotes the lactase persistence phenotype.

^b Mann-Whitney U test

^c *t*-test

^d Chi-square test

^e Binary Logistic Regression OR=0.47

Capítulo 5: DISCUSSÃO

Considerando-se que os resultados foram discutidos em detalhe nos capítulos correspondentes aos artigos, nesta discussão final serão abordados os resultados de forma ampla inserindo-os no contexto geral de conhecimentos gerados até o momento para o gene *LCT* e para os fenótipos da lactase.

É interessante ressaltar que a persistência da lactase tem despertado o interesse em diferentes campos da ciência, seja para estudos populacionais, evolução ou com ênfase médica.

5.1 Estudos populacionais e evolutivos

Os estudos populacionais visam compreender como os genes e seus alelos se distribuem na população no intuito de entender como esta é estruturada, sua história e as relações deles com as características de determinada população.

A variante da PL pode ser empregada na exploração da existência de estratificação de populações aparentemente homogêneas e essa informação pode ser útil para estudos de associação, de farmacogenética, etc.

O polimorfismo -13910C>T pode ser utilizado para a detecção da subdivisão da população visto que, a frequência do alelo T se dá na forma de gradiente em alguns países, apresentando frequências bastante divergentes de acordo com a região, como na Inglaterra (Smith *et al*, 2009), em Portugal (Manco *et al*, 2013), na Índia (Romero *et al*, 2012) ou então no continente europeu como um todo onde a frequência do alelo T decresce gradativamente de norte a sul e de oeste a leste (Ingram *et al*, 2009a). No Brasil, a distribuição do alelo T também não é homogênea, assim como a presença de outras variantes de PL, apresentando diferenças em frequência entre as regiões, como apresentado nos resultados do Capítulo 3. A maior frequência do alelo T é no Rio Grande do Sul e a menor em Belém.

Além da PL, os haplótipos do gene *LCT*, ou então alguns SNPs que são utilizados para derivar estes haplótipos, podem ser utilizados nos estudos populacionais pois apresentam distribuição e frequências características às populações (Hollox *et al*, 2001). Os dados apresentados nos Capítulos 2 e 3 são fontes de informações que podem auxiliar na utilização do *LCT* para os fins acima

citados. Por exemplo, o haplótipo E foi encontrado em frequência elevada nos ameríndios (Capítulo 2) em comparação com outras populações. Nas populações eurodescendentes e afrodescendentes do RS e miscigenadas do Norte e Nordeste, a frequência desse haplótipo foi de aproximadamente 6% (Capítulo 3), similar a frequência do sul da Europa (Hollox *et al*, 2001).

É importante ressaltar que as variantes -946A>G e -942C>G são monomórficas na população brasileira e o alelo derivado de -875G>A é extremamente raro ou ausente (Anexo 2). Esses dados devem ser levados em consideração quando os SNPs do gene *LCT* forem usados em estudos populacionais, visto que essas três variantes não serão muito informativas.

Com relação aos estudos evolutivos, a PL é bastante informativa para a história humana. A partir dela, são gerados indícios sobre hábitos, migrações, contato entre populações. Por exemplo, a ocorrência dos alelos -13910*T, -13915*G, -13907*G e -14010*C no mesmo grupo étnico pode significar contato passado entre populações que tomavam leite durante as migrações, através de práticas culturais compartilhadas (Ingram *et al*, 2009b). Outro exemplo são os resultados combinados do SNP -13910C>T e de análises do cromossomo Y na população de Camarões que sugerem uma história demográfica complexa para esta parte da África, que inclui ao menos uma grande introdução de genes de fora do continente, pois é provável que a presença do alelo T tenha ocorrido por introgressão (Mulcare *et al*, 2004).

5.2 Considerações sobre a saúde

A contribuição dos estudos das variantes genéticas relacionadas à PL para a saúde ainda é incerta pois se cogita a possibilidade de que os SNPs descritos não sejam causais, mas sim marcadores da PL. Além disso, os resultados dos estudos de associação da PL com outras condições são contraditórios. No Capítulo 4 relatamos que a presença do alelo -13910*T está associada ao menor risco do desenvolvimento de síndrome metabólica, já que o grupo de portadores deste alelo apresentou menor incidência desta condição. Por outro lado, os poucos estudos similares publicados apresentam os dados na direção oposta, de

portadores do alelo T terem um risco maior de desenvolver a síndrome metabólica (Lawlor *et al*, 2005; Almon *et al*, 2010).

Mais estudos sobre o consumo de leite e derivados e genótipos de PL devem ser realizados para se chegar a um denominador comum sobre o possível benefício ou ainda a ausência de influência destes alimentos sobre a síndrome metabólica, variáveis metabólicas, pressão arterial, obesidade, etc. Pois, se o leite realmente traz benefícios, deve-se ponderar se estes benefícios também existirão para os intolerantes, sendo que os sintomas causados pelo consumo podem ser mais prejudiciais do que os possíveis benefícios. Dessa forma, os futuros estudos sobre hábitos alimentares deveriam levar em consideração o fenótipo de intolerância a lactose, característica essa que pode ser um fator confundidor nos estudos sobre consumo de leite e derivados fazendo com que haja os resultados contraditórios presentes na literatura.

Um dado que pode auxiliar no entendimento do mecanismo de ação da PL sobre a obesidade é que indivíduos não persistentes possuem contagens diferentes de bifidobactérias e lactobacilos, que são consideradas bactérias benéficas promotoras de saúde, quando comparado aos persistentes (Corella *et al*, 2011). Além disso, já foi demonstrado que a lactose é capaz de modificar a microbiota intestinal aumentando a contagem fecal total de bifidobactérias e lactobacilos e diminuindo o número de *Bacteroides*, *Prevotella* e *Clostridium* (Francavilla *et al*, 2012). Também já foi demonstrado que obesos e não obesos apresentam diferenças na microbiota intestinal e estas estariam envolvidas nas diferenças de extração calórica da comida (Corella *et al*, 2011). Assim sendo, as diferenças na microbiota intestinal induzidas pelo fenótipo de NPL apresentariam um papel no desenvolvimento da obesidade. Entretanto, as evidências do papel protetor do leite sobre o risco de sobrepeso e obesidade são sugestivas mas não consistentes, tornando as conclusões difíceis (Louie *et al*, 2011).

Com relação a osteoporose, sendo esta influenciada pela quantidade de cálcio disponibilizado na dieta, seria interessante estudar se a PL na nossa população influencia o consumo de leite e se está relacionada ao desenvolvimento da osteoporose, assim como vem sendo estudado em diversos países (Enattah *et al*, 2004; Enattah *et al*, 2005; Laaksonen *et al*, 2009; Agueda *et al*, 2010).

5.3 O uso de -13910C>T como diagnóstico molecular

O oferecimento do diagnóstico molecular de tolerância à lactose é uma realidade mundial. No Brasil, já foram realizados dois estudos de validação da determinação do genótipo -13910C>T com relação ao fenótipo, no sudeste (Mattar *et al*, 2008) e no Rio Grande do Sul (Bulhões *et al*, 2007).

Nossos resultados apresentados no Capítulo 3 demonstram que é necessária a adequação das técnicas à população, pois quatro outras variantes, além da -13910C>T foram encontradas na população brasileira. Consequentemente, para as populações afrodescendentes e miscigenadas é recomendado o sequenciamento dos íntrons do *MCM6* aos invés de somente pesquisar a presença do polimorfismo -13910C>T. Para as populações brancas do Rio Grande do Sul, a detecção do alelo -13910*T pode ser utilizada como diagnóstico molecular da persistência da lactase, de acordo com os resultados do Capítulo 3, pois não foram encontradas outras variantes nessa população.

5.4 A importância dos estudos funcionais

O estudo funcional de novas variantes que vêm sendo descritas se faz necessário pois nem toda mutação que é encontrada na região *enhancer* do *LCT* apresenta funcionalidade.

A afirmação acima decorre dos resultados obtidos com o estudo funcional da mutação -13937G>A apresentados no Anexo 1. Brevemente, um fragmento de aproximadamente 200 pb da região *enhancer* foi clonado em um vetor no qual o gene da luciferase estava sob o controle da região promotora da lactase de ratos. A mutação em estudo foi gerada por mutagênese sítio dirigida. O vetor foi então transfectado nas células Caco-2. Os ensaios de luciferase que se seguiram mostraram que tanto o alelo ancestral quanto o derivado nesta posição apresentam a mesma taxa de indução da expressão do gene repórter. Além disso, a expressão proporcionada pelos dois alelos de -13937 foi similar a do alelo -13910*C, o de não persistência da lactase, ou seja, aparentemente esta mutação

não apresenta expressão diferencial, portanto, não estando envolvida no fenótipo persistência da lactase.

Entretanto, é importante considerar que estudos de correlação fenótipo-genótipo, que não foram abordados neste trabalho, também geram dados muito informativos e seriam de extrema valia à nossa população.

5.5 Causalidade da persistência da lactase

Embora a PL seja atribuída aos SNPs presentes nos íntrons do *MCM6* pela frequência dos alelos estarem associadas aos fenótipos da lactase e pelos estudos funcionais já realizados, a região regulatória do *LCT* continua incompletamente entendida. Não se pode afirmar que esses SNPs afetam diretamente a expressão da lactase *in vivo* ou se são simplesmente marcadores da PL que estão em desequilíbrio de ligação com outra variante funcional ainda não observada (Montgomery *et al*, 2007; Gerbault *et al*, 2011). Além disso, há os casos de indivíduos tolerantes que não apresentam nenhuma variante no *enhancer*, sugerindo a presença de mudanças genéticas adicionais fora dessa região ou de fatores modificadores do fenótipo PL (Tishkoff *et al*, 2007; Ingram *et al*, 2009b). Por esses motivos, é importante que se continue a pesquisar as possíveis causas da PL.

A contribuição para a dúvida sobre a causalidade da PL estar no *enhancer* foi reforçada por um estudo recente que demonstrou que um fragmento de 3,3 Kb correspondente a sequência 5' flanqueadora do *LCT* de um indivíduo persistente foi suficiente para promover a PL em camundongos, ou seja, esse fragmento que não contém nenhum SNP relacionado a PL promoveu a manutenção da expressão da lactase ao longo do tempo. Porém, para obter uma resposta mais conclusiva, o mesmo estudo deve ser realizado com a sequência flanqueadora de um indivíduo não persistente esperando-se que a expressão decline ao longo do tempo, a exemplo do fenótipo de NPL, o que não foi feito por estes autores (Baffour-Awuah *et al*, 2011).

Adicionalmente, uma característica dessa região cromossômica que poderia ser investigada é a presença de diversas ilhas CpG nos genes *LCT* e *MCM6*.

Talvez um controle epigenético também atue ao longo do desenvolvimento gerando os fenótipos da lactase. Um dado interessante relacionado é que as células Caco-2, que possuem o genótipo -13910TT, apresentam um estado de cromatina relaxado dessas ilhas nessa região, significando transcrição ativa (Epigenomics NCBI). Pensando em controle epigenético, a atuação da proteína Oct-1 no fenótipo da lactase talvez tenha relação com a sua atuação em diversos tecidos, pois esta age como um estabilizador de estados de indução ou de repressão, recrutando o complexo modelador de cromatina NuRD ou Jmjd1a (Shakya *et al*, 2011). Assim como também seria interessante que as sequências *Alu* presentes na região 5' flanqueadora do *LCT* somente em humanos fosse investigada quanto a uma possível relação com a PL.

5.6 Considerações finais

Além de tudo que foi citado anteriormente, o entendimento do funcionamento da lactase pode ser utilizado para outras finalidades, como por exemplo, a utilização desta como gene repórter para estudos *in vivo* de expressão por longos períodos. A lactase apresenta diversas vantagens no desempenho dessa função: não é tóxica quando expressa ectopicamente, não gera produtos metabolicamente ativos e como possui expressão restrita ao intestino delgado, pode ser usada como gene repórter em qualquer parte do corpo (Salehi *et al*, 2009).

Por fim, a prevalência da persistência da lactase deveria ser levada em consideração quando da elaboração de programas de intervenção nutricional e políticas públicas de combate a fome, tanto para crianças provenientes de grupos com alta frequência desta característica (que poderiam se beneficiar do consumo de leite e derivados), quanto em grupos com baixa frequência como, por exemplo, os indígenas para os quais o leite não seria o alimento mais adequado visto que, com base nos dados do Capítulo 2, eles possuem uma frequência baixa ou extremamente rara de alelos de PL.

REFERÊNCIAS

- Abreu S, Santos R, Moreira C, Santos PC, Vale S, Soares-Miranda L, Mota J and Moreira P (2012) Milk intake is inversely related to body mass index and body fat in girls. *Eur J Pediatr* 171(10):1467-1474.
- Agueda L, Urreiziti R, Bustamante M, Jurado S, Garcia-Giralt N, Díez-Pérez A, Nogués X et al (2010) Analysis of three functional polymorphisms in relation to osteoporosis phenotypes: replication in a Spanish cohort. *Calcif Tissue Int* 87(1): 14-24.
- Almon R, Álvarez-Leon EE, Engfeldt P, Serra-Majem L, Magnuson A and Nilsson TK (2010) Associations between lactase persistence and the metabolic syndrome in a cross-sectional study in the Canary Islands. *Eur J Nutr* 49:141-146.
- Anagnostou P, Battaggia C, Coia V, Capelli C, Fabbri C, Pettener D, Destro-Bisol G and Luiselli D (2009) Tracing the distribution and evolution of lactase persistence in Southern Europe through the study of the T₋₁₃₉₁₀ variant. *Am J Hum Biol* 21:217–219.
- Azadbakht L, Mirmiran P, Esmailzadeh A and Azizi F (2005) Dairy consumption is inversely associated with the prevalence of the metabolic syndrome in Tehranian adults. *Am J Clin Nutr* 82:523-530.
- Baffour-Awuah NY, Delemarre E, Fujiwara Y, Heijmans J, Boerwinkel D, Algra S, Fialkovich JJ et al (2011) Characterization of expression in mice of a transgene containing 3.3 Kb of the human Lactase-Phlorizin Hydrolase (LPH) 5' flanking sequence. *Dig Dis Sci* 56(1):59-69.
- Battochio APR, Santos AG e Coelho CAR (2003) Leite materno: considerações sobre nutrientes específicos e seus benefícios. *Rev Bras Nutr Clin* 18(3):136-141.
- Beja-Pereira A, Luikart G, England PR, Bradley DG, Jonn OC, Bertorelle G, Chamberlain AT et al (2005) Gene-culture coevolution between cattle milk protein genes and human lactase gene. *Nat Genet* 35(4):311-313.
- Berkey CS, Colditz GA, Rockett HR, Frazier AL and Willett WC (2009) Dairy consumption and female height growth: prospective cohort study. *Cancer Epidemiol Biomarkers Prev* 18:1881-1887.
- Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE and Hirschhorn JN (2004) Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* 74:1111-1120.

- Boll W, Wagner P and 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.
- Bosse T, van Wering HM, Gielen M, Dowling LN, Fialkovich JJ, Piaseckyj CM, Gonzalez FJ et al (2006) Hepatocyte nuclear factor-1 α is required for expression but dispensable for histone acetylation of the lactase-phlorizin hydrolase gene in vivo. *Am J Physiol Gastrointest Liver Physiol* 290:G1016-G1024.
- Brüssow H (2013) Nutrition, population growth and disease: a short history of lactose. *Environ Microbiol*, *In press*.
- Bulhões AC, Goldani HAS, Oliveira FS, Matte US, Mazzuca RB and Silveira TR (2007) Correlation between lactose absorption and the C/T-13910 and G/A-22018 mutations of the lactase-phlorizin hydrolase (LCT) gene in adult-type hypolactasia. *Braz J Med Biol Res* 40(11):1441-1446.
- Burger J, Kirchner M, Bramanti B, Haak W and Thomas MG (2007) Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *PNAS* 104:3736-3741.
- Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, Altshuler D, Ardlie KG and Hirschhorn JN (2005) Demonstrating stratification in a European American population. *Nat Genet* 37(8):686-872.
- Coelho M, Luiselli D, Bertorelle G, Lopes AI, Seixas S, Destro-Bisol G and Rocha J (2005) Microsatellite variation and evolution of human lactase persistence. *Hum Genet* 117:329-339.
- Coelho M, Sequeira F, Luiselli D, Beleza S and Rocha J (2009) On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evol Biol* 9:80.
- Corella D, Arregui M, Coltell O, Portolés O, Guillem-Sáiz P, Carrasco P, Sorlí JV et al (2011) Association of the LCT-13910C>T polymorphism with obesity and its modulation by dairy products in a Mediterranean population. *Obesity* 19(8):1707-1714.
- Dalmeijer GW, Struijk EA, van der Schouw YT, Soedamah-Muthu SS, Verschuren WM, Boer JM, Geleijnse JM and Beulens JW (2012) Dairy intake and coronary heart disease or stroke - A population-based cohort study. *Int J Cardiol*, *In press*.
- Després JP and Pérusse L (2008) Genetic and nutritional determinants of the metabolic syndrome: introduction. *J Nutrigenet Nutrigenomics* 1:97-99.
- Dudd SN and Evershed RP (1998) Direct demonstration of milk as an element of archaeological economics. *Science* 282:1478-1481.

- Eadala P, Matthews SB, Waud JP, Green JT and Campbell AK (2011) Association of lactose sensitivity with inflammatory bowel disease – demonstrated by analysis of genetic polymorphism, breath gases and symptoms. *Aliment Pharmacol Ther* 34:735-746.
- Elwood PC, Pickering JE and Fehily AM (2007) Milk and dairy consumption, diabetes and the metabolic syndrome: the Caerphilly prospective study. *J Epidemiol Community Health* 61:695-698.
- Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L and Järvelä I (2002) Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 30:233-237.
- Enattah NS, Välimäki VV, Välimäki MJ, Löyttyniemi E, Sahi T and Järvelä I (2004) Molecularly defined lactose malabsorption, peak bone mass and bone turnover rate in young Finish men. *Calcif Tissue Int* 75:488-493.
- Enattah NS, Sulkava R, Halonen P, Kontula K and Järvelä I (2005) Genetic variant of lactase-persistent C/T₋₁₃₉₁₀ is associated with bone fractures in very old age. *J Am Geriatr Soc* 53:79-82.
- Enattah NS, Trudeau A, Pimenoff V, Maiuri L, Auricchio S, Greco L, Rossi M, Lentze M et al (2007a) Evidence of still-ongoing convergence evolution of the lactase persistence T₋₁₃₉₁₀ alleles in humans. *Am J Hum Genet* 81:615-625.
- Enattah NS, Kuokkanen M, Forsblom C, Nataraj S, Oksanen A, Järvelä I, Peltonen L and Savilahti E (2007b) Correlation of intestinal disaccharidase activities with the C/T₋₁₃₉₁₀ variant and age. *World J Gastroenterol* 13(25):3508-3512.
- Enattah NS, Jensen TGK, Nielsen M, Lewinsky R, Kuokkanen M, Rasinpera H, El-Shanti H, Seo JK et al (2008) Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am J Hum Genet* 82:57-72.
- Fang R, Santiago NA, Olds LC and Sibley E (2000) The homeodomain protein Cdx-2 regulates lactase gene promoter activity during enterocyte differentiation. *Gastroenterology* 118:115-127.
- Fang R, Olds LC, Santiago NA and Sibley E (2001) GATA family transcription factors activate lactase gene promoter in intestinal Caco-2 cells. *Am J Physiol* 280:G58-G67.
- Fang R, Olds LC and Sibley E (2006) Spatio-temporal patterns of intestine-specific transcription factor expression during postnatal mouse gut development. *Gene Expr Patterns* 6(4):426-432.
- Feldman MW and Cavalli-Sforza LL (1989) On the theory of evolution under genetic and cultural transmission with application to the lactose absorption

- problem. In: Feldman MW (ed) *Mathematical evolutionary theory*. Princeton University Press, Princeton, NJ pp 145-173.
- 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-G324.
- FitzGerald RJ, Murray BA and Walsh DJ (2004) Hypotensive peptides from milk proteins. *J Nutr* 134:980S–988S.
- FrancaVilla R, Calasso M, Calace L, Siragusa S, Ndagijimana M, Vernocchi P, Brunetti L, Mancino G et al (2012) Effect of lactose on gut microbiota and metabolome of infants with cow's milk allergy. *Pediatr Allergy Immunol* 23(5):420-427.
- Gerbault P, Moret C, Currat M and Sanchez-Mazas A (2009) Impact of selection and demography on the diffusion of lactase persistence. *PLoS ONE* 4(7): e6369.
- Gerbault P, Liebert A, Itan Y, Powell A, Currat M, Burger J, Swallow DM and Thomas MG (2011) Evolution of lactase persistence: an example of human niche construction. *Phil Trans R Soc B* 366:863-877.
- Giovannucci E, Pollak M, Liu Y, Platz EA, Majeed N, Rimm EB and Willett WC (2003) Nutritional predictors of insulin-like growth factor I and their relationships to cancer in men. *Cancer Epidemiol Biomarkers Prev* 12:84-89.
- Grand RJ, Montgomery RK, Chitkara DK and Hirschhorn JN (2003) Changing genes; losing lactase. *Gut* 52:617-619.
- Grundmann SA, Stratmann E, Brehler R, Luger TA and Ständer S (2011) Lactase deficiency: a potential novel aetiological factor in chronic pruritus of unknown origin. *Acta Derm Venereol* 91:698-703.
- Gugatschka M, Dobnig H, Fahrleitner-Pammer, Pietschmann P, Kudlacek S, Strele A and Obermayer-Pietsch B (2005) Molecularly-defined lactose malabsorption, milk consumption and anthropometric differences in adult males. *Q J Med* 98:857-863.
- Hartmann S, Lacorn M and Steinhart H (1998) Natural occurrence of steroid hormones in food. *Food Chem* 62(1):7-20.
- Harvey CB, Hollox EJ, Poulter M, Wang Y, Rossi M, Auricchio S, Iqbal TH, Cooper BT et al (1998) Lactase haplotype frequencies in Caucasians: association with the lactase persistence/non-persistence polymorphism. *Ann Hum Genet* 62:215-223.
- Hauri HP, Sander B and Naim H (1994) Induction of lactase biosynthesis in the human intestinal epithelial cell line Caco-2. *Eur J Biochem* 219:539-546.

- He T, Venema K, Priebe MG, Welling GW, Brummer RJM and Vonk RJ (2008) The role of colonic metabolism in lactose intolerance. *Eur J Clin Invest* 38(8):541-547.
- Hollox EJ, Poulter M, Wang Y, Krause A and 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 and Swallow DM (2001) Lactase haplotype diversity in the old world. *Am J Hum Genet* 68:160-172.
- Hoppe C, Molgaard C, Juul A, Michaelsen KF (2004) High intakes of skimmed milk, but not meat, increase serum IGF-I and IGFBP-3 in eight-years-old boys. *Eur J Nutr* 58:1211-1216.
- Imtiaz F, Savilahti E, Sarnesto A, Trabzuni D, Al-Kahtani K, Kagevi I, Rashed MS, Meyer BF and Järvelä I (2007) The T/G -13915 variant upstream of the lactase gene (LCT) is the founder allele of lactase persistence in an urban Saudi population. *J Med Genet* 44(10): e89.
- Ingram CJE, Elamin MF, Mulcare CA, Waelw ME, Tarekegn A, Raga TO, Bekele E, Elamin FM et al (2007) A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Hum Genet* 120:779-788.
- Ingram CJE, Mulcare CA, Itan Y, Thomas MG and Swallow DM (2009a) Lactose digestion and the evolutionary genetics of lactase persistence. *Hum Genet* 124:579-591.
- Ingram CJE, Raga TO, Tarekegn A, Browning SL, Elamin MF, Bekele E, Thomas MG et al (2009b) Multiple rare variants as a cause of a common phenotype: several different lactase persistence associated alleles in a single ethnic group. *J Mol Evol* 69:579-588.
- Ingram CJE and Swallow DM (2009) Lactose Malabsorption. In: *Advanced Dairy Chemistry*. Vol 3. pp 203-229.
- Itan Y, Powell A, Beaumont MA, Burger J and Thomas MG (2009) The origins of lactase persistence in Europe. *PLoS Comput Biol* 5(8): e1000491.
- Jensen TGK, Liebert A, Lewinsky R, Swallow DM, Olsen J and Troelsen JT (2011) The -14010°C variant associated with lactase persistence is located between an Oct-1 and HNF1 α binding site and increases lactase promoter activity. *Hum Genet* 130(4):483-493.

- Kettunen J, Silander K, Saarela O, Amin N, Müller M, Timpson N, Surakka I et al (2010) European lactase persistence genotype shows evidence of association with increase in body mass index. *Hum Mol Genet* 19(6):1129-1136.
- Khabarova Y, Torniaainen S, Nurmi H, Järvelä I, Isokoski M and Mattila K (2009) Prevalence of lactase persistent/non-persistent genotypes and milk consumption in a young population in north-west Russia. *World J Gastroenterol* 15(15):1849-1853.
- Kolars JC, Levitt MD, Aouji M, Savaiano DA (1984) Yogurt – an autodigestive source of lactose. *N Engl J Med* 310:1-3.
- Krasinski SD, Van Wering HM, Tannemaat MR and Grand RJ (2001) Differential activation of intestinal gene promoters: functional interactions between GATA-5 and HNF-1 alpha. *Am J Physiol* 281:G69-G84.
- Krawczyk M, Wolska M, Schwartz S, Gruenhagen F, Terjung B, Portincasa P, Sauerbruch T and Lammert F (2008) Concordance of genetic and breath tests for lactose intolerance in a tertiary referral centre. *J Gastrointest Liver Dis* 17(2):135-139.
- Kruse TA, Bolund L, Grzeschik K-H, Ropers HH, Sijstrijm H, Norh O, Mantei N and Semenza G (1988) The human lactase-phlorizin hydrolase gene is located on chromosome 2. *FEBS Letters* 240(1,2):123-126.
- Kuchay RAH, Anwar M, Thapa BR, Mahmood A and Mahmood S (2013) Correlation of G/A -22018 single nucleotide polymorphism with lactase activity and its usefulness in improving the diagnosis of adult-type hypolactasia among North Indian children. *Genes Nutr* 8:145-151.
- Kuokkanen M, Enattah NS, Oksanen A, Savilahti E, Orpana A and Järvelä I (2003) Transcriptional regulation of the lactase-phlorizin hydrolase gene by polymorphisms associated adult-type hypolactasia. *Gut* 52:647-625.
- Kuokkanen M, Butzow R, Rasinperä H, Medrek K, Nilbert M, Malander S, Lubinski J and Järvelä I (2005) Lactase persistence and ovarian carcinoma risk in Finland, Poland and Sweden. *Int J Cancer* 117:90-94.
- Laaksonen MML, Impivaara O, Sievänen H, Viikari JSA, Lehtimäki TJ, Lamberg-Allardt CJE, Kärkkäinen MUM et al (2009) Associations of genetic lactase non-persistence and sex with bone loss in young adulthood. *Bone* 44:1003-1009.
- Lawlor DA, Ebrahim S, Timpson N and Smith GD (2005) Avoiding milk is associated with a reduced risk of insulin resistance and the metabolic syndrome: findings from the British Women's Heart and Health Study. *Diabet Med* 22:808–811.

- Lee MF and Krasinski SD (1998) Human adult-onset lactase decline: an update. *Nutr Rev* 56:1-8.
- Lee SY, Wang Z, Lin CK, Contag CH, Olds LC, Cooper AD and Sibley E (2002) Regulation of intestine-specific spatiotemporal expression by the rat lactase promoter. *J Biol Chem* 277(15):13099-13105.
- Lehtimäki T, Hemminki J, Rontu R, Mikkilä V, Räsänen L, Laaksonen M, Hutri-Kähönen N et al (2006) The effects of adult-type hypolactasia on body height growth and dietary calcium intake from childhood into young adulthood: A 21-year follow-up study - The Cardiovascular Risk in Young Finns Study. *Pediatrics* 118:1553-1559.
- Lewinsky RH, Jensen TGK, Moller J, Stensballe A, Olsen J and Troelsen JT (2005) T₋₁₃₉₁₀ DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity *in vitro*. *Hum Mol Genet* 14(24):3945-3953.
- Lokki AI, Järvelä I, Israelsson E, Maiga B, Troye-Blomberg M, Dolo A, Doumbo OK et al (2011) Lactase persistence genotypes and malaria susceptibility in Fulani of Mali. *Malar J* 10:9.
- Louie JC, Flood VM, Hector DJ, Rangan AM and Gill TP (2011) Dairy consumption and overweight and obesity: a systematic review of prospective cohort studies. *Obes Rev* 12(7):e582-92.
- Malek AJ, Klimentidis YC, Kell KP and Fernández JR (2013) Associations of the lactase persistence allele and lactose intake with body composition among multiethnic children. *Genes Nutr*, *In press*.
- Malmström H, Linderholm A, Lidén K, Storå J, Molnar P, Holmlund G, Jakobsson M and Götherström A (2010) High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. *BMC Evol Biol* 10:89.
- Manco L, Pires S, Lopes AI, Figueiredo I, Albuquerque D, Alvarez M, Rocha J and Abade A (2013) Distribution of the -13910C>T polymorphism in the general population of Portugal and in subjects with gastrointestinal complaints associated with milk consumption. *Ann Hum Biol* 40(2):205-208.
- Mantei N, Villa M, Enzler T, Wacker H, Boll W, James P, Hunziker W and 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(9):2705-2713.
- Mattar R, Monteiro MS, Villares CA, Santos AF and Carrilho FJ (2008) Single nucleotide polymorphism C/T-13910, located upstream of the lactase gene, associated with adult-type hypolactasia: Validation for clinical practice. *Clin Biochem* 41:628-630.

- Mattar R, Monteiro MS, Villares CA, Santos AF, Silva JMK and Carrilho FJ (2009) Frequency of LCT -13910C>T single nucleotide polymorphism associated with adult-type hypolactasia/lactase persistence among Brazilians of different ethnic groups. *Nutr J* 8:46.
- Matthews SB and Campbell AK (2000a) When sugar is not so sweet. *Lancet* 355:1309.
- Matthews SB and Campbell AK (2000b) Neuromuscular symptoms associated with lactose intolerance. *Lancet* 356:511.
- Matthews SB, Waud JP and Roberts AG (2005) Systemic lactose intolerance: a new perspective on an old problem. *Prostgrad Med J* 81:167-173.
- Meloni GF, Colombo C, La Vecchia C, Ruggiu G, Mannazzu MC, Ambrosini G and Cherchi PL (1999) Lactose absorption in patients with ovarian cancer. *Am J Epidemiol* 150(2):183-186.
- Meloni GF, Colombo C, La Vecchia C, Pacifico A, Tomasi P, Ogana A, Marinaro AM and Meloni T (2001) High prevalence of lactose absorbers in Northern Sardinian patients with type 1 and type 2 diabetes mellitus. *Am J Clin Nutr* 73:582-585.
- Mirmiran P, Esmailzadeh A and Azizi F (2005) Dairy consumption and body mass index: an inverse relationship. *Int J Obes* 29:115-121.
- Montgomery RK, Büller HA, Rings EHHM and Grand RJ (1991) Lactose intolerance and the genetic regulation of the intestinal lactase-phlorizin hydrolase. *FASEB J* 5:2824-2832.
- Montgomery RK, Krasinski SD, Hirschhorn JN and Grand RJ (2007) Lactose and lactase – Who is lactose intolerant and why? *J Pediatr Gastroenterol Nutr* 45:S131-S137.
- Morales E, Azocar L, Maul X, Perez C, Chianale J and Miquel JF (2011) The European lactase persistence genotype determines the lactase persistence state and correlates with gastrointestinal symptoms in the Hispanic and Amerindian Chilean population: a case-control and population-based study. *BMJ Open* 1:e000125.
- Mulcare CA, Weale ME, Jones AL, Connell B, Zeitlyn D, Tarekgen A, Swallow DM, Bradman N and Thomas MG (2004) The T allele of a single –nucleotide polymorphism 13.9 kb upstream of the lactase gene (*LCT*) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. *Am J Hum Genet* 74:1102-1110.
- Nelson DL and Cox MM (2002) *Lehninger: Princípios de bioquímica*. 3ª edição. Sarvier Editora, São Paulo, pp 232; 424–427.

- Olds LC and Sibley E (2003) Lactase persistence DNA variant enhances lactase promoter activity *in vitro*: functional role as a *cis* regulatory element. *Hum Mol Genet* 12(18):2333-2340.
- Olds LC, Ahn JK and Sibley E (2011) -13915*G DNA polymorphism associated with lactase persistence in Africa interacts with Oct-1. *Hum Genet* 129:111-113.
- Peng M-S, He J-D, Zhu C-L, Wu S-F, Jin J-Q and Zhang Y-P (2012) Lactase persistence may have an independent origin in Tibetan populations from Tibet, China. *J Hum Genet* 57:394-397.
- Pereira MA, Jacobs DR, Horn LV, Slattery ML, Kartashov AL and Ludwig DS (2002) Dairy consumption, obesity, and the insulin resistance syndrome in young adults. *JAMA* 287(16):2081-2089.
- Pfeuffer M and Schrezenmeir J (2006) Milk and the metabolic syndrome. *Obesity Reviews* 8:109-118.
- Potter J, Ho M-W, Bolton H, Furth AJ, Swallow DM and Griffiths B (1985) Human lactase and molecular basis of lactase persistence. *Biochem Genet* 23:423-439.
- Poulter M, Hollox E, Harvey CB, Mulcare C, Peuhkuri K, Kajander K *et al* (2003) The causal element for the lactase persistence / non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. *Ann Hum Genet* 67:298-311.
- Pretto FM, Silveira TR, Menegaz V and Oliveira J (2002) Má absorção de lactose em crianças e adolescentes: diagnóstico através do teste do hidrogênio expirado com o leite de vaca como substrato. *J Pediatr (Rio J)* 78(3):213-218.
- Rajpathak SN, Rimm EB, Rosner B, Willett WC and Hu FB (2006) Calcium and dairy intakes in relation to long-term weight gain in US men. *Am J Clin Nutr* 83:559-566.
- Rasinerä H, Forsblom C, Enattah NS, Halonen P, Salo K, Victorzon M, Mecklin J-P *et al* (2005) The C/C₋₁₃₉₁₀ genotype of adult-type hypolactasia is associated with an increased risk of colorectal cancer in Finnish population. *Gut* 54:643-647.
- Romero IG, Mallick CB, Liebert A, Crivellaro F, Chaubey G, Itan Y, Metspalu M *et al.* (2012) Herders of Indian and European cattle share their predominant allele for lactase persistence. *Mol Biol Evol* 29(1):249-260.
- Salehi S, Eckley L, Sawyer GJ, Zhang X, Dong X, Freund J-N and Fabre JW (2009) Intestinal lactase as an autologous β -Galactosidase reporter gene for *in vivo* gene expression studies. *Hum Gene Ther* 20:21-30.

- Shakya A, Kang J, Chumley J, Williams MA and Tantin D (2011) Oct1 is a switchable, bipotential stabilizer of repressed and inducible transcriptional states. *J Biol Chem* 286:450-459.
- Smith GD, Lawlor DA, Timpson NJ, Baban J, Kiessling M, Day INM and Ebrahim S (2009) Lactase persistence-related genetic variant: population substructure and health outcomes. *Eur J Hum Genet* 17:357-367.
- Snijder MB, van Dam RM, Stehouwer CDA, Hiddink GJ, Heine RJ and Dekker JM (2008) A prospective study of dairy consumption in relation to changes in metabolic risk factors: The Hoorn Study. *Obesity* 16:706-709.
- Stancliffe RA, Thorpe T and Zemel MB (2011) Dairy attenuates oxidative and inflammatory stress in metabolic syndrome. *Am J Clin Nutr* 94:422-430.
- Tag CG, Schiffers MC, Mohnen M, Gressner AM, Weiskirchen R (2007) A novel proximal -13914G>A base replacement in the vicinity of the common -13910C/T lactase gene variation results in an atypical LightCycler melting curve in testing with the MutaREAL lactase test. *Clin Chem* 53:146-148.
- Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM et al (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* 39:31-40.
- Torniaainen S, Hedelin M, Autio V, Rasinperä H, Balter KA, Klint A, Bellocco R et al (2007) Lactase persistence, dietary intake of milk, and the risk for prostate cancer in Sweden and Finland. *Cancer Epidemiol Biomarkers Prev* 16(5):956-961.
- Torniaainen S, Parker MI, Holmberg V, Lahtela E, Dandara C and Järvelä I (2009) Screening of variants for lactase persistence/non-persistence in populations from South Africa and Ghana. *BMC Genetics* 10:31.
- Torres EM, Prieto LLV, Camacho JLV and Torregroza DAV (2011) Diagnosis of adult-type hypolactasia/lactase persistence: genotyping of single nucleotide polymorphism (SNP C/T₋₁₃₉₁₀) is not consistent with breath test in Colombian Caribbean population. *Arq Gastroenterol* 49(1):5-8.
- Travis RC, Appleby PN, Siddiq A, Allen NE, Kaaks R, Canzian F, Feller S et al (2013) Genetic variation in the lactase gene, dairy product intake and risk for prostate cancer in the European Prospective Investigation into Cancer and Nutrition. *Int J Cancer* 132:1901-1910.
- Treudler R, Tebbe B, Steinhoff M and Orfanos CE (2002) Familial aquagenic urticaria associated with familial lactose intolerance. *J Am Acad Dermatol* 47:611-613.

- Troelsen JT, Olsen J, Norén O and Sjöström H (1992) A novel intestinal trans-factor (NF-LPH1) interacts with the lactase-phlorizin hydrolase promoter and co-varies with enzymatic activity. *J Biol Chem* 267:20407-20411.
- Troelsen JT, Mitchelmore C, Spodsberg N, Jensen AM, Norén O and Sjöström H (1997) Regulation of lactase-phlorizin hydrolase gene expression by the caudal-related homeodomain protein Cdx-2. *Biochem J* 322:833-838.
- Troelsen JT, Olsen J, Moller J and Sjöström H (2003) An upstream polymorphism associates with lactase persistence has increased enhancer activity. *Gastroenterology* 125:1686-1694.
- Van Wering HM, Huibregtse IL, Van der Zwan SM, De Bie MS, Dowling LN, Grand RJ and Krasinski SD (2002) Physical interaction between GATA-5 and HNF-1 alpha results in synergistic activation of the human lactase-phlorizin hydrolase promoter. *J Biol Chem* 277:27659-27667.
- Van Wering HM, Bosse T, Musters A, de Jong E, de Jong N, Esch CEH, Boudreau F *et al* (2004) Complex regulation of the lactase-phlorizin hydrolase promoter by GATA-4. *Am J Physiol Gastrointest Physiol* 287:G899-G909.
- Wang Y, Harvey CB, Pratt WS, Sams VR, Sarnar M, Rossi M, Auricchio S and Swallow DM (1995) The lactase persistence / non-persistence polymorphism is controlled by a cis-acting element. *Hum Molec Genet* 4:657-662.
- Wang Y, Harvey CB, Hollox EJ, Phillips AD, Poulter M, Clay P, Walker-Smith JA and Swallow DM (1998) The genetically programmed down-regulation of lactase in children. *Gastroenterology* 114:1230-1236.
- Wang Z, Fang R, Olds LC and Sibley E (2004) Transcriptional regulation of the lactase-phlorizin hydrolase promoter by PDX-1. *Am J Physiol Gastrointest Liver Physiol* 287:G555-G561.
- Zemel MB (2003) Role of dietary calcium and dairy products in modulating adiposity. *Lipids* 38:139-146.

REFERÊNCIAS ELETRÔNICAS

Epigenomics/NCBI, <http://www.ncbi.nlm.nih.gov/epigenomics> (14 de maio de 2013).

Entrez database, <http://www.ncbi.nlm.nih.gov/Entrez/> (abril de 2013).

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/omim> (28 de abril de 2013).

Wikipedia – A enciclopédia livre,
http://upload.wikimedia.org/wikipedia/commons/6/6b/Lactose_etc.png (30 de abril
de 2013).

ANEXOS

Anexo 1: Análise funcional da variante -13937G>A em células Caco-2.

Conforme o artigo constante no Capítulo 3 desta tese, além do polimorfismo -13910C>T, foram identificados na população brasileira as mutações -13779G>C, -13937G>A, -14010G>C e -14011C>T. O estudo funcional da variante -14010G>C foi publicado em 2011 por Jensen e colaboradores. Dados de estudo funcional dos polimorfismos -13779G>C e -14011C>T foram apresentados pelo grupo da Dra Dallas Swallow, da University College London, na forma de pôster no 12th *International Congress of Human Genetics*, Montreal, 2011. Assim, dentre as variantes encontradas na população brasileira, somente a -13937G>A não havia sido estudada. Então, após concessão de bolsa doutorado sanduíche do Programa Ciência sem Fronteiras - CNPq, a análise funcional desta mutação foi desenvolvida no laboratório de Gastroenterologia Pediátrica na Stanford University sob supervisão de Eric Sibley, MD, PhD. O objetivo deste projeto foi determinar se -13937G>A possui a capacidade de dirigir a expressão diferencial de um gene repórter numa construção transfectada na linhagem celular humana Caco-2, o que daria indícios de associação com o fenótipo de persistência da lactase.

Materiais e métodos

a) Mutagênese sítio dirigida

Para gerar a mutação de interesse nos plasmídeos p3K-13910C direto e p3K-13910C reverso o kit GENEART® Site-Directed Mutagenesis System (Life Technologies) foi utilizado.

Os plasmídeos p3K-13910C direto e reverso (Fang *et al*, 2000) são construções contendo um fragmento de 3 Kb da região promotora do gene da lactase de ratos clonado no vetor pGL3 (Promega) onde esta região promotora está controlando o gene da luciferase de vaga-lume (*firefly luciferase*) (Figura A1). *Upstream* ao promotor, foi adicionada a região *enhancer* do gene da lactase humana, um fragmento de 218 pb, das posições -14017 a -13800, que foi clonado no sentido direto e no reverso.

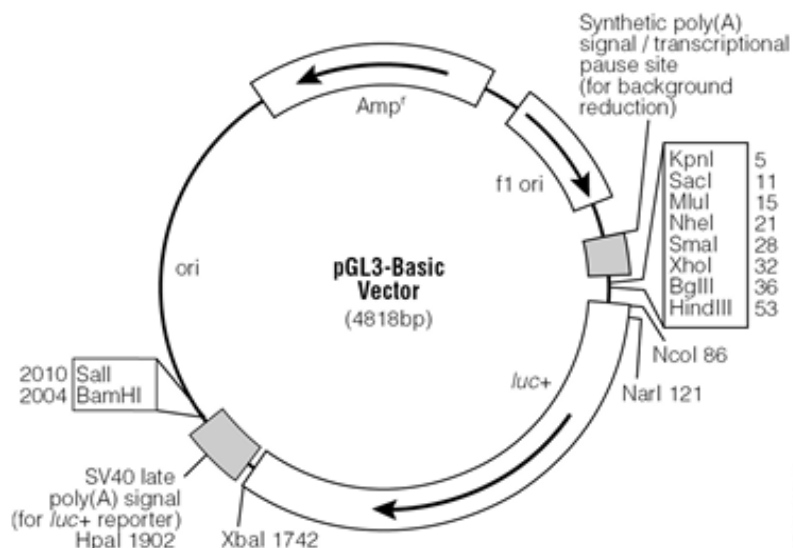


Figura A1: Plasmídeo pGL3 usado nos estudos de expressão de regiões controladoras de mamíferos. Imagem retirada de www.promega.com.

Para a mutagênese, os seguintes *primers* direto e reverso (base mutante sublinhada) foram utilizados:

MutagenesisF: ACTAGTAGGCCTCTGCACTGGCAATACAGATAA

MutagenesisR: TTATCTGTATTGCCAGTGCAGAGGCCTACTAGT

A sequência de passos (Figura A2) da mutagênese sítio dirigida foi realizada de acordo com o protocolo do kit e foi iniciada com uma reação de PCR, utilizando-se os plasmídeos p3K-13910C direto e p3K-13910C reverso como molde e a enzima AccuPrime™ Pfx DNA Polymerase (Life Technologies), uma Taq-polimerase de alta fidelidade. A reação está descrita na Tabela A1.

Tabela A1: Reação de PCR da mutagênese sítio dirigida.

	Volume	Concentração final
10X AccuPrime Pfx reaction mix	5 µL	1X
10X Enhancer	5 µL	1X
Mix de primers	1,5 µL	0,3 µM cada
DNA plasmidial (40 ng/µL)	0,5 µL	20 ng
DNA metilase (4 U/µL)	1 µL	4 U

25X SAM	2 μ L	1X
AccuPrime Pfx (2,5 U/ μ L)	0,4 μ L	1 U
Água	34,6 μ L	

A Tabela A2 descreve o programa utilizado na reação de PCR, no termociclador Veriti 96 well Thermal Cycller (Applied Biosystems).

Tabela A2: Programa utilizado no termociclador para a reação de mutagênese

Temperatura	Duração	Número de ciclos
37°C	20 min	1
94°C	2 min	1
94°C	20 seg	
57°C	30 se g	18
68°C	4 min	
68°C	5 min	1
4°C	∞	

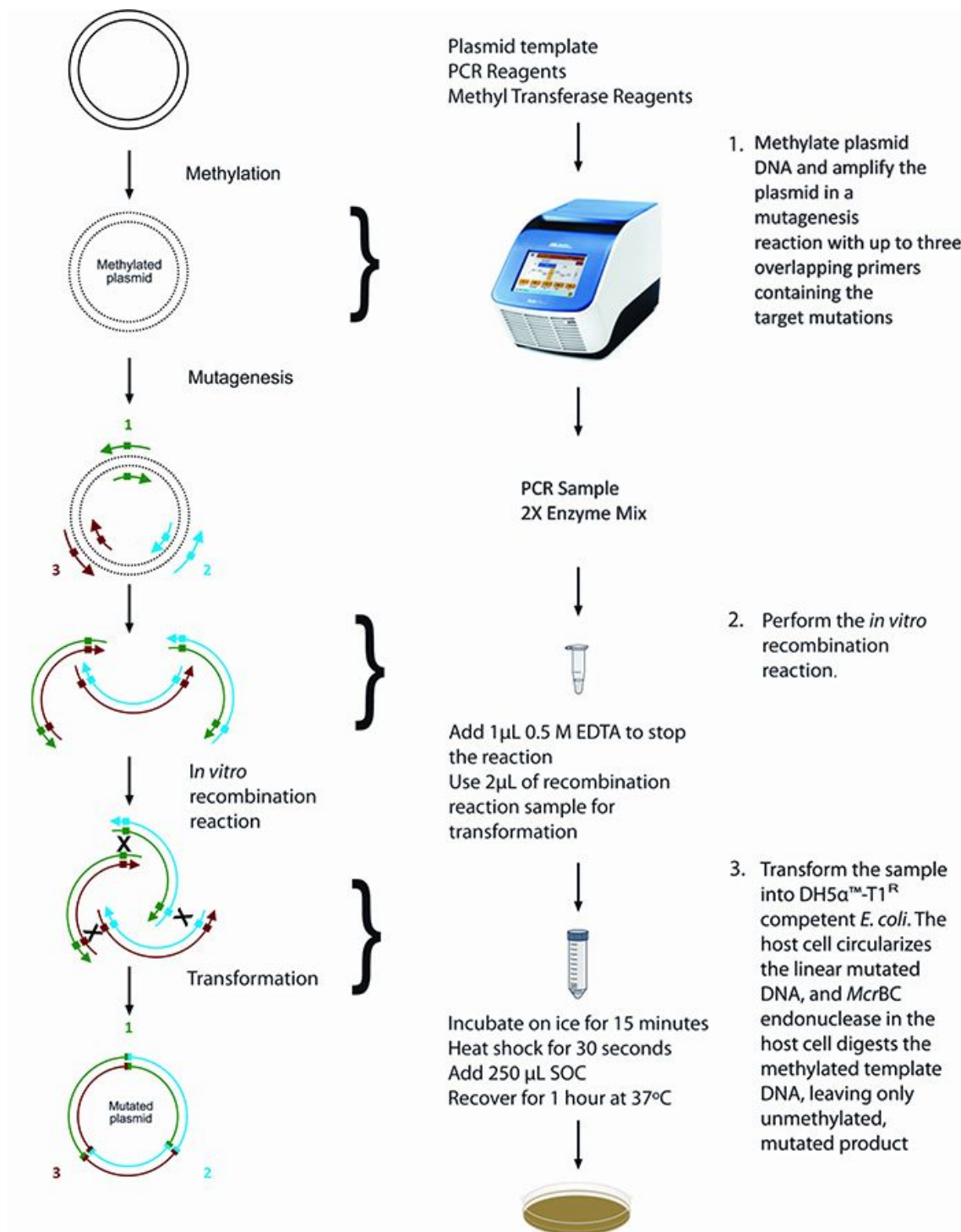


Figura A2: Esquema dos passos da técnica de mutagênese sítio dirigida. Imagem extraída de www.invitrogen.com.

O resultado da reação foi visualizado em gel de agarose 0,8%. A etapa que se seguiu foi a reação de recombinação *in vitro*, conforme a Tabela A3.

Tabela A3: Reação de recombinação.

	Volume	Concentração final
5X tampão de reação	4 µL	1X
Água	10 µL	
Produto da reação de PCR	4 µL	
10X Enzymer mix	2 µL	1X

A reação de recombinação foi incubada em temperatura ambiente por 10 minutos. Após, 1 µL de EDTA 0,5M foi adicionado para parar a reação, que em seguida foi colocada no gelo.

O passo seguinte foi a transformação nas células *E. coli* (One Shot® MAX Efficiency® DH5α™-T1) fornecidas com o kit e de acordo com as instruções do fabricante. Resumidamente, 2 µL da reação de recombinação foi adicionado em 50 µL de células descongeladas. A mistura foi incubada no gelo por 12 minutos, após foi realizado um choque térmico de 42°C por 30 segundos e, novamente, foi incubado no gelo por 2 minutos. Após, foi adicionado meio SOC e incubado por uma hora a 37°C com agitação. A transformação foi transferida para placas de cultura com meio ágar LB contendo 100 µg/mL de ampicilina. As placas foram incubadas de 16 a 20 horas a 37°C.

b) Confirmação da mutação gerada

Algumas colônias isoladas foram selecionadas para serem sequenciadas. Para tanto, as colônias foram inoculadas em meio líquido LB contendo 100 µg/mL de ampicilina, para após serem isolados os plasmídeos através do kit QIAprep® Spin Miniprep Kit (QIAGEN), conforme o protocolo. A qualidade e a concentração do DNA foram analisados pelo equipamento Glomax Multi Detection System (Promega).

Os plasmídeos foram sequenciados pelo serviço especializado da universidade, a PAN Facility. O *primer* utilizado para sequenciamento foi o RVprimer3 (Promega). Este primer anela no plasmídeo pGL3, perto do sítio de clonagem.

Os resultados do sequenciamento foram analisados no programa CodonCode Aligner.

c) Preparação dos plasmídeos para a transfecção nas células Caco-2

Os plasmídeos contendo a mutação confirmada, os selvagens e os plasmídeos utilizados como controles foram transformados por choque térmico em células competentes *E. coli* DH5 α (Invitrogen) para obtenção da quantidade necessária de DNA plasmidial para a transfecção. Após a transformação, uma colônia isolada foi inoculada em meio LB líquido. O DNA plasmidial foi isolado através do kit Purelink™ HiPure Plasmid Filter Maxiprep Kit (Invitrogen) e Purelink™ HiPure Precipitation Module (Invitrogen).

d) Transfecção transiente

As células Caco-2 foram cultivadas em meio DMEM (*Dulbecco's modified Eagle's medium*) modificado pela adição de 10% de soro fetal bovino, 1% de Penicilina/Estreptomicina/Glutamina e 1% de amino ácidos não essenciais. Ao atingir aproximadamente 80% de confluência, as placas de cultura com 6 poços de 35 mm foram preparadas para serem usadas na transfecção.

As placas foram semeadas com 2×10^5 células, em meio DMEM modificado, 48 horas antes da transfecção, 3 poços para cada construção a ser testada.

No dia da transfecção, uma mistura de DNA consistindo de 0,4 pmol da construção *enhancer*-promotor-repórter e 10 ng de pRL-CMV (Promega) foi preparada para cada transfecção. pRL-CMV é um vetor que contém o gene da luciferase de *Renilla reniformis*, que é cotransfectado para ser utilizado como controle interno.

As misturas de DNA foram transfectadas nas células (50 a 70% confluentes) com auxílio de lipofectamina (LipofectAMINE™, Life Technologies), de acordo com as instruções do fabricante.

48 horas após a transfecção, as células foram colhidas e a atividade de luciferase medida utilizando o kit Dual-Luciferase™ Reporter Assay System (Promega), no leitor Monolight 3010. Duas leituras de 10 segundos foram feitas, a

primeira detectando a luciferase da construção *enhancer*-promotor-repórter e a segunda detectando a luciferase do vetor pRL-CMV.

A atividade da luciferase da construção em teste foi normalizada pela atividade do controle interno pRL-CMV.

e) Plasmídeos utilizados na transfecção e ensaio de luciferase

Foram utilizados os plasmídeos contendo a mutação de interesse, com o fragmento humano clonado no sentido direto e reverso, que foram chamados de MutF e MutR, respectivamente. Além dos mutantes, os seguintes plasmídeos foram utilizados como controle:

- pGL3: plasmídeo sem o promotor e sem a região *enhancer*, gera o controle da emissão da luz de fundo, do próprio plasmídeo.
- gLac3K: vetor contendo a região promotora de 3 Kb de ratos
- p3K-13910C direto: construção *enhancer*-promotor-repórter sendo que a região *enhancer* foi clonada no sentido direto e representa o alelo -13910*C
- p3K-13910C reverso: construção *enhancer*-promotor-repórter sendo que a região *enhancer* foi clonada no sentido reverso e representa o alelo -13910*C
- p3K-13910T direto: construção *enhancer*-promotor-repórter sendo que a região *enhancer* foi clonada no sentido direto e representa o alelo -13910*T
- p3K-13910T reverso: construção *enhancer*-promotor-repórter sendo que a região *enhancer* foi clonada no sentido reverso e representa o alelo -13910*T

Resultados

Os plasmídeos contendo a região *enhancer* humana clonada no sentido direto e reverso foram utilizados na mutagênese sítio dirigida para a geração da mutação -13937G>A. A mutação foi confirmada por sequenciamento. Os plasmídeos foram transfectados, juntamente com controles, nas células humanas Caco-2.

Como pode ser visto na Figura A3, a expressão do mutante direto é similar à do alelo -13910*C, menor que do alelo -13910*T, dessa forma não dando indícios de

que esta mutação esteja envolvida na persistência da lactase. Os resultados para os fragmentos no sentido reverso foram similares.

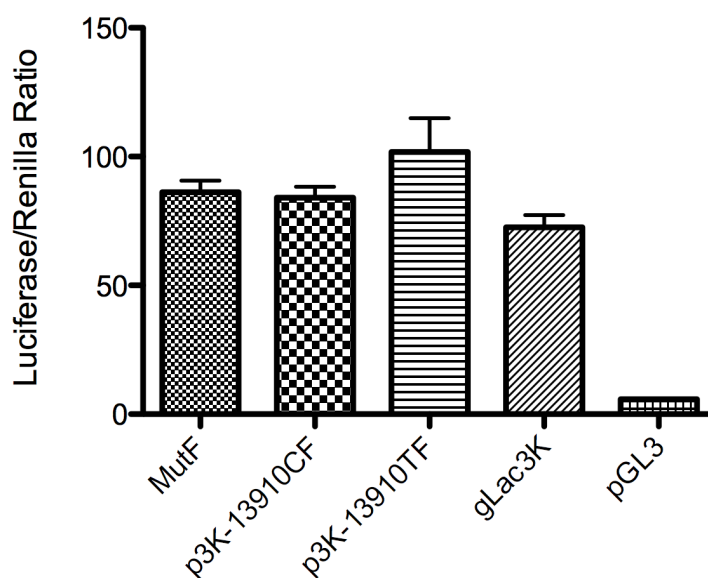


Figura A3: Histograma da expressão da luciferase da construção *enhancer*-promotor-repórter normalizada pela expressão da luciferase de *Renilla*. O plasmídeo contendo a mutação -13937*A apresenta os mesmos níveis de expressão do que o plasmídeo contendo o genótipo de não persistência da lactase em sentido direto (p3K-13910CF).

Referências do Anexo 1

Fang R, Santiago NA, Olds LC and Sibley E (2000) The homeodomain protein Cdx2 regulates lactase gene promoter activity during enterocyte differentiation. *Gastroenterology* 118:115–127.

Jensen TG, Liebert A, Lewinsky R, Swallow DM, Olsen J and Troelsen JT (2011) The -14010*C variant associated with lactase persistence is located between an Oct-1 and HNF1 α binding site and increases lactase promoter activity. *Hum Genet* 130(4):483-493.

Anexo 2: Tabela de frequências (%) dos alelos derivados dos polimorfismos do gene *LCT* nas populações brasileiras estudadas.

	Guarani- Ñandeva	Guarani- Kaiowá	Kaingang	Xavante	Belém	Recife	Negros RS	Branços RS
-958T	6	11,3	9,7	10	17,6	18,8	14,8	22,7
-946G	0	0	0	0	0	0	0	0
-942G	0	0	0	0	0	0	0	0
-942del	6,9	11,9	6,9	7,5	8,8	12,2	14,6	4,4
-875A	0,8	0	0	0	2,5	0,7	1	2,8
-678G	48,3	39,3	51,4	72,3	30,2	24	28,3	23,5
-552/9A ₉	45,7	56,6	38	16,5	47,4	49,8	49,2	51,1
666A	13,5	23,3	16,7	18,8	29,7	33,8	34,6	29,2
5579C	51,7	60,7	42,4	14,3	46,2	45,8	42,3	53,6
6236del	6,9	11,3	11,8	11,4	22,7	22,3	22,8	24,9