

Guilherme Debortoli

**Elaboração de um painel de marcadores de ancestralidade  
para estudos caso-controle e organização das informações em banco  
de dados local**

Dissertação submetida ao Programa de Pós-Graduação em Biologia Celular e do Desenvolvimento da Universidade Federal de Santa Catarina, como requisito parcial para a obtenção do Grau de Mestre em Biologia Celular e do Desenvolvimento.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Ilíada Rainha de Souza

Coorientadora: Dr<sup>a</sup>. Andrea Rita Marrero

Florianópolis  
2013

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Debortoli, Guilherme

Elaboração de um painel de marcadores de ancestralidade para estudos caso-controle e organização das informações em banco de dados local / Guilherme Debortoli ; orientadora, Iliáda Rainha de Souza ; co-orientadora, Andrea Rita Marrero. - Florianópolis, SC, 2013.

112 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências Biológicas. Programa de Pós-Graduação em Biologia Celular e do Desenvolvimento.

Inclui referências

1. Biologia Celular e do Desenvolvimento. 2. Biologia Celular e do Desenvolvimento. 3. Psoríase. 4. Marcadores Informativos de Ancestralidade. 5. Banco de Dados Local. I. Souza, Iliáda Rainha de. II. Marrero, Andrea Rita. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Biologia Celular e do Desenvolvimento. IV. Título.





## AGRADECIMENTOS

Primeiramente gostaria de agradecer a Universidade Federal de Santa Catarina pelo suporte. Obrigado a CAPES pelo auxílio financeiro concedido através da bolsa.

A Dra. Ilíada Rainha de Souza pela orientação, confiança, e pelos ensinamentos dos quais recordarei sempre com carinho. Lembro-me da primeira conversa que tivemos e posteriormente a sensação de felicidade ao me concederes a carta de aceite para orientação. Novamente deixo aqui meu obrigado e meus profundos sentimentos de respeito a uma pessoa que muito contribuiu para minha formação profissional e pessoal.

Um agradecimento mais do que especial a Dra Andrea Rita Marrero pela coorientação e por ter me ensinado valores que ultrapassam o meio acadêmico/profissional. És um exemplo a ser seguido em todos os sentidos e, sinceramente me faltam palavras para descrever o sentimento de carinho e respeito que tenho e com certeza sempre terei por você. Saiba que se hoje realizo este sonho, é porque tive a quem me espelhar. Mais uma vez, muito obrigado do fundo do meu coração!

A Dra. Yara Costa Netto Muniz por contribuir para minha formação acadêmica, pelo conselhos e conversas. Obrigado por me ensinar a ser sempre positivo e enfrentar a vida como ela é. É difícil encontrar palavras suficientes que contemplem tudo o que eu aprendi ao conviver contigo nestes dois anos. Muito obrigado!

Agradeço aos companheiros do LAPOGE com os quais convivi neste tempo: Leili, Amanda, Gabriela, Maria Luíza, Larissa, Emiliana, Mariáh, Bibiana, Ticiane, Renan, Emily, Bruna. Sucesso!

Agradecimentos ao Günther pela ajuda nas genotipagens e por todo o esforço em me ajudar a concluir esta dissertação! Desejo-lhe muito sucesso! Obrigado!

Agradeço a Bibiana por ter realizado as coletas dos pacientes de Psoríase, estendendo este agradecimento a todos os envolvidos no Hospital Universitário da UFSC pela ajuda.

Obrigado as pessoas que doaram suas amostras para que este trabalho fosse realizado. Do contrário nada disso seria possível!

Obrigado a Dra. Angélica Francesca Maris, por me deixar utilizar o termociclador de seu laboratório quando o termociclador do LAPOGE se “recusou” a amplificar minhas amostras.

Aos Professores da Pós Graduação em Biologia Celular e do Desenvolvimento pelas disciplinas muito bem ministradas e pelos

conhecimentos passados que muito contribuíram para minha formação!  
Obrigado!

Obrigado ao Dr. Aguinaldo Luiz Simões do Programa de Pós Graduação em Genética da Faculdade de Medicina de Ribeirão Preto - FMRP - USP, por ceder alíquotas de primers para a padronização destes no LAPOGE.

Agradeço ao Dr. Paulo Roberto PetersenHofmann por me proporcionar o prazer de assistir as suas aulas de genética clássica na turma de biologia da UFSC. Aulas essas, que contribuíram e muito na minha formação.

A Msc. Dalva Marques, por me ensinar genética durante a graduação e por ter me apresentado a Professora Ilíada.

Obrigado a Dra. Maíra Ribeiro Rodrigues por ter renovado meu interesse em bioinformática e banco de dados!

Um agradecimento mais do que especial para o meus pais e irmãos (Jorge Dimas Debortoli, Lioni da Silveira Debortoli, Fabiano Debortoli e Rodrigo Debortoli), que sempre me apoiaram em minhas decisões! Este sonho torna-se realidade porque vocês sempre estiveram ao meu lado! Amo vocês!

Esta dissertação jamais teria sido finalizada sem ajuda de meu irmão Fabiano Debortoli e seus conselhos e ajuda com o banco de dados! Horas e horas do seu fim de semana me ajudando e sempre muito bem disposto! Você é um exemplo de dedicação e para mim sempre será a pessoa mais inteligente do mundo! Amo você!

Não posso deixar de agradecer a pessoas que muito estimo (Gercino Bertol, Maria de Lourdes Lourenço Bertol, Bruna Cristina Bertol e Priscila Paula Bertol). Vocês são uma família para mim! Obrigado pelo apoio incondicional!!!

Obrigado a minha companheira Bruna Cristina Bertol, que me apoia em todas as minhas decisões. Você sempre acreditou em mim quando eu mesmo não acreditava e sempre deseja o melhor para mim! Estamos a quase seis anos juntos e meu amor por você cresce a cada dia que passa! Muito obrigado por seres está pessoa maravilhosa e me fazeres seguir em frente sempre! Amo você!

Jamais poderia deixar de agradecer a Marlene Martins da Cunha e sua família, por me acolherem em sua casa ao longo destes dois anos em Florianópolis! Muito obrigado!

Obrigado aos amigos do grupo de biologia molecular da Universidade da Região de Joinville - UNIVILLE. Vocês contribuíram para minha formação de cientista e, a cada visita minha ao laboratório, minhas expectativas referentes a carreira que escolhi se renovam!

Agradeço em especial aos membros da banca, Dr<sup>a</sup> Yara Costa Netto Muniz, Dr<sup>a</sup> Renata Toscano Simões, Dr Erick da Cruz Castelli e DrGeison de Souza Izídio, por cordialmente aceitarem corrigir esta dissertação. Considerações que com certeza serão muito bem-vindas para a melhoria deste trabalho.

Obrigado a todos que de certa forma contribuíram para a realização deste trabalho



“O êxito da vida não se mede pelo caminho que você conquistou,  
mas sim pelas dificuldades que superou no caminho.”  
Abraham Lincoln

Dedico a meus amados pais Jorge e Lioni  
Aos meus queridos irmãos Fabiano e Rodrigo  
Para minha amada Bruna

"Para estar junto, não é preciso estar perto. Mas sim, do lado de dentro"  
(Leonardo Da Vinci)



## RESUMO

DEBORTOLI, G. **Elaboração de um painel de marcadores de ancestralidade para estudos caso-controle e organização das informações em banco de dados local.** 2013. 112p. Dissertação de Mestrado – Universidade Federal de Santa Catarina, Florianópolis, SC.

A população brasileira é produto da miscigenação entre três principais grupos étnicos: Ameríndios (que já habitavam estas terras à época do descobrimento), Europeus (principalmente portugueses) e Africanos (trazidos de forma forçada como mão de obra escrava). Estimar as diferentes proporções de ancestralidade em populações miscigenadas é de extrema importância para estudos genéticos que buscam encontrar alelos de suscetibilidade a doenças, uma vez que a não correção da estratificação populacional pode ocasionar desvios levando a associações espúrias. Para o controle da estratificação populacional, a utilização de marcadores informativos de ancestralidade tem sido amplamente utilizada por apresentar um alto diferencial de frequência entre populações de regiões geograficamente distintas. O objetivo deste estudo foi genotipar quatro marcadores informativos de ancestralidade adequados para amostras da população brasileira visando estender esta abordagem para entender a estruturação populacional em estudos de caso-controle, utilizando como modelo a doença psoríase. Adicional ao principal objetivo, um banco de dados biológico (BDB) foi diagramado e implementado, para auxiliar na organização e armazenamento de informações a respeito das amostras aqui utilizadas. As frequências dos quatro marcadores (AT3, Sb19.3, APO e PV92) foram estimadas em amostras de pacientes (PSR) diagnosticados com psoríase (n=100) e controles saudáveis (PSC) sem histórico de doença (n=100, PSC) coletadas na cidade de Florianópolis - SC. Os perfis genéticos foram obtidos através de PCR convencional e eletroforese em gel de agarose. As análises estatísticas empregaram programas tais como GENEPOP, GDA e ADMIX 3. O BDB proposto foi desenvolvido e estruturado em sistema de gestão de bases de dados relacionais MySQL, e a interface gráfica foi desenvolvida em linguagem PHP e HTML. Por meio de análises moleculares e estatísticas, os resultados obtidos evidenciaram que os dois grupos de indivíduos aqui estudados (PSR e PSC) apresentaram semelhanças quanto a sua ancestralidade genética, não sendo observado grandes diferenças populacionais. As estimativas de mistura revelaram que as duas populações analisadas são tri-híbridas, com alta preponderância do componente europeu, seguido de africano e ameríndio respectivamente. Apesar dos poucos marcadores utilizados

foi possível discriminar a contribuição genética dos grupos ancestrais indicando que mesmo um número reduzido de marcadores pode ser suficiente, desde que apropriadamente selecionados, para responder uma pergunta específica. O BDB construído e diagramado neste trabalho recebeu o nome de LAPOGEDb (LAPOGE Database - [www.lapoge.sites.ufsc](http://www.lapoge.sites.ufsc)) e, mostrou-se uma ferramenta eficaz para o armazenamento estruturado de informações que variam entre dados epidemiológicos e genéticos referentes a amostras utilizadas ao longo de 19 anos de funcionamento do Laboratório de Polimorfismos Genéticos da Universidade Federal de Santa Catarina (LAPOGE-UFSC).

**Palavras-chave:**AIMs. Estratificação Populacional. Epidemiologia Molecular.

## ABSTRACT

DEBORTOLI, G. **Elaboration of a panel of ancestry-informative markers for case-control studies and organization of information in a local database.**2013. 112p. Dissertação de Mestrado – Universidade Federal de Santa Catarina, Florianópolis, SC.

The Brazilian population is the product of interbreeding between the three main ethnic groups: Amerindians (who inhabited these lands during the discovery), Europeans (mainly Portuguese) and Africans (brought forcefully as slave labor). Estimating the different proportions of ancestry in admixed populations is of utmost importance for genetic studies that seek to find disease susceptibility alleles, since no correction of population stratification, can cause deviations leading to spurious associations. To control the population stratification, the use of ancestry informative markers have been widely used for its feature of presenting a high frequency differential between geographically distinct populations. The aim of this study was to genotype four ancestry informative markers suitable for samples of the Brazilian population in order to extend this approach to understand the population structure in case-control studies, using as a model the psoriasis disease. Additional to the main objective, a biological database (BD) was diagrammed and implemented through programming languages, to assist in organizing and storing information about the samples used here. The frequencies of the four markers (AT3, Sb19.3, APO and PV92) were estimated in samples of patients (PSR) diagnosed with psoriasis (n = 100) and healthy controls (PSC) with no history of disease (n = 100, PSC) collected in the city of Florianópolis - SC. The genetic profiles were obtained by standard PCR and visualization with agarose gel and electrophoresis. The statistical analysis employed programs such as GENEPOP, GDA and ADMIX 3. The proposed BD was developed and structured in MySQL relational database management system, while the general user interface (GUI) was developed in PHP and HTML. The results obtained from the molecular analysis and statistics, showed that the two groups studied here (PSR and PSC) had some similarities in their genetic ancestry, not observing large population differences. Estimates of mixture revealed that the two populations are tri-hybrid, with high preponderance of European component, followed by African and Amerindian respectively. Despite the few markers used it was able to discriminate the genetic contribution of ancestral groups indicating, that even a small number of markers may be sufficient, when appropriately selected, to answer a specific question. The BD built and

diagrammed in this work was named LAPOGEdb (LAPOGE Database - [www.lapoge.sites.ufsc](http://www.lapoge.sites.ufsc)) and proved to be an effective tool for storing structured information, ranging from genetic and epidemiological data, related to the samples used throughout 19 years of operation of the Laboratory of Genetic Polymorphisms of the Federal University of Santa Catarina (UFSC-LAPOGE).

**Keywords:** AIMS. Population Stratification. Molecular Epidemiology.

## LISTA DE FIGURAS

- Figura 1: Estrutura de um Banco de Dados. A estrutura de um BD é dividida em três partes: A primeira consiste de um sistema de gerenciamento de banco de dados; a segunda de um local (servidor), onde são armazenados os dados; a terceira é a interface gráfica para o usuário do banco realizar as consultas e gerar relatórios das pesquisas realizada no banco de dados. Adaptada de: STEIN, 2003.....<sup>26</sup>
- Figura 2: Gel de agarose corado com GelRed para visualização de fragmentos de Inserção Alu. O gel digitalizado ilustra a presença de uma inserção Alu homozigota (raia 4), inserção/deleção (raias 1, 2 e 3), deleção homozigota (raias 6 e 7).....<sup>38</sup>
- Figura 3: Gel de agarose corado com GelRed para visualização de fragmentos de Indel. O gel digitalizado ilustra a presença de uma inserção homozigota (raia 3), inserção/deleção (raias 1 e 5) e deleção homozigota (raias 2, 4 e 6).....<sup>38</sup>
- Figura 4: Especificações para o servidor da UFSC para gerenciamento de Banco de Dados.....<sup>44</sup>
- Figura 5: Página inicial do site LAPOGE com o link no menu à direita da página (destacado em vermelho).....<sup>49</sup>
- Figura 6: Página inicial do LAPOGEdb. Links e campo de Identificação para usuário e senha.....<sup>50</sup>
- Figura 7: Página do LAPOGEdb com duas opções: uma para o cadastro de novos questionários e outra opção para a edição de formulários.....<sup>51</sup>

Figura 8: Formulário de Dados Pessoais do LAPOGEDb. Caixa de seleção para o tipo de formulário a ser cadastrado.....	52
Figura 9: Formulário de Dados Familiais do LAPOGEDb. Em destaque (cor verde e marrom), o novo código de identificação gerado automaticamente pelo banco e o código antigo, respectivamente.....	53
Figura 10: Formulário de Dados Familiais do LAPOGEDb. Caixas de seleção para histórico de doenças na família.....	54
Figura 11: Formulário de Dados Genéticos do LAPOGEDb. Caixas de seleção de marcadores e seus possíveis genótipos.....	55
Figura 12: Formulário de Informações do LAPOGEDb. Caixas de seleção para organização e localização física de amostras.....	56
Figura 13: Formulário de Confirmação do LAPOGEDb. Todas as variáveis de todos os formulários são mostrados na tela informando os dados preenchidos e dados não informados (não preenchidos).....	57
Figura 14: Formulário de Confirmação do LAPOGEDb. Todas as variáveis de todos os formulários são mostrados na tela informando os dados preenchidos e dados não informados (não preenchidos).....	58
Figura 15: Edição de formulários no LAPOGEDb. Formulários de PSR (100) disponíveis para edição de informações.....	59
Figura 16: Comparação das frequências alélicas dos AIMs (alelo *1) entre as populações de casos (PSR) e controles (PSC) com as populações parentais europeia (Portugueses; TOMÁS et al., 2002), ameríndia (LUIZON et al., 2008) e africana (TOMÁS et al., 2002).....	60

## LISTA DE TABELAS

Tabela 1: Exemplo hipotético de três AIMs com seus respectivos diferenciais de frequência ( $\delta$ ) entre populações parentais: africanas (AFR), europeias (EUR) e ameríndias (AMZ).....	20
Tabela 2: Condições de PCR dos quatro AIMs analisados.....	36
Tabela 3: Comprimentos dos fragmentos, localização citogenética e sequência dos primers dos quatro AIMs analisados.....	37
Tabela 4: Comprimento dos fragmentos, localização citogenética e sequência dos primers dos quatro AIMs analisados.....	45 - 46
Tabela 5: Valores de p para o Equilíbrio de Hardy-Weinberg. Valores em equilíbrio para $p > 0,05$ .....	47
Tabela 6: Diferenciação alélica e genotípica entre as populações de casos (PSR) e controles (PSC).....	48
Tabela 7: Valores de $F_{is}$ e $F_{st}$ nas populações de casos (PSR) e controles (PSC), adotando-se intervalo de confiança de 95% (IC 95%).....	48
Tabela 8: Estimativa de mistura étnica nas populações de casos (PSR) e controles (PSC) considerando os quatro AIMs analisados.....	48



## LISTA DE ABREVIATURAS

AF	Africano
AIMs	<i>AncestryInformativeMarkers</i>
ALD	<i>AdmixtureLinkageDisequilibrium</i>
AM	Ameríndio
ART	Artrite Reumatoide
BD	Banco de Dados
BDB	Banco de Dados Biológico
BDR	Banco de Dados Relacional
CMA	Câncer de Mama
DBA	<i>Database Administrator</i>
DF	Dados Familiais
DG	Dados Genéticos
DP	Dados Pessoais
EHW	Equilíbrio de Hardy-Weinberg
ER	Entidade Relacional
EU	Europeu
HTML	<i>HyperTextMarkupLanguage</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
InDels	Inserção/Deleção
LAPOGE	Laboratório de Polimorfismos Genéticos
LD	<i>LinkageDisequilibrium</i>
LES	Lúpus Eritematoso Sistêmico
PGH	Projeto Genoma Humano
PHM	Projeto HapMap
PHP	<i>Hypertext Preprocessor</i>
PSAs	Population-SpecificAlleles
PSC	Controles
PSR	Psoríase
SGBD	Sistemas de Gerenciamento de Banco de Dados
SNPs	<i>Single NucleotidePolymorphisms</i>
SQL	<i>Structured Query Language</i>
TA	Temperatura ambiente
TCLE	Termos de Consentimento Livre e Esclarecidos
UFSC	Universidade Federal de Santa Catarina
VNTRs	<i>Variable Number of Tandem Repeats</i>



## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>29</b>
<b>1.1 EPIDEMIOLOGIA GENÉTICA E A VARIABILIDADE HUMANA .....</b>	<b>29</b>
<b>1.2 ANCESTRALIDADE EM ESTUDOS DE ASSOCIAÇÃO GENÉTICA .....</b>	<b>30</b>
1.2.1 Marcadores Informativos de Ancestralidade .....	32
<b>1.3 A POPULAÇÃO BRASILEIRA COMO MODELO TRI-HÍBRIDA .....</b>	<b>34</b>
<b>1.4 GENÉTICA EPIDEMIOLOGICA DAS DOENÇAS AUTOIMUNES .....</b>	<b>36</b>
<b>1.5 BIOINFORMÁTICA E A GERAÇÃO DE DADOS BIOLÓGICOS .....</b>	<b>36</b>
1.5.1 Banco de Dados Biológicos .....	39
<b>2 JUSTIFICATIVA .....</b>	<b>41</b>
<b>OBJETIVOS .....</b>	<b>43</b>
<b>3.1 OBJETIVO GERAL .....</b>	<b>43</b>
<b>3.2 OBJETIVOS ESPECÍFICOS .....</b>	<b>43</b>
<i>Reagentes e Soluções</i> .....	46
<i>Procedimentos</i> .....	46
<b>4.4 MARCADORES INFORMATIVOS DE ANCESTRALIDADE UTILIZADOS ..</b>	<b>47</b>
4.4.1 Determinação fenotípica (nomenclatura) .....	47
4.4.2 Inserção/deleção (AT3-I/D) .....	47
4.4.3 Inserções Alu (SB19.3, APO e PV92) .....	48
<b>4.5 GENOTIPAGEM DOS MARCADORES INFORMATIVOS DE ANCESTRALIDADE .....</b>	<b>48</b>
4.5.1 Reação em Cadeia da Polimerase .....	48
<i>Reagentes e Soluções</i> .....	48
<i>Procedimento</i> .....	49
<i>Condições de PCR</i> .....	49
4.5.2 Análise dos produtos amplificados .....	50
<i>Reagentes e Soluções</i> .....	50

<i>Procedimento</i> .....	50
<i>Leitura dos resultados no gel de agarose</i> .....	51
<b>4.6 TRATAMENTO DOS DADOS E ANÁLISES ESTATÍSTICAS</b> .....	<b>53</b>
4.6.1 Aderência ao equilíbrio de Hardy-Weinberg .....	53
4.6.2 Estimativas das frequências alélicas e genotípicas.....	54
4.6.3 Associações par-a-par entre loci.....	54
4.6.4 Diferenciação genética das populações .....	55
4.6.5 Diversidade interpopulacional (estatísticas F) .....	55
4.6.6 Estimativas de mistura étnica .....	56
<b>4.7 BANCO DE DADOS BIOLÓGICOS</b> .....	<b>56</b>
4.7.1 Organização dos dados físicos e virtuais .....	56
4.7.2 Diagramação e programação .....	56
<b>5 RESULTADOS</b> .....	<b>59</b>
<b>5.1 FREQUÊNCIAS GENOTÍPICAS, ALÉLICAS E EQUILÍBRIO DE HARDY-WEINBERG</b> .....	<b>59</b>
<b>5.4 DIVERSIDADE INTER-POPULACIONAL (ESTATÍSTICAS F)</b> .....	<b>62</b>
5.4.1 Coeficiente de endogamia ( $F_{IS}$ ) e diversidade inter-populacional ( $F_{ST}$ ).....	62
<b>5.5 ESTIMATIVAS DE MISTURA ÉTNICAS</b> .....	<b>63</b>
<b>5.1 BANCO DE DADOS: LAPOGEdb</b> .....	<b>63</b>
5.1.1 Acesso e definições de usuário e senha .....	63
5.1.2 Cadastro de questionários .....	66
5.2.3 Edição de formulários .....	74
<b>6 DISCUSSÃO</b> .....	<b>75</b>
<b>6.1 FREQUÊNCIAS ALÉLICAS E EQUILÍBRIO DE HARDY-WEINBERG</b> .....	<b>75</b>
<b>6.2 ASSOCIAÇÃO ENTRE LOCI NÃO LIGADOS</b> .....	<b>77</b>
<b>6.3 DIFERENCIAÇÃO POPULACIONAL</b> .....	<b>77</b>
<b>6.4 DIVERSIDADE INTER-POPULACIONAL</b> .....	<b>77</b>

<b>6.5 ESTIMATIVAS DE MISTURA ÉTNICA .....</b>	<b>77</b>
<b>7 CONSIDERAÇÕES FINAIS .....</b>	<b>79</b>
<b>REFERÊNCIAS .....</b>	<b>81</b>
<b>APÊNDICE A – POPULAÇÕES ESTUDADAS PELO PROJETO HAPMAP .....</b>	<b>93</b>
<b>APÊNDICE B – POPULAÇÕES ESTUDADAS PELO PROJETO 1000 GENOMES.....</b>	<b>94</b>
<b>APÊNDICE C – QUESTIONÁRIO PACIENTES COM PSORÍASE.....</b>	<b>95</b>
<b>APÊNDICE D – QUESTIONÁRIO INDIVÍDUOS-CONTROLES.....</b>	<b>98</b>
<b>APÊNDICE E – TERMO DE CONSENTIMENTO LIVRE ESCLARECIDO .....</b>	<b>103</b>
<b>APÊNDICE F – TERMO DE CONSENTIMENTO LIVRE ESCLARECIDO INDIVÍDUOS-CONTROLES.....</b>	<b>105</b>
<b>APÊNDICE G – MODELO ENTIDADE RELACIONAL DO BDB .....</b>	<b>107</b>
<b>APÊNDICE H – PÁGINA DO FORMULÁRIO DE DADOS PESSOAIS DO LAPOGEDB .....</b>	<b>108</b>
<b>APÊNDICE I – PÁGINA DO FORMULÁRIO DE DADOS FAMILIAIS DO LAPOGEDB.....</b>	<b>109</b>
<b>APÊNDICE J – PÁGINA DO FORMULÁRIO DE DADOS GENÉTICOS DO LAPOGEDB.....</b>	<b>110</b>
<b>APÊNDICE L – PÁGINA DO FORMULÁRIO DE INFORMAÇÕES DO LAPOGEDB .....</b>	<b>111</b>
<b>APÊNDICE M – FREQUÊNCIAS DOS AIMS (AT3, SB19.3, APO, PV92) EM DOIS ESTUDOS BRASILEIROS DO TIPO CASO-CONTROLE.....</b>	<b>112</b>



# 1 INTRODUÇÃO

## 1.1 EPIDEMIOLOGIA GENÉTICA E A VARIABILIDADE HUMANA

Com a finalização da primeira etapa de sequenciamento do Projeto Genoma Humano (PGH) a comunidade científica contemporânea deparou-se com uma grande quantidade de informações valiosas relacionadas à variabilidade genética. Algumas das principais contribuições foram (1) realizar estudos para determinar a estrutura genética das populações humanas a partir de polimorfismos genéticos (LI et al., 2008; QIN et al., 2013); (2) estudar os padrões de diversidade de genes de interesse biomédico (PACKER et al., 2006; BEATY et al., 2010; DUCHARME et al., 2013) e (3) desenvolver um mapa de desequilíbrio de ligação (LD, do inglês Linkage Disequilibrium) entre regiões genômicas das populações humanas (FRAZER et al., 2007) que revolucionou os estudos de epidemiologia genética de doenças complexas (TODD et al., 2007; MCARTHY et al., 2008).

O Projeto HapMap (PHM) possui objetivos que vão além do desenvolvimento de mapas de desequilíbrio de ligação. O objetivo principal foi determinar os padrões comuns de variação da sequência de DNA no genoma humano, caracterizando estas variações, bem como suas frequências e correlações entre elas, em amostras de DNA inicialmente provenientes de populações com ascendência de partes da África, Ásia e Europa e posteriormente da América do Norte (Apêndice A) (ALTSHULER et al., 2010). Assim como o PHM, o Projeto 1000Genomes beneficia os estudos de epidemiologia genética com a produção de um catálogo de variações genéticas (polimorfismos), utilizando tecnologias de sequenciamento avançadas para refinar ainda mais o sequenciamento genômico na era pós PGH, dando um passo à frente para a criação de um mapa em alta resolução das principais diferenças entre os seres humanos. O projeto piloto lançado em 2008 contou com o sequenciamento do genoma de mil indivíduos de diferentes populações (Apêndice B) (ABECASIS et al., 2010), aumentando para 1.092 indivíduos em um estudo em 2012 (ABECASIS et al., 2012).

Outro grande projeto, denominado The Genographic Project, consiste em um esforço internacional que visa traçar a história evolutiva e migratória da espécie humana. Embora a primeira fase do projeto tenha focado em marcadores uniparentais (Cromossomo Y e DNA mitocondrial), a fase atual foca em marcadores espalhados através de todo o genoma para obter uma compreensão mais completa da variação

genética humana. Mesmo sem ter a premissa de levar em consideração a detecção de variações que possam estar associadas a doenças, os resultados do “Genographic” podem ajudar a delinear estudos que necessitam de um aprofundamento sobre a ascendência das populações humanas (ELHAIK et al., 2013).

Estes quatro grandes projetos são exemplos de estudos que tem levado em consideração a ascendência genética (ou ancestralidade genética), das populações humanas. Embora a ancestralidade explique apenas uma pequena fração da variação genética humana, existem variantes genéticas que possuem uma frequência elevada em populações com determinadas ancestralidades, podendo estar ou não associada a algum traço ou característica complexa (por exemplo, doenças).

## 1.2 ANCESTRALIDADE EM ESTUDOS DE ASSOCIAÇÃO GENÉTICA

Estudos de associação genética procuram analisar tanto fenótipos (tais como doenças) e genótipos (variações no DNA) de um dado conjunto de indivíduos. O objetivo dos estudos de associação é identificar variantes de DNA que susceptibilizam o risco de desenvolver determinadas doenças ou outras características de interesse (DENG, 2001).

O estudo do tipo caso-controle é considerado um dos mais simples e mais utilizado para testar associação, onde casos afetados com a doença de interesse são coletados juntamente com uma série de indivíduos controles (LEWIS & KNIGHT, 2012). No entanto, estudos de associação podem ser distorcidos por diferenças na ancestralidade levando a associações espúrias. Por exemplo, os resultados dúbios podem surgir se os indivíduos selecionados como casos possuem ascendência diferente, em média, daqueles considerados controles saudáveis. Se houver uma maior prevalência da doença na população amostrada de casos, isto pode representar uma associação espúria, pois indivíduos com a ancestralidade da população parental onde a doença for mais comum podem estar sobre-representados na amostragem de casos e, desta forma, qualquer alelo pode acusar associação com fenótipo, mesmo este não sendo um alelo de suscetibilidade (FREEDMAN et al., 2004; HEGALSON et al., 2005; PRICE et al., 2008; CHOUDHRY et al., 2006).

Essas associações são influenciadas por fatores como idade, gênero, estrutura populacional e miscigenação recente (RISCH, 2000). Fatores que implicam em falsas associações podem ocorrer quando: (I) um alelo é falsamente associado ao fenótipo, caso este seja mais

frequente na população de casos (pacientes); (II) penetrância diferencial devido a variações ambientais diversas; (III) Diferenças na amostragem (erros de amostragem por fatores não genéticos) (SUAREZ-KURTZ et al., 2005). Entretanto, associações espúrias não serão observadas quando: (I) a heterogeneidade for equivalente em casos e controles (ARDLIE et al., 2002); (II) a frequência alélica do loco em estudo for homogênea entre diferentes subgrupos populacionais; (III) a frequência da doença for a mesma nos subgrupos (ZIV & BURCHARD, 2003).

Tais associações irreais devido a problemas de estratificação populacional são apontadas como a principal explicação para repetidas falhas na tentativa de replicar achados positivos em estudos de genética epidemiológica (THOMAS et al., 2002; CARDON & PALMER, 2003). Um modo efetivo de controlar estas variações alélicas devido a fatores ancestrais (estratificação populacional) tem sido o foco de pesquisadores ao longo dos anos.

Pesquisadores apontam que o controle da estratificação populacional pode ser obtido ao equiparar etnicamente os grupos de casos e controles (HINDS et al., 2004; PRITCHARD & DONNELLY, 2001). Porém, quanto maior o nível de miscigenação entre as subpopulações de uma população, mais laboriosa torna-se a tarefa de equiparar o mais próximo e cuidadosamente possível os casos e controles de um estudo de associação genética. Esta dificuldade em nivelar os grupos étnicos pode ser observada na América Latina, sobretudo no Brasil, cuja população é considerada majoritariamente tri-híbrida (com populações parentais advindas de ameríndios, europeus e africanos) (LOPES et al., 2011).

Marcadores moleculares têm sido utilizados para estimar a miscigenação populacional e individual. Eles podem ser classificados quanto à sua localização: cromossomos autossômicos (biparentais), DNA mitocondrial e cromossomo Y (uniparentais); ou ainda quanto às características moleculares: SNPs (Single Nucleotide Polymorphisms), repetições em tandem de número variável (VNTRs, do inglês Variable Number of Tandem Repeats), que consistem de micro e minissatélites e InDels (inserções/deleções e Inserções Alu).

Marcadores uniparentais como aqueles do cromossomo Y e do DNA mitocondrial são especialmente informativos para traçar a história genética paterna ou materna (respectivamente) das populações, mas não eficientes em estimar a ancestralidade individual.

Os SNPs possuem características importantes, principalmente pela sua abundante e ampla distribuição no genoma, sendo capazes de fornecer informações de diversas regiões gênicas e extragênicas e, desta

forma, apontados como um dos tipos de marcadores mais informativos para estudos de ancestralidade genética (ALTSCHULER et al., 2010).

Os microssatélites são multialélicos (alguns com mais de 30 alelos diferenciados nas populações europeias, africanas e ameríndias; DUNCAN et al., 1996) e sua caracterização dá-se pelo número de repetições em tandem de seus nucleotídeos.

InDels, inserções ou deleções de nucleotídeos em um determinado locus, também contribuem para a estimativa de ancestralidade individual (PENA et al., 2011; MANTA et al., 2013).

As inserções Alu são repetições intercalantes curtas (SINE, do inglês Short Interspersed Repetitive Elements) de aproximadamente 300pb e representam 10% do genoma humano. Um número variado de inserções Alu foram identificadas (BATZER et al., 1995, 1996; BLONDEN et al., 1994; HAMMER, 1994; MILEWICZ et al., 1996), incluindo algumas que são únicas para determinadas ancestralidades, sendo passadas por descendência (WALLACE et al., 1991; TERREROS et al., 2009). Os polimorfismos de inserções Alu são considerados marcadores ideais para estudos evolucionários humanos porque os eventos de inserção são irreversíveis e amplamente distribuídos, cada qual com o estado ancestral conhecido (WATKINS et al., 2001).

### **1.2.1 Marcadores Informativos de Ancestralidade**

Uma das formas mais utilizadas atualmente para o controle da estratificação populacional é o uso de Marcadores Informativos de Ancestralidade (AIMs, do inglês Ancestry Informative Markers; BONILLA et al., 2004). Primeiramente chamados de alelos de populações específicas (PSAs, do inglês Population-Specific Alleles; SHRIVER et al., 1997), AIMs são loci (SNPs, Indels, Microssatélites, Inserções Alu, entre outros) que apresentam alelos com significantes diferenças de frequência entre populações de regiões geograficamente distintas (MARTINEZ-FIERRO et al., 2009).

A eficiência na determinação dos níveis de miscigenação em uma população, indivíduo ou região cromossômica depende do número de AIMs genotipados e principalmente do diferencial de frequência ( $\delta$ ) superior a 30% entre quaisquer duas populações parentais (SILVA et al., 2010). Alguns autores defendem um diferencial de frequência ( $\delta$ ) superior a 50% para considerar um determinado marcador como AIM (YEAGER et al., 2008; GALANTER et al., 2012). A tabela 1 exemplifica de forma hipotética três AIMs com alto diferencial de frequência entre populações parentais.

Tabela 1: Exemplo hipotético de três AIMS com seus respectivos diferenciais de frequência ( $\delta$ ) entre populações parentais: africanas (AFR), europeias (EUR) e ameríndias (AMZ).

AIMs*	Diferencial de Frequência ( $\delta$ )		
	AFR/ EUR	AFR/ AMZ	EUR/ AMZ
A	0,993	1,000	0,007
B	0,670	0,151	0,780
C	0,077	0,748	0,825

\*A = Distingue Africanos de Europeus ou Ameríndios; B = Distingue Europeus de Africanos ou Ameríndios; C = Distingue Ameríndios de Europeus ou Africanos.

Os AIMS já foram empregados em estimativas de mistura étnica em populações das Américas do Norte, Central e do Sul. Em um estudo realizado por Parra et al. (2001), observou-se que a mistura africana é preponderante nas populações afro-americanas. Salari et al. (2005) e Choudhry et al. (2006) observaram que a mistura europeia é predominante nas estimativas de mexicanos-americanos e porto-riquenhos (hispânicos). A ancestralidade ameríndia e europeia é maior dentre os mexicanos e equatorianos (MARTINEZ-MARIGNAC et al., 2007; BRYC et al., 2010) e quase completa entre os nativos americanos do México (BONILLA et al., 2005) e da Amazônia Brasileira (LUIZON et al., 2008). Um estudo na Bahia mostrou prevalência das ancestralidades europeias e africanas (TELÓ, 2010). No estado de Santa Catarina a contribuição europeia (sobretudo de portugueses), prevalece em comunidades consideradas isoladas e semi-isoladas na ilha de Florianópolis (MUNIZ, 2008).

Estes marcadores vêm sendo cada vez mais utilizados para verificar a homogeneidade genética de populações, bem como para identificação da estratificação populacional em estudos de caso-controle com populações miscigenadas (HOGGART et al., 2003; ENOCH et al., 2006; LINS et al., 2011). Apesar da grande quantidade de marcadores que podem ser utilizados como AIMS, estudos recentes demonstram que a informação de ancestralidade é suficientemente confiável para as inferências estatísticas (GALANTER et al., 2012; PEREIRA et al., 2012). Do ponto de vista econômico, a utilização de AIMS é desejável para a redução dos esforços de genotipagem, por permitir a diminuição do número de marcadores necessários para

estimar a miscigenação em testes de estratificação de população e estudos de epidemiologia genética, mantendo a confiança da análise (CIRULLI & GOLDSTEIN, 2010).

### 1.3 A POPULAÇÃO BRASILEIRA COMO MODELO TRI-HÍBRIDA

A população brasileira é o produto da miscigenação entre três principais grupos étnicos: Ameríndios, Europeus e Africanos (LOPES et al., 2011). O processo de miscigenação iniciou entre os anos de 1500 e 1530 com a chegada dos europeus (portugueses) e dos negros africanos, estes últimos trazidos como mão de obra escrava, em um continente habitado por aproximadamente 2.000.000 de nativos americanos (ALMEIDA, 1997; SALZANO & BORTOLINI, 2002).

A forte crise de subsistência que assolava Portugal devido à densidade populacional elevada obrigou a busca por novas terras e à formação de novas colônias portuguesas. O estímulo do rei de Portugal às famílias a emigrarem deixava claro o propósito de ocupação do território, através de uma colonização de base estável no então recém-descoberto Brasil (FLORES, 2000). Além de portugueses, o Brasil recebeu imigrantes de outros países como: italianos, alemães, espanhóis, poloneses, japoneses e sírio-libaneses que saíram de seus países em busca de melhores condições de vida nas Américas. Após a abertura dos portos em 1808, o fluxo de imigrantes europeus foi intensificado com a chegada de italianos, alemães e espanhóis (SALZANO & BORTOLINI, 2002; VENÂNCIO, 2000). Estima-se que cerca de 15 milhões de europeus tenham cruzado o Atlântico até 1880 e que entre 1881 a 1915 o número de migrantes tenha alcançado seu apogeu, chegando a cerca de 31 milhões de indivíduos (FAUSTO, 1999).

O componente africano na população brasileira foi atribuído à prática escravagista que teve início por volta de 1538 e se manteve por aproximadamente 350 anos (SALZANO & FREIRE-MAIA, 1970). Durante este período, milhares de escravos foram embarcados no litoral atlântico da África, em diversos locais da costa que se estende do Senegal ao sul de Angola e ao norte da Namíbia e também no litoral da África Oriental, de Moçambique à Somália. Estes entraram no país principalmente pelo Maranhão, Bahia, Rio de Janeiro e Pernambuco, sendo posteriormente distribuídos para as demais regiões (TAVARES, 1988).

Estima-se que até 1880, tenham sido trazidos para o continente americano pelo menos 10 milhões de africanos. A migração em massa de europeus e de africanos promoveu padrões genéticos diferentes, sendo observados apenas nas gerações seguintes, havendo o

estabelecimento de populações com vários tipos e graus de mistura em toda América. Além disso, a migração em massa promoveu também mudanças demográficas, havendo redução da população nativa em aproximadamente 90% do que existia à época do início da colonização (BEDOYA et al., 2006).

Os grupos indígenas encontrados no litoral brasileiro pelos portugueses eram principalmente tribos do tronco Tupi. Somavam cerca de um milhão de índios, divididos em dezenas de tribos e aproximadamente 300 a 2000 habitantes em cada aldeia (RIBEIRO, 1997). Os primeiros habitantes das terras que hoje correspondem a Santa Catarina foram indígenas, identificados em relação ao tronco linguístico em dois grandes grupos: Tupi (Guarani) e Macro-Jê (Kaingang e Xokleng) (RODRIGUES, 2000; 2011).

A bibliografia relativa aos Guarani é provavelmente a mais vasta das tribos brasileiras, uma vez que desde o século XVI, inúmeras obras descritivas compõem um dos mais numerosos registros de povos indígenas sul-americanos. Dominavam as florestas subtropicais do Rio Grande do Sul, Santa Catarina, Paraná e Misiones (Argentina), assim como as florestas tropicais de São Paulo, Mato Grosso do Sul, Paraguai e Bolívia (SCHIAVETTO, 2003). Os Guarani do Brasil Central podem ser divididos de acordo com diferenças linguísticas e peculiaridades da cultura material em três grandes grupos: Nandeva, Kaiowá e M'Byá (apud MARRERO et al., 2007). O declínio da população Guarani foi devido a causas não muito diferentes daquelas que vitimaram outras populações nativas: novas doenças de origem europeia ou africana, associadas a ataques de bandeirantes, entre outras causas (DERENGOSKI, 2002).

Este processo migratório que envolveu o Brasil desde a época de seu descobrimento e que perdura há mais de cinco séculos culminou na atual composição da população brasileira e implicou não somente alterações nas composições genéticas inerentes à miscigenação, como também na inserção de diferentes culturas e costumes.

Além de ser uma população com o histórico de formação recente, estudos em genética molecular e de populações tornam-se de grande contribuição na tentativa de caracterizar as bases genéticas de uma população sabidamente tri-híbrida como a brasileira, que possui um confiável registro histórico de miscigenação.

## 1.4 GENÉTICA EPIDEMIOLOGICA DAS DOENÇAS AUTOIMUNES

Identificar genes e variantes causais de doenças complexas representa um importante passo para compreensão dos mecanismos genéticos envolvidos na patogênese dessas doenças e, em alguns casos, na melhora do tratamento, diagnóstico e prevenção (ALTSHULER et al., 2008). A busca por respostas satisfatórias nesta área tem levado cientistas ao rastreamento de diferentes polimorfismos genéticos, principalmente pela sua ampla distribuição no genoma, sendo capazes de fornecer informações de diversas regiões gênicas e extragênicas.

Em doenças autoimunes complexas, por exemplo, são esperadas muitas variantes com pequenas contribuições individuais em relação aos fatores de risco. Além disso, deve existir um grande número de alelos de susceptibilidade para um dado gene, portanto, o risco global de um gene será a combinação entre o risco individual para cada alelo de susceptibilidade em um locus. Interpretações sobre uma possível associação entre determinado gene e uma doença dependem do conhecimento do desequilíbrio de ligação entre os SNPs genotipados e os alelos funcionais (DESHMUKH et al., 2008).

O estudo da ancestralidade genética em doenças autoimunes possui alguns relatos na literatura, principalmente para Lúpus Eritematoso Sistêmico. Contudo, a maioria destes estudos se restringe a populações europeias e norte americanas (ALONSO-PEREZ et al., 2011, 2012; MOLINEROS et al., 2009), com poucos estudos para populações latinas (SANCHEZ et al., 2010; SELDIN et al., 2008).

Portanto, é de extrema importância testar se alelos de susceptibilidade a doenças autoimunes estão relacionados com componentes étnicos, levando em consideração que determinadas etnias possuem diferenças nas ocorrências destas doenças.

## 1.5 BIOINFORMÁTICA E A GERAÇÃO DE DADOS BIOLÓGICOS

As origens da bioinformática, tanto como termo e disciplina, são difíceis de datar (ATTWOOD et al., 2011). Durante a década de 1960, grupos de pesquisadores das mais diversas áreas das ciências biológicas viram-se diante da necessidade de dedicar esforços em aprender aspectos computacionais para aplicar em seus objetos de estudo. Posteriormente, o início do PGH, na década de 1980, proporcionou o grande impulso para o desenvolvimento da bioinformática e o seu estabelecimento como disciplina. Além disso, o desenvolvimento e o uso cada vez mais difundido da internet na década de 1990 tornaram

possíveis o acesso instantâneo, intercâmbio e divulgação de dados biológicos (HIGGINS & TAYLOR, 2001).

O grande volume de dados biológicos gerados com a finalização da primeira etapa do PGH evidenciou o que até então era sabidamente esperado e, de certa forma, temido pelos pesquisadores envolvidos no projeto: O PGH estaria fadado a falhar caso a tecnologia computacional (principalmente no campo da sistematização de anotações biológicas e armazenamento de informações) não evoluísse de forma concomitante e progressiva com o projeto (ROBBINS, 1994). Ademais, o sucesso do PGH dependia não apenas da evolução das tecnologias informacionais, como também da capacitação dos pesquisadores e habilidade dos mesmos em interpretar dados de sequenciamento e variações genéticas, até então uma novidade para muitos da área da pesquisa aplicada (JORDAN, 2010).

Um Banco de Dados (BD) é um arquivo informatizado utilizado para armazenar e organizar os dados de tal forma que as informações podem ser recuperadas facilmente através de uma variedade de critérios de pesquisa. O objetivo principal do desenvolvimento de um BD é organizar dados em um conjunto de registros estruturados para permitir a fácil recuperação de informações. Cada registro, também chamado de entry, deve conter um número de campos que possuem os itens de dados reais, por exemplo, campos para nomes, números de telefone, endereços, datas. Para recuperar um determinado registro do BD, o usuário pode especificar uma determinada informação, chamada de value, para ser encontrado em um campo particular e esperar o computador para recuperar todo o registro de dados para esta informação. Este processo é chamado de making a query (do inglês, fazendo uma consulta) (WILTON & COLBY, 2005).

De modo geral, pequenos laboratórios de pesquisas têm utilizado arquivos comuns ou planilhas (por exemplo, Excel®) como BD. Este tipo de abordagem torna-se interessante para armazenar pequenas informações (com um número finito de informações a serem armazenadas), exigindo menor poder de trabalho do computador. Entretanto, à medida que aumenta o tamanho do BD ou se os tipos de dados se tornam mais complexos, o estilo de BD através de arquivos comuns ou planilhas pode tornar a recuperação da informação muito difícil, frequentemente causando falhas no sistema devido à falta de memória ou processador para executar a tarefa (DATE, 2003).

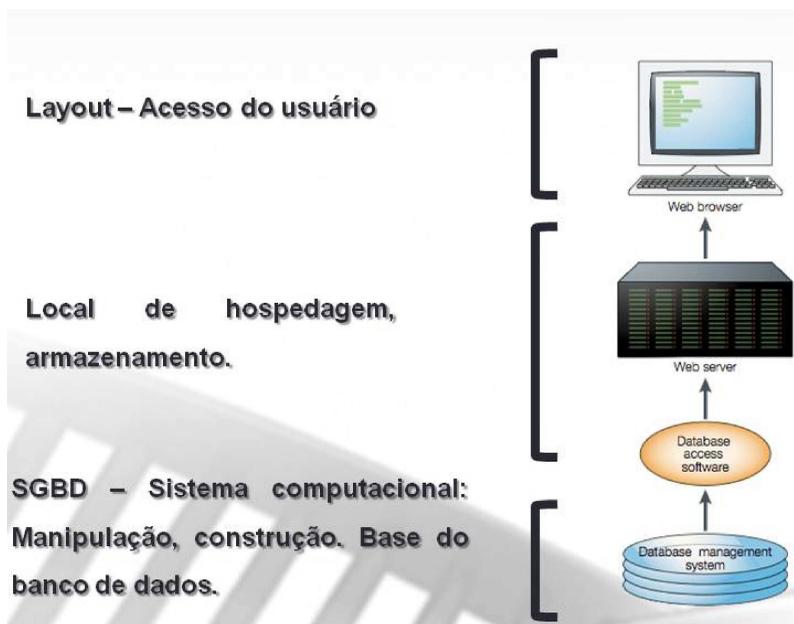
Para facilitar o acesso e recuperação de dados, sofisticados programas de computador para organizar, pesquisar e acessar dados têm sido desenvolvidos. Eles são chamados de Sistemas de Gerenciamento

de Banco de Dados (SGBD). Esses sistemas contêm não somente os registros de dados, mas também as instruções operacionais para ajudar a identificar as conexões entre os registros de dados. O objetivo da criação de uma estrutura de dados por um SGBD é facilitar a execução de pesquisas e combinar diferentes registros para formar relatórios de pesquisa final. Um dos modelos de SGBD mais utilizados é o modelo de Banco e Dados Relacionais (BDR) (ELMASRI & NAVATHE, 2010). A figura 1 demonstra de forma simplificada a arquitetura de um BD padrão com o uso de um SGBD (STEIN, 2003).

Em vez de usar uma única tabela como em uma planilha, BDR usam um conjunto de tabelas para organizar os dados. Cada tabela, também chamada de “relação”, é composta de linhas e colunas. As colunas representam os campos individuais. As linhas representam os valores nos campos de registros. As colunas de uma tabela são indexadas de acordo com uma característica comum chamado “atributo”, para que eles possam ter uma referência cruzada em outras tabelas. Para executar uma consulta em um BDR, o sistema seleciona os itens de dados ligados de diferentes tabelas e combina as informações em um único relatório. Portanto, informações específicas podem ser encontradas mais rapidamente a partir de um BDR que a partir de um BD de arquivo simples ou de uma única tabela (STEIN, 2013).

Os BDR podem ser criados usando uma linguagem de programação chamada Structured Query Language (SQL). A criação deste tipo de bases de dados é trabalhosa e depende de um grande planejamento durante a fase de construção e modelagem do banco. Após a criação da base de dados uma nova categoria de dados pode ser facilmente adicionada sem exigir que todas as tabelas existentes tenham que ser também modificadas. Fatores tais como código aberto, alta velocidade de processamento e disponibilidade para vários sistemas operacionais, devem ser levados em consideração para escolher o SGBD a ser utilizado na construção de BD. O SGBD MySQL utiliza a linguagem SQL e, sua rapidez, capacidade de manusear grandes quantidades de dados e por atender todas as especificações citadas acima, o MySQL pode ser considerado como um dos BDR de maior sucesso no mundo atual (TAHAGHOGHI & WILLIAMS, 2007).

**Figura 1:** Estrutura de um Banco de Dados. A estrutura de um BD é dividida em três partes: A primeira consiste de um sistema de gerenciamento de banco de dados; a segunda de um local (servidor), onde são armazenados os dados; a terceira é a interface gráfica para o usuário do banco realizar as consultas e gerar relatórios das pesquisas realizada no banco de dados. Fonte: STEIN. 2003.



### 1.5.1 Banco de Dados Biológicos

Bancos de Dados Biológicos (BDB) são considerados essenciais para a experimentação e análise biológica. Como tal, eles representam uma importante base nas descobertas científicas (ELTABAKH et al., 2006).

Uma forma adequada e organizada de armazenar dados biológicos evidenciou um papel central em bioinformática, sobretudo no que diz respeito aos BDB. Eles oferecem aos cientistas a oportunidade de acessar uma grande variedade de dados biologicamente relevantes, incluindo as seqüências genômicas de uma gama cada vez mais ampla de organismos (BAXEVANIS, 2011).

Com base em seu conteúdo, BDB podem ser divididos em três categorias: bancos de dados primários, secundários e especializados.

Bancos de dados primários contêm os dados biológicos originais. Eles são arquivos de sequência ou dados estruturais apresentados pela comunidade científica. GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>), EMBL (<http://www.ebi.ac.uk/embl/>) e DDBJ (<http://www.ddbj.nig.ac.jp/>) são exemplos de bases de dados primários. Bancos de dados secundários contêm informações computacionalmente processadas com base na informação original a partir de bancos de dados primários. Bases de dados de sequências de proteínas traduzidas contendo anotação funcional pertencem a esta categoria. Swiss-Prot (<http://www.ebi.ac.uk/uniprot>) é um exemplo de banco de dados secundário. Bases de dados especializadas são aquelas que atendem a um interesse de investigação particular. Por exemplo, Flubase (<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>) e banco de dados de sequência do HIV (<http://www.hiv.lanl.gov/content/index>), são bases de dados que se especializaram em um organismo particular ou a um tipo particular de dados. A maioria dos laboratórios de pesquisas acadêmicas utilizam os bancos de dados especializados, tendo em vista que, cada laboratório possui suas particularidades em suas pesquisas (XIONG, 2006).

Os BDB se tornaram componentes essenciais no cotidiano da vida científica atual. Tal afirmação remete aos grandes BDB, permitindo aos cientistas, o fácil acesso (público) a uma base de dados estável para as suas análises. Grandes BDB com informações gerais de diversos organismos, como GenBank, EMBL e DDBJ, são considerados referências para modelos de BDB. A integração de dados é, talvez, uma das funcionalidades mais marcantes destes bancos, garantindo que não haja por exemplo, a sobreposição de informações disponíveis. Isto permite a criação e implementação de bancos menores e especializados que atendam as necessidades de cada grupo de pesquisa (STEIN, 2003).

## 2 JUSTIFICATIVA

Dados sobre a ancestralidade genômica utilizando AIMs são importantes para avaliar as implicações da estrutura genética da população em estudos de associação entre genótipo e fenótipo, do tipo caso-controle, especialmente em uma população heterogênea como a brasileira. Desta forma, a padronização de AIMs aplicado a populações brasileiras auxiliará no delineamento de estudos de ancestralidade e epidemiologia genética, visando evitar associações espúrias devido a diferenças nas frequências alélicas entre indivíduos provenientes de diferentes regiões geográficas, uma vez que o risco de obter falsos positivos em estudos de associação genótipo/fenótipo depende do nível de estruturação (ou estratificação) da população.

Uma vez que marcadores de ancestralidade promovem uma estimativa de mistura étnica confiável, a hipótese é que estes podem ser utilizados para genotipagem, verificando de forma segura a estruturação populacional, bem como determinar se diferentes grupos (com e sem determinada doença) podem ser utilizados em estudos de caso-controle.

Anterior a proposta da genotipagem de AIMs o Laboratório de Polimorfismos Genéticos da Universidade Federal de Santa Catarina (LAPOGE – UFSC), buscava determinar a ancestralidade fenotipicamente entre as amostras de casos e controles. Entretanto, tal abordagem mostrou-se insuficiente, sendo a utilização da metodologia aqui apresentada uma possível solução para contornar possíveis problemas gerados pela estratificação populacional.

O grande volume de dados gerados pela genotipagem de AIMs e marcadores de outras naturezas em estudos de epidemiologia genética dá suporte a hipótese de que a construção e implementação de um BDB informacional e relacional é vantajosa para garantir a adequada organização dos dados obtidos, bem como, de grande valia no auxílio para análises e metanálises que venham a ser feitas a posteriori em um determinado estudo.



### 3 OBJETIVOS

#### 3.1 OBJETIVO GERAL

Através de análises moleculares, genotipar quatro AIMs, adequados para amostras da população brasileira, bem como estender esta abordagem para entender a estruturação populacional em estudos de caso-controle, utilizando como modelo a doença psoríase.

#### 3.2 OBJETIVOS ESPECÍFICOS

- Genotipar os marcadores selecionados em amostras de indivíduos diagnosticados com psoríase (casos) e indivíduos sem histórico de doenças (controles).
- Calcular as frequências alélicas obtidas, visando quantificar a mistura étnica nos diferentes grupos amostrais (casos e controles) e estabelecer comparações entre eles, avaliando a contribuição dos componentes africano, europeu e ameríndio na formação da referida população.
- Verificar a dinâmica do processo de mistura étnica entre os diferentes grupos.
- Construir e implementar um BDB utilizando linguagens de programação de alto nível para o armazenamento das informações genéticas, clínicas e epidemiológicas, que inclua todas as amostras depositadas no LAPOGE, visando concatenar informações e inter cruzá-las, para futuras análises.



## 4 MATERIAL E MÉTODOS

### 4.1 ASPECTOS ÉTICOS

Este trabalho é parte de projetos maiores, intitulados: (1) “Estudo genético e epidemiológico em pacientes com psoríase no estado de Santa Catarina”, (2) “Genética da autoimunidade: polimorfismos em Lúpus Eritematoso Sistêmico e Artrite Reumatoide em pacientes de Santa Catarina” e (3) “Câncer de Mama: avaliação de parâmetros informativos para diagnóstico, prognóstico e tratamento de pacientes na população do estado de Santa Catarina”, todos coordenados pela Dra Líada Rainha de Souza e aprovados pelo Comitê de Ética em Pesquisa com Seres Humanos da UFSC (CEPSH/UFSC): (1) em 25 de outubro de 2010, parecer no 1034/12, (2) em 26 de junho de 2006, parecer nº 172/06, prorrogado pelo ofício de no 9/CEPSH/PRPE/11 de 02 de março de 2011, e (3) em 26 de junho de 2006, parecer no 349/05, com adendo citado pelo ofício de número 107/CEPSH/PRPE/09. O terceiro projeto também foi aprovado pela Comissão Nacional de Ética em Pesquisa (CONEP) sob parecer nº. 027/2006 e registro 13312, de 24 de janeiro de 2007.

### 4.2 CARACTERIZAÇÃO DA AMOSTRA

As amostras de casos utilizadas neste trabalho como modelo foram pacientes diagnosticados com psoríase, que inicialmente foram utilizadas em um projeto/dissertação de mestrado realizada no Laboratório de Polimorfismos Genéticos (LAPOGE) da UFSC, disponível no acervo de teses e dissertações da biblioteca central da UFSC sob o título de “Polimorfismos do gene HLA-G em pacientes com psoríase”.

A amostra foi constituída de 100 pacientes diagnosticados com psoríase, atendidos pelo Ambulatório de Dermatologia/Psoríase do Hospital Nereu Ramos, no período de dezembro de 2010 a novembro de 2011 e de 100 indivíduos-controles, sem diagnóstico nem histórico familiar para a patologia em estudo, que são participantes de projetos realizados paralelamente pela equipe do LAPOGE-UFSC, citados em tópico anterior.

Os dados epidemiológicos e familiares dos pacientes e dos indivíduos-controles, foram obtidos através de entrevistas realizadas com questionários estruturados (Apêndices C e D), após a assinatura dos Termos de Consentimento Livre e Esclarecidos (TCLEs) pelos voluntários (Apêndices E e F). Após as entrevistas, amostras de sangue

periférico (cerca de 8 ml) foram coletadas através de punção venosa, para extração do DNA genômico e obtenção dos dados genéticos, e armazenadas e transportadas a 4°C para o posterior processamento no LAPOGE.

As amostras biológicas e os questionários obtidos nessa pesquisa foram catalogados e constituem um banco de dados e amostras, de pacientes com psoríase e de indivíduos-controles, armazenado no LAPOGE.

Alguns desses dados obtidos através dos questionários, como a idade de manifestação da psoríase e o possível acometimento das articulações (artrite psoriásica) foram utilizados para correlacionar com os dados genéticos obtidos no trabalho (dissertação de mestrado), mencionado anteriormente.

#### 4.3 EXTRAÇÃO DO DNA GENÔMICO

As amostras de sangue periférico foram centrifugadas (Centrifuge 206 BL Excelsa II®) a 3000g (1835rpm) durante 15 min a temperatura ambiente (TA) para a separação dos componentes sanguíneos. Após a centrifugação, foram separados o plasma, o concentrado de hemácias e a camada de leucócitos. Esses componentes sanguíneos foram aliquotados, identificados e estocados a -20°C. Os leucócitos foram utilizados para a extração do DNA genômico. O plasma e o concentrado de hemácias não foram utilizados nesse estudo.

##### **Reagentes e Soluções**

- 1) Solução de Lise I (Tris-HCl 0,01M, Amresco®; Sacarose 0,32M, Merck®; MgCl<sub>2</sub> 0,0025M; Triton X 100 – 1%, Nuclear®);
- 2) Solução de Lise II (Tris-HCl 0,01M, Amresco®; KCl 0,05M, Vetec®; MgCl<sub>2</sub> 0,0025M, Nuclear®; Nonidet – 1%, Amresco®; TWEEN 20 – 1%, Amresco®);
- 3) SDS 10%(Amresco®);
- 4) Solução de Perclorato de Sódio 5,0M (Vetec®);
- 5) Solução Saturada de NaCl 6,0M (Nuclear®);
- 6) TE (Tris-HCl 1M, Amresco®; EDTA 0,5M, Vetec®);
- 7) Álcool Isopropílico Absoluto, TA- Temperatura Ambiente (Merck®);
- 8) Etanol 70%, TA (Merck®).

##### **Procedimentos**

A extração de DNA foi realizada através de um método de Salting Out modificado, baseado em Miller et al. (1988). Para cada amostra, foram colocados 100µL da camada de leucócitos em

microtubos de polipropileno de 1,5mL (tipo eppendorf), utilizando-se uma micropipeta e ponteiros estéreis. Em seguida, adicionou-se 1,0mL Solução de Lise I em cada um desses microtubos. As amostras foram homogeneizadas e centrifugadas (Centrifuge 5415D, Eppendorf®) a 13400g (12000 rpm) durante 4 min a TA. O sobrenadante foi descartado e, este procedimento foi repetido (de 3 a 4 vezes) até que os glóbulos vermelhos fossem removidos e o precipitado apresentasse cor branca, indicando a presença dos glóbulos brancos. Posteriormente, foi acrescentado ao precipitado de leucócitos 300µL de Solução de Lise II, 10µL de SDS 10% e 75µL de Perclorato de Sódio 5M. As amostras foram agitadas em um agitador de tubos, do tipo vórtex, e a cada tubo foi acrescentado 130µL de NaCl 6M e, a seguir, as amostras foram centrifugadas a 13400g por 5 min a TA. Novos microtubos de 1,5mL foram identificados e, para esses, foram transferidos os sobrenadantes resultantes da centrifugação. Ao sobrenadante foram adicionados 300µL de Álcool Isopropílico Absoluto e as amostras foram, novamente, centrifugadas a 13400g por 15 min. O sobrenadante foi descartado e ao precipitado foi acrescentado 300µL de Etanol 70%. As amostras foram centrifugadas a 13400g por 5 min, o sobrenadante foi descartado e o precipitado foi seco, a temperatura ambiente, overnight. Após a secagem dos precipitados, foi adicionado a cada tubo 100 µL de TE. As amostras foram colocadas no banho-maria a 56°C, por 30 min e, posteriormente, armazenadas a -20°C.

#### 4.4 MARCADORES INFORMATIVOS DE ANCESTRALIDADE UTILIZADOS

##### 4.4.1 Determinação fenotípica (nomenclatura)

Para os AIMS analisados no presente estudo a denominação dos alelos foi feita seguindo a convenção estabelecida em PARRA et al. (1998), onde o alelo \*1 (por ex., AT3\*1) é o fragmento correspondente à banda de maior peso molecular observada no gel de agarose, devido à presença de uma inserção ou inserção Alu (para os loci AT3-I/D, Sb19.3, APO e PV92).

##### 4.4.2 Inserção/deleção (AT3-I/D)

O gene da AT3 localiza-se no cromossomo 1 (1q25.1), possui 19kb e sete éxons. No presente estudo foi analisado a presença do polimorfismo de comprimento de 76pb na região 5' do éxon 1 (LIU et

al., 1995). A presença desta inserção gera um fragmento de 572pb e caracteriza o alelo AT3\*1, mais frequente na população africana.

#### **4.4.3 Inserções Alu (SB19.3, APO e PV92)**

A inserção Alu Sb19.3 está localizada no cromossomo 19p12 (ARCOT et al., 1998). A presença da inserção Alu gera um fragmento de aproximadamente 457pb que caracteriza o alelo Sb19.3\*1. Este alelo apresenta frequência elevada em europeus e nativo-americanos. O locus Alu APO está próximo ao complexo de genes da apolipoproteína AI-CIII-AIV no braço longo do cromossomo 11 (KARATHANANISIS et al., 1986). A presença da inserção Alu gera um fragmento de aproximadamente 409pb e caracteriza o alelo APO\*1, frequente em nativos americanos e europeus. O gene Alu PV92 localiza-se no cromossomo 16 (BATZER et al., 1994) e a caracterização do alelo PV92\*1 dá-se pela presença da inserção Alu que gera um fragmento de aproximadamente 400pb, mais prevalente em populações nativo americanas (LUIZON et al., 2008).

### **4.5 GENOTIPAGEM DOS MARCADORES INFORMATIVOS DE ANCESTRALIDADE**

A genotipagem dos quatro AIMs selecionados foi realizada através da amplificação dos fragmentos desejados pela Reação em Cadeia da Polimerase (PCR, do inglês Polymerase Chain Reaction).

#### **4.5.1 Reação em Cadeia da Polimerase**

##### ***Reagentes e Soluções***

- 1) *Água Ultrapura;*
- 2) *dNTPs solução estoque: quatro soluções separadas de 100mM de cada base (dATP, dCTP, dGTP, dTTP), pH 8,3 (Invitrogen®);*
- 3) *dNTP solução de trabalho (20mM): obtida diluindo-se com água a solução estoque (100mM) de cada dNTP para uma solução única de concentração 20mM (160 µL de água de MiliQ autoclavada mais 10 µL da solução estoque de cada dNTP);*
- 4) *MgCl<sub>2</sub> 1,5mM (50mM, Invitrogen®);*
- 5) *Tampão de PCR 10X (0,2M Tris-HCl pH 8,5; 0,5M KCl; Invitrogen®);*
- 6) *Iniciadores (Primers) específicos – solução estoque (100 µM): os primers liofilizados (IDT®) Forward e Reverse específicos para cada AIM foram diluídos em água auto-clavada, e estocados separadamente.*

7) *Iniciadores (Primers) – solução trabalho (2,5 µM): 10 µL do primer Forward, 10 µL do primer Reverse e 180 µL de água auto-clavada.*

8) *Taq DNA Polymerase Platinum (0,5U/µL; Invitrogen®).*

### ***Procedimento***

Os ensaios da PCR foram realizados em concentração final de 25 µL. Todos os reagentes (Água, Tampão, dNTP – solução trabalho, primers – solução trabalho, MgCl<sub>2</sub> e a TaqPlatinum DNA polimerase), com exceção do DNA, foram misturados para cada locus em um único tubo de 0,6 µL (mistura de reação) para garantir a homogeneidade das reações. O cálculo para uma reação foi: 16,15 µL de água; 2,5 µL de Tampão de PCR; 0,25 µL de dNTP (solução trabalho); 1,00 µL de MgCl<sub>2</sub>; 2,00 µL de Primers (solução trabalho) e 0,1 µL de TaqPlatinum Polimerase. Após feito o mix de reação, 22 µL da solução de reação são pipetados e distribuídos em tubos de 0,2 µL.

Em cada microtubo de 0,2 µL foram pipetados 3 µL do DNA genômico, previamente extraído. Para cada análise foi usado como controle negativo contendo água no lugar do DNA genômico. Em cada microtubo, sob a amostra, foram pipetados 22 µL da mistura de reação. Os termocicladores utilizados foram o Mastercycler, Eppendorf® e MyCycler, Bio-Rad®. A este passo seguiu-se o programa correspondente a cada locus (Tabela 2). Após o término da reação de PCR o produto da PCR foi guardado em geladeira (4°C) até sua utilização.

### ***Condições de PCR***

Após um passo inicial de desnaturação de 5 min a 94°C, as amostras foram amplificadas por 30 ciclos nas temperaturas de desnaturação/pareamento/extensão especificadas (ToC PCR) para cada loci, seguidos por um passo final de extensão por 5 min a 72°C. A menos que indicado em PCR - NOTAS, os passos de desnaturação/pareamento/extensão foram respectivamente: 45 seg; 1min e 1 min.

**Tabela 2:** Condições de PCR dos quatro AIMS analisados.

Locus	Tipo	PCR Notas	Referência
AT3-I/D	76pb indel	1' ext.	
APO	Inserção Alu	1' des./2' par./2' ext.	LUIZON et al., 2007
PV92	Inserção Alu	1' des./2' par./2' ext.	
Sb19.3	Inserção Alu	1' des./2' par./2' ext.	

#### 4.5.2 Análise dos produtos amplificados

##### *Reagentes e Soluções*

- 1) *UltraPure Agarose (Invitrogen®)*
- 2) *Tampão TBE 1X (100mL do Tampão TBE 10X; 900mL H<sub>2</sub>O);*
- 3) *Água Ultrapura;*
- 4) *Peso Molecular de 100pb (BioLabs®).*
- 5) *GelRed® Nucleic Acid Gel Stain (Biotium Inc.)*
- 6) *Tampão de Aplicação de Amostras (900 µL de Bromofenol, Merck®; 900 µL Xilenocianol, Sigma-Aldrich®; 900 µL TBE; 4,5mL Ficoll 30% diluído em H<sub>2</sub>O, Sigma-Aldrich®; 1,8mL EDTA 0,5M pH 8,0, Vetec®; 3,6g Sacarose, Merck®);*

##### *Procedimento*

Os produtos amplificados (Tabela 3) foram visualizados através de eletroforese em gel de Agarose a 1,5%. Estes geis foram feitos adicionando: 40,0 mL de TBE 0,5%; 0,6 g de Agarose. Essa solução foi colocada em suporte específico para a gelificação do gel e, após este processo, a placa contendo o gel foi encaixada na cuba de eletroforese modelo BIOAHS-12G (Bioamerica Inc.). As amostras foram preparadas da seguinte forma: 6,0µL de produto de PCR; 1µL de mix de aplicação contendo 0,3 µL de GelRed; 0,3 µL de Tampão de Aplicação de Amostras e 4,0 µL de Água Ultrapura); e, posteriormente, aplicadas no gel. No primeiro poço foi colocado o padrão de peso molecular de 100pb para, posteriormente, ser comparado com as bandas amplificadas. A corrida foi realizada, durante 50min, e a fonte de eletroforese (CBS, ScientificCompany, Modelo EPS 4000) foi regulada de maneira que, a

Voltagem (V) ficou fixa em 85 e, a Amperagem (mA) e a Potência (W) ficaram livres.

**Tabela 3:** Comprimentos dos fragmentos, localização citogenética e sequência dos *primers* dos quatro AIMs analisados.

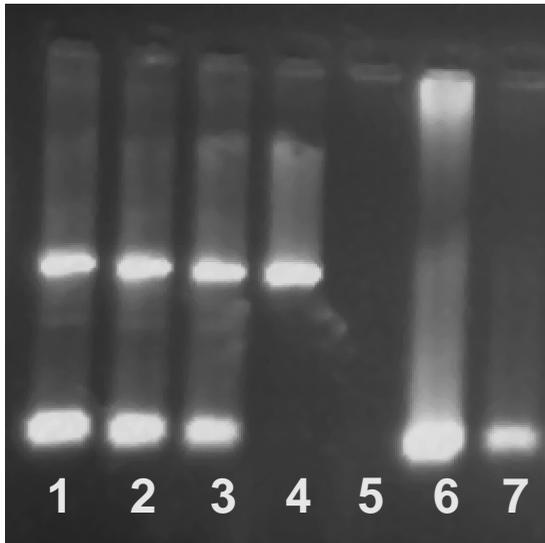
Locus	Tipo	Localização	Sequência dos Primers (5'-3')*	Alelo e Tamanho dos fragmentos (pb)*
AT3-I/D	Indel	1q25.1	CCACAGGTGTAACATTGT	
			GT	1: 572
			GAGATAGTGTGATCTGAG	2: 496
			GC	
APO	Inserção Alu	11q23	AAGTGCTGTAGGCCATTT	1: ~ 400
			AGATTAG	2: ~ 100
PV92	Inserção Alu	16q23.3	AGTCTTCGATGACAGCGT	
			ATACAGA	
			AACTGGGAAAATTTGAAG	1: ~ 400
			AGAAAGT	2: ~ 130
Sb19.3	Inserção Alu	19p12	TGAGTTCTCAACTCCTGTG	
			TGTTAG	
			TCTAGCCCCAGATTTATG	1: ~ 450
			GTAACGT	2: ~ 150
			AAGCACAATTGGTTATTT	
			TCTGAC	

\*LUIZON et al., 2007

### ***Leitura dos resultados no gel de agarose***

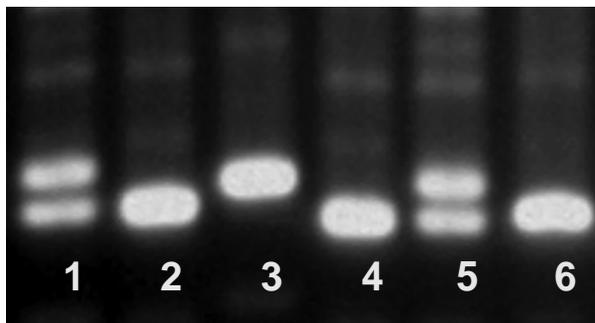
O modo de visualização dos fragmentos de DNA para a detecção e anotação dos AIMs do tipo Inserção Alu (APO, PV92, Sb19.3) está representado na figura 2. Nessa figura, é possível observar a presença de uma inserção Alu homozigota (raia 4), inserção/deleção (raias 1, 2 e 3), deleção homozigota (raias 6 e 7). Na raia 5 não aparece amplificação.

**Figura 2:** Gel de agarose corado com GelRed para visualização de fragmentos de Inserção Alu. O gel digitalizado ilustra a presença de uma inserção Alu homozigota (raia 4), inserção/deleção (raias 1, 2 e 3), deleção homozigota (raias 6 e 7).



A figura 3 representa o modo de visualização dos fragmentos de DNA para a determinação dos genótipos do AIM AT3-I/D (indel). É possível observar a presença de uma inserção homozigota (raia 3), inserção/deleção (raias 1 e 5) e deleção homozigota (raias 2, 4 e 6).

**Figura 3:** Gel de agarose corado com GelRed para visualização de fragmentos de Indel. O gel digitalizado ilustra a presença de uma inserção homozigota (raia 3), inserção/deleção (raias 1 e 5) e deleção homozigota (raias 2, 4 e 6).



## 4.6 TRATAMENTO DOS DADOS E ANÁLISES ESTATÍSTICAS

### 4.6.1 Aderência ao equilíbrio de Hardy-Weinberg

Segundo o teorema de Hardy-Weinberg as frequências genotípicas esperadas no equilíbrio podem ser estimadas a partir da expansão do seguinte binômio:

$$(x_i + x_j)^2 = x_i^2 + 2x_i x_j + x_j^2$$

Onde:

$x_i^2$  é a frequência esperada dos homozigotos do alelo  $i$ ;

$2x_i x_j$  é a frequência esperada do heterozigoto  $ij$ ;

$x_j^2$  é a frequência esperada dos homozigotos para o alelo  $j$ .

A aderência das frequências genotípicas observadas às proporções teóricas de Hardy-Weinberg foi verificada através do programa GENEPOP®, versão 3.4 (RAYMOND; ROUSSET, 1995). Foram realizados três testes baseados na hipótese nula de união aleatória

dos gametas: teste exato de probabilidade, teste para detecção da deficiência e para detecção do excesso de heterozigotos.

No teste exato de probabilidade, o valor de  $p$  corresponde à soma de probabilidades de todas as tabelas com probabilidade menor ou igual ao observado. O segundo e o terceiro são testes mais sensíveis do que o de probabilidade e utilizam uma hipótese alternativa ( $H_1$ ) de excesso ou de deficiência de heterozigotos, respectivamente.

#### 4.6.2 Estimativas das frequências alélicas e genotípicas

As frequências alélicas ( $x_i$ ) e genotípicas ( $X_{ii}$ ) de cada locus, em casos e controles foram estimadas utilizando-se o programa GENEPOP® (RAYMOND; ROUSSET, 1995) versão 3.4 (disponível em <http://wbiomed.curtin.edu.au/genepop>), de acordo com as seguintes equações:

$$x_i = \frac{2n_{ii} + \sum n_{ij}}{2n} \quad \text{e} \quad X_{ii} = \frac{n_{ii}}{n}$$

Onde:

$x_i$  é a frequência do alelo “i”;

$X_{ii}$  é a frequência do genótipo “ii”;

$n_{ii}$  e  $n_{ij}$  correspondem ao número de homozigotos e heterozigotos observados para o alelo i, respectivamente;

$n$  corresponde ao número de indivíduos analisados.

#### 4.6.3 Associações par-a-par entre loci

A análise de LD foi realizada utilizando-se o programa GENEPOP®, versão 3.4 (RAYMOND; ROUSSET, 1995). A hipótese nula é a de que a distribuição genotípica em um locus é independente da distribuição em outro locus. A verificação da ocorrência não aleatória de certas combinações alélicas foi realizada pela seguinte fórmula:

$$\Delta_{ab} = p_{ab} - p_a p_b$$

Onde:

$\Delta_{ab}$  é o valor do desequilíbrio de ligação;

$p_{ab}$  é a frequência observada de um haplótipo;

$p_a$  é a frequência observada do alelo a de um dado locus;

$p_b$  é a frequência observada do alelo b de um outro locus;

$p_a p_b$  é a frequência esperada de um haplótipo ab.

O desequilíbrio de ligação pode assumir valores no intervalo de -1,0 a +1,0, sendo que os valores negativos indicam a repulsão entre os alelos em questão, os valores positivos indicam o acoplamento e, o valor zero indica a ausência de desequilíbrio de ligação.

#### 4.6.4 Diferenciação genética das populações

Os testes exatos para diferenciação populacional foram realizados com o uso do programa GENEPOP® 3.4 (RAYMOND & ROUSSET, 1995a). Este utiliza Tabelas de contingência RxC geradas automaticamente para cada locus, em que R é o número de populações e C é o número de alelos no locus.

Este procedimento compara cada locus em pares de populações, para determinar se existem diferenças nas frequências alélicas e genotípicas observadas, onde a hipótese nula testada é a de que a distribuição alélica é idêntica entre as populações (RAYMOND & ROUSSET, 1995b).

#### 4.6.5 Diversidade interpopulacional (estatísticas F)

As estatísticas F foram obtidas como descrito por WEIR (1984), usando o software GDA® (Genetic Data Analysis Package; LEWIS & ZAYKIN, 1997). Nesta dissertação foram definidos 2 parâmetros:

$F_{IS}$ : Coeficiente de endogamia dentro de uma população (déficit de heterozigotos dentro de uma população);

$F_{ST}$ : Coeficiente de endogamia em um par de populações, dado pela probabilidade de dois alelos tomados ao acaso em duas populações serem idênticos por descendência (déficit de heterozigotos entre populações).

Para testar se os valores de  $F_{ST}$  e  $F_{IS}$  diferiam significativamente de zero, foi realizado o procedimento de bootstrap com 1000 replicações. Se o intervalo de confiança de 95% obtido não incluía o

zero, a estimativa foi considerada significativamente diferente de zero com  $\alpha= 5\%$ .

#### **4.6.6 Estimativas de mistura étnica**

As estimativas das proporções étnicas foram obtidas segundo o método de identidade gênica (CHAKRABORTY, 1985) e foram realizadas com o uso do programa ADMIX 3®.

A estimativa foi realizada primeiro utilizando as frequências nas três populações parentais admitindo-se um modelo tri-híbrido de mistura. As frequências parentais europeias foram representadas por Portugueses (TOMÁS et al., 2002) e as africanas pela frequência observada na cidade de Santana, cidade da Ilha de São Tomé (TOMÁS et al., 2002), tal como realizado por Parra et al. (2003). Esta Ilha serviu como entreposto durante o tráfico negreiro pelo Atlântico e, por isso, possui uma população teve origem em várias regiões Africanas (TOMÁS et al., 2002). A frequência média de sete aldeias da Amazônia Brasileira foi utilizada como parental ameríndia (LUIZON et al., 2008).

### **4.7 BANCO DE DADOS BIOLÓGICOS**

#### **4.7.1 Organização dos dados físicos e virtuais**

Os questionários (dados físicos) estão catalogados em fichários (cada qual para sua doença e/ou população específica), cujo conteúdo estão apenas os questionários e os TCLEs. O LAPOGE possui questionários de doenças autoimunes como Psoríase (PSR), Lúpus Eritematoso Sistêmico (LES) e Artrite Reumatoide (ART), além de questionários referentes aos pacientes diagnosticados com Câncer de Mama (CMA) e, por fim, questionários de populações semi-isoladas e híbrida atual (que não possuem caso clínico conhecido) do estado de Santa Catarina. Os dados virtuais estão armazenados em tabelas (planilhas) assim como os questionários físicos, cada planilha representa um BD referente aos pacientes, controles e populações semi-isoladas.

#### **4.7.2 Diagramação e programação**

Para a diagramação e programação do BDB, dois passos essenciais foram seguidos:

- Identificar as informações (perguntas) comuns em todos os formulários utilizados pelo LAPOGE para a coleta de informações de pacientes, controles e populações em geral.
- Dividir as informações identificadas em categorias: Informações

peçoais, familiares, epidemiológicas, genéticas e de prontuários clínicos.

Após a identificação e divisão das categorias, um diagrama chamado de Entidade Relacional (ER) foi criado para visualizar as informações de cada categoria (formulário) (Apêndice G). O modelo ER é utilizado como uma ferramenta para construir e visualizar um modelo conceitual de BD, auxiliando na definição da estrutura e das restrições, durante a fase de análise de desenvolvimento de um BD (HOFFER et al., 2011).

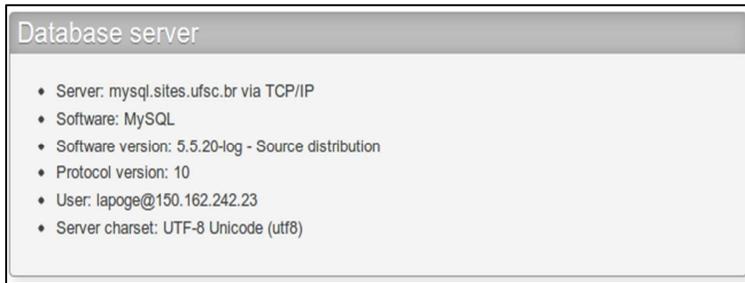
Definido o modelo de ER, o BDB foi programado em linguagem SQL (MySQL), tendo sua interface gráfica desenvolvida em linguagem HTML, com implementações em PHP estabelecendo parâmetros de conexão entre os formulários e MySQL. HTML é a abreviação para HyperTextMarkupLanguage ou Linguagem de Marcação HiperTexto, uma linguagem simples composta de marcações de formatação e diagramação de hipertexto/hipermídia (informações em texto, imagens, sons e ações ligadas umas às outras) e que se destina a escrever documentos que possam ser lidos por browsers (navegadores) (SILVA, 2008). A linguagem de programação PHP (Hypertext Preprocessor) é de ampla utilização, interpretada e de livre acesso, que é especialmente interessante para desenvolvimento de websites, podendo ser mesclada dentro do código HTML. Além disso, possui compatibilidade com a linguagem de BD MySQL para criar páginas web dinâmicas que interagem com os visitantes. Com a combinação PHP/MySQL, pode-se coletar dados de seus usuários e criar conteúdos específicos em tempo real.

A programação foi realizada com o emprego do editor de texto Sublime Text 2, disponível de forma livre na web (<http://www.sublimetext.com/>). A hospedagem local do BDB encontra-se em um computador desktop localizado no LAPOGE. As correções de erros, updates e aprimoramentos do BDB foram realizadas com o emprego do phpMyAdmin (uma ferramenta de software livre escrito em PHP, destinado a lidar com a administração do MySQL pela internet ou localmente) ([www.phpmyadmin.net](http://www.phpmyadmin.net)).

A hospedagem do BDB localiza-se em um servidor (próprio) da UFSC (Figura 4), com senha criptografada e intransferível, para a seguridade dos dados, sendo realizadas cópias de segurança três vezes ao dia para garantir a recuperação dos dados armazenados. Eventuais alterações de correções ou updates foram primeiramente realizadas localmente no computador desktop do LAPOGE e posteriormente

enviadas ao Servidor da UFSC com o auxílio do software livre FileZilla ([www.filezilla-project.org](http://www.filezilla-project.org)).

**Figura 4:** Especificações para o servidor da UFSC para gerenciamento de Banco de Dados.



## 5 RESULTADOS

### 5.1 FREQUÊNCIAS GENOTÍPICAS, ALÉLICAS E EQUILÍBRIO DE HARDY-WEINBERG

As frequências alélicas e genotípicas de cada um dos quatro AIMs (AT3, Sb19.3, APO e PV92) estão descritas na tabela 4. O tamanho amostral para cada locus variou, pois alguns indivíduos não amplificaram.

**Tabela 4:** Comprimento dos fragmentos, localização citogenética e sequência dos *primers* dos quatro AIMs analisados.

Locus	Casos (PSR)	Controles (PSC)
	n (frequência)	
AT3		
Ins	70 (0,361)	64 (0,320)
Del	124 (0,639)	136 (0,680)
Total de Alelos	194	200
Ins/Ins	14 (0,144)	16 (0,160)
Ins/Del	42 (0,433)	32 (0,320)
Del/Del	41 (0,423)	52 (0,520)
Total de indivíduos	97	100

Locus	Casos (PSR)	Controles (PSC)
n (frequência)		
Sb19.3		
Ins	177 (0,903)	170 (0,850)
Del	19 (0,097)	30 (0,150)
Total de Alelos	196	200
Ins/Ins	81 (0,827)	73 (0,730)
Ins/Del	15 (0,153)	24 (0,240)
Del/Del	2 (0,020)	3 (0,030)
Total de indivíduos	98	100
Locus	Casos (PSR)	Controles (PSC)
n (frequência)		
APO		
Ins	189 (0,945)	192 (0,960)
Del	11 (0,055)	8 (0,040)
Total de Alelos	200	200
Ins/Ins	90 (0,900)	93 (0,930)
Ins/Del	9 (0,090)	6 (0,060)
Del/Del	1 (0,010)	1 (0,010)
Total de indivíduos	100	100

Locus	Casos (PSR)	Controles (PSC)
n (frequência)		
PV92		
Ins	46 (0,232)	58 (0,290)
Del	152 (0,768)	142 (0,710)
Total de Alelos	198	200
Ins/Ins	4 (0,040)	12 (0,120)
Ins/Del	38 (0,384)	34 (0,340)
Del/Del	57 (0,576)	54 (0,540)
Total de indivíduos	99	100

As frequências genótípicas dos quatro AIMs foram comparadas aos valores esperados pelo Equilíbrio de Hardy-Weinberg, onde constata-se o equilíbrio nas populações de casos (PSR) e controles (PSC) para os locus APO, Sb19.3 e PV92. Somente um desvio do equilíbrio foi encontrado na população de PSC (locus AT3) (Tabela 5).

**Tabela 5:** Valores de p para o Equilíbrio de Hardy-Weinberg. Valores em equilíbrio para  $p > 0,05$ .

População	AT3	Sb19.3	APO	PV92
PSR	p=0,513	p=0,210	p=0,253	p=0,586
PSC	p=0,011*	p=0,463	p=0,137	p=0,092

\*Valor fora do Equilíbrio de Hardy-Weinberg.

## 5.2 ASSOCIAÇÃO ENTRE LOCI NÃO LIGADOS

Associações significativas entre loci não ligados são indicativas de desequilíbrio de ligação por mistura recente (ALD, do inglês *AdmixtureLinkageDisequilibrium*) (PFAFF et al., 2001) e sugerem a

presença de estrutura genética, razão pela qual foi analisado o padrão de associações par a par entre os AIMs não ligados.

Em cada uma das populações, os quatro AIMs considerados par-a-par permitiram 12 comparações. Não foram observadas associações entre os loci para a população de casos, porém, a população de controles apresentou associação entre os loci APO/Sb19.3 ( $p=0,004$ ) e APO/AT3 ( $p=0,008$ ).

### 5.3 DIFERENCIAÇÃO POPULACIONAL

A diferenciação genética (alélica e genotípica) entre as populações de casos (PSR) e controles (PSC) está descrita na tabela 6. Não houve diferenciação alélica e genotípica significativa entre as duas populações aqui estudadas.

**Tabela 6:** Diferenciação alélica e genotípica entre as populações de casos (PSR) e controles (PSC).

Loci	Diferenciação alélica	Diferenciação genotípica
	PSR X PSC	PSR X PSC
AT3	0,390	0,428
Sb19.3	0,134	0,157
APO	0,637	0,679
PV92	0,215	0,226
Total	0,274	0,330

### 5.4 DIVERSIDADE INTER-POPULACIONAL (ESTATÍSTICAS F)

#### 5.4.1 Coeficiente de endogamia ( $F_{IS}$ ) e diversidade inter-populacional ( $F_{ST}$ )

Ambos, o coeficiente de endogamia ( $F_{IS}$ ) e a diversidade inter-populacional ( $F_{ST}$ ) foram significativos em casos (PSR) e controles (PSC) (Tabela 7).

**Tabela 7:** Valores de Fis e Fst nas populações de casos (PSR) e controles (PSC), adotando-se intervalo de confiança de 95% (IC 95%).

Populações	Fis	IC 95%
PSR	0,700	-0,969 a -0,339
PSC	0,190	0,100 a 0,262
Fst	0,191	0,003 a 0,344

## 5.5 ESTIMATIVAS DE MISTURA ÉTNICAS

As estimativas de mistura étnica apresentaram valores preponderantes de contribuição européia em ambas as populações. A contribuição africana foi a segunda maior nas duas populações, com contribuições muito semelhantes, seguida da contribuição ameríndia (Tabela 8).

**Tabela 8:** Estimativa de mistura étnica nas populações de casos (PSR) e controles (PSC) considerando os quatro AIMs analisados.

População	Europeu	Africano	Ameríndio	r <sup>2</sup>
PSR	87,1 ± 4,7	8,6 ± 2,3	4,3 ± 4,6	87,4
PSC	75,1 ± 4,0	8,4 ± 1,9	16,5 ± 3,8	93,9

## 5.1 BANCO DE DADOS: LAPOGEDb

Os scripts de programação deste BDB não estão incluídos nos Anexos e Apêndices desta dissertação devido ao volume e à quantidade de linhas de programação tornar inviável a visualização em formato impresso. Desta forma, os scripts estão armazenados em um repositório online, podendo ser visualizado mediante requerimento ao e-mail lapoge.mail@gmail.com.

### 5.1.1 Acesso e definições de usuário e senha

O BDB proposto nesta dissertação recebeu o nome de LAPOGEDb (LAPOGE database) e atualmente encontra-se online podendo ser acessado diretamente no endereço eletrônico

www.lapoge.sites.ufsc.br. O site encontra-se hospedado no domínio [ufsc.br]. A solicitação para a utilização do domínio foi realizada pela Dra. Ilíada Rainha de Souza junto a Superintendência de Governança Eletrônica e Tecnologia da Informação e Comunicação (SETIC) da UFSC. A figura 5 mostra página inicial do site do LAPOGE (www.lapoge.ufsc.br), realçando a opção de link para o LAPOGEdb.

**Figura 5:** Página inicial do site LAPOGE com o link no menu à direita da página (destacado em vermelho).

The screenshot shows the homepage of the LAPOGE website. At the top, there is a yellow header with the logo of the Ministério da Educação and the Universidade Federal de Santa Catarina. Below this, a navigation menu includes links for 'Estudante.ufsc.br', 'Professor.ufsc.br', 'STAE.ufsc.br', 'Comunidade.ufsc.br', 'Estrutura.ufsc.br', and 'Geral'. The main content area is titled 'Laboratório de Polimorfismos Genéticos' and features a 'Página do Lapoge' section. This section includes a 'Publicado em 28/11/2012 às 10:18' timestamp and a 'Graduanda: Amanda Fimino' section. The 'Banca Examinadora' section lists several individuals, including Dra. Yara Costa Netto Muniz, Dra. Ilíada Rainha de Souza, Msc. Bibiana Sporia de Almeida, Dr. Erick da Cruz Castelli, and Dra. Andréa Rita Marreto. A 'Lista de Links' sidebar on the right highlights 'LAPOGEdb' in red. The footer contains the text '1960 - 2010 - Universidade Federal de Santa Catarina (UFSC) | Central Telefônica - (48) 3721-9000' and 'Última atualização do site foi em 18 de dezembro 2012 - 10:17:00'.

A página inicial do LAPOGEdb (figura 6), mostra os campos de login de usuário e senha, bem como links úteis relacionados ao LAPOGE e a estudos de genética epidemiológica. Os logins e senhas foram definidos de acordo com os participantes ativos do laboratório (professores, pós-doutores, alunos de doutorado, mestrado e iniciação científica). A revogação do acesso aos participantes é definida ao final de um projeto de pesquisa, caso não haja mais renovação do vínculo destes com o LAPOGE.

A utilização de um sistema de login de usuário e senha torna-se vital para a segurança dos dados, além de garantir que o acesso aos dados seja exclusivo de integrantes ativos do LAPOGE. O controle de acesso ao LAPOGEdb é verificado utilizando um sistema de rastreamento através de data e horário dos usuários que por último atualizaram ou editaram o banco (ELMASRI & NAVATHE, 2010). Além disso, é dado um acesso de controle geral do banco para um dos usuários, definindo-o como "Administrador do Banco de Dados" (DBA, do inglês Database

Administrator). Este usuário é capaz de realizar alterações na estrutura do LAPOGEDb podendo, por exemplo, incluir, alterar ou remover categorias dos formulários.

**Figura 6:** Página inicial do LAPOGEDb. Links e campo de Identificação para usuário e senha.

**Atualização**  
Investigação da variabilidade humana através da análise de polimorfismos genéticos.

**Linhas de Pesquisa**

- ▶ Ancestralidade
- ▶ Artrite Reumatoide
- ▶ Câncer de Mama
- ▶ Lupus Eritematoso Sistêmico
- ▶ Psoríase

**Links**

- ▶ Divergenome
- ▶ NCBI
- ▶ Mendeleey
- ▶ CAPES
- ▶ CNPq
- ▶ Lattes

**IDENTIFICAÇÃO**  
Insira seu nome e sua senha para acessar os dados:

Usuário

Senha

**ENTRAR**

### 5.1.2 Cadastro de questionários

Depois de realizar o login o usuário é direcionado a página de cadastro (Cadastrar Novo Formulário) e edição de formulários (Editar um Formulário) (Figura 7).

**Figura 7:** Página do LAPOGEdb com duas opções: uma para o cadastro de novos questionários e outra opção para a edição de formulários.



A opção de “Cadastrar Novo Formulário” redireciona o usuário à página de Dados Pessoais (DP) (Apêndice H). Neste formulário o usuário preenche informações relacionadas à identificação pessoal, endereço e algumas informações pessoais referentes a amostra. Na caixa de “IDENTIFICAÇÃO PESSOAL”, o tipo de formulário a ser cadastrado no banco é especificado pela opção “CÓDIGO”, onde o usuário pode escolher entre as doenças ou populações estudadas no LAPOGE (Figura 8). Logo abaixo, a opção “COD. ANTIGO (questionário)” remete-se ao código de identificação da amostra anotado no questionário físico a ser cadastrado. Uma das justificativas para a diagramação e construção do LAPOGEdb foi a necessidade de organizar e padronizar a nomenclatura das amostras coletadas pelo LAPOGE ao longo de seus 19 anos de funcionamento. Questionários de uma mesma doença ou população acumularam ao longo destes anos mais de um tipo de nomenclatura (por exemplo, CMA, CaMa e CAMA para designar amostras de pacientes diagnosticados com Câncer de Mama). Utilizar diferentes nomenclaturas para identificar uma mesma variável pode

levar a erros de conflito na busca pela variável em um BD (seja ele construído em tabelas simples ou relacional com o auxílio de um SGBD). Portanto, códigos novos são gerados para cada tipo de questionário cadastrado, obedecendo a nomenclatura de três letras e quatro números (por exemplo, PSR0001 para identificar a primeira amostra a ser cadastrada na variável Psoríase), garantido assim a padronização das nomenclaturas. Estes códigos são gerados automaticamente a medida que um questionário físico é cadastrado e portanto, torna-se extremamente importante o preenchimento da variável que leva a nomenclatura do código antigo, pois desta forma, a correlação entre o código antigo anotado no questionário físico e o novo código gerado automaticamente pelo banco fica estabelecida.

Após o preenchimento do formulário de DP, o usuário pode escolher passar para o próximo formulário e continuar a cadastrar as informações ou limpar os dados do atual formulário. A opção “Salvar / Próxima Etapa” avança ao próximo formulário, enquanto que a opção “Limpar Formulário” apaga todos os campos preenchidos da página de DP.

**Figura 8:** Formulário de Dados Pessoais do LAPOGEdb. Caixa de seleção para o tipo de formulário a ser cadastrado.

Logado como Guilherme DeBortoli ... [Home](#) [Desconectar](#)

**LAPOGE**  
UFSC

**Laboratório de Polimorfismos Genéticos**  
**UFSC CCB BEG LAPOGE**

« voltar Home **Dados Pessoais** Dados Familiais Dados Genéticos Informações Confirmar

**IDENTIFICAÇÃO INICIAL**

CÓDIGO	<input type="text" value="-- seleccione --"/>
COD. ANTIGO (questionário)	<input type="text"/>
Data desta Entrevista	<input type="text"/>
Nome do Entrevistado	<input type="text"/>

**ENDEREÇO**

CEP	<input type="text"/>	Não sabe o CEP? PESQUISE AQUI
Logradouro	<input type="text"/>	
Numero	<input type="text"/>	
Complemento	<input type="text"/>	

O formulário de Dados Familiais (DF) apresenta informações sobre a origem familiar (paterna, materna e irmãos) e históricos de doenças na família (doenças autoimunes e tipos de câncer) (Apêndice I). A parte superior do formulário mostra o novo código gerado automaticamente e o seu respectivo código antigo (preenchido no formulário de DP) (figura 9), informando ao usuário o número de registros atuais cadastrados no banco para está doença/ população, uma vez que, conforme mencionado, o novo código é gerado automaticamente de acordo com o último cadastro realizado no banco. A programação realizada na seção “ANTECEDENTES” com o emprego de caixas de seleção permite o cadastro de infinitos registros de histórico familiar de doenças autoimunes e tipos de cânceres (Figura 10).

**Figura 9:** Formulário de Dados Familiais do LAPOGEdb. Em destaque (cor verde e marrom), o novo código de identificação gerado automaticamente pelo banco e o código antigo, respectivamente.

Logado como Guilherme DeBortoli... [Home](#) [Desconectar](#)

**LAPOGE** UFSC  
Laboratório de Polimorfismos Genéticos  
UFSC CCB BEG LAPOGE

« voltar Home **Dados Pessoais** **Dados Familiais** Dados Genéticos Informações Confirmar

**IDENTIFICAÇÃO INICIAL**

NOVO CÓDIGO	PSR-0001
COD. ANTIGO (questionário)	PSR 01

**DADOS DO PAI**

Nome	<input type="text"/>
Estado/Cidade onde nasceu	-- seleccione -- -- Primeiro seleccione o estado --
Profissão	-- seleccione -- <input type="checkbox"/> Aposentado
Onde nasceu o avô paterno do probando	-- seleccione -- -- Primeiro seleccione o estado --
Onde nasceu a avó paterna do probando	-- seleccione -- -- Primeiro seleccione o estado --

**Figura 10:** Formulário de Dados Familiais do LAPOGEdb. Caixas de seleção para histórico de doenças na família.

**ANTECEDENTES**

Histórico de câncer na família?  Não  Sim

Grau de Parentesco	Avó Paterna Pulmões <a href="#">+ um parente &gt;&gt;</a>
	Irmão Biológico Cólón e Reto <a href="#">Excluir &gt;&gt;</a>
	-- seleccione o parente -- -- seleccione o tipo -- <a href="#">Excluir &gt;&gt;</a>
	-- seleccione o parente -- -- seleccione o tipo -- <a href="#">Excluir &gt;&gt;</a>
Observação	<div style="border: 1px solid black; padding: 5px;"> <ul style="list-style-type: none"> <li>Mama</li> <li>Próstata</li> <li>Pulmões</li> <li>Cólón e Reto</li> <li>Pâncreas</li> <li>Leucemia</li> <li>Linfoma</li> <li>Fígado</li> <li>Ovário</li> <li>Esófago</li> </ul> </div>

Histórico de doença autoimune na família?  Não  Sim

Grau de Parentesco	Tio Paterno Artrite Reumatoide <a href="#">+ um parente &gt;&gt;</a>
Observação	<div style="border: 1px solid gray; height: 100px;"></div>

O formulário de Dados Genéticos (DG) agrega informações de diferentes marcadores genéticos utilizados no LAPOGE (Apêndice J). O objetivo principal deste formulário é obter a informação de quais marcadores a amostra possui genotipado e, quais marcadores ela não possui resultado. A visão geral do quadro de marcadores genotipados para cada uma das amostras cadastradas no LAPOGEDb permite estudos mais complexos do ponto de vista das análises e meta-análises (figura 11). Concentrar os dados genéticos em apenas um BD evita a redundância causada ao armazenar o mesmo tipo de dado várias vezes em diferentes arquivos e, desta forma, a perda de dados (ELMASRI & NAVATHE, 2010).

**Figura 11:** Formulário de Dados Genéticos do LAPOGEDb. Caixas de seleção de marcadores e seus possíveis genótipos.

IDENTIFICAÇÃO INICIAL	
NOVO CÓDIGO	PSR-0001
COD. ANTIGO (questionário)	PSR 01

MARCADORES GENÉTICOS	
14bp (INS/DEL)	INS Homozigoto
HLA-G 3'UTR +3003 (T/C)	INS Homozigoto
HLA-G 3'UTR +3010 (C/G)	INS/DEL
HLA-G 3'UTR +3027 (A/C)	DEL Homozigoto
HLA-G 3'UTR +3035 (C/T)	AA
HLA-G 3'UTR +3142 (G/C)	TT
HLA-G 3'UTR +3187 (A/G)	GC
HLA-G 3'UTR +3196 (G/C)	AG
AT3 (INS/DEL)	CC
APO (INS/DEL)	INS/DEL
PV92 (INS/DEL)	INS Homozigoto
Sb19.3 (INS/DEL)	DEL Homozigoto
	INS/DEL

O formulário de Informações (Apêndice L) está destinado a registrar dados acerca de informações físicas da amostra. Dados como a data da coleta do material biológico, armazenamento e localização dos materiais pós-processamento (separação de sangue e seus constituintes) e métodos de extração de DNA. Um aspecto fundamental da pesquisa biológica gira em torno da coleta de amostras e sua preservação para exame e análises posteriores. Uma vez que o tempo decorre entre o momento da coleta e quando a amostra é de fato analisada, amostras biológicas muitas vezes estão sujeitas a degradar ao longo do tempo. Portanto, é imperioso que haja um processo de armazenamento (curto e

longo prazo) eficiente e que preserve a integridade da amostra (HOLLAND et al., 2003). A catalogação informacional através de um BDB das amostras em seus acondicionamentos (geladeiras e freezers) prevê a organização e o fácil rastreo das mesmas garantindo, assim, o rápido acesso ao material desejado para realização de experimentos (Figura 12).

**Figura 12:** Formulário de Informações do LAPOGEdb. Caixas de seleção para organização e localização física de amostras.

ARMAZENAMENTO	
<input checked="" type="checkbox"/> Sangue Total	Freezer TRÊS
<input type="checkbox"/> Leucócitos	Não se aplica
<input type="checkbox"/> Plasma	Freezer UM
<input type="checkbox"/> Hemácias	Freezer DOIS
<input type="checkbox"/> Soro	Freezer TRÊS
<input type="checkbox"/> Tecido	Freezer QUATRO
	Geladeira
	Não se aplica
Observação	

EXTRAÇÃO DE DNA			
<input checked="" type="checkbox"/> Método da Primeira Extração	Salting Out	Data de Extração	24/10/2011
<input type="checkbox"/> Método da Segunda Extração	Não se aplica	Data de Extração	
<input type="checkbox"/> Método da Terceira Extração	Não se aplica	Data de Extração	

O cadastro das informações termina com o formulário de Informações no botão “CONFIRMAR”. Este botão redireciona o usuário a uma tela de confirmação dos dados, onde todas as variáveis (independente destas estarem ou não preenchidas), são exibidas na tela. Desta forma, o usuário pode conferir em apenas uma página todas as informações cadastradas ao longo dos 4 formulários (Figuras 13 e 14). Após conferir os dados, o botão “CONCLUIR CADASTRO” encerra o cadastro e redireciona todas as informações para o LAPOGEdb.

**Figura 13:** Formulário de Confirmação do LAPOGEDb. Todas as variáveis de todos os formulários são mostrados na tela informando os dados preenchidos e dados não informados (não preenchidos).

IDENTIFICAÇÃO INICIAL	
NOVO CÓDIGO	PSR-0001
COD. ARTIGO (questionário)	PSR 01
Data desta Entrevista	20/07/2011
Nome do Entrevistado	Guilherme Debortoli

ENDEREÇO	
CEP	89227-055
Logradouro	Dado não informado
Numero	0
Complemento	Dado não informado
Bairro	Dado não informado
Local onde Mora	Joinville - SC

DADOS DE CONTATO	
Telefone Residencial	Dado não informado
Telefone Comercial	Dado não informado
Telefone Celular	Dado não informado
E-mail	g.debortoli@gmail.com
Horário para Contato	das 00:00 até 00:00

INFORMAÇÕES PESSOAIS	
Data de Nascimento	20/07/1988
Local de Nascimento	Joinville - SC
Peso	78 Kg
Altura	1,83 m
Tipo Sanguíneo	A
Fator Rh	Positivo
Gênero	Masculino
Profissão	Biólogo
Etnicidade (pelo IBGE)	Branco
Etnicidade (pelo pesquisador)	Branco
Estado Civil	Solteiro
Escolaridade	Dado não informado - Incompleto
Entrevistador	Guilherme Debortoli

**Figura 14:** Formulário de Confirmação do LAPOGEDb. Todas as variáveis de todos os formulários são mostrados na tela informando os dados preenchidos e dados não informados (não preenchidos).

DADOS DO PAI	
Nome	Dado não informado
Local onde nasceu	Dado não informado
Profissão	Dado não informado
Onde nasceu o avô paterno do probando	Dado não informado
Onde nasceu a avó paterna do probando	Dado não informado

DADOS DA MÃE	
Nome	Dado não informado
Local onde nasceu	Dado não informado
Profissão	Dado não informado
Onde nasceu o avô materno do probando	Dado não informado
Onde nasceu a avó materna do probando	Dado não informado

DADOS DOS IRMÃOS	
Possui Irmãos?	Sim

ANTECEDENTES	
Histórico de câncer na família?	Sim
Grau de Parentesco	Desconhece = Próstata Filho = Leucemia
Histórico de doença autoimune na família?	Sim
Grau de Parentesco	Avô Paterno = Artrite Reumatoide Pai = Artrite Reumatoide Irmão Biológico = Artrite Reumatoide Tia Paterna = Artrite Reumatoide
Histórico de outra doença?	Sim
Grau de Parentesco	Pai = Hepatite B

MARCADORES GENÉTICOS	
14bp (INS/DEL)	INS Homozigoto
+3003 (T/C)	TT
+3010 (C/G)	GG
+3027 (A/C)	AA
+3035 (C/T)	
+3142 (G/C)	
+3187 (A/G)	
+3196 (G/C)	GG
AT3 (INS/DEL)	INS/DEL
APO (INS/DEL)	INS Homozigoto
PV92 (INS/DEL)	DEL Homozigoto
Sb19.3 (INS/DEL)	DEL Homozigoto

ARMAZENAMENTO	
Data da Coleta da Amostra	23/10/2011

[« voltar Home](#)

Dados Pessoais

Dados Familiais

Genética

Informações

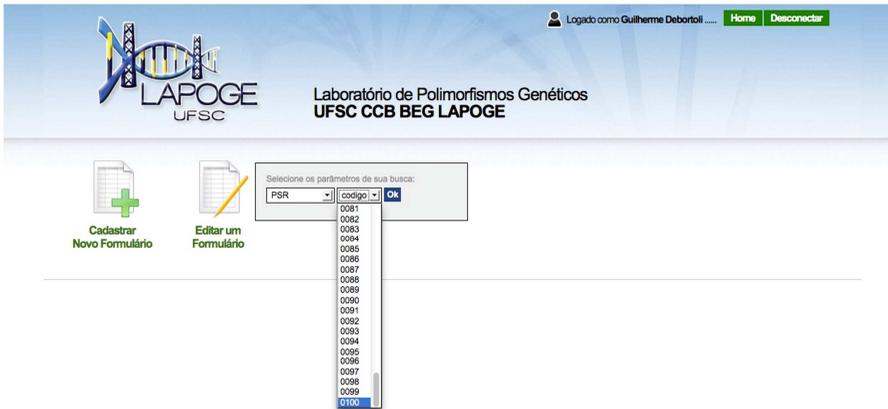
Confirmar

CONCLUIR CADASTRO

### 5.2.3 Edição de formulários

A opção de edição de um formulário abre uma caixa de *select* (seleção), retornando as opções de formulários disponíveis para edição. O retorno dos formulários disponíveis para a edição é realizado com base nos formulários que se encontram cadastrados no banco. Portanto, caso uma determinada doença ou população não possua nenhum registro cadastrado, não haverá retorno desta, pois a mesma não consta cadastrada no banco. A figura 15 mostra o retorno de formulários de PSR (casos de pacientes diagnosticados com psoríase), que foram cadastrados no LAPOGEdb para a realização desta dissertação.

**Figura 15:** Edição de formulários no LAPOGEdb. Formulários de PSR (100) disponíveis para edição de informações.

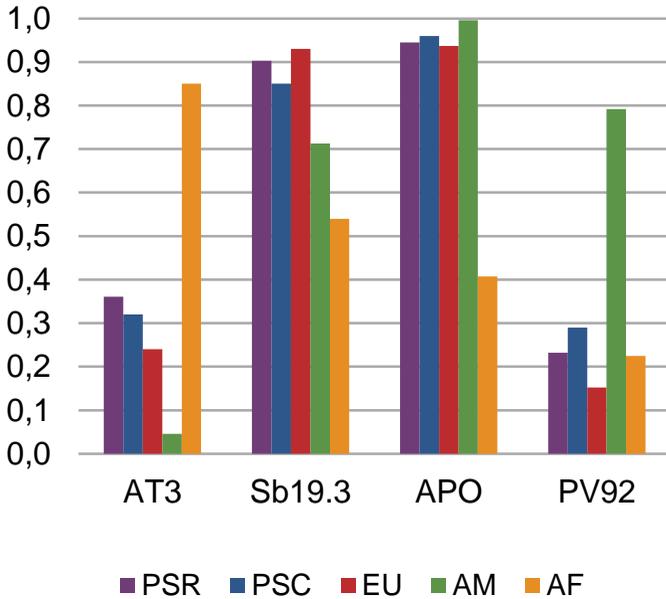


## 6 DISCUSSÃO

### 6.1 FREQUÊNCIAS ALÉLICAS E EQUILÍBRIO DE HARDY-WEINBERG

Das frequências alélicas dos quatro AIMs aqui analisados, três deles (AT3, Sb19.3 e APO) apresentam frequências semelhantes entre europeus e ameríndios (AT3, EU=0,240 e AM=0,046; Sb19.3, EU=0,930 e AM=0,713; APO, EU=0,937 e AM=0,996), porém contrastantes daquelas encontradas em populações africanas (AT3, AF=0,850; Sb19.3, AF=0,540; APO, AF=0,407) (Figura 16) (PARRA et al., 1998; MUNIZ et al., 2008; BRUM et al., 2013). Os marcadores PV92 e AT3 apresentam alta frequência em ameríndios (PV92) e africanos (AT3), porém baixa frequência na população européia e africana (para PV92, EU=0,152 e AF=0,225) e européia e ameríndia (para AT3, EU=0,240 e AM=0,046) (SHRIVER et al., 2003; LUIZON et al., 2008; MUNIZ, 2008; BRUM et al., 2013). Portanto, os marcadores PV92 (capaz de diferenciar ameríndios de europeus e/ou africanos) e AT3 (capaz de diferenciar africanos de europeus e/ou ameríndios) são especialmente importantes no estudo da composição tri-híbrida da população brasileira que, como nos informam os relatos históricos, foi constituída pela união de europeus (principalmente portugueses), africanos (trazidos como mão de obra escrava) e indígenas autóctones (RIBEIRO, 1995).

**Figura 16:** Comparação das frequências alélicas dos AIMs (alelo \*1) entre as populações de casos (PSR) e controles (PSC) com as populações parentais europeia (Portugueses; TOMÁS et al., 2002), ameríndia (LUIZON et al., 2008) e africana (TOMÁS et al., 2002).



Quando comparadas as frequências alélicas dos quatro AIMs nas populações de casos e controles desta dissertação (PSR e PSC, respectivamente) foram encontradas frequências semelhantes em outros dois estudos brasileiros (Apêndice M), sendo o primeiro envolvendo pacientes com Doença Arterial Coronariana e controles sem histórico de doença (ZEMBRZUSKI et al., 2006) e o segundo estudo envolvendo dois grupos de pacientes (um grupo diagnosticado com Neuromielite Óptica e o outro com Esclerose Múltipla) e seus controles sem histórico de doença (BRUM et al., 2013).

Considerando as duas populações de casos e controles aqui analisadas (PSR e PSC) e os loci estudados, apenas um desvio para o EHW ( $p < 0,050$ ) foi observado para o marcador AT3 ( $p = 0,011$ ) na população de controles (PSC). O desvio observado em PSC poderia ser explicado pelo fluxo gênico, uma vez que, os indivíduos pertencem a uma população exogâmica. Outra hipótese

## 6.2 ASSOCIAÇÃO ENTRE LOCI NÃO LIGADOS

O fluxo gênico entre populações geneticamente distintas gera ALD entre os loci (ligados e não ligados) que possuem diferentes frequências alélicas (HARTL & CLARK, 2010). Associações entre loci não ligados sugere miscigenação recente, sendo considerada tão mais recente quanto maior for ALD. Em cada uma das populações, os quatro AIMs considerados par a par permitiram 12 comparações. Duas comparações na população de controles (PSC) apresentaram desvios significativos ( APO/Sb19.3,  $p=0,004$  e APO/AT3,  $p=0,008$ ). Pode-se admitir que estas duas associações seriam indicativo de mistura recente ou até mesmo atribuídas ao acaso.

## 6.3 DIFERENCIAÇÃO POPULACIONAL

A partir de testes exatos, não foram encontradas diferenças alélicas e genóticas entre as duas populações aqui estudadas. Portanto, aceita-se a hipótese de que os alelos e genótipos possuem a mesma distribuição em casos e controles.

## 6.4 DIVERSIDADE INTER-POPULACIONAL

Ambas populações de casos e controles (PSR e PSC, respectivamente) apresentam Fis significativo, ou seja, o intervalo de confiança de 95% obtido não incluiu o zero, considerando a estimativa significativamente diferente de zero com  $\alpha=5\%$ .

Entretanto, as duas populações apresentam história demográfica de mistura étnica característica de populações exogâmicas. Os AIMs analisados neste estudo indicam que ambas as populações são tri-híbridas (Tabela 7 pag 48) e, portanto, tal parâmetro seria provavelmente decorrente da sub-estrutura populacional e não de endogamia.

O Fst obteve um resultado significativo considerado baixo (0,191). Um baixo valor de Fst, significa que as frequências alélicas nas populações de casos e controles são similares, porém com uma pequena diferenciação (HOLSINGER & WEIR, 2009).

## 6.5 ESTIMATIVAS DE MISTURA ÉTNICA

Os resultados aqui observados mostraram que ambas as populações de casos e controles são consideradas tri-híbridas (Tabela 8, pag 48). A preponderante contribuição européia nas duas populações estudadas estão de acordo com outras estimativas obtidas para populações da região Sul do Brasil a partir de AIMs. Parra et al. (2003) observou a predominância europeia na região Sul do Brasil através de

um conjunto de dez AIMs, assim como Santos et al. (2009), com uma estimativa de 94% de contribuição europeia em 81 indivíduos da região Sul do Brasil com um painel de 48 AIMs indels e, Pena et al. (2011), que verificou um componente europeu de 77% em um estudo envolvendo 238 indivíduos do Rio Grande do Sul e Santa Catarina com um painel de 40 AIMs indels. Destaca-se que as estimativas de mistura ancestral entre as populações de casos (PSR) e controles (PSC) são semelhantes, tornando-as aptas para estudos de associação.

A contribuição africana em ambas as populações de casos e controles (8,6% e 8,4%, respectivamente), são similares à estimativa observada em uma amostra Euro-derivada (6,5%) formada por indivíduos controles saudáveis do Rio Grande do Sul utilizando dez AIMs, incluindo três AIMs analisados nesta dissertação (APO, AT3 e Sb19.3) (ZEMBRZUSKI et al., 2006). Foram encontradas estimativas semelhantes em controles saudáveis em dois estudos realizados em Florianópolis, Santa Catarina (14,7% MUNIZ, 2008; 7,8% LUIZON, 2007), com ambos os estudos utilizando o mesmo conjunto de oito marcadores, sendo os quatro AIMs aqui analisados também inclusos nestes. Entretanto, estimativas de mistura contrastantes foram encontradas para europeus e africanos em um estudo no estado da Bahia, onde Teló (2010), observou estimativas de mistura africana (45,08%), europeia (45,16%) e ameríndias (9,75%) (este também utilizando os marcadores aqui descritos, além de outros marcadores autossômicos).

A população de casos (PSR) apresenta um desvio-padrão maior que a própria estimativa de mistura para a população parental ameríndia, o que se deve possivelmente ao número de marcadores utilizados e até mesmo as populações parentais escolhidas uma vez que, há escassez de frequências alélicas das populações ameríndias que possivelmente poderiam ter participado na formação das populações aqui estudadas.

## 7 CONSIDERAÇÕES FINAIS

Os quatro AIMs autossômicos aqui analisados foram capazes de estimar a ancestralidade genômica das populações (casos e controles) deste estudo (Tabela 8 pag), sugerindo a efetividade destes marcadores para o propósito de controle da estratificação populacional. As estimativas de mistura étnica mostraram que as duas populações são tri-híbridas (Tabela 8 pag), com preponderância do componente europeu, seguido do componente africano e ameríndio, respectivamente.

Para conhecimento, este é o primeiro estudo no Brasil que utiliza a psoríase (um dos modelos de doenças autoimune) para estudos de controle da estratificação populacional com o uso de AIMs. A afirmação de que a psoríase (e muitas outras doenças) é predominantemente associada com uma ou outra ascendência genética baseia-se principalmente na percepção visual das características fenotípicas dos pacientes e não necessariamente do background genético. No presente estudo pode-se observar que as contribuições destes grupos ancestrais só apresentam pequenas diferenças entre o grupo de pacientes (PSR) e controles (PSC), e que a contribuição europeia predomina tanto em pacientes como em controles.

Apesar dos poucos AIMs, bem como o pequeno número de indivíduos utilizados neste trabalho, foi possível discriminar a contribuição genética dos grupos ancestrais indicando, que mesmo um número reduzido de marcadores pode ser suficiente, quando apropriadamente selecionados, para responder uma pergunta específica.

As perspectivas futuras incluem a incorporação de um número maior de AIMs para refinar as estimativas aqui apresentadas. Estimar a ancestralidade individual de pessoas diagnosticadas com psoríase poderá auxiliar na compreensão das bases genéticas envolvidas no aparecimento da doença, como também evidenciar a real prevalência desta doença em indivíduos com ascendências europeias, africanas e ameríndias.

O LAPOGEdb mostrou-se uma ferramenta eficaz para o armazenamento estruturado de informações que variam entre dados epidemiológicos e genéticos referentes a amostras utilizadas ao longo de 19 anos de funcionamento do LAPOGE-UFSC.

Os próximos passos serão a implantação de um sistema de busca eficiente, que saia do abstracionismo das linhas de comandos implementadas por programação e torne a busca amigável ao usuário. Desta forma, os usuários do LAPOGEdb poderão realizar suas consultas

de dados sem a necessidade de conhecimento prévio de programação e linhas de comando.

Diversos programas de análises estatísticas para estudos genéticos humanos (populacionais e epidemiológicos) possuem arquivos de entrada de dados distintos. Isto demanda tempo dos pesquisadores para montar arquivos e adequar seus dados para cada uma das análises necessárias. A implementação de uma ferramenta de conversão de dados para os mais variados programas utilizados em estudos genéticos, auxiliaria de forma incisiva os pesquisadores que não demandam de tempo para converter seus dados para diferentes aplicações. Levando isto em consideração, o LAPOGEdb propõe como perspectiva futura, a utilização de um único arquivo com formatação pré-definida, para converter os dados nos formatos que mais se adequam aos principais softwares disponíveis para análises em genética populacional e epidemiológica humana.

Destaca-se também a ideia de que o layout do LAPOGEdb foi programado para a possibilidade do uso deste diretamente nas entrevistas através de tablets, reduzindo a margem de erros durante a coleta de dados. Desta forma, o uso de tablets substituiria a necessidade de realizar questionários físicos para a coleta de dados, exceto o TCLE. As informações serão diretamente cadastradas e enviadas ao LAPOGEdb, semelhante ao que ocorre com os cadastros do censo demográfico brasileiro realizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE), onde as informações coletadas em 2010 pelo agentes do censo são automaticamente cadastradas com o auxílio de um aparelho portátil para tal finalidade.

## REFERÊNCIAS

- ABECASIS, G. R.; ALTSHULER, D.; AUTON, A. et al. A map of human genome variation from population-scale sequencing. *Nature*, v. 467, n. 7319, p. 1061–73, 2010.
- ABECASIS, G. R.; AUTON, A.; BROOKS, L. D. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*, v. 491, n. 7422, p. 56–65, 2012.
- ALMEIDA, R. H. O diretório dos índios: um projeto de “civilização” no Brasil do século XVIII. Brasília Ed UnB 370p, 1997.
- ALONSO-PEREZ, E.; SUAREZ-GESTAL, M.; CALAZA, M. et al. Association of systemic lupus erythematosus clinical features with European population genetic substructure. *PloS one*, v. 6, n. 12, p. e29033, 2011.
- ALONSO-PEREZ, E.; SUAREZ-GESTAL, M.; CALAZA, M. et al. Further evidence of subphenotype association with systemic lupus erythematosus susceptibility loci: a European cases only study. *PloS one*, v. 7, n. 9, p. e45356, 2012.
- ALTSHULER, D.; DALY, M. J.; LANDER, E. S. Genetic mapping in human disease. *Science (New York, N.Y.)*, v. 322, n. 5903, p. 881–8, 2008.
- ALTSHULER, D. M.; GIBBS, R. A.; PELTONEN, L. et al. Integrating common and rare genetic variation in diverse human populations. *Nature*, v. 467, n. 7311, p. 52–8, 2010.
- ARCOT, S. S.; ADAMSON, A. W.; RISCH, G. W. et al. High-resolution cartography of recently integrated human chromosome 19-specific Alu fossils. *Journal of molecular biology*, v. 281, n. 5, p. 843–56, 1998.
- ARDLIE, K. G.; LUNETTA, K. L.; SEIELSTAD, M. Testing for population subdivision and association in four case-control studies. *American journal of human genetics*, v. 71, n. 2, p. 304–11, 2002.
- ATTWOOD, T. K. Concepts, historical milestones and the central place of bioinformatics in modern biology: a European perspective. 2011.
- BATZER, M. A.; ARCOT, S. S.; PHINNEY, J. W. et al. Genetic variation of recent Alu insertions in human populations. *Journal of molecular evolution*, v. 42, n. 1, p. 22–9, 1996.

BATZER, M. A.; RUBIN, C. M.; HELLMANN-BLUMBERG, U. et al. Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. *Journal of molecular biology*, v. 247, n. 3, p. 418–27, 1995.

BATZER, M. A.; STONEKING, M.; ALEGRIA-HARTMAN, M. et al. African origin of human-specific polymorphic Alu insertions. *Proceedings of the National Academy of Sciences of the United States of America*, v. 91, n. 25, p. 12288–92, 1994.

BAXEVANIS, A. D. The importance of biological databases in biological discovery. *Current protocols in bioinformatics*, v. Chapter 1, p. Unit 1.1, 2011.

BEATY, T. H.; MURRAY, J. C.; MARAZITA, M. L. et al. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nature genetics*, v. 42, n. 6, p. 525–9, 2010.

BEDOYA, G.; MONTOYA, P.; GARCÍA, J. et al. Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proceedings of the National Academy of Sciences of the United States of America*, v. 103, n. 19, p. 7234–9, 2006.

BLONDEN, L. A.; TERWINDT, G. M.; DEN DUNNEN, J. T. et al. A polymorphic STS in intron 44 of the dystrophin gene. *Human genetics*, v. 93, n. 4, p. 479–80, 1994.

BONILLA, C.; PARRA, E. J.; PFAFF, C. L. et al. Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. *Annals of human genetics*, v. 68, n. Pt 2, p. 139–53, 2004.

BONILLA, C.; GUTIÉRREZ, G.; PARRA, E. J. et al. Admixture analysis of a rural population of the state of Guerrero, Mexico. *American journal of physical anthropology*, v. 128, n. 4, p. 861–9, 2005.

BRUM, D. G.; LUIZON, M. R.; SANTOS, A. C. et al. European ancestry predominates in neuromyelitis optica and multiple sclerosis patients from Brazil. *PloS one*, v. 8, n. 3, p. e58925, 2013.

BRYC, K.; AUTON, A.; NELSON, M. R. et al. Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proceedings of the National Academy of Sciences of the United States of America*, v. 107, n. 2, p. 786–91, 2010.

CAMPBELL, C. D.; OGBURN, E. L.; LUNETTA, K. L. et al. Demonstrating stratification in a European American population. *Nature genetics*, v. 37, n. 8, p. 868–72, 2005.

CARDON, L. R.; PALMER, L. J. Population stratification and spurious allelic association. *Lancet*, v. 361, n. 9357, p. 598–604, 2003.

CHAKRABORTY, R. Gene identity in racial hybrids and estimation of admixture rates. In Ahuja YR, Neel JV, editors. *Genetic differentiation in human and others animal populations*. Indian Anthropological Association. 1985.

CHOUDHRY, S.; COYLE, N. E.; TANG, H. et al. Population stratification confounds genetic association studies among Latinos. *Human genetics*, v. 118, n. 5, p. 652–64, 2006.

CIRULLI, E. T.; GOLDSTEIN, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature reviews. Genetics*, v. 11, n. 6, p. 415–25, 2010.

CLAYTON, D. G.; WALKER, N. M.; SMYTH, D. J. et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nature genetics*, v. 37, n. 11, p. 1243–6, 2005.

DATE, C. J. *An Introduction to Database Systems (8th Edition)*. Addison-Wesley, 2003.

DENG, H. W. Population admixture may appear to mask, change or reverse genetic effects of genes underlying complex traits. *Genetics*, v. 159, n. 3, p. 1319–23, 2001.

DERENGOSKI, P. R. *A Saga dos Guarani: Guerreiros, Gaúchos e Gaudérios*. Insular, 2002.

DESHMUKH, H.; KIM-HOWARD, X.; NATH, S. K. Replication of recently identified associated single-nucleotide polymorphisms from six autoimmune diseases in Genetic Analysis Workshop 16 rheumatoid arthritis data. *BMC proceedings*, v. 3 Suppl 7, p. S31, 2009.

DUCHARME, V.; GUAUQUE-OLARTE, S.; GAUDREAU, N. et al. NOTCH1 genetic variants in patients with tricuspid calcific aortic valve stenosis. *J Heart Valve Dis*, v. 22, n. 2, p. 142–9, 2013.

DUNCAN, G.; THOMAS, E.; GALLO, J. C. et al. Human phylogenetic relationships according to the DIS80 locus. *Genetica*, v. 98, n. 3, p. 277–87, 1996.

ELHAIK, E.; GREENSPAN, E.; STAATS, S. et al. The GenoChip: A New Tool for Genetic Anthropology. *Genome biology and evolution*, v. 5, n. 5, p. 1021–31, 2013.

ELMASRI, R.; NAVATHE, S. B. *Sistemas de Banco de Dados*. Pearson, 2010.

ELTABAKH, M. Y. *bdbms – A Database Management System for Biological Data*. 2006.

ENOCH, M.; SHEN, P.; XU, K. et al. Using ancestry-informative markers to define populations and detect population stratification. *Journal of psychopharmacology (Oxford, England)*, v. 20, n. 4 Suppl, p. 19–26, 2006.

FAUSTO, B. *Fazer a América: a imigração em Massa para a América Latina*. 1999.

FLORES, M. B. R. *Povoadores da Fronteira: Os casais açorianos rumo ao sul do Brasil*. 2000.

FRAZER, K.; BALLINGER, D. G.; COX, D. R. et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature*, v. 449, n. 7164, p. 851–61, 2007.

FREEDMAN, M. L.; REICH, D.; PENNEY, K. L. et al. Assessing the impact of population stratification on genetic association studies. *Nature genetics*, v. 36, n. 4, p. 388–93, 2004.

GALANTER, J. M.; FERNANDEZ-LOPEZ, J. C.; GIGNOUX, C. R. et al. Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS genetics*, v. 8, n. 3, p. e1002554, 2012.

HAMMER, M. F. A recent insertion of an alu element on the Y chromosome is a useful marker for human population studies. *Molecular biology and evolution*, v. 11, n. 5, p. 749–61, 1994.

HARTL, D. L.; CLARK, A. G. *Principles of population genetics*. 4th ed. Sinauer Associates, Inc. Publishers. Sunderland, Massachusetts, 2010.

HELGASON, A.; YNGVADÓTTIR, B.; HRAFNKELSSON, B. et al. An Icelandic example of the impact of population structure on association studies. *Nature genetics*, v. 37, n. 1, p. 90–5, 2005.

- HIGGINS, D.; TAYLO, W. Bioinformatics: Sequence, Structure and Databanks. Briefings in Bioinformatics, v. 2, n. 2, p. 202–204, 2001.
- HINDS, D.; STOKOWSKI, R. P.; PATIL, N. et al. Matching strategies for genetic association studies in structured populations. American journal of human genetics, v. 74, n. 2, p. 317–25, 2004.
- HOFFER, J. A.; HAMESH, V.; TOPI, H. Modern Database Management. 2011.
- HOGGART, C. J.; PARRA, E. J.; SHRIVER, M. D. et al. Control of confounding of genetic associations in stratified populations. American journal of human genetics, v. 72, n. 6, p. 1492–1504, 2003.
- HOLLAND, N. T.; SMITH, M. T.; ESKENAZI, B. et al. Biological sample collection and processing for molecular epidemiological studies. Mutation research, v. 543, n. 3, p. 217–34, 2003.
- HOLSINGER, K. E.; WEIR, B. S. Genetics in geographically structured populations: defining, estimating and interpreting  $F_{ST}$ . Nature reviews. Genetics, v. 10, n. 9, p. 639–50, 2009.
- INTERNATIONAL HAPMAP PROJECT. Integrating common and rare genetic variation in diverse human populations. v. 467, n. September, 2010.
- JORDAN, D. M.; RAMENSKY, V. E.; SUNYAEV, S. R. Human allelic variation: perspective from protein function, structure, and evolution. Current opinion in structural biology, v. 20, n. 3, p. 342–50, 2010.
- KARATHANASIS, S. K.; OETTGEN, P.; HADDAD, I. A. et al. Structure, evolution, and polymorphisms of the human apolipoprotein A4 gene (APOA4). Proceedings of the National Academy of Sciences of the United States of America, v. 83, n. 22, p. 8457–61, 1986.
- LEWIS, C. M.; KNIGHT, J. Introduction to genetic association studies. Cold Spring Harbor protocols, v. 2012, n. 3, p. 297–306, 2012.
- LEWIS, P. O.; ZAYKIN, D. Genetic data analysis: software for the analysis of discrete genetic data. Version 1.0, 1997.
- LI, J. Z.; ABSHER, D. M.; TANG, H. et al. Worldwide human relationships inferred from genome-wide patterns of variation. Science (New York, N.Y.), v. 319, n. 5866, p. 1100–4, 2008.
- LINS, T. C.; VIEIRA, R. G.; ABREU, B. S. et al. Genetic Heterogeneity of Self-Reported Ancestry Groups in an Admixed Brazilian Population. Journal of Epidemiology, v. 21, n. 4, p. 240–245, 2011.

LIU, Y.; SAHA, N.; LOW, P. S. et al. Linkage disequilibrium between two loci (5' untranslated exon 1 and intron 5-DdeI) of the antithrombin III gene in three ethnic groups in Singapore. *Human heredity*, v. 45, n. 4, p. 192–8, 1995.

LOPES, L. G. M.; RIBEIRO, E. M. R.; CARNEIRO, N. P. S. et al. Afro-derived Amazonian populations: inferring continental ancestry and population substructure. *Human biology*, v. 83, n. 5, p. 627–36, 2011.

LUIZON, M. R.; MENDES-JUNIOR, C. T.; DE OLIVEIRA, S. F. et al. Ancestry informative markers in Amerindians from Brazilian Amazon. *American journal of human biology : the official journal of the Human Biology Council*, v. 20, n. 1, p. 86–90, 2008.

LUIZON, M. R. Dinâmica da Mistura Étnica em Comunidades Remanescentes de Quilombo Brasileiras. Tese de Doutorado, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2007.

MANTA, F. S. N.; PEREIRA, R.; CAIAFA, A. et al. Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers. *Annals of human biology*, v. 40, n. 1, p. 94–8, 2013.

MARRERO, A. R.; SILVA-JUNIOR, W. A.; BRAVI, M. et al. Demographic and Evolutionary Trajectories of the Guarani and Kaingang Natives of Brazil. *American journal of physical anthropology*, 2007.

MARTINEZ-FIERRO, M. L.; BEUTEN, J.; LEACH, R. J. et al. Ancestry informative markers and admixture proportions in northeastern Mexico. *Journal of human genetics*, v. 54, n. 9, p. 504–9, 2009.

MARTINEZ-MARIGNAC, V. L.; VALLADARES, A.; CAMERON, E.; et al. Admixture in Mexico City: implications for admixture mapping of type 2 diabetes genetic risk factors. *Human genetics*, v. 120, n. 6, p. 807–19, 2007.

MCCARTHY, M. I.; ABECASIS, G. R.; CARDON, L. R. et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature reviews. Genetics*, v. 9, n. 5, p. 356–69, 2008.

MILEWICZ, D. M.; BYERS, P. H.; REVEILLE, J. et al. A dimorphic Alu Sb-like insertion in COL3A1 is ethnic-specific. *Journal of molecular evolution*, v. 42, n. 2, p. 117–23, 1996.

MILLER, S. A.; DYKES, D. D.; POLESKY, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic acids research*, v. 16, n. 3, p. 1215, 1988.

MOLINEROS, J. E.; KIM-HOWARD, X.; DESHMUKH, H. et al. Admixture in Hispanic Americans: its impact on ITGAM association and implications for admixture mapping in SLE. *Genes and immunity*, v. 10, n. 5, p. 539–45, 2009.

MUNIZ, Y. C. N. Marcadores Genéticos de Ancestralidade em Comunidades Fundadas por Açorianos na Ilha de Santa Catarina. Tese de Doutorado, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2008.

PACKER, B. R.; YEAGER, M.; BURDETT, L. et al. SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. *Nucleic acids research*, v. 34, n. Database issue, p. D617–21, 2006.

PARRA, E. J.; KITTLES, R.; ARGYROPOULOS, G. et al. Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *American journal of physical anthropology*, v. 114, n. 1, p. 18–29, 2001.

PARRA, E. J.; MARCINI, A.; AKEY, J. et al. Estimating African American admixture proportions by use of population-specific alleles. *American journal of human genetics*, v. 63, n. 6, p. 1839–51, 1998.

PARRA, F. C.; AMADO, R. C.; LAMBERTUCCI, J. R. et al. Color and genomic ancestry in Brazilians. *Proceedings of the National Academy of Sciences of the United States of America*, v. 100, n. 1, p. 177–82, 2003.

PENA, S. D. J.; DI PIETRO, G.; FUCHSHUBER-MORAES, M. et al. The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PloS one*, v. 6, n. 2, p. e17063, 2011.

PEREIRA, R.; PHILLIPS, C.; PINTO, N. et al. Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PloS one*, v. 7, n. 1, p. e29684, 2012.

PFAFF, C. L.; PARRA, E. J.; BONILLA, C. et al. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *American journal of human genetics*, v. 68, n. 1, p. 198–207, 2001.

PRICE, A. L.; BUTLER, J.; PATTERSON, N. et al. Discerning the ancestry of European Americans in genetic association studies. *PLoS genetics*, v. 4, n. 1, p. e236, 2008.

PRITCHARD, J. K.; DONNELLY, P. Case-control studies of association in structured or admixed populations. *Theoretical population biology*, v. 60, n. 3, p. 227–37, 2001.

QIN, P.; LI, Z.; JIN, W. et al. A panel of ancestry informative markers to estimate and correct potential effects of population stratification in Han Chinese. *European journal of human genetics : EJHG*, 2013.

RAYMOND, M.; ROUSSET, F. GENEPOP (version 1.2): population genetics software for exact test and ecumenism. *J Hered* 86: 248-249, 1995a.

RAYMOND, M.; ROUSSET, F. An exact test for population differentiation. *Evolution* 49: 1280-1283, 1995b.

RIBEIRO, D. O povo brasileiro: a formação e o sentido do Brasil. Companhia das Letras, São Paulo, 1995.

RIBEIRO, M.A. C-Reactive Protein: Re-evaluation of a Diagnostic Laboratory Test. *The Brazilian journal of infectious diseases*, v. 1, n. 5, p. 212–225, 1997.

RISCH, N. J. Searching for genetic determinants in the new millennium. *Nature*, v. 405, n. 6788, p. 847–56, 2000.

RODRIGUES, A. D. Flexão relacional no tronco Macro-Jê. *Boletim da Associação Brasileira de Lingüística*, v. 25, p. 219-231, 2000.

RODRIGUES, A. D. Classificação do tronco linguístico Tupi. *Revista Brasileira de Linguística Antropológica*, Volume 3, Número 2, 2011.

ROBBINS, R. J. *Biological Databases: A New Scientific Literature*. 1994.

SALARI, K.; CHOUDHRY, S.; TANG, H. et al. Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. *Genetic epidemiology*, v. 29, n. 1, p. 76–86, 2005.

SALZANO, F. M.; FREIRE-MAIA, N. *Problems in Human Biology: A Study of Brazilian Populations*. Wayne State University Press, 1970.

SALZANO, F. M.; BORTOLINI, M. C. *The Evolution and Genetics of Latin American Populations*. Cambridge University Press, 2002.

SANCHEZ, E.; WEBB, R. D.; RASMUSSEN, A. et al. Genetically determined Amerindian ancestry correlates with increased frequency of risk alleles for systemic lupus erythematosus. *Arthritis and rheumatism*, v. 62, n. 12, p. 3722–9, 2010.

SANTOS, N. P. C.; RIBEIRO-RODRIGUES, E. M.; RIBEIRO-DOS-SANTOS, A. K. C. et al. Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Human mutation*, v. 31, n. 2, p. 184–90, 2010.

SCHIAVETTO, S. N. *A arqueologia Guarani : construção e desconstrução da identidade indígena*. 2003.

SELDIN, M. F.; QI, L.; SCHERBARTH, H. R. et al. Amerindian ancestry in Argentina is associated with increased risk for systemic lupus erythematosus. *Genes and immunity*, v. 9, n. 4, p. 389–93, 2008.

SHRIVER, M. D.; SMITH, M. W.; JIN, L. et al. Ethnic-affiliation estimation by use of population-specific DNA markers. *American journal of human genetics*, v. 60, n. 4, p. 957–64, 1997.

SHRIVER, Mark. D.; PARRA, E. J.; DIOS, S. et al. Skin pigmentation, biogeographical ancestry and admixture mapping. *Human genetics*, v. 112, n. 4, p. 387–99, 2003.

SILVA, M. C. F.; ZUCCHERATO, L. W.; SOARES-SOUZA, G. B. et al. Development of two multiplex mini-sequencing panels of ancestry informative SNPs for studies in Latin Americans: an application to populations of the State of Minas Gerais (Brazil). *Genetics and molecular research : GMR*, v. 9, n. 4, p. 2069–85, 2010.

SILVA, M. S. *Construindo Sites com CSS e (X)HTML*. Novatec, 2008.

STEIN, L. D. *Creating databases for biological information: an introduction*. *Current protocols in bioinformatics*, v. Chapter 9, p. Unit9.1, 2013.

STEIN, L. D. Integrating biological databases. *Nature reviews. Genetics*, v. 4, n. 5, p. 337–45, 2003.

SUAREZ-KURTZ, G. Pharmacogenomics in admixed populations. *Trends in pharmacological sciences*, v. 26, n. 4, p. 196–201, 2005.

TAHAGHOGHI, S.; WILLIAMS, H. *Learning MySQL*. O'Reilly, 2007

TAVARES, L. *Comércio Proibido de Escravos. Ática*, São Paulo, 1988.

THOMAS, D. C.; WITTE, J. S. Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer epidemiology, biomarkers & prevention*, v. 11, n. 6, p. 505–12, 2002.

TELÓ, E. P. Estimativa de mistura étnica avaliada por Marcadores Informativos de Ancestralidade (AIMs) e Microsatélites (STRs). Dissertação de Mestrado, Fiocruz Bahia, Salvador, 2010.

TERREROS, M. C.; ALFONSO-SÁNCHEZ, M. A.; NOVICK, G. E. et al. Insights on human evolution: an analysis of Alu insertion polymorphisms. *Journal of human genetics*, v. 54, n. 10, p. 603–11, 2009.

THOMAS, D. C.; WITTE, J. S. Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer epidemiology, biomarkers & prevention*, v. 11, n. 6, p. 505–12, 2002.

TODD, J. A.; WALKER, N. M.; COOPER, J. D. et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nature genetics*, v. 39, n. 7, p. 857–64, 2007.

TOMÁS, G.; SECO, L.; SEIXAS, S. et al. The peopling of São Tomé (Gulf of Guinea): origins of slave settlers and admixture with the Portuguese. *Human biology*, v. 74, n. 3, p. 397–411, 2002.

VENÂNCIO, R. P. Presença portuguesa: de colonizadores a imigrantes. IBGE. Brasil: 500 anos de povoamento. Rio de Janeiro: IBGE, 2000. p. 61-77.

WALLACE, M. R.; ANDERSEN, L. B.; SAULINO, A. M. et al. A de novo Alu insertion results in neurofibromatosis type 1. *Nature*, v. 353, n. 6347, p. 864–6, 1991.

WATKINS, W. S.; RICKER, C. E.; BAMSHAD, M. J. et al. Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. *American journal of human genetics*, v. 68, n. 3, p. 738–52, 2001.

WEIR, B. S.; COCKERHAM, C. C. Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6):1358-1370, 1984.

WILTON, P.; COLBY, J. *Beginning SQL*. Wrox, 2005.

XIONG, J. *Essential Bioinformatics*. Cambridge University Press, 2006.

YAEGER, R.; AVILA-BRONT, A.; ABDUL, K. et al. Comparing genetic ancestry and self-described race in african americans born in the United States and in Africa. *Cancer epidemiology, biomarkers & prevention*, v. 17, n. 6, p. 1329–38, 2008.

ZEMBRZUSKI, V. M.; CALLEGARI-JACQUES, S. M.; HUTZ, M. H. Application of an African Ancestry Index as a genomic control approach in a Brazilian population. *Annals of human genetics*, v. 70, n. Pt 6, p. 822–8, 2006.

ZIV, E.; BURCHARD, E. G. Human population structure and genetic association studies. *Pharmacogenomics*, v. 4, n. 4, p. 431–41, 2003.



## APÊNDICE A – Populações estudadas pelo Projeto HapMap

Populações estudadas pelo Projeto HapMap.	
Sigla da População	População
ASW	African ancestry in Southwest USA
CEU	Utah residents with Northern and Western European ancestry from the CEPH
CHB	Han Chinese in Beijing, China
CHD	Chinese in Metropolitan Denver, Colorado
GIH	Gujarati Indians in Houston, Texas
JPT	Japanese in Tokyo, Japan
LWK	Luhya in Webuye, Kenya
MEX	Mexican ancestry in Los Angeles, California
MKK	Maasai in Kinyawa, Kenya
TSI	Toscans in Italy
YRI	Yoruba in Ibadan, Nigeria
Adaptada de: ALTSHULER et al., 2010.	

## APÊNDICE B – Populações estudadas pelo Projeto 1000 Genomes

Populações estudadas pelo Projeto 1000 Genomes.	
Sigla da População	População
CHB	Han Chinese in Beijing, China
JPT	Japanese in Tokyo, Japan
CHS	Southern Han Chinese
CDX	Chinese Dai in Xishuangbanna, China
KHV	Kinh in Ho Chi Minh City, Vietnam
CEU	Utah Residents (CEPH) with Northern and Western European ancestry
TSI	Toscani in Italia
FIN	Finnish in Finland
GBR	British in England and Scotland
IBS	Iberian population in Spain
YRI	Yoruba in Ibadan, Nigeria
LWK	Luhya in Webuye, Kenya
GWD	Gambian in Western Divisions in The Gambia
MSL	Mende in Sierra Leone
ESN	Esan in Nigeria
ASW	Americans of African Ancestry in SW USA
ACB	African Caribbeans in Barbados
MXL	Mexican Ancestry from Los Angeles USA
PUR	Puerto Ricans from Puerto Rico
CLM	Colombians from Medellin, Colombia
PEL	Peruvians from Lima, Peru
GIH	Gujarati Indian from Houston, Texas
PJL	Punjabi from Lahore, Pakistan
BEB	Bengali from Bangladesh
STU	Sri Lankan Tamil from the UK
ITU	Indian Telugu from the UK

Adaptada de: ABECASIS et al., 2012.

## APÊNDICE C – Questionário pacientes com psoríase

Universidade Federal de Santa Catarina  
 Departamento de Biologia Celular, Embriologia e Genética/CCB  
 Ambulatório de Dermatologia/Psoríase – Hospital Nereu Ramos  
 Estudo Genético e Epidemiológico em Pacientes com Psoríase no Estado de  
 Santa Catarina

### QUESTIONÁRIO PACIENTES COM PSORÍASE

#### DADOS PESSOAIS

Nome: \_\_\_\_\_

Data: \_\_\_\_\_

Idade: \_\_\_\_\_ anos SEXO: ( ) F ( ) M Ocupação: \_\_\_\_\_

Cidade onde mora: \_\_\_\_\_ Natural de: \_\_\_\_\_

Telefone \_\_\_\_\_ para

Contato: \_\_\_\_\_

Cor da Pele: ( ) B ( ) N ( ) M ( ) P ( ) A

#### DADOS FAMILIARES

Nome do pai: \_\_\_\_\_

Cidade onde nasceu: \_\_\_\_\_

Profissão: \_\_\_\_\_

Ascendência: \_\_\_\_\_ Materna \_\_\_\_\_

Paterna \_\_\_\_\_

Nome da mãe: \_\_\_\_\_

Cidade onde nasceu: \_\_\_\_\_

Profissão: \_\_\_\_\_

Ascendência: \_\_\_\_\_ Materna \_\_\_\_\_

Paterna \_\_\_\_\_

#### HISTÓRICO DA DOENÇA

Tempo de doença diagnosticada: \_\_\_\_\_

Fator desencadeante: \_\_\_\_\_

Local que surgiram as primeiras lesões: \_\_\_\_\_

Classificação da doença na consulta:

( ) ausente ( ) leve ( ) moderada ( ) grave ( )

muito grave

Classificação PASI: \_\_\_\_\_

Padrão Morfológico: \_\_\_\_\_

( ) Vulgar ( ) Gutata ( ) Pustulosa ( ) Eritrodérmica ( ) Palmoplantar ( )  
Invertida

\_\_\_\_\_

Tratamento Atual:

\_\_\_\_\_

**HISTÓRICO DE DOENÇAS PESSOAIS E/OU NA FAMÍLIA**

Psoríase ( )S ( )N Parentesco:

Doença de Crohn ( )S ( )N Artrite ( )S ( )N DM ( )S ( )N HAS ( )S ( )N

Dislipidemia ( )S ( )N Depressão ( )S ( )N Hepatite ( )S ( )N HIV ( )S ( )N

Dermatite Tópica ( )S ( )N Dermatite de Contato ( )S ( )N

Outras ( )S ( )N

Quais? \_\_\_\_\_

Observações: \_\_\_\_\_

**HÁBITOS**

Álcool: ( )S ( )N

Frequência \_\_\_\_\_

Cigarro: ( )S ( )N Cigarros/dia: \_\_\_\_\_ Já fumou? ( )S ( )N

Drogas Ilícitas: ( )S ( )N Já usou? ( )S ( )N Qual? \_\_\_\_\_

Observações: \_\_\_\_\_

**OUTRAS INFORMAÇÕES**

TTT Medicamentoso Antes do Diagnóstico: ( )S ( )N Quais? \_\_\_\_\_

TTT Hormonal Antes do Diagnóstico: ( )S ( )N Qual? ( )AC ( ) Outro \_\_\_\_\_

Idade da Menarca: \_\_\_\_\_ anos Menopausa: ( )S ( )N

Idade: \_\_\_\_\_ anos

Nº de Gestações: \_\_\_\_\_ Nº de

Filhos: \_\_\_\_\_

**PAINEL LABORATORIAL**

Hemograma c/ plaquetas:

ALT/AST: \_\_\_\_\_ γGT: \_\_\_\_\_ Fosfatase Alcalina: \_\_\_\_\_ Uréia: \_\_\_\_\_

Creatinina: \_\_\_\_\_

Ácido Úrico: \_\_\_\_\_ Colesterol Total: \_\_\_\_\_ HDL: \_\_\_\_\_ Triglicerídeos: \_\_\_\_\_

Glicose: \_\_\_\_\_

Na, K, Ca, Mg: \_\_\_\_\_ VHS: \_\_\_\_\_ Proteína C Reativa: \_\_\_\_\_

ASLO: \_\_\_\_\_ HBSAg: \_\_\_\_\_ Anti-HBS: \_\_\_\_\_ Anti-HCV: \_\_\_\_\_ Anti-HIV 1 e 2: \_\_\_\_\_

FAN: \_\_\_\_\_ IgE total: \_\_\_\_\_

Observações: \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

**APÊNDICE D – Questionário indivíduos-controles**

Universidade Federal de Santa Catarina  
 Departamento de Biologia Celular, Embriologia e Genética/CCB  
 Laboratório de Polimorfismos Genéticos

QUESTIONÁRIO  
 INDIVÍDUOS-CONTROLES

Dados Pessoais:

Data: \_\_/\_\_/\_\_

Nome: \_\_\_\_\_

Endereço: \_\_\_\_\_

Cidade: \_\_\_\_\_ Telefone \_\_\_\_\_

Residencial: \_\_\_\_\_

Telefone \_\_\_\_\_ Trabalho: \_\_\_\_\_ Celular: \_\_\_\_\_

Idade: \_\_\_\_\_ Sexo: ( ) M ( ) F Data de nascimento: \_\_\_\_\_

Estado Civil: \_\_\_\_\_ Tipo de sangue: \_\_\_\_\_

Profissão: \_\_\_\_\_ Aposentado: ( ) Sim ( ) Não

Escolaridade:

( ) analfabeto ( ) 1º grau incompleto ( ) 1º grau completo

( ) 2º grau incompleto ( ) 2º grau completo ( ) superior incompleto

( ) superior completo ( ) pós graduação

Peso: \_\_\_\_\_ Altura: \_\_\_\_\_

Cidade \_\_\_\_\_ onde nasceu: \_\_\_\_\_

Ascendência:

Materna \_\_\_\_\_ Paterna \_\_\_\_\_

Etnia: ( ) Euro descendente ( ) Afro descendente

( ) Asiático descendente ( ) Indígena descendente

Cor da pele: ( ) negra ( ) mulata ( ) amarela ( ) branca

Observações:

\_\_\_\_\_

Dados Familiares:

Nome \_\_\_\_\_ do \_\_\_\_\_ pai:

Cidade onde nasceu: \_\_\_\_\_

Descendência \_\_\_\_\_ do \_\_\_\_\_ pai:

Materna \_\_\_\_\_ Paterna \_\_\_\_\_

Profissão: \_\_\_\_\_

Nome \_\_\_\_\_ da \_\_\_\_\_ mãe:

Cidade onde nasceu: \_\_\_\_\_

Descendência \_\_\_\_\_ da \_\_\_\_\_ mãe:

Materna \_\_\_\_\_ Paterna \_\_\_\_\_

Profissão: \_\_\_\_\_

Possui Irmãos: ( ) Sim ( ) Não Quantos: \_\_\_\_\_

Possui filhos: ( ) Sim ( ) Não Quantos: \_\_\_\_\_

Ingere BEBIDA ALCOÓLICA? ( ) Sim ( ) Não

Frequência: ( ) Todos os dias ( ) Fim de semana

( ) Esporadicamente (Festas)

Quantidade \_\_\_\_\_ (copos) 200ml):

Que tipo de bebida alcoólica ingere mais frequentemente?

( ) Cerveja ( ) Vinho ( ) Cachaça ( ) Outro

Que tipo de bebida alcoólica nunca ingere?

( ) Cerveja ( ) Vinho ( ) Cachaça ( ) Outro

Pratica EXERCÍCIOS FÍSICOS? ( ) Sim ( ) Não

Tipo:

Quantidade: ( ) menos de 30 min ( ) 30 min ( ) 1h ( ) mais de 1 h

Frequência: ( ) 1x semana ( ) 2-3x semana ( ) 4-6x semana

( ) Todo os dias ( ) Menos de 1x semana

Você FUMA? ( ) Sim ( ) Não Você já FUMOU? ( ) Sim ( ) Não

Tipo: ( ) Cigarro ( ) Charuto ( ) Cachimbo ( ) Outro

Quantidade e Frequência (nº de cigarros por dia): \_\_\_\_\_

Tempo que fuma ou fumou: \_\_\_\_\_

Há quanto tempo parou: \_\_\_\_\_

Histórico Hormonal e Reprodutivo

Idade da MENARCA: \_\_\_\_\_

MENOPAUSA: ( ) Sim ( ) Não Idade: \_\_\_\_\_

HISTERECTOMIA: ( ) Sim ( ) Não

PARIDADE:

Nº de gestações \_\_\_\_\_ Idade da 1ª  
gestação \_\_\_\_\_

Nº de filhos ( ) nulípara N: \_\_\_\_\_

Abortos ( ) P ( ) E N: \_\_\_\_\_

Amamentou: ( ) Sim ( ) Não Tempo total (meses):  
\_\_\_\_\_

TRAT. HORMONAL:

Utiliza AC? ( ) Sim ( ) Não

Já utilizou AC? ( ) Sim ( ) Não

Nome e tipo (oral, adesivo, injetável) do  
AC: \_\_\_\_\_Tempo que usa ou usou  
AC: \_\_\_\_\_Há quanto tempo  
parou? \_\_\_\_\_

Faz TRH? ( ) Sim ( ) Não

Já fez TRH? ( ) Sim ( ) Não

Nome do  
Hormônio: \_\_\_\_\_Tempo que faz ou fez  
TRH: \_\_\_\_\_Há quanto tempo  
parou? \_\_\_\_\_( ) Outros hormônios \_\_\_\_\_ Tempo total:  
\_\_\_\_\_Observações \_\_\_\_\_  
\_\_\_\_\_

Histórico Médico

Caso de CÂNCER pessoal? ( ) Sim ( ) Não

Tipo:  
\_\_\_\_\_

Casos de CÂNCER DE MAMA na família? ( ) Sim ( ) Não

Grau de Parentesco: ( ) filha ( ) irmã ( ) mãe ( ) avó  
( ) tia materna 1º grau ( ) tia paterna 1º grau  
( ) prima materna 1º grau ( ) prima paterna 1º grau  
( ) \_\_\_\_\_ )

Outros \_\_\_\_\_

Casos de CÂNCER de outro tipo na família? ( ) Sim ( ) Não

Grau de Parentesco e tipo:

---



---

Caso de TUMOR BENIGNO pessoal? ( ) Sim ( ) Não

Local: \_\_\_\_\_

Caso de DOENÇA AUTOIMUNE pessoal? ( ) Sim ( ) Não

Qual? \_\_\_\_\_

Tempo de diagnóstico: \_\_\_\_\_

Casos de DOENÇA AUTOIMUNE na família? ( ) Sim ( ) Não

Grau de Parentesco e tipo: \_\_\_\_\_

Você tem alguma DOENÇA CARDIOVASCULAR?: ( ) Sim ( ) Não

Qual?

(s) \_\_\_\_\_

HIPERTENSÃO ARTERIAL: ( ) Sim ( ) Não

HIPERCOLESTEROLEMIA: ( ) Sim ( ) Não

OSTEOPOROSE: ( ) Sim ( ) Não

DOENÇA REUMÁTICA: ( ) Sim ( ) Não

DIABETES: ( ) Sim ( ) Não

ASMA: ( ) Sim ( ) Não

HIV: ( ) Sim ( ) Não ( ) Nunca fez exame

HEPATITE: ( ) Sim ( ) Não ( ) Nunca fez exame

DENGUE: ( ) Sim ( ) Não

TUBERCULOSE: ( ) Sim ( ) Não

DISTÚRPIO RENAL: ( ) Sim ( ) Não

DISTÚRPIO PULMONAR: ( ) Sim ( ) Não

DISTÚRPIO HEPÁTICO: ( ) Sim ( ) Não

Casos de DOENÇA DE ALZHEIMER na família: ( ) Sim ( ) Não

Grau de parentesco: \_\_\_\_\_

Casos de DOENÇA DE PARKINSON na família: ( ) Sim ( ) Não

Grau de parentesco: \_\_\_\_\_

OUTRAS

DOENÇAS?: \_\_\_\_\_

Alérgico a algum medicamento? \_\_\_\_\_

medicamento? \_\_\_\_\_

Alérgico a algum alimento? \_\_\_\_\_

Teve DEPRESSÃO? \_\_\_\_\_

Utilizou ou utiliza alguma medicação por longo tempo? ( ) Sim ( ) Não  
Nome do medicamento (dosagem e frequência) e tempo que utilizou:

\_\_\_\_\_  
\_\_\_\_\_

## APÊNDICE E – Termo de consentimento livre esclarecido

Universidade Federal de Santa Catarina  
Departamento de Biologia Celular, Embriologia e Genética/CCB  
Ambulatório de Dermatologia/Psoríase – Hospital Nereu Ramos  
Estudo Genético e Epidemiológico em Pacientes com Psoríase no Estado de  
Santa Catarina

### TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PACIENTES COM PSORÍASE

Projeto de Pesquisa: “Estudo genético e epidemiológico em pacientes com psoríase no estado de Santa Catarina”.

Informações: Pesquisadores da Universidade Federal de Santa Catarina estão desenvolvendo um projeto de pesquisa para avaliação de fatores genéticos, doenças e hábitos alimentares e pessoais que podem estar associados com a psoríase. Para isto pedimos sua permissão para coletarmos uma amostra de sangue sua (cerca de 10 ml) e colaboração para responder um questionário que dura aproximadamente quinze minutos. Com essa amostra, nós iremos extrair uma quantidade pequena de DNA (molécula que contém os genes, que são as informações de suas características biológicas). O DNA será analisado no laboratório para tentarmos descobrir se há relação entre alguns de seus genes, propostos no atual projeto (genes relacionados a células de defesa do organismo) e esta doença. A amostra coletada nesta ocasião poderá ser utilizada em possíveis futuros projetos que envolvam testes genéticos, aprovados pelo sistema CEP/CONEP, desde que receba novamente sua autorização, após um novo contato. Deixamos claro que sua participação é voluntária, não influenciando no seu atendimento e tratamento. A equipe agradece antecipadamente sua colaboração e se coloca à sua disposição para responder qualquer pergunta que você queira fazer, e esclarecer quaisquer dúvidas que porventura apareçam. Para isso você pode telefonar para o número (48) 3721-9804 e conversar com a Prof<sup>a</sup> Dra. Ilíada Rainha de Souza ou com qualquer outro participante do estudo.

Riscos: A coleta de sangue é um procedimento normal para o tratamento da sua doença. O aparecimento de mancha roxa ou dor no local da espetada da agulha podem ocorrer sem representar maiores preocupações. As informações coletadas e os resultados das análises genéticas serão mantidos em sigilo e serão utilizados somente pela equipe da pesquisa.

Custos: Como já dissemos a sua participação nessa pesquisa é voluntária, não havendo, portanto, nem custos nem pagamento para isso.

Benefícios: Você não terá nenhum benefício direto ao participar desta pesquisa, mas os resultados deste estudo poderão no futuro proporcionar novas alternativas para tratamento dessa doença, bem como para identificação de pessoas que tem risco de desenvolvê-la.

Desistência: Caso você desista de participar desse estudo, mesmo já tendo assinado esse termo, você só precisa entrar em contato com a equipe realizadora, através do telefone (48) 3721 9804, e informar a sua desistência. Nenhum tipo de penalidade lhe será aplicada.

#### DECLARAÇÃO DE CONSENTIMENTO

Eu, \_\_\_\_\_, fui esclarecido (a) sobre a pesquisa “Estudo genético e epidemiológico em pacientes com psoríase no estado de Santa Catarina”, e concordo que meus dados sejam utilizados na realização da mesma e autorizo a guarda de meu material biológico para o caso de futuras pesquisas, sendo eu contatado (a) para fornecer nova autorização caso forem realizadas novas pesquisas não mencionadas neste projeto.

Florianópolis, \_\_\_\_\_

Assinatura: \_\_\_\_\_

RG: \_\_\_\_\_

## **APÊNDICE F – Termo de consentimento livre esclarecido indivíduos-controles**

Universidade Federal de Santa Catarina  
Departamento de Biologia Celular, Embriologia e Genética/CCB  
Laboratório de Polimorfismos Genéticos

### **TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO INDIVÍDUOS-CONTROLES**

Projetos de Pesquisa: “Câncer de mama: avaliação de parâmetros informativos para diagnóstico e prognóstico na população do estado de Santa Catarina”, “Genética da autoimunidade: polimorfismos em Lúpus Eritematoso Sistêmico e Artrite Reumatóide em pacientes de Santa Catarina” e “Estudo Genético e Epidemiológico em Pacientes com Psoríase no Estado de Santa Catarina”

Informações: Pesquisadores da Universidade Federal de Santa Catarina estão desenvolvendo projetos de pesquisa para avaliação de fatores genéticos, doenças e hábitos alimentares e pessoais que podem estar associados ao aparecimento do câncer de mama e doenças autoimunes. Para isto pedimos sua colaboração e permissão para fazer parte do grupo controle e para extrairmos de parte de seu material biológico, uma quantidade pequena de DNA (molécula que contém os genes, que são as informações de suas características biológicas). O DNA será analisado no laboratório para tentarmos descobrir se há relação entre alguns genes, propostos nos atuais projetos (ligados ao metabolismo de hormônios sexuais, de substâncias estranhas ao organismo, relacionados ao reparo de DNA e sistema imune) e o aparecimento destas doenças. A amostra coletada nesta ocasião poderá ser utilizada em possíveis futuros projetos que envolvam testes genéticos, aprovados pelo sistema CEP/CONEP, desde que receba novamente sua autorização, após um novo contato. Deixamos claro que sua participação é voluntária. A equipe agradece antecipadamente sua colaboração e se coloca à sua disposição para responder qualquer pergunta que você queira fazer, e esclarecer quaisquer dúvidas que porventura apareçam. Para isso você pode telefonar para o número (48) 3721-9804 e conversar com a Profa. Dra. Ilfada Rainha de Souza ou seus orientandos.

Procedimentos: Caso você concorde em participar, você irá preencher um questionário para sabermos seus dados pessoais (como nome, endereço e telefone) e irá assinar um termo de consentimento livre e esclarecido para que possamos utilizar seus dados pessoais e material biológico nestas pesquisas. Também precisaremos tirar um pouco de sangue numa seringa. O DNA extraído das amostras coletadas será guardado no Laboratório sob responsabilidade da coordenadora do projeto.

Entraremos em contato pelo telefone fornecido o mais breve possível para realizarmos um novo questionário de duração máxima de 20 minutos. Este questionário irá conter dados como seus hábitos alimentares e pessoais,

histórico reprodutivo e histórico clínico, essenciais para o desenvolvimento das pesquisas.

Riscos: A coleta de sangue é um procedimento normal para a realização de vários exames. O aparecimento de mancha roxa ou dor no local da espetada da agulha podem ocorrer sem representar maiores preocupações. As informações coletadas, bem como os resultados das análises genéticas serão mantidas em sigilo e serão utilizadas somente pela equipe da pesquisa.

Custos: Você não precisará pagar nada para fazer parte deste estudo.

Benefícios: Você não terá nenhum benefício direto ao participar desta pesquisa, mas os resultados deste estudo poderão no futuro proporcionar novas alternativas para prevenção do câncer de mama e doenças autoimunes, e para identificação de pessoas que tem risco de desenvolver essas doenças, podendo beneficiar muitas outras pessoas.

Assinaturas:

Pesquisador

responsável

Florianópolis, \_\_\_/\_\_\_/\_\_\_\_\_

#### DECLARAÇÃO DE CONSENTIMENTO

Eu, \_\_\_\_\_, fui esclarecido(a) sobre as pesquisas “Câncer de mama: avaliação de parâmetros informativos para diagnóstico e prognóstico na população do estado de Santa Catarina”, “Genética da autoimunidade: polimorfismos em lúpus eritematoso sistêmico e artrite reumatóide em pacientes de Santa Catarina” e “Estudo Genético e Epidemiológico em Pacientes com Psoríase no Estado de Santa Catarina” e concordo que meus dados sejam utilizados na realização das mesmas.

Florianópolis, \_\_\_\_\_

Assinatura: \_\_\_\_\_

RG: \_\_\_\_\_

## APÊNDICE G – Modelo Entidade Relacional do BDB



# APÊNDICE H – Página do formulário de Dados Pessoais do LAPOGEdb


Logado como: Guilherme Deborral ...
[Home](#)
[Desconectar](#)

**Laboratório de Polimorfismos Genéticos**  
**UFSC CCB BEG LAPAGE**

[« voltar Home](#)
 [Dados Pessoais](#)
 [Dados Familiais](#)
 [Dados Genéticos](#)
 [Informações](#)
 [Confirmar](#)

### IDENTIFICAÇÃO INICIAL

CÓDIGO	== seletore ==
COD. ANTIGO (questionário)	<input type="text"/>
Data desta Entrevista	<input type="text"/>
Nome do Entrevistado	<input type="text"/>

### ENDEREÇO

CEP	<input type="text"/> Não sabe o CEP? PESQUISE AQUI
Logradouro	<input type="text"/>
Numero	<input type="text"/>
Complemento	<input type="text"/>
Bairro	<input type="text"/>
Cidade/Estado	== seletore == --> == Primeiro seletore o estado -->

### DADOS DE CONTATO

Telefone Residencial	<input type="text"/>
Telefone Comercial	<input type="text"/>
Telefone Celular	<input type="text"/>
E-mail	<input type="text"/>
Horário para Contato	de <input type="text"/> às <input type="text"/> (usar padrão de 24 horas)

### INFORMAÇÕES PESSOAIS

Data de Nascimento*	<input type="text"/>
Local de Nascimento	== seletore == --> == Primeiro seletore o estado -->
Peso	<input type="text"/> (em quilos, com uma casa decimal)
Altura	<input type="text"/> (com duas casas decimais)
Tipo Sanguíneo	<input type="radio"/> A <input type="radio"/> B <input type="radio"/> AB <input type="radio"/> O
Fator Rh	<input type="radio"/> Positivo <input type="radio"/> Negativo
Gênero	<input type="radio"/> Masculino <input type="radio"/> Feminino
Profissão	== seletore == <input type="checkbox"/> Aposentado
ETNICIDADE (pelo IBGE)	<input type="radio"/> Branco <input type="radio"/> Indio <input type="radio"/> Preto <input type="radio"/> Pardo <input type="radio"/> Amarelo
ETNICIDADE (pelo pesquisador)	<input type="radio"/> Branco <input type="radio"/> Mulato <input type="radio"/> Amarelo <input type="radio"/> Negro
ESTADO CIVIL	<input type="radio"/> Solteiro <input type="radio"/> Casado/ União Estável <input type="radio"/> Divorçado <input type="radio"/> Viúvo
ESCOLARIDADE	<input type="radio"/> Analfabeto <input type="radio"/> Completo <input type="radio"/> Ensino Fundamental <input type="radio"/> Incompleto <input type="radio"/> Ensino Médio <input type="radio"/> Graduação <input type="radio"/> Especialização <input type="radio"/> Pós-Graduação
Observação	<input style="width: 100%; height: 50px;" type="text"/>
Entrevistador	Não Identificado

Limpar Tudo
Salvar / Próxima Etapa

# APÊNDICE I – Página do formulário de Dados Familiais do LAPOGEdb

Logado como: Guilherme Deborral ... [Home](#) | [Desconectar](#)



**Laboratório de Polimorfismos Genéticos  
UFSC CCB BEG LAPOGE**

[< voltar Home](#)
[Dados Pessoais](#)
[Dados Familiais](#)
[Dados Genéticos](#)
[Informações](#)
[Confirmar](#)

---

**IDENTIFICAÇÃO INICIAL**

Nome do usuário	PSR-0001
COD. ANTIGO (questionário)	PSR 01

**DADOS DO PAI**

Nome	<input type="text"/>
Estado/Cidade onde nasceu	-- seletione -- >   -- Primeiro seletione o estado -- >
Profissão	-- seletione -- >   <input type="checkbox"/> Apertado
Quem nasceu e avô paterno do probando	-- seletione -- >   -- Primeiro seletione o estado -- >
Quem nasceu e avô materno do probando	-- seletione -- >   -- Primeiro seletione o estado -- >
Observação	<div style="border: 1px solid #ccc; height: 40px;"></div>

**DADOS DA MÃE**

Nome	<input type="text"/>
Estado/Cidade onde nasceu	-- seletione -- >   -- Primeiro seletione o estado -- >
Profissão	-- seletione -- >   <input type="checkbox"/> Apertado
Quem nasceu e avô materno do probando	-- seletione -- >   -- Primeiro seletione o estado -- >
Quem nasceu e avô paterno do probando	-- seletione -- >   -- Primeiro seletione o estado -- >
Observação	<div style="border: 1px solid #ccc; height: 40px;"></div>

**DADOS DOS IRMÃOS**

Possui irmãos?	<input type="radio"/> Não <input type="radio"/> Sim
Irmãos Biológicos	Não possui irmãos >
Irmãos Distintos	Não possui irmãos >
Irmãos Adotivos	Não possui irmãos adotivos >
Irmãos Adotivos	Não possui irmãos adotivos >

**ANTECEDENTES**

Histórico de alcool na família?	<input type="radio"/> Não <input type="radio"/> Sim
Observação	<div style="border: 1px solid #ccc; height: 40px;"></div>
Histórico de doença autoimune na família?	<input type="radio"/> Não <input type="radio"/> Sim
Observação	<div style="border: 1px solid #ccc; height: 40px;"></div>
Histórico de outra doença?	<input type="radio"/> Não <input type="radio"/> Sim
Observação	<div style="border: 1px solid #ccc; height: 40px;"></div>

Salvar / Próxima Etapa

## APÊNDICE J – Página do formulário de Dados Genéticos do LAPOGEdb



Logado como Guilherme Debartoli ..... [Home](#) [Desconectar](#)

**Laboratório de Polimorfismos Genéticos**  
**UFSC CCB BEG LAPOGE**

« voltar Home
Dados Pessoais
Dados Familiares
Dados Genéticos
Informações
Confirmar

**IDENTIFICAÇÃO INICIAL**

Novo código	PSR-0001
COD. ANTIGO (questionário)	PSR 01

**MARCADORES GENÉTICOS**

14bp (INS/DEL)	-- seleccione -- ▾
HLA-G 3'UTR +3003 (T/C)	-- seleccione -- ▾
HLA-G 3'UTR +3010 (C/G)	-- seleccione -- ▾
HLA-G 3'UTR +3027 (A/C)	-- seleccione -- ▾
HLA-G 3'UTR +3035 (C/T)	-- seleccione -- ▾
HLA-G 3'UTR +3142 (G/C)	-- seleccione -- ▾
HLA-G 3'UTR +3187 (A/G)	-- seleccione -- ▾
HLA-G 3'UTR +3196 (G/C)	-- seleccione -- ▾
AT3 (INS/DEL)	-- seleccione -- ▾
APO (INS/DEL)	-- seleccione -- ▾
PV92 (INS/DEL)	-- seleccione -- ▾
Sb19.3 (INS/DEL)	-- seleccione -- ▾

Salvar / Próxima Etapa

## APÊNDICE L – Página do formulário de Informações do LAPOGEdb



Logado como Guilherme Debortoli ...
Home
Desconectar

**Laboratório de Polimorfismos Genéticos**  
**UFSC CCB BEG LAPOGE**

« voltar Home
Dados Pessoais
Dados Familiais
Dados Genéticos
Informações
Confirmar

### IDENTIFICAÇÃO E DATA DE COLETA

NOVO CÓDIGO	PSR-0001
ODD_ ANTIGO (questionário)	PSR 01
DATA DE COLETA DA AMOSTRA	<input type="text"/>

### ARMAZENAMENTO

■ Sangue Total	<input type="text" value="Não se aplica"/>
■ Leucócitos	<input type="text" value="Não se aplica"/>
■ Plasma	<input type="text" value="Não se aplica"/>
■ Hemácias	<input type="text" value="Não se aplica"/>
■ Soro	<input type="text" value="Não se aplica"/>
■ Tecido	<input type="text" value="Não se aplica"/>
Observação	<div style="border: 1px solid #ccc; height: 100px;"></div>

### EXTRAÇÃO DE DNA

■ Método da Primeira Extração	<input type="text" value="Não se aplica"/>	Data da Extração	<input type="text"/>
■ Método da Segunda Extração	<input type="text" value="Não se aplica"/>	Data da Extração	<input type="text"/>
■ Método da Terceira Extração	<input type="text" value="Não se aplica"/>	Data da Extração	<input type="text"/>
Observação	<div style="border: 1px solid #ccc; height: 100px;"></div>		

### ALÍQUOTAS DE DNA

■ Concentração Estoque	<input type="text" value="Não se aplica"/>
■ Concentração de Uso	<input type="text" value="Não se aplica"/>
■ Diluição	<input type="text" value="Não se aplica"/>
Observação	<div style="border: 1px solid #ccc; height: 100px;"></div>

Limpar Tudo
Confirmar Dados

**APÊNDICE M – Frequências dos AIMs (AT3, Sb19.3, APO, PV92) em dois estudos brasileiros do tipo caso-controle**

<b>Zembrzusi <i>et al.</i>, 2006*</b>	<b>Casos (DAC)**</b>	<b>Controles</b>
AT3	0,306	0,235
Sb19.3	0,854	0,819
APO	0,942	0,951

\*O marcador PV92 não foi estudado em Zembrzusi *et al.*, 2006.

\*\*DAC = Doença Arterial Coronariana.

<b>Brum <i>et al.</i>, 2013</b>	<b>Casos (NO)*</b>	<b>Casos (EM)**</b>	<b>Controles</b>
AT3	0,483	0,310	0,421
Sb19.3	0,698	0,794	0,804
APO	0,888	0,883	0,874
PV92	0,310	0,258	0,336

\*NO= Neuromielite Óptica

\*\*EM = Esclerose Múltipla

