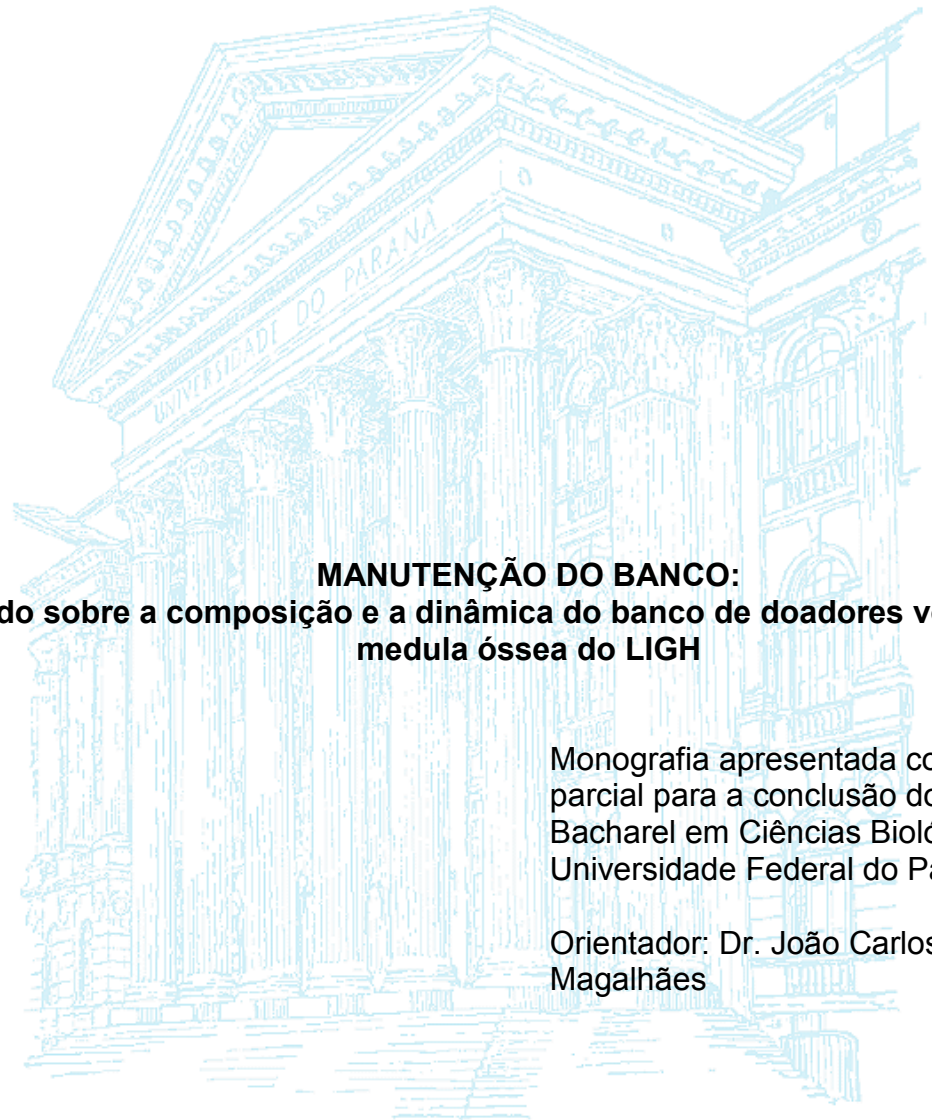


MARCO ÂNGELO GOLANOVSKI

**MANUTENÇÃO DO BANCO:
Estudo sobre a composição e a dinâmica do banco de doadores voluntários de
medula óssea do LIGH**

**CURITIBA
2009**

MARCO ÂNGELO GOLANOVSKI



**MANUTENÇÃO DO BANCO:
Estudo sobre a composição e a dinâmica do banco de doadores voluntários de medula óssea do LIGH**

Monografia apresentada como requisito parcial para a conclusão do Curso de Bacharel em Ciências Biológicas da Universidade Federal do Paraná.

Orientador: Dr. João Carlos Marques Magalhães

**CURITIBA
2009**

AGRADECIMENTOS

A todos que contribuíram para este trabalho, os doadores incógnitos pelos dados, os funcionários do LIGH pelo apoio, em especial ao Waldir por ajudar com o levantamento dos dados e pela paciência de refazer isto tantas vezes, e à Professora Doutora Maria da Graça quanto aos toques sobre a constituição do banco e a oportunidade de participar da família LIGH, aos meus pais por ajudarem nas horas difíceis e por fazerem parte das horas felizes, à minha namorada por se preocupar por mim e em especial ao professor João Carlos pela perseverança, dedicação e seus sutis puxões de orelha.

SUMÁRIO

1. INTRODUÇÃO	1
1.1 HLA	1
1.2 Transplantes de Medula Óssea (TMO)	2
1.3 REDOME	2
1.4 LIGH	3
1.5 Constituição Racial da População Brasileira	4
1.6 Problemas na Autodeclaração de Cor/Raça	5
2. OBJETIVOS	5
2.1 Gerais	5
2.2 Específicos	5
3. JUSTIFICATIVAS	5
4. MATERIAIS E MÉTODOS	6
4.1 Banco de Dados	6
4.2 Software Arlequin	6
4.3 Manutenção do Banco	6
4.3.1 Amostras e Procedimento	6
4.4 Composição do Banco	7
4.4.1 Amostra	7
4.4.2 Índices de Fixação (F_{ST} , F_{IS} , F_{IT})	8
4.4.3 Calculando os Índices de Fixação através da Análise de Variância	9
4.4.4 Análise Molecular de Variância	10
4.4.5 Software Arlequin	10
4.4.6 Teste de Significância para AMOVA no software Arlequin	11
4.4.7 Procedimentos	11
5. RESULTADOS E DISCUSSÃO	12
5.1 Manutenção do Banco	12
5.2 Composição do Banco	16
5.2.1 Resultados de F_{ST}	17
5.2.2 Resultados de F_{SC}	18
5.2.3 Resultados de F_{CT}	18
6. CONCLUSÃO	18
6.1 Manutenção do Banco	18
6.2 Composição do Banco	19
7. REFERÊNCIAS	20

RESUMO

Utilizando o banco de dados do Sistema LIGH, foram retiradas amostras com o objetivo de visualizar a dinâmica do banco de doadores voluntários de medula óssea, levando em conta a idade de ingresso e de saída dos mesmos. Foi feita uma projeção para o tamanho do banco, bem como a distribuição de idade de seus integrantes, dependendo da taxa de ingresso e da idade de novos doadores cadastrados. Também foram retiradas amostras para investigar uma possível estruturação populacional dos indivíduos do banco. Tal procedimento foi realizado através da análise molecular de variância entre os indivíduos do banco agrupados em regiões e raças. Os resultados obtidos mostram a importância das campanhas de adesão de novos doadores. Também mostram que existe uma estruturação populacional entre as raças e entre as regiões de doadores, além de uma estruturação entre raça e região, quando consideradas as diferenças entre dois estados, Paraná e Piauí.

1. INTRODUÇÃO

1.1. HLA

Todas as espécies de organismos vivos da Terra estão expostas a dezenas de patógenos, que invariavelmente acabam por entrar em seu organismo, podendo causar-lhes algum mal. Os vertebrados, por exemplo, dispõem de vários mecanismos de defesa, tanto mecânicos, como a pele, quanto químicos, além de mecanismos mais sofisticados envolvidos na resposta imune adaptativa da qual faz parte o MHC (*Major Histocompatibility Complex*), que na espécie humana recebe o nome de Antígenos Leucocitários Humanos (HLA), responsáveis pela ativação da resposta imune celular e humoral.

Os genes do MHC são considerados altamente polialélicos (<http://www.ebi.ac.uk/imgt/hla/stats.html>) e polimórficos na grande maioria das espécies animais, inclusive na espécie humana. Este complexo gênico localiza-se no braço curto do cromossomo seis e concentra-se numa região genômica denominada 6p21.3, contando com aproximadamente 4 milhões de pares de bases.

Os genes HLA clássicos e as proteínas por eles codificadas são classificadas em dois grupos. As de classe I estão presentes em todas as células nucleadas e são responsáveis por apresentar antígenos aos linfócitos T CD8+. Alguns exemplos de genes HLA de classe I são HLA-A, HLA-B e HLA-C. As proteínas de classe II, por sua vez, estão presentes apenas em macrófagos, células dendríticas, linfócitos B e linfócitos T ativados, apresentando antígenos exógenos aos linfócitos T CD4+. Como exemplo tem: HLA-DR, HLA-DP e HLA-DQ. A quantidade de alelos destes genes e o número de variações das proteínas por eles codificadas encontradas até outubro de dois mil e nove encontram-se na Tabela 1.

TABELA 1: NÚMERO DE ALELOS E PROTEÍNAS HLA DESCRITOS ATÉ OUTUBRO/2009

Gene	A	B	C	DRA	DRB	DQA1	DQB1	DPA1	DPB1
Alelos	893	1.431	569	3	814	35	106	28	136
Proteínas	681	1.165	431	2	637	26	77	16	118

Tal diversidade genética observada nos locis MHC tem sido associada à mecanismos seletivos associados: a) à parasitas, seja através da seleção sobredominante ou seleção dependente de frequência, atuando sobre os produtos deste sistema e b) escolha por parceiro MHC dissimilar. A origem destes alelos se deve principalmente a trocas de segmentos pré-existentes dentre os alelos, fenômeno denominado “colcha de retalhos” e, em menor grau, às mutações de ponto ocorridas nesta região. As frequências dos alelos são numericamente semelhantes, sugerindo uma seleção balanceada, apesar de existirem trabalhos mostrando uma tendência ao aumento da frequência de alguns alelos em alguns grupos de organismos da mesma espécie.

1.2. Transplantes de Medula Óssea (TMO)

O TMO pode ser definido como procedimento que envolve a infusão endovenosa de células-tronco, retiradas geralmente da medula óssea de um doador compatível, previamente selecionado, em um receptor previamente condicionado, ou seja, que teve suas células do sistema imunológico reduzidas drasticamente por radiação e quimioterapia, impedindo o organismo de executar uma reação imune ao receber o transplante. Tem a finalidade de restituir a integridade ao órgão hematopoiético debilitado. Este procedimento aplica-se no tratamento de desordens hematológicas malignas, quando descartadas outras formas de tratamento, e está entre um dos mais intensos tratamentos aplicados no homem, uma vez que desencadeia um estado de comprometimento múltiplo de órgãos, tecidos e depressão imunológica severa, com maior predisposição para infecções sistêmicas, efeitos esse decorrente dos tratamentos prévios de quimioterapia, radioterapia e uso de drogas hepatotóxicas e nefrotóxicas.

Os doadores compatíveis podem ser encontrados dentro da própria família do receptor (doador relacionado), fato que ocorre em aproximadamente 30% dos casos (BICALHO, 2005). Quando isto não acontece, o paciente deve recorrer a bancos de dados com doadores voluntários, na esperança de encontrar algum doador não-aparentado compatível (doador não-relacionado).

1.3. REDOME

O Registro Brasileiro de Doadores de Medula Óssea (REDOME) centraliza as informações de doadores voluntários de medula óssea (DVMO) registrados no Brasil, e está instalado no Instituto Nacional do Câncer (INCA), na cidade do Rio de Janeiro. Os dados nele contidos resultam do recrutamento de potenciais doadores nos hemocentros distribuídos em todo o país, cujas amostras sanguíneas são enviadas aos laboratórios credenciados pelo SUS/MS, que repassam os resultados de tipagem para o REDOME. Esta última instituição unifica todas as informações relacionadas aos doadores e receptores, iniciando a busca de doador compatível para os receptores (pacientes) em lista de espera. Em 2009, o banco conta com aproximadamente 1.312.334 doadores, conforme Boletim N° 15/MedulaNET, do INCA (novembro/dezembro - 2009 / janeiro - 2010) (http://www.inca.gov.br/conteudo_view.asp?id=2596). O número de indivíduos cadastrados no REDOME entre os anos de dois mil e três e dois mil e nove e pacientes transplantados entre dois mil e um e dois mil e nove estão representados nas Figuras 1 e 2.

FIGURA 1: NÚMERO DE DOADORES CADASTRADOS NO REDOME ATÉ OUTUBRO/2009

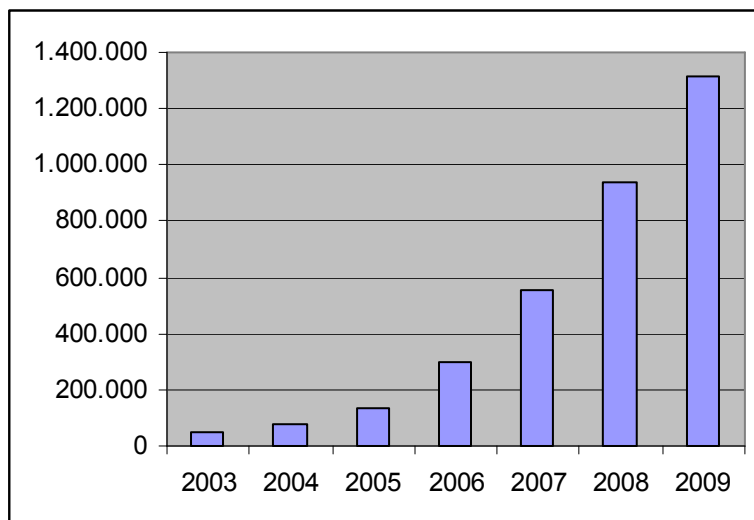
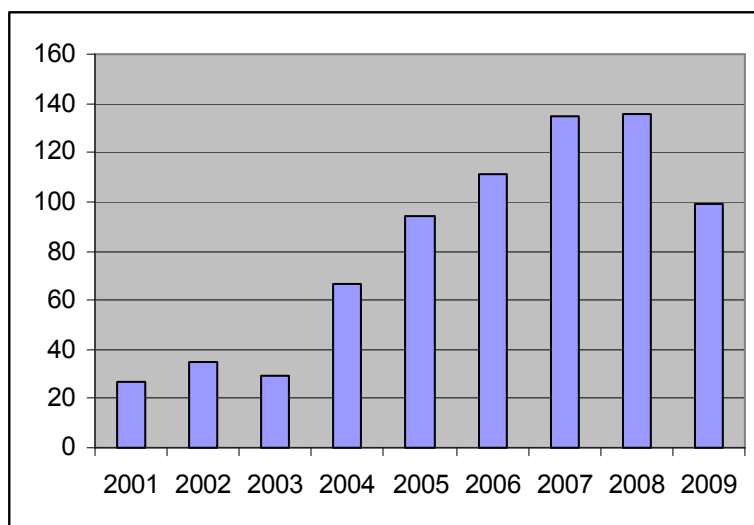


FIGURA 2: NÚMERO DE PACIENTES TRANSPLANTADOS VIA REREME/REDOME ATÉ OUTUBRO/2009



Para ser escolhido para um possível transplante, o receptor deve estar cadastrado no Sistema REDOME/REREME (Registro Brasileiro de Receptores de Medula Óssea), podendo realizar o transplante tão logo um doador compatível seja encontrado tanto no REDOME quanto nos registros internacionais de potenciais doadores (NMDP, CRIR e NYBC), integrados num único sistema de buscas para pacientes que aguardam o transplante.

1.4. LIGH

O Laboratório de Imunogenética e Histocompatibilidade do Departamento de Genética da Universidade Federal do Paraná (LIGH-UFPR) é um dos laboratórios que

integra a rede de colaboradores do REDOME. Faz parte de uma rede institucional que no Paraná integra diversas unidades da Rede Hemepar (Instituto de Hematologia e Hemoterapia do Paraná) na capital e no interior do estado e a Central de Transplantes do Paraná/CESA. Utilizando-se do Sistema LIGH, um aplicativo de cadastramento *on-line* desenvolvido pelo LIGH (BICALHO *et. al*, 2005), que faz uma interface com o sistema REDOMENET e remete as tipagens HLA realizadas neste laboratório para o REDOME.

Totalmente desenvolvido a partir de softwares livres (Linux, PostgreSQL, PHP), o Sistema LIGH pode ser instalado em qualquer unidade cadastradora, sem nenhum custo adicional para a mesma, bastando possuir uma impressora e acesso a Internet para o funcionamento ideal do programa.

O Sistema LIGH possui uma ficha única a ser preenchida no momento do cadastro. Os principais dados coletados do doador são seu nome e sobrenome, a data de nascimento, a cidade de origem, o local de coleta, o sexo, o estado civil, se é fumante ou não, a raça (por autodeclaração) e os alelos HLA-A, HLA-B e HLA-DRB1 possuídos pelo doador, tipados em baixa resolução. Os dados de tipagem e as informações pessoais seguem caminhos diferentes no cadastro, podendo ser cruzadas apenas por pessoas habilitadas, garantindo que os dados do doador fiquem em sigilo.

Além do principal objetivo, que é facilitar o encontro de doadores compatíveis, o Sistema LIGH tem um potencial científico enorme, por reunir uma grande quantidade de informações da população paranaense, tanto demográficas quanto genotípicas, permitindo o cálculo de frequências alélicas e haplotípicas dos alelos genotipados, bem como estudar relações entre as características do doador e seus alelos, como verificar os alelos possuídos por indivíduos fumantes e sua significância, por exemplo.

1.5. Constituição Racial da População Brasileira

A alta taxa de polimorfismo encontrada nos genes HLA torna a tarefa de encontrar um doador não relacionado bastante difícil. Esta tarefa torna-se ainda mais complicada quando se leva em conta às características raciais do receptor, uma vez que alguns alelos são mais freqüentes em alguns grupos raciais e raros ou inexistentes em outros (BICALHO, 2005; RUIZ, 2008). Por isto torna-se necessária uma descrição da constituição racial da população brasileira e as conseqüências que derivam desta variedade.

O Brasil é um país que apresenta uma alta taxa de miscigenação, o que culmina em uma grande taxa de hibridização em populações não-isoladas. A população do Paraná compõe-se principalmente de etnias européias, por ter sido colonizado por países deste local, o que gerou um grande número de imigrantes que vieram a instalar-se em todo o seu litoral e posteriormente no interior. Também possui grupos negróides, descendentes em sua maioria dos escravos trazidos da África durante a época do tráfico negreiro, mas sem região definida devido ao caráter informal desta atividade. E por fim os nativos (ameríndios), que eram encontrados em grande número, que é estimado entre um e dez milhões (KROEBER, 1934; DOBYNS, 1966), mas que sofreram severas baixas com a ocupação européia, com doenças, morte em guerras e

no trabalho escravo, e desestruturação sócio-cultural (RIBEIRO, 1996; RICARDO, 1996).

1.6. Problemas na Autodeclaração de Cor/Raça

A grande miscigenação da população brasileira causa certa confusão quando o indivíduo é perguntado quanto a sua origem. Uma vez que geralmente utiliza-se de traços raciais (cor da pele, aspecto do cabelo, etc), acaba por denominar-se de um grupo étnico, ou racial, ao qual não pertence. Um exemplo bastante ilustrativo é quando um indivíduo de descendência negra, mas cujo tom de pele é amarronzado, denomina-se como índio, devido a esta cor de pele ser normalmente atribuída a esta raça como senso comum.

Para a tipagem laboratorial tal fator não é relevante, no entanto, quando estes dados são utilizados para estudos de genética de populações, eles podem constituir um problema. Uma vez que, como dito anteriormente, existem diferenças genéticas entre os grupos, inclusive alelos exclusivos de certas raças, este tipo de dado (cor etc.) pode levar a equívocos. O critério de atribuição de raça pode diferir em algumas outras populações, por exemplo, nos Estados Unidos, onde um indivíduo com uma ancestralidade negróide, mesmo que pequena, é considerado negro (SALZANO & FREIRE-MAIA, 1970).

Entretanto, a autodeclaração de cor/raça é utilizada no cadastramento do Sistema LIGH, assim como em todo o REDOME. Embora vago tal fator reflete em parte a diversidade genética da população e, assim, pode fornecer informações úteis e deve ser considerado na pesquisa.

2. OBJETIVOS

2.1. Geral

Obter um quadro geral da composição genética, étnica e demográfica do banco de doadores de medula óssea cadastrados no Sistema LIGH.

2.2. Específicos

Investigar a dinâmica do banco, projetando a quantidade de indivíduos que entram e saem do banco em função das campanhas de coleta e da idade.

Investigar a possível estrutura populacional, em função da procedência e origem racial do DVMO, inclusive comparando estes fatores.

Comparar os dados referentes a composição étnica e genética da população do Estado do Paraná com os dados da população do Piauí, disponíveis no LIGH.

3. JUSTIFICATIVAS

Atualmente, a probabilidade de se encontrar doadores não relacionados de medula óssea ainda é relativamente baixa. Uma das causas disto é a própria natureza

do sistema HLA, mas também deriva do número de doadores voluntários disponíveis no REDOME, bem como sua origem racial. Para alcançar valores satisfatórios de probabilidade de se encontrar um doador compatível, faz-se necessário o aumento do número de DVMO, considerando a diversidade racial destes, de modo a contemplar a diversidade genética dos pacientes que aguardam o transplante. Por isto a necessidade de se conhecer a composição racial do banco de doadores, bem como de pacientes em lista de espera para o transplante, o que daria uma idéia da diversidade genética pouco representada no registro, de tal forma a canalizar os esforços de recrutamento para tais grupos, possibilitando a descoberta de carência de voluntários de certas raças e auxiliando no planejamento de campanhas de coleta das mesmas.

Também levando em conta um número aceitável de doadores, deve-se ter em mente a dinâmica do banco, uma vez que existe uma idade mínima e uma idade máxima exigida para que o indivíduo seja um doador, aqueles cadastrados que ultrapassam a idade máxima de doar devem ser repostos. Por isto, uma projeção de quando e quantos destes indivíduos serão retirados do banco é importante para a sua manutenção.

4. MATERIAIS E MÉTODOS

4.1. Banco de Dados

O Sistema LIGH foi o banco de dados utilizado para a retirada das amostras do presente estudo. Possui uma plataforma *on-line* onde o usuário pode, com um *login* e uma senha, extrair dados populacionais dos doadores voluntários cadastrados ou, inserindo seu próprio CPF, atualizar seus dados no sistema.

Além deste método, alguns dados, por serem mais específicos, foram retirados diretamente do console do banco dos dados, através de *queries* (buscas ou, mais literalmente, questionamentos) específicas.

4.2. Software Arlequin

Durante este trabalho, o programa Arlequin foi utilizado para o cálculo da análise molecular de variância (AMOVA). Foram utilizados os parâmetros *Default* do programa, ou seja, sem nível individual, com mil permutações para o teste de significância e a matriz de distância computada a partir do número de diferentes alelos.

4.3. Manutenção do Banco

4.3.1. Amostras e Procedimento

Para o estudo da dinâmica do banco, foram retiradas duas amostras do Sistema LIGH, uma com a idade de entrada dos doadores (amostra **A**) e outra com a idade atual (amostra **B**). Ao todo, foram amostrados sessenta e um mil setecentos e oitenta indivíduos, quantidade existente no banco até agosto de 2009. Nas amostras foi feita uma separação em grupos de cinco em cinco anos, ou seja, a partir da idade de

entrada do indivíduo (amostra **A**) ou de sua idade atual (amostra **B**), ele era colocado no grupo de 16-20 anos, 21-25, 26-30 e assim por diante até o grupo de 61-65 anos. Para estas duas amostras, o banco inteiro foi considerado. A taxa de entrada de doadores em cada faixa etária pré-estabelecida (cinco em cinco anos) foi então calculada conforme a porcentagem de cada uma correspondente na amostra **A**. Tais valores foram usados para ponderar a distribuição da entrada dos novos doadores, e foram denominados de “Idade de Ingresso (%)”. O número de doadores em cada faixa etária foi então tabelado, bem como o total atual de indivíduos no banco e foram criadas variáveis que indicam a quantidade de novos indivíduos que ingressariam no banco dentro de um período futuro de cinco, dez, quinze, vinte e vinte e cinco anos, tais variáveis foram denominadas “Ingresso”, e podem possuir qualquer valor, onde 1 representaria 100% de ingresso do banco. Os períodos de tempo foram denominados arbitrariamente de “Intervalos”. Tais valores e variáveis foram então tabelados em uma planilha do programa EXCEL®.

4.4. Composição do Banco

4.4.1. Amostra

Para os estudos da composição do banco, foi retirada uma amostra do banco contendo as informações dos alelos A, B e DR e a raça do doador voluntário, bem como o local onde ocorreu a sua coleta. Foi estabelecido um período de três anos (2005 até 2008) para a retirada dos dados, totalizando cinquenta e um mil, trezentos e setenta e três indivíduos. Os indivíduos foram então separados em populações conforme o local de coleta e a raça e os dados foram convertidos para o formato aceito no programa Arlequin. Como existem seis grupos raciais, amarelo, pardo, preto, branco, índio e não informado, e dezessete locais de coleta, incluindo uma amostra do Estado do Piauí, obteve-se ao todo setenta e sete grupos, relacionados no Quadro 1. Vale ressaltar que algumas raças não foram encontradas em todos os locais de coleta, por isto alguns destes possuem menos populações que outros.

QUADRO 1: POPULAÇÕES DE DOADORES CONFORME RAÇA E LOCAL DE DOAÇÃO

erasto b*	heme a	paran pa	ponta a	pato a	guarap a	palmas b
						palmas
erasto a	heme i	paran pr	ponta i	pato i	guarap i	pr
						palmas
erasto i	heme pr	paran ni	ponta pa	pato pa	guarap pa	ni
erasto pa	heme ni	apu b	ponta pr	pato pr	guarap pr	irati b
erasto pr	belt b	apu a	ponta ni	pato ni	guarap ni	irati a
ligh b	belt ni	apu pr	casca b	umu b	foz b	irati pa
ligh a	belt a	apu ni	casca a	umu a	foz pa	irati pr
ligh pa	belt pa	cia b	casca pa	umu pa	foz pr	ufpi b
ligh pr	belt pr	cia a	casca pr	umu pr	foz ni	ufpi pa
ligh ni	paran b	cia pa	casca ni	umu ni	toledo b	ufpi pr
heme b	paran a	ponta b	pato b	guarap b	toledo pa	ufpi ni

*O nome do grupo é formado por uma abreviação do laboratório e uma abreviação da raça. b: Branco; a: Amarelo; pa: Pardo; pr: Preto; ni: Não Informado; i: Indígena; erasto: Banco de Sangue

Erasto Gaertner S.C. Ltda; irati: Unidade de Coleta de Irati; ufpi: Unidade de Coleta LIB-UFPI; ligh: LIGH; heme: HEMEPAR; belt: Hemonúcleo Regional de Francisco Beltrão; paran: Hemonúcleo Regional de Paranaguá; apu: Hemonúcleo de Apucarana; cia: Unidade de Coleta e Transfusão de Cianorte; ponta: Hemonúcleo de Ponta Grossa; casca: Hemocentro Regional de Cascavel; pato: Unidade de Coleta e Transfusão de Pato Branco; umu: Hemonúcleo de Umuarama; guara: Hemonúcleo de Guarapuava; foz: Hemonúcleo de Foz do Iguaçu; toledo: Unidade de Coleta e Transfusão de Toledo; palmas: Unidade de Coleta e Transfusão de Palmas.

Três laboratórios possuem características que devem ser consideradas: o HemePar e o LIGH são os que possuem a maior quantidade de indivíduos, e estes são catalogados dentro do Estado do Paraná como um todo; as populações “ufpi” são formadas por amostras de indivíduos tipados no laboratório da Universidade Federal do Piauí, não sendo, portanto, representativas de nenhuma localidade do Paraná.

4.4.2 Índices de Fixação (F_{ST} , F_{IS} , F_{IT})

Para se obter um valor que mede as relações entre as populações, possibilitando uma comparação entre elas, foram utilizados os Índices de Fixação (F), atribuídos para cada comparação. Também chamados de estatísticas F de Wright (WRIGHT, 1951, 1965), os índices de fixação medem a taxa de perda de genótipos heterozigotos por uma população, atribuindo esta taxa ao endocruzamento que ocorre na população (cf. HARTL, 1989).

Os valores de F dependem de três variáveis: H_I , a heterozigosidade de um indivíduo em uma subpopulação; H_S , a heterozigosidade esperada em um cruzamento aleatório dentro de uma subpopulação e; H_T , a heterozigosidade esperada em um cruzamento aleatório dentro da população. Cada índice é definido da seguinte maneira:

$$F_{IS} = \frac{H_S - H_I}{H_S}$$

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

$$F_{IT} = \frac{H_T - H_I}{H_T}$$

Cada índice define o coeficiente de endocruzamento de maneira hierárquica, variando as populações de referência. Logo, F_{IS} é relativo ao endocruzamento entre indivíduos (H_I) em relação à sub-população à que pertencem (H_S); F_{ST} refere-se ao endocruzamento entre indivíduos (H_I) comparado com a população total (H_T) e F_{IT} compara os efeitos de cruzamentos não-aleatórios dentro das subpopulações (F_{IS}) e os efeitos da subdivisão da população (F_{ST}).

Os índices variam numericamente de zero a um, sendo que quanto maior o valor, maior o endocruzamento dentro da hierarquia observada. A hierarquia dentre os índices é: $F_{IT} > F_{ST} > F_{IS}$, e dentre os agrupamentos é população > subpopulação > indivíduo.

4.4.3 Calculando os Índices de Fixação através da Análise de Variância

Um método para calcular os índices de fixação esperados e que permitisse o teste de hipóteses foi desenvolvido por Cockerham em 1969 e 1973 (cf. HARTL, 1989), assinalando valores para cada alelo e calculando sua frequência esperada, obtendo uma variância associada para cada índice e, através destas, calculando um nível de significância para os valores de F. Para este método, os índices são calculados da seguinte maneira:

$$F_{IT} = \frac{\sigma^2_w + \sigma^2_b}{\sigma^2}$$

$$F_{ST} = \frac{\sigma^2_a}{\sigma^2}$$

$$F_{IS} = \frac{\sigma^2_b}{\sigma^2_b + \sigma^2_w}$$

Onde σ^2_a é a variância associada a subpopulação, σ^2_b é associada aos indivíduos dentro das populações, σ^2_w está associada a alelos entre indivíduos e

$$\sigma^2 = \sigma^2_a + \sigma^2_b + \sigma^2_w.$$

Este método é utilizado, de maneira adaptada por EXCOFFIER (2005), para o cálculo de Análise Molecular de Variância (AMOVA) em populações estruturadas, substituindo o nível individual (σ^2_w) por um nível adicional de agrupamento (σ^2_c). E define o índice F_{CT} como a relação entre os haplótipos randômicos dentro de um grupo de populações em relação àqueles retirados da amostra como um todo, e F_{SC} como a relação da diversidade molecular dos haplótipos escolhidos ao acaso dentro das populações aos haplótipos escolhidos aleatoriamente dentro do grupo. Como um novo nível hierárquico foi adicionado, os índices de fixação passam a ser considerados, quanto ao cálculo das variâncias e suas interpretações, da seguinte forma:

$$F_{ST} = \frac{\sigma^2_c + \sigma^2_b}{\sigma^2}$$

$$F_{CT} = \frac{\sigma^2_a}{\sigma^2}$$

$$F_{SC} = \frac{\sigma^2_b}{\sigma^2_b + \sigma^2_c}$$

Como os índices, neste caso, são inferidos pela manipulação algébrica de variâncias, e a σ_a^2 é obtida por diferença, é possível encontrar valores negativos para os mesmos (para detalhes, ver EXCOFFIER, 2005). Estes índices, também neste caso, não representam apenas o endocruzamento, podendo indicar uma estruturação populacional, isto é, a forma como a variação genética está organizada em grupos na população, ou outros fatores que alteram a frequência de heterozigotos.

A hierarquia dentre os índices calculados é: $F_{ST} > F_{CT} > F_{SC}$. E dentre os grupos é: amostra total > grupo > população, onde a amostra total corresponde a todos os valores amostrados, o grupo ao ajuntamento das populações em conjuntos dos quais serão levantadas as hipóteses e população às amostras com algum tipo de semelhança. Por exemplo, em uma amostra genotipada de doadores voluntários brancos, alguns pertencentes a hemocentros (locais de coleta) do Estado do Paraná e outros a hemocentros do Estado do Piauí, pode-se definir a hierarquia da análise como: a amostra total > hemocentros (Estados) > populações de brancos em cada hemocentro, ou esta hierarquia pode ser definida como amostra total > populações de brancos em cada hemocentro > hemocentros (Estado), dependendo se as hipóteses a serem testadas referem-se à região (primeira hierarquia, entendendo-se a região, neste caso, como o local de coleta), ou à população branca nos hemocentros (segunda hierarquia). Ainda neste último exemplo, o F_{CT} seria a relação entre os haplótipos de todos os indivíduos brancos em cada Estado e a amostra total, independente do hemocentro; o F_{SC} entre os indivíduos brancos de cada hemocentro e o total de indivíduos brancos e o F_{ST} mostra a relação entre os indivíduos brancos de cada hemocentro para com os indivíduos brancos do Estado a que este pertence.

4.4.4 Análise Molecular de Variância

A Análise Molecular de Variância (EXCOFFIER, 2005) é um método que permite o cálculo dos índices de F para dados multialélicos e com dois ou mais locus. É calculado através de *bootstraps* computacionais, ou seja, uma série de instruções executadas por um programa que, neste caso, obtém um valor empírico para as estatísticas F, bem como uma análise de variância destes coeficientes dentro das populações, entre as populações dentro dos grupos e entre os grupos.

4.4.5 Software Arlequin

O programa Arlequin consiste em um pacote integrado de softwares para análise de dados genéticos, capaz de manipular grandes quantidades de dados tanto moleculares (RFLPs, seqüências de DNA, microsátélites) quanto dados convencionais (frequências alélicas, multi-locus) e realizar diversos cálculos genéticos como Equilíbrio Hardy-Weinberg, Desequilíbrio de Ligação, Teste de Neutralidade, AMOVA, etc.

Foi desenvolvido pelo Instituto de Zoologia da Universidade de Berne, na Suíça, em 1998, e teve sua mais atual versão (3.11) lançada em 2007.

4.4.6 Teste de Significância para AMOVA no software Arlequin

Para encontrar um valor para o teste de significância (*P-Value*), o software realiza um número pré-estabelecido de permutações, por amostragem aleatória, dentro do nível hierárquico que está sendo testado, sendo que cada nível possui um modo diferente de permutação. No índice F_{ST} são retirados dois alelos ao acaso da amostra, com reposição, e verifica-se se estes são idênticos, chegando a um valor P indicando as chances deste fato ocorrer, ponderado pelo número de permutações. Para a obtenção do F_{SC} , assume-se que os grupos são reais, mas que as populações dentro deles não, permutando indivíduos entre os grupos, independentemente de sua população de origem e calculando o valor P conforme estes indivíduos sejam semelhantes. Finalmente, para obter F_{CT} , o programa considera as populações como reais, mas os agrupamentos como artificiais, permutando populações inteiras entre os grupos e contabilizando o valor P conforme estas populações sejam semelhantes ao resto do grupo em que foram colocadas. Todos os valores de P, em seu respectivo nível hierárquico, admitem como hipótese nula a formação genotípica do grupo estudado ocorrendo ao acaso, ou seja:

H_0 = A formação genotípica do grupo estudado ocorre de forma aleatória.

4.4.7. Procedimentos

Os grupos foram reunidos de diferentes formas e então comparados entre si através de uma série de AMOVAs (Análise de Variância Molecular) efetuadas no programa Arlequin, em busca de relações entre estes. As populações foram agrupadas e comparadas das seguintes formas:

- (a) Um grupo comparado entre si, com todas as populações;
- (b) Populações separadas por raças, totalizando seis grupos;
- (c) Populações separadas por raças, mas excluídas as populações originadas do LIGH e do Hemepar, totalizando seis grupos;
- (d) Populações separadas por regiões, totalizando dezessete grupos;
- (e) Populações separadas por regiões, mas excluídas as populações originadas do LIGH, do Hemepar e do Laboratório da UFPI, totalizando quatorze grupos;
- (f) Todas as populações originadas no Paraná comparada com os dados do Piauí (“ufpi”);
- (g) Populações de doadores Brancos, comparadas com doadores Brancos do Piauí (“ufpi br”);
- (h) Populações de doadores Pardos, comparadas com doadores pardos do Piauí (“ufpi pa”);
- (i) Populações de doadores Pretos, comparadas com doadores Pretos do Piauí (“ufpi pr”);

5. RESULTADOS E DISCUSSÃO

5.1. Manutenção do Banco

Os dados tabelados durante o procedimento (Tabela 1) foram computados em um gráfico (Figura 1), que permite visualizar a dinâmica do banco e sua composição em comparação com a atual. Observando a altura das colunas, que indica o total de indivíduos em cada geração, pode-se perceber uma tendência à estabilidade deste total. No entanto, verificando a quantidade de saída de indivíduos do banco, presente na Tabela 2, nota-se que este valor cresce a cada geração, ou seja, após alguns anos, a quantidade de indivíduos que saem do banco pode ser maior que aquela que é cadastrada, gerando um déficit de doadores voluntários. Vale ressaltar que estes dados supõem um ingresso de doadores no banco de dez por cento a cada geração, mas se colocados valores maiores, esta inclinação continua, apesar do decréscimo de doadores disponíveis tornar-se menos acentuado. Pode-se observar isto na Figura 2, que simula um ingresso de vinte por cento por geração e na Figura 3, que simula uma quantidade decrescente de entrada de indivíduos no banco, de cinquenta, quarenta, trinta, vinte e dez por cento de ingresso por geração.

TABELA 2: EXPECTATIVAS DOS INTERVALOS*

Idade	Atual	% (idade de ingresso no banco)	1 Geração	2 Intervalos	3 Intervalos	4 Intervalos	5 Intervalos	Ingresso
16-20	3340	0,16	968,15	1063,63	1158,58	1228,92	1268,94	1° geração
21-25	12530	0,21	4644,96	2401,80	2625,27	2815,03	2939,31	0,10
26-30	12299	0,17	13579,01	5797,42	3657,15	3956,83	4189,95	2° geração
31-35	9445	0,14	13181,50	14548,55	6853,51	4777,36	5113,51	0,10
36-40	8409	0,12	10205,36	14016,85	15458,47	7818,67	5773,95	3° geração
41-45	6976	0,10	9015,31	10871,47	14742,43	16228,09	8613,36	0,10
46-50	5078	0,06	7374,13	9452,71	11347,91	15247,79	16749,91	4° geração
51-55	2891	0,03	5272,63	7587,95	9685,62	11594,96	15502,88	0,10
56-60	727	0,00	2903,86	5286,76	7603,34	9701,94	11611,82	5° geração
61-65	85	0,00	728,10	2905,07	5288,07	7604,73	9703,38	0,10
SAÍDA			85,00	728,10	2905,07	5288,07	7604,73	
TOTAL	61780		67873,00	73932,20	78420,36	80974,32	81467,02	

*A primeira coluna define as faixas etárias de cada grupo, a segunda a quantidade atual de doadores em cada faixa, a terceira a porcentagem de entrada no banco em cada faixa (variável "Idade de Ingresso (%)"), a quarta, quinta, sexta, sétima e oitava a expectativa da quantidade de doadores em cada faixa etária dentro de cinco, dez, quinze, vinte e vinte e cinco anos, considerando um crescimento do banco conforme as variáveis "Ingresso" em cada geração. No final de cada coluna com a expectativa da geração, existe uma entrada denominada "SAÍDA", indicando o número de indivíduos que sairiam do sistema naquela geração, e outra entrada que indica o total de indivíduos daquela geração ("TOTAL").

FIGURA 1: GRÁFICO EXPECTATIVA DOS INTERVALOS CONSIDERANDO INGRESSO DO BANCO DE 10% POR INTERVALO (variável Ingresso= 0,1)

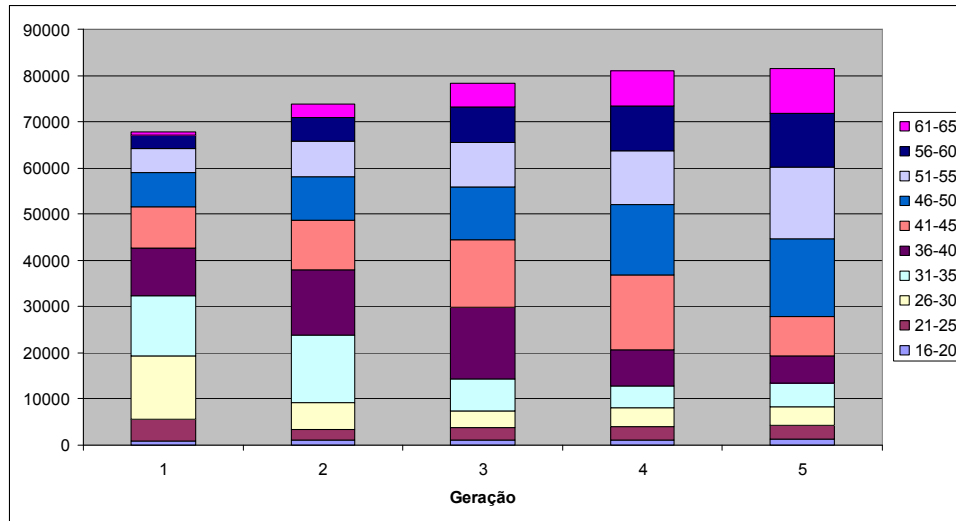


FIGURA 2: GRÁFICO EXPECTATIVA DOS INTERVALOS CONSIDERANDO INGRESSO DO BANCO DE 20% POR INTERVALO (variável Ingresso= 0,20)

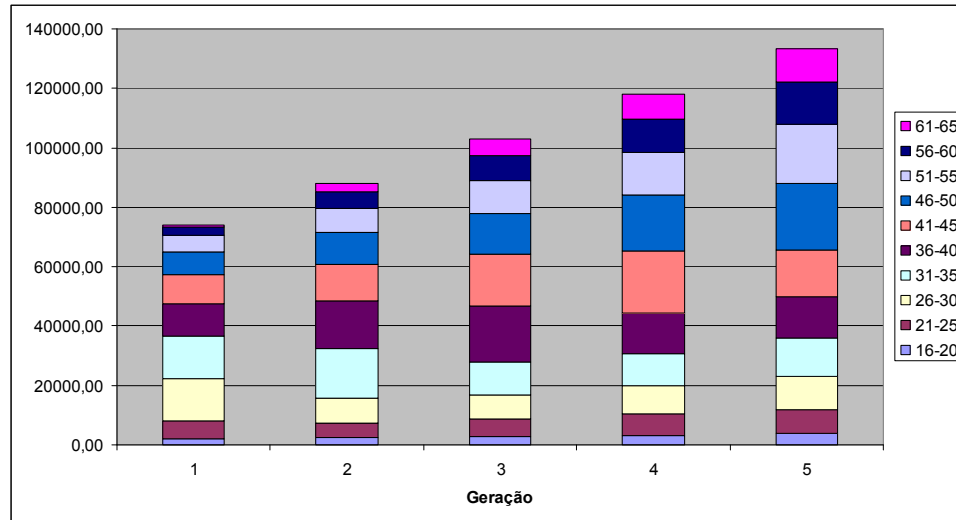
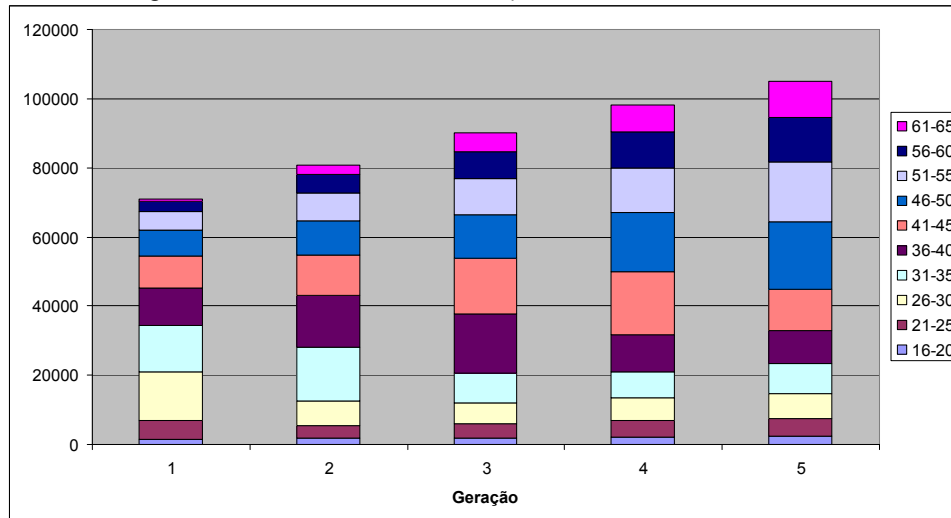


FIGURA 3: GRÁFICO EXPECTATIVA DOS INTERVALOS CONSIDERANDO INGRESSO DO BANCO DE 50, 40, 30, 20 E 10% POR INTERVALO, RESPECTIVAMENTE (variável Ingresso= 0,5; 0,4; 0,3; 0,2 e 0,1).



Uma vez que a porcentagem de indivíduos que ingressam no banco por faixa etária foi inferida a partir dos dados já existentes no banco de dados, deve-se considerar que a simulação não é realista, assim a paisagem dos gráficos não seria de estabilidade se este valor fosse diferente. As Figuras 4 e 5 mostram como ficaria o banco se houvesse uma porcentagem de entrada igual para todas as faixas etárias (dez por cento), com crescimento por intervalo de dez e de vinte por cento, respectivamente.

FIGURA 4: GRÁFICO EXPECTATIVA DOS INTERVALOS CONSIDERANDO INGRESSO DO BANCO DE 10% POR INTERVALO E FREQUÊNCIA DE ENTRADA IGUAL A 10% POR FAIXA ETÁRIA (variável Ingresso= 0,1)

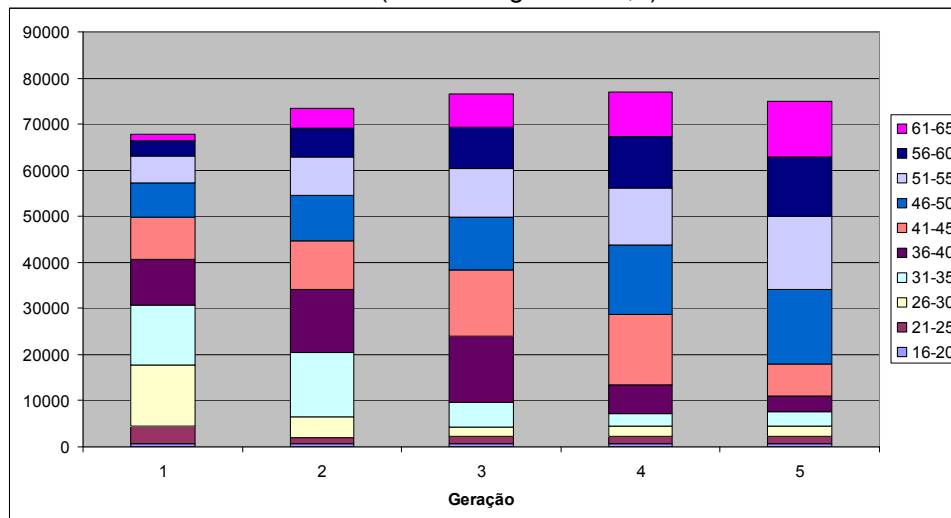
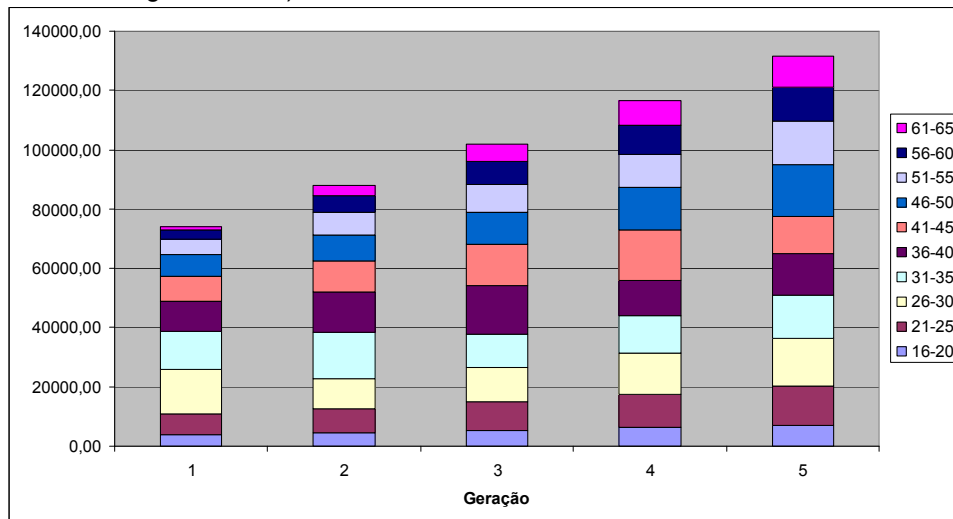


FIGURA 7: GRÁFICO EXPECTATIVA DOS INTERVALOS CONSIDERANDO INGRESSO DO BANCO POR FAIXA ETÁRIA IGUAL A 30, 30, 20, 5, 5, 2, 2, 2, 2 E 2, RESPECTIVAMENTE E QUANTIDADE DE ENTRADA IGUAL A 20% (variável Ingresso= 0,2).



Percebe-se que a Figura 6 não é muito diferente das anteriores, no entanto, a Figura 7 mostra uma tendência ao crescimento, uma vez que todas as faixas etárias aumentam em quantidade e a entrada maior de jovens repõe a saída dos mais idosos e sustenta a quantidade de doadores voluntários em todas as faixas.

É importante ressaltar que esta é uma estimativa que tem como parâmetros apenas a porcentagem de entrada do banco em cada idade e a quantidade de doadores que ingressa no mesmo, e é projetada até vinte e cinco anos no futuro. Uma vez que existem variações não contabilizadas no experimento, como mudanças de políticas ou nas estratégias de coleta, estas estatísticas podem variar ano a ano, fazendo com que esta projeção seja apenas um indicativo acerca da dinâmica do banco.

5.2. Composição do Banco

No total foram realizadas nove AMOVAs, mostradas na tabela a seguir (Tabela 3) com as estatísticas F e seus respectivos valores p .

TABELA 3: RESULTADOS DAS AMOVAS

AMOVA	F _{ST}	Valor p	F _{SC}	Valor p	F _{CT}	Valor p
(a)	0.00098	0.00000*	-	-	-	-
(b)	0.00231	0.00000*	0.00035	0.00000*	0.00196	0.00000*
(c)	0.00193	0.00000*	0.00049	0.00000*	0.00144	0.00000*
					-	0.86901+-
(d)	0.00083	0.00000*	0.00255	0.00000*	0.00172	0.00988
				0.00196+-	-	0.99316+-
(e)	0.00056	0.00000*	0.00203	0.00136*	0.00147	0.00231
						0.10850+-
(f)	0.00289	0.00000*	0.00080	0.00000*	0.00210	0.00886
						0.05083+-
(g)	0.00305	0.00000*	0.00017	0.00000*	0.00288	0.00708*
	-				-	0.78201+-
(h)	0.00313	0.28348+-0.01352	0.00034	0.23656+-0.01196	0.00347	0.01443
		0.01271+-		0.03617+-		0.19257+-
(i)	0.03436	0.00392*	0.00180	0.00495*	0.03262	0.01315

* = Valor p significativo para $\alpha = 0,05$.

- (a) Um grupo comparado entre si, com todas as populações;
- (b) Populações separadas por raças, totalizando seis grupos;
- (c) Populações separadas por raças, mas excluídas as populações originadas do LIGH e do Hemepear, totalizando seis grupos;
- (d) Populações separadas por regiões, totalizando dezessete grupos;
- (e) Populações separadas por regiões, mas excluídas as populações originadas do LIGH, do Hemepear e do Laboratório da UFPI, totalizando quatorze grupos;
- (f) Todas as populações originadas no Paraná comparadas com os dados do Piauí ("ufpi");
- (g) Populações de doadores Brancos, comparadas com doadores Brancos do Piauí ("ufpi br");
- (h) Populações de doadores Pardos, comparadas com doadores pardos do Piauí ("ufpi pa");
- (i) Populações de doadores Pretos, comparadas com doadores Pretos do Piauí ("ufpi pr");

5.2.1. Resultados de F_{ST}

Observando todos os valores de significância do valor p associado a este índice, percebe-se que quase todos são significativos, o que pode ser interpretado como o ajuntamento das populações dentro de cada grupo como sendo real, ou seja, existe uma estruturação da população em relação às raças, visível nas AMOVAs (b), (c), (g), e (i) e em relação às regiões de coleta, nas AMOVAs (d), (e) e (f). No entanto, a AMOVA (h) não apresenta um valor significativo, o que indica que os indivíduos considerados pardos no Paraná são semelhantes aos indivíduos considerados pardos no Piauí.

5.2.2. Resultados de F_{SC}

A significância dos valores p para o F_{SC} das AMOVAS realizadas reforça a idéia da estruturação da população em regiões e raças ser real. E revela mais uma vez a disparidade encontrada no F_{ST} quanto à AMOVA (h). Como o valor p é alto, assume-se que para este conjunto de dados a chance de se obter os grupos definidos aleatoriamente é grande, o que demonstra que não houve nenhum fenômeno que levasse a estruturação destes grupos. Tal evento mostra que os indivíduos pardos não possuem diferenças genéticas suficientes entre os estados do Paraná e Piauí de forma a caracterizar uma estruturação dentro deste grupo.

5.2.3. Resultados de F_{CT}

Neste índice, percebe-se que os valores p significativos foram aqueles em que existe a separação por raça, ou seja, quando populações de uma região são comparadas com populações de outra região, considerando que estas populações estão agrupadas por raça, as chances de elas serem semelhantes ao acaso é bastante baixa, indicando uma estruturação dentro destas raças. No entanto, quando se considera o agrupamento por região, os valores de p não são significativos, o que demonstra que provavelmente não existe uma diferenciação das raças em relação às regiões em que estas foram amostradas. É importante notar que exceto a AMOVA (g), as outras AMOVAs que comparam os resultados do Paraná com o Piauí [(h) e (i)] não são significativas, apesar de serem relativas às raças. Tal fator deve-se, provavelmente, ao método usado pelo programa para inferir um valor de p para o F_{CT} , pois, como este compara uma população com outras populações, permutando o grupo a que pertencem e o grupo da amostra do Piauí possui apenas uma população, as chances das populações do Paraná, que diferem, neste caso, apenas na região, serem comparadas entre si aumenta e, como estas regiões não apresentam uma diferenciação significativa, característica indicada no F_{SC} , acabam por superestimar o valor p para o F_{CT} . O aumento do número de populações do Piauí, ou o aumento no número de permutações talvez alterem este resultado, indicando um valor p mais confiável.

6. CONCLUSÃO

6.1. Manutenção do Banco

As inferências feitas através dos dados retirados do banco de dados do Sistema LIGH mostram que o número de doadores pode chegar a um valor crítico, onde o número de doadores cadastrados durante certo período pode ser quase equivalente ao número de doadores que é excluído do banco. No entanto, como mostrado na Figura 7, quando ocorre uma entrada maior de indivíduos de faixas etárias menores, o banco diminui a velocidade com que alcança a estabilidade, pois além de repor os indivíduos que são excluídos, os indivíduos jovens mantêm-se por mais tempo no registro.

Isto leva-nos a crer que, para continuar fornecendo doadores voluntários para transplante com uma frequência satisfatória, os laboratórios precisam manter

campanhas contínuas de cadastramento, tentando chamar a atenção dos indivíduos mais jovens para que se tornem voluntários. Atenta também para o fato que, mesmo que por razões práticas o tamanho do banco chegue a um limite, espera-se que haja uma quantidade de doadores voluntários que maximize as chances de um receptor encontrar um doador. Assim, as campanhas devem continuar em um ritmo que pode ser avaliado por simulações como as aqui apresentadas.

6.2. Composição do Banco

Os índices de fixação e seus valores de significância mostram que a população de indivíduos cadastrados no Sistema LIGH encontra-se estruturada entre suas respectivas raças, exceto no caso dos pardos, em regiões e, quando a comparação é feita entre Estados, existe uma estruturação da mesma raça entre as regiões com relação aos brancos e pretos. No entanto, não existe qualquer sinal de estruturação da mesma raça entre as regiões do Paraná. É possível, considerando este último fator, que a diferenciação encontrada entre as regiões do Paraná se deva apenas a uma estruturação racial. Para confirmar tal hipótese, seriam necessários estudos posteriores que comparem as populações regionais duas a duas, obtendo-se desta forma valores mais refinados de suas semelhanças, revelando se existe mesmo a estruturação regional. Por outro lado, os dados de procedência regional podem ser melhor avaliados se considerarmos o local de nascimento do DVMO que possivelmente deverá refletir melhor a estruturação geográfica, uma vez que os dados referentes aos postos de coleta incluem indivíduos que migraram ou que pertencem a outras populações regionais. Em relação à estruturação dos grupos raciais entre os Estados, é provável que esta realmente exista, embora não tenha sido evidenciado entre os pardos. Isto pode ter ocorrido por erro na amostragem, possivelmente no momento da autodeclaração de cor, diminuindo a confiabilidade dos dados. Estudos posteriores, com mais dados dos Estados, seriam necessários para verificar e quantificar esta estruturação.

7. REFERÊNCIAS

ROBINSON J, WALLER MJ, PARHAM P, de GROOT N, BONTROP R, KENNEDY LJ, STOEHR P, MARSH SGE; "IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex." *Nucleic Acids Research* (2003), **31**:311-314.

Disponível em <<http://www.ebi.ac.uk/imgt/hla/stats.html>> Acesso em : 10 dez, 2009

Boletim MedulaNet Número 15.

Disponível em <http://www.inca.gov.br/conteudo_view.asp?id=2596> Acesso em 10 dez, 2009.

ALBERTON, U. P.; **Estimativas das Probabilidades de Encontrar Doadores de Medula Óssea Compatíveis, Não Relacionados, em Função do Tamanho e da Composição do Banco de Dados.** Trabalho de Conclusão de Curso. Curitiba: Ciências Biológicas – UFPR, 2009

BICALHO, M. G. & SANTOS, P. S. C. & SILVA, W. A. & RUIZ, T. M. **Sistema LIGH: Disponibilizar para o Redome em Tempo Real as Informações do Doador Voluntário de Medula Óssea.** Carta ao Editor. *Rev. bras. hematol. hemoter.* 2005;27(2):138-147

BOMPEIXE, E. P. & RUIZ, T. M. & MAGALHÃES, J. C. M. & SILVA, W. A. & GOLANOVSKI, M. A. & BICALHO, M. G. **Distribution of HLA Haplotype Frequencies of Hematopoietic Stem-cells Potential Donors in Paraná State, Brazil.** Supplement to *Human Immunology* Vol. 70, Sup. 1, p 94.

BOUZAS, L. F. S. **Os avanços da Sociedade Brasileira de Transplante de Medula Óssea.** *Rev. bras. hematol. hemoter.* 2004;26(3):153-154

CLARK, A. G. & HARTL, D. L. **Principles of Population Genetics.** Editora Sinauer Associates, INC. Segunda Edição. 683p. 1989.

EXCOFFIER, L. G. Laval, and S. Schneider (2005). **Arlequin ver. 3.0: An integrated software package for population genetics data analysis.** *Evolutionary Bioinformatics Online* 1:47-50.

GILLESPIE, J. H. **Population Genetics: A Concise Guide.** Editora The Johns Hopkins University Press. Primeira Edição. 169p. 1998.

MATSUBARA, T. C. & CARVALHO, E. C. de & CANINI, S. R. M. da S. & SAWADA, N. O. **A Crise Familiar no Contexto do Transplante de Medula Óssea (TMO): Uma Revisão Integrativa.** *Revista Latino-Americana de Enfermagem* 2007 julho-agosto; 15(4). Disponível em: www.eerp.usp.br/rlae > Acesso em 12 de dez. de 2009

PROBST, C. M.; **Diversidade Alélica e Haplótipica dos Genes HLA de Classe II em Populações Paranaenses**. Dissertação de mestrado. Curitiba: Programa de Pós-Graduação em Genética, UFPR, 2000.

RUIZ, T. M. & MAGALHÃES, J. C. M. & BOMPEIXE, E. P. & GOLANOVSKI, M. A. & SILVA, W. A. & BICALHO, M. G. **Distribuição das Frequências Haplóticas HLA e Distâncias Genéticas em Doadores Voluntários de Medula Óssea conforme seu Grupo Étnico**. Rev. bras. hematol. hemoter., 2008;30(3):6-13

RUUTU, T. **Allogeneic Bone Marrow Transplantation for Malignant Haematological Disorders**. Acta Oncologica Vol. 36, No. 6, pp. 547-548, 1997.

SALZANO, F. M. & FREIRE-MAIA, N. **Problems in Human Biology**. Editora Wayne State University Press. Primeira Edição. 199p. 1970.

Natural history of the major histocompatibility complex

J Klein - 1986 - John Wiley & Sons