Renan Barbosa Lemes

Estimativa de parâmetros genético-populacionais de
interesse em isolados populacionais do Vale do Ribeira
(remanescentes de quilombos).

*Estimation of population genetic parameters in human
isolates from Vale do Ribeira, São Paulo ("quilombo"
populations)*

São Paulo

2013

Renan Barbosa Lemes

Estimativa de parâmetros genético-populacionais de
interesse em isolados populacionais do Vale do Ribeira
(remanescentes de quilombos).

*Estimation of population genetic parameters in human
isolates from Vale do Ribeira, São Paulo ("quilombo"
populations)*

Dissertação apresentada ao
Instituto de Biociências da
Universidade de São Paulo,
para a obtenção de Título de
Mestre em Ciências, na Área de
Biologia/Genética.

Orientador: Prof. Dr. Paulo
Alberto Otto

São Paulo

2013

Ficha Catalográfica

Lemes, Renan Barbosa

Estimativa de parâmetros genético-populacionais de interesse em isolados populacionais do Vale do Ribeira (remanescentes de quilombos).

120 páginas

Dissertação (Mestrado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.

1. Endocruzamento; 2. Isolados populacionais; 3. Remanescentes de Quilombos. I. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof. Dr.
Paulo Alberto Otto
Orientador

À Juliana Carnavalli
e à minha família,
pelo apoio, afeto e
companhia.

"A mente que se abre a uma nova idéia
jamais voltará ao seu tamanho original"

Albert Einstein

AGRADECIMENTOS

Ao meu orientador Prof. Dr. Paulo A. Otto pela orientação segura e extrema paciência.

À Profa. Dra. Regina Célia Mingroni Netto e ao Prof. Dr. Diogo Meyer pela colaboração.

Às colegas Lilian Kimura e Kelly Nunes pelo auxílio nas técnicas moleculares e amizade.

À Juliana Carnavalli pelo auxílio nas regenotipagens.

Aos colegas de laboratório Ana Carla, Bárbara, Dayane, Debora, Fábio, Leandro, Magnolia, Maria Helena, Ramalho, Renata, Rodrigo, Uirá e Vitor, pelas estimulantes discussões.

À chefia do Depto. de Genética e Biologia Evolutiva e à direção do Instituto de Biociências da Universidade de São Paulo, pela infra-estrutura que permitiu a realização desse projeto.

À Profa. Dra. Mayana Zatz, coordenadora do CEPID (Centro de Estudos do Genoma Humano).

Aos funcionários do Centro de Estudos do Genoma Humano, pelo auxílio técnico com as genotipagens.

À FAPESP e ao CNPq pelo auxílio financeiro.

Aos Profs. Drs. João Farias Guerreiro, Sidney Emanuel Batista dos Santos, Andrea Ribeiro dos Santos e Eduardo José Melo dos Santos da Universidade Federal do Pará, pela cessão da genealogia de Arara do Iriri.

Aos funcionários e amigos Israel, Maraisa, Maria Teresa e Paulo Rogério pelo suporte técnico.

Às prefeituras dos municípios de Eldorado, Iporanga e Barra do Turvo pelo apoio.

Às irmãs Angela Biagioni e Maria Sueli Berlanga, da Casa Paroquial de Eldorado e a Antônio Carlos Nicomedes, do MOAB, pela ajuda no contato com as comunidades.

Aos habitantes das comunidades quilombolas, especialmente aos seus líderes e agentes de saúde.

Ao Prof. Fabio M. do Nascimento pelo incentivo.

Aos amigos Carolina, Juliana Jordão, Marco, Rodrigo, Adam, Felipe, Luis Gustavo, Daniele, Michael, Renê, Julia e Rafael pelo apoio.

Aos meus familiares, meu pai Sebastião, Inês, minha mãe Tania, Ivo, Bianca, Mauricio, Rafael, Amanda, tio Roney, avó Neyde, tios Reginaldo e Jussara, Aquiles, Átila e a toda minha família, pelo carinho.

ÍNDICE

---

1. INTRODUÇÃO	1
1.1. Sistemas de cruzamentos.....1	
1.1.1. Casamentos aleatórios.....1	
1.1.2. Casamentos preferenciais.....2	
1.1.3. Casamentos seletivos.....4	
1.1.4. Casamentos endogâmicos.....4	
1.2. Efeito Wahlund.....6	
1.3. Remanescentes de quilombos do Vale do Ribeira.....7	
2. OBJETIVOS	14
3. MATERIAIS E MÉTODOS	16
3.1. Estimativas do coeficiente médio de endocruzamento por meio de análise de genealogias.....18	
3.1.1. Confecção de genealogias.....18	
3.1.2. Metodologia de análise.....19	
3.2. Obtenção do valor do coeficiente médio de endocruzamento por meio da análise da frequência de pessoas e casais com o mesmo sobrenome.....23	
3.3. Obtenção do valor do coeficiente médio de endocruzamento por meio da análise de marcadores moleculares.....25	
3.3.1. Marcadores moleculares previamente genotipados.....25	
3.3.2. Extração de DNA genômico.....26	
3.3.3. Genotipagem de microssatélites genômicos.....27	
3.3.4. Análise computacional dos resultados obtidos a partir dos marcadores moleculares.....29	
4. RESULTADOS	34
4.1. Estimativas do coeficiente médio de endocruzamento por meio de análise de genealogias.....34	
4.1.1. Teste da metodologia de análise de genealogias.....37	
4.1.1.1. Isolado indígena Arara do Iriri.....38	
4.1.1.2. Isolado quilombola de Valongo (SC) 39	

4.2. Correção do método de estimativa do valor do coeficiente médio de endocruzamento por meio da análise da frequência de pessoas e casais com o mesmo sobrenome.....	41
4.3. Análise dos marcadores moleculares.....	43
4.3.1. Obtenção dos coeficientes médios de endocruzamento.....	43
4.3.2. Obtenção da variância de F.....	46
4.3.3. Análise de subestruturação populacional.....	47
4.4. Evolução do valor de F em sistemas regulares de endocruzamento.....	50
5. DISCUSSÃO	51
5.1. Estimativa do coeficiente médio de endocruzamento por meio da análise de genealogias.....	51
5.2. Obtenção do coeficiente de endocruzamento por meio da frequência de casamentos de pessoas com o mesmo sobrenome.....	55
5.3. Obtenção do coeficiente de endocruzamento por meio da análise de marcadores moleculares.....	57
5.4. Análise comparativa dos valores de F.....	59
5.5. Análise de subestruturação populacional.....	60
6. CONCLUSÃO	64
7. RESUMO	66
8. ABSTRACT	68
9. REFERÊNCIAS BIBLIOGRÁFICAS	70
10. ANEXOS	79
10.1. Anexo 1.....	79
10.2. Anexo 2.....	80
10.3. Anexo 3.....	81
10.4. Anexo 4.....	89
10.5. Anexo 5.....	92
10.5.1. Autofecundação.....	92
10.5.2. Casamentos exclusivos entre irmãos.....	93

10.5.3.	Casamentos exclusivos entre meio-irmãos.....	107
10.5.4.	Casamentos exclusivos entre primos duplos em primeiro grau.....	110
10.5.5.	Casamentos circulares entre meio-irmãos.....	112
10.5.6.	Análise comparativa dos sistemas de casamentos exclusivos.....	115

ÍNDICE DE FIGURAS

Figura 1.1: Localização das comunidades quilombolas do Vale do Ribeira.....	9
Figura 1.2: Rio Ribeira do Iguape.....	11
Figura 1.3: Casa quilombo de pau-a-pique.....	12
Figura 1.4: A Capela de Nossa Senhora do Rosário dos Homens Pretos (Ivaporanduva).....	12
Figura 1.5: Balsa de acesso ao quilombo Pedro Cubas.....	12
Figura 1.6: Moradia quilombola, ilustrando o isolamento em meio à Mata Atlântica.....	12
Figura 1.7: Uma das vilas quilombolas inseridas no quilombo Nhunguara.....	13
Figura 3.1: Genealogia representando um casamento entre irmãos com destaque para cada possível caminho de transmissão de um alelo.....	21
Figura 3.2: Genealogia representando as possíveis trajetórias de um alelo no caso em que os genitores de um indivíduo endocruzado possuem um ascendente também endocruzado.....	22
Figura 4.1: Heredogramas simplificados de casamentos consanguíneos encontrados nas genealogias das populações quilombolas.....	35
Figura 4.2: Heredogramas resultantes do desmembramento da genealogia de Arara do Iriri (Anexo 1)	38
Figura 4.3: Heredogramas resultantes do desmembramento da genealogia de Valongo (Anexo 2).....	40
Figura 4.4: Em destaque, as combinações possíveis de genitores de um casal qualquer de primos em primeiro grau.....	42
Figura 4.5: Gráfico correlacionando as estimativas da variância de F (presente trabalho x literatura).....	47
Figura 10.1: Sistema de casamentos exclusivos entre pares de irmãos.....	94

Figura 10.2: Representação dos caminhos possíveis de transmissão de alelo para a geração dois.....	99
Figura 10.3: Representação dos caminhos possíveis de transmissão de alelo para a geração três.....	100
Figura 10.4: Representação dos caminhos possíveis de transmissão de alelo para a geração quatro.....	100
Figura 10.5: Representação dos caminhos possíveis de transmissão de alelo para a geração cinco, nos quais um ancestral comum a ambos os genitores da geração cinco também é endocruzado.....	101
Figura 10.6: Sistema de casamentos exclusivos entre pares de irmãos.....	103
Figura 10.7: Sistema de casamentos exclusivos entre meio-irmãos.....	107
Figura 10.8: Sistema de casamentos exclusivos entre primos duplos em primeiro grau.....	110
Figura 10.9: Sistema de casamentos circulares entre meio-irmãos.....	112

ÍNDICE DE TABELAS

Tabela 3.1: Número estimado aproximado do total de habitantes de cada comunidade, número amostrado de indivíduos quilombolas entrevistados, número máximo de indivíduos cujas amostras de DNA foram genotipadas em cada comunidade e respectivas porcentagens aproximadas de cobertura em relação à população total de cada comunidade.....	17
Tabela 3.2: Localização cromossômica, sequências dos <i>primers</i> dos 16 microssatélites analisados no presente estudo e respectivas marcações.....	28
Tabela 4.1: Número de heredogramas analisados em cada comunidade quilombola.....	34
Tabela 4.2: Coeficiente de endocruzamento individual e número de indivíduos endocruzados em cada população, de acordo com os heredogramas representados na Figura 4.1.....	36
Tabela 4.3: Número de indivíduos, coeficiente médio de endocruzamento e porcentagem de indivíduos provenientes de casamentos consanguíneos de cada população e em relação à população total.....	37
Tabela 4.4: Valores do coeficiente médio de endocruzamento F de cada loco por população.....	44
Tabela 4.5: Estimativas de F_m em cada uma das populações para os conjuntos de locos de microssatélites e de todos os demais locos.....	45
Tabela 4.6: Dados de frequências genotípicas e alélicas, coeficiente médio de endocruzamento (com respectiva variância) em cada uma das populações, referentes ao loco <i>HBB</i>	46
Tabela 4.7: Estimativas de F_{IT} , F_{ST} e F_{IS} e dos valores de suas medianas baseados em 1000 simulações e respectivos intervalos de confiança a 95%.....	48

Tabela 5.1: Estimativas do coeficiente médio de endocruzamento e da frequência de casamentos consanguíneos em diversos isolados e semi-isolados populacionais encontrados na literatura.....	52

Tabela 5.2: Valores de médios de F obtidos por análise genealógica e pela análise de marcadores moleculares.....	59

1. INTRODUÇÃO

Os assuntos básicos de genética de populações teórica apresentados nos itens 1.1 e 2.1 abaixo já constituem uma espécie de senso comum e podem ser encontrados, em níveis de complexidade variáveis, em qualquer livro texto da especialidade (Crow e Kimura, 1970; Spiess, 1977; Weir, 1996; Beiguelman, 2005; Otto, 2008; Hartl e Clark, 2010) e diversos artigos científicos publicados em periódicos da área (Hardy, 1908; Wright, 1921a, 1921b, 1921c, 1921d, 1921e; Morton *et al.*, 1956; Lewontin *et al.*, 1968; Peltonen *et al.*, 1995; Jorde *et al.*, 2000; Arcos-Burgos e Muenke, 2002; Stulp *et al.*, 2013). Devido a isso, dispensamos (a não ser em casos especiais) as referências bibliográficas correspondentes aos trabalhos originais que os criaram.

1.1. Sistemas de cruzamentos

A teoria da genética de populações admite a existência de quatro tipos distintos de cruzamentos: aleatórios (panmíticos), preferenciais, seletivos e endogâmicos.

1.1.1. Casamentos aleatórios

Os casamentos aleatórios (em regime de pan-mixia) ocorrem quando a probabilidade de um indivíduo de genótipo qualquer se cruzar com outro de um genótipo particular é igual à frequência desse genótipo na população. Se, para um

loco autossômico com dois alelos (**A** e **a**), os genótipos ocorrerem com frequências **d**, **h** e **r**, os diversos tipos de cruzamentos (**AAxAA** , **AAxAa** , ... , **Aaxaa** , **aaxaa**) ocorrerão com frequências **d**², **2dh**, ..., **2hr**, **r**² , que correspondem à expansão de (**d** + **h** + **r**)². Definindo-se **p** = **d** + **h**/2 como sendo a frequência do alelo **A** e **q** = 1-**p** = **h**/2 + **r** a frequência do alelo **a**, vem que

$$\begin{aligned}
 (\mathbf{d} + \mathbf{h} + \mathbf{r})^2 &= [(\mathbf{d} + \mathbf{h}/2) + (\mathbf{h}/2 + \mathbf{r})]^2 \\
 &= (\mathbf{p} + \mathbf{q})^2 = \mathbf{p}^2 + 2\mathbf{pq} + \mathbf{q}^2 .
 \end{aligned}$$

Essa formulação simples demonstra imediatamente que cruzamentos ao acaso correspondem à união ao acaso de gametas; além disso, mostra que, quaisquer que sejam as frequências **d**, **h** e **r** de indivíduos **AA**, **Aa** e **aa**, as frequências genotípicas estabilizam-se a partir da geração filial nas proporções **p**², **2pq** e **q**² (proporções de Hardy-Weinberg), admitindo-se como desprezíveis os efeitos da seleção, deriva, mutação e migração.

1.1.2. Casamentos preferenciais

Trata-se de um sistema de cruzamentos no qual a escolha do parceiro, em vez de ser aleatória, é orientada pelo fenótipo. Quando pares de indivíduos com fenótipo semelhante ocorrem com uma frequência superior à esperada sob regime de pan-mixia, diz-se que os cruzamentos são preferenciais positivos; no caso contrário (quando pares de indivíduos com fenótipos diferentes ocorrem com uma

frequência superior à esperada sob regime de pan-mixia), recebem o nome de preferenciais negativos. Em todos os sistemas possíveis de cruzamentos preferenciais positivos não ocorre alteração das frequências alélicas (observadas as mesmas restrições mencionadas em relação aos cruzamentos aleatórios). Na maioria dos sistemas de cruzamentos preferenciais negativos, no entanto, geralmente ocorre alteração de frequências alélicas e essas se mantêm constantes apenas em alguns poucos modelos, geralmente associados a condições iniciais especiais (como, por exemplo, frequências alélicas idênticas). Enquanto o efeito principal dos cruzamentos preferenciais positivos é o aumento da homozigose dentro do loco que alberga os alelos que determinam a característica fenotípica que é levada em conta na escolha do parceiro, os negativos determinam um aumento (restrito) da taxa de heterozigose do loco ao nível populacional. Estudos feitos ao longo das últimas décadas sugerem que os indivíduos de algumas espécies de mamíferos escolhem parceiros com o sistema imunológico mais diverso possível por meio do odor emitido. Esse tipo de casamento resulta no aumento da frequência de heterozigotos na população (Brown e Eklund, 1994; Ober et al., 1997; Roberts et al., 2008 Penn e Potts, 2013).

1.1.3. Casamentos seletivos

Toda vez que o sistema de cruzamentos determina alteração de frequências alélicas ele é dito seletivo. A maioria dos sistemas de cruzamentos antigamente conhecidos como preferenciais negativos são atualmente classificados nesta categoria, uma vez que geralmente (a não ser em condições iniciais especiais associadas a alguns poucos modelos) determinam alterações drásticas de frequências alélicas e geralmente a eliminação do alelo menos favorecido da população. Para exemplificar o casamento seletivo, podemos supor um modelo simples no qual 90% dos indivíduos de uma população apresentam um dado fenótipo e 10% o outro; no caso de cruzamentos preferenciais negativos grande parte da classe mais frequente não vai conseguir se reproduzir devido à baixa frequência da outra classe.

1.1.4. Casamentos endogâmicos

Diz-se que num sistema de cruzamentos endogâmicos a escolha do parceiro é orientada pelo grau de parentesco. A consequência mais óbvia é que os indivíduos endocruzados frequentemente herdam o mesmo alelo (idêntico por descendência) de ambos os genitores, esse fenômeno sendo o responsável direto pelo aumento da taxa de homozigose populacional. Como um parente é fenotipicamente mais semelhante ao indivíduo que exerce a escolha, as definições de cruzamentos preferenciais e endogâmicos são ambíguas,

pois claramente um tipo determina sempre um resíduo do outro. Devido a isso, a literatura contém vários trabalhos que tentam diferenciar os dois sistemas de uma maneira tanto heurística como prática (Crow e Felsenstein, 1968; Lewontin *et al.*, 1968).

Define-se o coeficiente de endocruzamento (**F**) como a probabilidade de que um par de alelos em dois gametas, os quais se unem para formar o zigoto, seja idêntico por descendência, ou, alternativamente, como a probabilidade de um indivíduo ser autozigoto para um determinado loco. Assim sendo, a probabilidade de um indivíduo ser homozigoto **AA** para um loco bialélico qualquer é obtida somando-se a probabilidade **Fp** do indivíduo ser autozigoto quanto ao alelo **A** e a probabilidade **(1-F)p²** do indivíduo ser homozigoto **AA** por origens distintas (alozigoto). Portanto,

$$P(AA) = Fp + (1-F)p^2 = p^2 + Fpq$$
$$P(aa) = Fq + (1-F)q^2 = q^2 + Fpq$$
$$P(Aa) = 2pq(1-F) = 2pq - 2Fpq .$$

Nas expressões acima a quantidade **2Fpq** indica o excesso de homozigose populacional (ou o decréscimo da taxa de heterozigose) devido ao endocruzamento, em relação às taxas existentes em pan-mixia. Obviamente o efeito principal da endogamia é o aumento da taxa de homozigose da população, em relação a todos os locos do indivíduo (em contraste com o efeito observado em um loco apenas no caso dos cruzamentos preferenciais).

Estimativas do valor de **F** podem ser obtidas tanto de forma individual (análise do heredograma de uma família) como para a população como um todo (coeficiente médio de endocruzamento). Neste último caso, pode ser obtido da ponderação de todos os valores **F_i** individuais, estimados a partir da análise de todos os heredogramas de uma comunidade ou população (v. item 3.1.2) ou a partir da fórmula **F = 1 - h/2pq**, obtida diretamente da frequência populacional observada **h** de heterozigotos quanto a dois alelos de frequências estimadas **p** e **q**, **h = 2pq(1-F)**.

1.2. Efeito Wahlund

As taxas de homozigose e de heterozigose das populações podem ser alteradas não apenas pelos diversos sistemas de cruzamentos descritos, como também pela presença de subestruturação populacional (subdivisão de uma população em isolados menores, dentro dos quais os cruzamentos ocorrem apenas entre indivíduos restritos ao mesmo isolado).

O efeito genético da fragmentação de uma grande população em isolados menores é o aumento da homozigose na população considerada como um todo, devido a diferenças nas frequências alélicas das subpopulações, geradas por deriva e outros mecanismos.

Com o fenômeno conhecido pelo nome de **quebra dos isolados** ou **efeito Wahlund**, as frequências gênicas da

população única continuam as mesmas que existiam na população considerada como um todo no período anterior à fragmentação, mas a frequência de homozigotos diminui.

Sob o ponto de vista prático, a análise de subestruturação compreende a estimativa dos seguintes indicadores: F_{ST} (índice de fixação devido à subdivisão da população), F_{IT} (índice de fixação total, devido tanto a diferenças de frequências alélicas entre as subpopulações como a endogamia dentro dessas mesmas subpopulações) e F_{IS} (índice de fixação devido à endogamia ocorrendo dentro das subpopulações). Esses valores são obtidos das fórmulas

$$F_i = 1 - h_i / 2p_i q_i ;$$

$$F_{ST} = (F_{IT} - F_{IS}) / (1 - F_{IS}) = \text{var}(p) / pq$$

$$= [\sum x_i p_i^2 - (\sum x_i p_i)^2] / (\sum x_i p_i \cdot \sum x_i q_i) ;$$

$$F_{IT} = F_{ST} + F_{IS} - F_{IS} \cdot F_{ST} = 1 - P(Aa) / 2pq$$

$$= 1 - \sum x_i p_i q_i (1 - F_i) / (\sum x_i p_i \cdot \sum x_i q_i) ;$$

$$F_{IS} = (F_{IT} - F_{ST}) / (1 - F_{ST}) = 1 - 2 \sum x_i p_i q_i (1 - F_i) / 2 \sum x_i p_i q_i ,$$

em que h_i é a frequência de heterozigotos na população, p_i e q_i são as frequências alélicas em cada subpopulação, x_i é um fator de ponderação e p e q são as frequências alélicas na população total.

1.3. Remanescentes de quilombos do Vale do Ribeira

A ocupação do Vale do Ribeira iniciou-se no período pré-colombiano, já que populações de ameríndios utilizavam

a região como rota de acesso ao litoral. Os colonizadores, principalmente portugueses, começaram a utilizar a região como forma de acesso ao interior apenas no século XVI. Já os africanos chegaram maciçamente à região durante o século XVII como mão-de-obra escrava atrelada à atividade de mineração. Com o fim da produção aurífera, ocorreu a gradativa saída da população branca e o início da formação dos primeiros quilombos (Oliveira Junior *et al.*, 2000).

O vale do Ribeira concentra a maior quantidade de comunidades remanescentes de quilombos do Estado de São Paulo. Essas comunidades foram fundadas por escravos alforriados, fugidos ou simplesmente abandonados (Santos e Tatto, 2008).

A porção paulista do Vale do Ribeira (Figura 1.1-B) abrange uma área de cerca de 10% do território do estado, contendo 22 municípios (Santos e Tattoo, 2008). Segundo dados recentes publicados pelo ITESP (Instituto de Terras do Estado de São Paulo), das 31 comunidades identificadas nesses municípios (**Biguazinho, Morro Seco, Engenho, Peropava, Mandira, Poça, São Pedro, Galvão, Ivaporanduva, Pedro Cubas, Nhunguara, André Lopes, Sapatu, Pedro Cubas de Cima, Abobral, Bananal Pequeno, Boa Esperança, Reginaldo, Terra Seca, Cedro, Pedra Preta, Cangume, Maria Rosa, Pilões, Praia Grande, Porto Velho, Bombas, Castelhanos, Piririca, Poço Grande e Jacumirim**), as 24 indicadas em

negrito já foram reconhecidas oficialmente ou estão em fase de reconhecimento como remanescentes de quilombos.

Figura 1.1: A. Mapa do Brasil, destacando o estado de São Paulo. B. Mapa do estado de São Paulo contendo a porção paulista do Vale do Ribeira (em cinza) e em seu interior os municípios (em preto) destacados ao lado. C. Mapa dos municípios de Eldorado (EL), Jacupiranga (JP), Iporanga (IP) e Barra do Turvo (BT) e das comunidades de Maria Rosa (MR), Pilões (PS), Galvão (GA), São Pedro (SP), Pedro Cubas (PC), Ivaporanduva (IV), Sapatu (TU), André Lopes (AN), Nhunguara (NH), Abobral margem esquerda (AB), Abobral margem direita (DR), Poça (PA) e Reginaldo (RE).

O modo de vida atual das comunidades quilombolas do Vale do Ribeira é tradicionalmente rural, mas está em transformação devido à sobreposição de terras com parques estaduais e áreas de proteção ambiental, já que o Vale do Ribeira está inserido na maior área contínua do bioma Mata Atlântica do território brasileiro (Santos e Tatto, 2008; Pasinato e Rettl, 2009).

A transformação no modo de vida que essas comunidades quilombolas vêm atravessando tem relação, também, com um

processo de abandono das práticas tradicionais de subsistência e a ênfase no cultivo de produtos com valor comercial, dando início a um processo de transição nutricional (Angeli, 2008; Crevelaro, 2009).

Esses remanescentes de quilombos são populações de transição também sob o ponto de vista epidemiológico, pois convivem ao mesmo tempo com doenças resultantes da sua falta de acesso a serviços básicos de saneamento e saúde (como por exemplo, as doenças parasitárias) e com doenças frequentes do mundo moderno, como a hipertensão e a obesidade, com prevalência muito alta nos países desenvolvidos e uma prevalência já considerada importante mesmo em países em desenvolvimento como o Brasil (Angeli *et al.*, 2011; Crevelaro, 2009; Kimura *et al.*, 2012).

As comunidades estudadas pelo grupo da Dra. Mingroni-Netto (Figura 1.1-C), as quais também foram utilizadas neste estudo, localizam-se nos municípios de Eldorado, Iporanga, Barra do Turvo e Jacupiranga; são elas: Abobral, André Lopes, Ivaporanduva, Galvão, Maria Rosa, Nhunguara, Pedro Cubas, Pilões, São Pedro, Sapatu, Poça e Reginaldo.

O grupo da Dra. Mingroni-Netto tem estudado, em projetos multidisciplinares da área de genética humana e médica, as comunidades remanescentes de quilombos do Vale do Ribeira desde o ano de 2000. Todas as informações obtidas a partir de estudos a respeito dessas comunidades

foram gentilmente disponibilizadas para o presente trabalho.

Destas comunidades, foram coletados, pelo grupo da Dra. Mingroni-Netto, dados antropométricos e clínicos, além de amostras de sangue periférico para extração de DNA, utilizados em diversos trabalhos (Mingroni-Netto *et al.*, 2002, 2009a, 2009b; Angeli, 2003, 2008; Cotrim, 2003; Macedo de Souza, 2003; Cotrim *et al.*, 2004; Angeli *et al.*, 2005, 2011; Auricchio *et al.*, 2007; Yeh *et al.*, 2008; Rincon, 2009; Kimura, 2010; Kimura *et al.*, 2012, 2013). As amostras de DNA foram disponibilizadas para o presente estudo, a fim de complementar os dados já existentes com a genotipagem de diversos marcadores de microssatélites, visando à análise dos parâmetros genético-populacionais de interesse.

Figura 1.2: Rio Ribeira do Iguape em meio ao remanescente de Mata Atlântica na região do quilombo Ivaporanduva (Fonte: arquivo do Laboratório de Genética Humana).

Figura 1.3: Casa de pau-a-pique, moradia comumente encontrada nas populações (Fonte: arquivo do Laboratório de Genética Humana).

Figura 1.4: A Capela de Nossa Senhora do Rosário dos Homens Pretos (construída em 1791) está inserida no quilombo de Ivaporunduva (Fonte: arquivo do Laboratório de Genética Humana).

Figura 1.5: Balsa de acesso ao quilombo Pedro Cubas, através do rio Ribeira do Iguape (Fonte: arquivo do Laboratório de Genética Humana).

Figura 1.6: Moradia quilombola, ilustrando o isolamento em meio à Mata Atlântica (Fonte: arquivo do Laboratório de Genética Humana).

Figura 1.7: Uma das vilas quilombolas inseridas no quilombo Nhunguara (Fonte: arquivo do Laboratório de Genética Humana).

2. OBJETIVOS

Por meio da análise conjunta de locos de inserção Alu autossômicos, de polimorfismos de regiões funcionais e intrônicas de genes autossômicos, marcadores clássicos como genes de grupos sanguíneos e alelos mutantes C e S das hemoglobinas, cujos resultados isolados foram apresentados em dissertações, teses e trabalhos publicados do grupo da Dra. Mingroni-Netto e também utilizando os resultados por nós obtidos da análise de microssatélites autossômicos, procuramos obter parâmetros globais de caracterização das comunidades: sistema de cruzamentos e medidas de subestruturação populacional.

O objetivo específico mais importante do presente trabalho é a determinação dos níveis populacionais de endogamia nos isolados acima, que apresentamos de maneira mais detalhada a seguir. Isso deverá ser realizado combinando-se três técnicas:

- (1) Obtenção do coeficiente médio de endocruzamento ponderando-se os coeficientes de endocruzamento de todos os indivíduos da população por meio da análise das genealogias que compõem os agregados populacionais (cada uma delas com um coeficiente de endocruzamento distinto) utilizando-se um método tradicional descrito em Crow e Kimura (1970).

- (2) Correção do valor do coeficiente médio de endocruzamento através da análise da frequência de pessoas e de casais com o mesmo sobrenome (Crow e Mange, 1965).
- (3) Determinação do coeficiente médio de endocruzamento por meio da genotipagem de amostras populacionais dos isolados, empregando-se para isso, além dos marcadores de DNA já estudados pelo grupo, os 16 listados abaixo, sugeridos pelo Dr. Diogo Meyer e já em uso pelo seu grupo: D5S816, D1S551, D7S3061, D4S3248, D13S317, D8S2324, D19S559, D6S1040, D16S539, D9S922, D10S1426, D7S821, D20S482, D21S1437, D9S301 e D18S535. A genotipagem foi feita utilizando-se as instalações e recursos dos laboratórios da Dra. Regina Célia Mingroni-Netto e do Dr. Diogo Meyer e do Centro de Estudos do Genoma Humano, coordenado pela Profa. Dra. Mayana Zatz.

3. MATERIAL E MÉTODOS

Os estudos nas comunidades quilombolas do Vale do Ribeira conduzidos pelo grupo da Dra. Mingronni-Netto iniciaram-se como parte de um projeto na área de genética médica sobre síndrome do cromossomo X Frágil, realizado entre 2000 e 2003. Coletaram-se durante as primeiras viagens amostras de sangue periférico de indivíduos adultos que consentiram em participar do estudo. A partir do ano de 2003, iniciou-se a segunda etapa do projeto voltada para o estudo da obesidade e da hipertensão. Essa etapa durou até 2009 e nela foram realizadas um total de 21 viagens, com o objetivo de reexaminar os indivíduos adultos, cujas amostras de DNA já haviam sido colhidas, e de coletar medidas para avaliar a pressão arterial e medidas antropométricas complementares. Também foram coletadas amostras de sangue periférico de novos indivíduos que aderiram à pesquisa (Kimura, 2010).

As comunidades que participaram dessas coletas (Figura 1.1-C) foram: Maria Rosa (MR), Pilões (PS), Galvão (GA), São Pedro (SP), Pedro Cubas (PC), Ivaporanduva (IV), Sapatu (TU), André Lopes (AN), Nhunguara (NH), Abobral margem esquerda (AB), Abobral margem direita (DR), Poça (PA) e Reginaldo (RE).

Utilizamos para o presente estudo os dados genealógicos provenientes de entrevistas realizadas com

cerca de 2000 indivíduos e dados moleculares obtidos de amostras de DNA de cerca de 1000 indivíduos das comunidades citadas. Cada um dos marcadores moleculares utilizados foram genotipados considerando quantidades diferentes de indivíduos. Os locos *HBB*, SNPs associados à obesidade, SNPs associados à hipertensão e marcadores de microssatélites foram genotipados, respectivamente, para aproximadamente 1000, 400, 700 e 300 indivíduos (Auricchio *et al.*, 2007; Angeli *et al.*, 2011; Kimura *et al.*, 2011).

Apresentamos na Tabela 3.1 a distribuição dos indivíduos amostrados utilizados em nosso estudo.

Tabela 3.1: Número estimado aproximado do total de habitantes de cada comunidade, número amostrado de indivíduos quilombolas entrevistados, número máximo de indivíduos cujas amostras de DNA foram genotipadas em cada comunidade e respectivas porcentagens aproximadas de cobertura em relação à população total de cada comunidade.

Comunidade	Nº estimado de indivíduos	Nº absoluto de indivíduos entrevistados	Porcentagem de indivíduos entrevistados	Nº absoluto de indivíduos genotipados	Porcentagem de indivíduos genotipados
AB/DR	400	364	91%	208	52%
AN	290	247	85%	133	46%
GA/SP	240	224	93%	126	52%
IV	290	217	75%	144	50%
MR/PS	185	148	80%	76	41%
NH	440	237	54%	128	29%
PC	265	263	99%	143	54%
TU	290	179	61%	102	35%
TOTAL	2400	1879	78%	1060	44%
PA	220	51	23%	50	23%
RE	250	73	29%	-	-

As comunidades de Abobral (junção de Abobral margem esquerda e margem direita), Galvão/São Pedro e Maria Rosa/Pilões são populações que, além de serem

geograficamente contíguas, são, também, extremamente relacionadas sob o ponto de vista genealógico, visto que os núcleos familiares mais representativos dessas populações são comuns dentro de cada um desses pares. Para tanto, o grupo da Dra. Mingroni Netto aglutinou as genealogias dessas seis comunidades em três genealogias. Portanto, em todas as nossas análises, tratamos essas comunidades da forma como foram agrupadas.

Na Tabela 3.1 excluímos as comunidades de Poça e Reginaldo, porque essas duas populações possuem uma origem ancestral diferente das outras.

Apresentamos nos itens a seguir a descrição detalhada da metodologia empregada neste estudo.

3.1. Estimativas do coeficiente médio de endocruzamento por meio de análise de genealogias

3.1.1. Confeção de genealogias

Utilizamos em nosso trabalho as genealogias desenhadas à mão ou em computador pelo grupo da Dra. Mingronni-Netto durante quase uma década. As genealogias desenhadas em computador foram digitalizadas por meio do programa GenoPro 2007 v2.00 (© GenoPro).

Durante este trabalho, colaboramos com a confeção digital da genealogia da comunidade de Pedro Cubas, também

com a utilização mesmo do programa, com a finalidade de corrigirmos algumas incongruências e repetições.

Contribuímos, também, com algumas correções nas genealogias de diversas comunidades, utilizando informações pessoalmente obtidas durante duas expedições a campo.

3.1.2. Metodologia de análise

Para obter o coeficiente médio de endocruzamento das comunidades remanescentes de quilombos do Vale do Ribeira por meio da análise das genealogias, padronizamos durante a iniciação científica dois métodos de análise.

Os dois métodos foram testados em uma comunidade com estrutura complexa, em regime total de endocruzamento. Trata-se de um isolado populacional amazônico de indígenas, denominado Arara do Iriri, composto por 38 indivíduos descendentes de um casal de irmãos provenientes da tribo Arara do Laranjal (isolado com um tamanho total, portanto, de 40 indivíduos).

O primeiro método padronizado consistiu em uma metodologia clássica segundo a qual são contados os passos genéticos contidos em cada genealogia e cuja descrição é encontrada em livros-texto básicos de genética de populações como o de Beiguelman (2005). O passo genético é a passagem de um gene de uma geração a outra, correspondendo à probabilidade de um dos dois alelos de um

loco diplóide ser transmitido ao indivíduo da geração seguinte (taxa de segregação com valor $1/2$).

O segundo método avaliado, o qual foi utilizado nas análises descritas neste trabalho, é encontrado também em diversos livros-texto, geralmente com conteúdo mais avançado ou aprofundado, como o de Crow e Kimura (1970). Esse método é bem mais eficiente que o método baseado na contagem de passos genéticos, visto que reduz o número de possíveis erros por meio da listagem dos possíveis caminhos de cada alelo. O método está descrito e exemplificado por meio da Figura 3.1. A inspeção da figura mostra que a probabilidade do alelo transmitido de **A** para **D** ser o mesmo já transmitido de **A** para **C** é $1/2$.

Analogamente, a probabilidade de que os alelos transmitidos de **C** para **E** e de **D** para **E** também sejam iguais ao alelo transmitido de **A** para **C** ou **D** é $1/2$ para ambos os casos. Ou seja, a probabilidade de que o trajeto **CAD** (a maneira pela qual um mesmo alelo de **A** é transmitido em dose dupla para **E**) seja percorrido por um mesmo alelo é $(1/2)^3 = 1/8$.

Figura 3.1: Genealogia representando um casamento entre irmãos com destaque para cada possível caminho de transmissão de um alelo.

O trajeto percorrido (CBD) por um alelo transmitido inicialmente por **B** é semelhante, correspondendo ao mesmo valor de probabilidade (1/8).

Assim, a probabilidade de o indivíduo **E** ser homozigoto por origem comum é composta pela soma das probabilidades das possíveis trajetórias do alelo, ou seja:

$$F_E = P(\underline{C}\underline{A}\underline{D}) + P(\underline{C}\underline{B}\underline{D}) = (1/2)^3 + (1/2)^3 = (1/2)^2 \\ = 1/4$$

Da mesma forma que o primeiro método citado, este método também leva em conta a possibilidade do ascendente comum dos genitores do indivíduo endocruzado também ser endocruzado. A situação é exemplificada pela Figura 3.2.

Figura 3.2: Genealogia representando as possíveis trajetórias de um alelo no caso em que os genitores (**G** e **H**) de um indivíduo endocruzado (**I**) possuem um ascendente também endocruzado (**E**).

Neste caso, os ascendentes comuns dos genitores do indivíduo **I** são os indivíduos **E** e **F**, porém o indivíduo **E** pode ser homozigoto por origem comum (com probabilidade F_E) ou não (com probabilidade $1-F_E$). Na hipótese de ser homozigoto por origem comum, o mesmo alelo é transmitido com certeza a **G** e **H**; na hipótese de não ser homozigoto por origem comum, a probabilidade de um mesmo alelo ser transmitido a **G** e **H** é $1/2$; portanto, dado que é endocruzado, a probabilidade de transmissão $\underline{GEH}$, em vez de reduzir-se a $(1/2)^3$, é

$$\begin{aligned}
 P(\underline{GEH}) &= [F_E \cdot 1 + (1-F_E) \cdot 1/2] \cdot (1/2)^2 \\
 &= (F_E + 1/2 - F_E/2) \cdot (1/2)^2 = (F_E/2 + 1/2) \cdot (1/2)^2 \\
 &= (1+F_E) \cdot (1/2)^3 .
 \end{aligned}$$

Nesse caso, portanto, F_I toma o seguinte valor:

$$\begin{aligned}
 F_I &= P(\underline{GEH}) + P(\underline{GFH}) = (1/2)^3 \cdot (1+F_E) + (1/2)^3 \\
 &= (1/2)^3 + (1/2)^3 \cdot (1+F_E) = 1/8 + 1/8 \cdot (1 + 1/4) \\
 &= 1/8 + 1/8 \cdot 5/4 = 1/8 + 5/32 \\
 &= 9/32
 \end{aligned}$$

Após a obtenção dos valores de F_i individuais, obtemos o valor de F para a população por meio da média dos valores de F_i de todos os indivíduos.

3.2. Obtenção do valor do coeficiente médio de endocruzamento por meio da análise da frequência de pessoas e casais com o mesmo sobrenome

Para se obter a estimativa de F por meio das frequências de sobrenomes tentamos adaptar o método de Crow e Mange (1965) às peculiaridades das populações em estudo, quais sejam as maneiras alternativas de transmissão dos sobrenomes parentais às progênies e a distribuição não-aleatória dos tipos de sexo entre os parentes progenitores dos casais consanguíneos.

Descrevemos nas linhas abaixo o método original desses autores, que foi baseado na aplicação de regras rigorosas de transmissão dos sobrenomes e na distribuição aleatória dos sexos entre os progenitores aparentados do casal consanguíneo.

Numa população qualquer, o coeficiente total de endocruzamento F (F_{IT} na notação de Wright) está

relacionado ao coeficiente de endocruzamento F_r (F_{ST} na notação de Wright) devido a fatores aleatórios (deriva ocorrendo em populações de tamanho efetivo pequeno, com cruzamentos ao acaso) e ao coeficiente de endocruzamento F_n (F_{IS} na notação de Wright) devido a fatores não-aleatórios (desvios da pan-mixia dentro das populações) através da equação $(1-F) = (1-F_n).(1-F_r)$ ou, explicitamente, $F = F_r + F_n - F_n F_r = F_n + (1-F_n)F_r = F_r + (1-F_r)F_n$.

A frequência populacional P observada de casais com o mesmo sobrenome indica que o coeficiente de endocruzamento total F é aproximadamente igual a $P/4$. Sendo p_k a proporção da subpopulação masculina com um determinado sobrenome e q_k a proporção correspondente à subpopulação feminina, vem que a frequência esperada, na população, de casais com o mesmo sobrenome é $\sum p_i q_i$; a contribuição para o coeficiente de endocruzamento aleatório F_r é $1/4$ dessa quantidade, ou seja, $F_r = \sum p_i q_i / 4$. De $F = F_r + F_n - F_n F_r$ obtemos $F_n = (F - F_r) / (1 - F_r)$; substituindo-se, nessa fórmula, F por $P/4$ e F_r por $\sum p_i q_i / 4$, obtemos imediatamente $F_n = (P/4 - \sum p_i q_i / 4) / (1 - \sum p_i q_i / 4) = (P - \sum p_i q_i) / (4 - \sum p_i q_i)$. No trabalho original desses autores a fórmula é apresentada por descuido como $F_n = (P - \sum p_i q_i) / 4(1 - \sum p_i q_i)$.

3.3. Obtenção do valor do coeficiente médio de endocruzamento por meio da análise de marcadores moleculares

3.3.1. Marcadores moleculares previamente genotipados

Os dados de alguns marcadores moleculares utilizados neste estudo foram gentilmente cedidos pelo Grupo da Dra. Mingroni-Netto. Esses marcadores são compostos por polimorfismos dos locos *ACE*, *NOS3*, *GNB3*, *AGT*, *ADD2*, *GRK4*, *PLIN1*, *INSIG2*, *LEP*, *LEPR*, *ADRB2*, *PPARG*, *RETN* e *HBB*.

Os locos *ACE* (rs1799752), *NOS3* (rs1799983), *GNB3* (rs5443 e rs5441), *AGT* (rs669), *ADD2* (rs3755351) e *GRK4* (rs1801058) foram genotipados por se tratarem marcadores presentes em genes candidatos possivelmente associados ao aumento de hipertensão essencial, já que estão relacionados direta ou indiretamente com a homeostase do volume sanguíneo e equilíbrio de Na⁺, por meio da participação do sistema renina-angiotensina-aldosterona (Kimura, 2010; Kimura et al., 2012).

Os dados do loco polimórfico *ACE* do tipo inserção de *Alu* já haviam sido genotipados inicialmente em indivíduos de seis comunidades quilombolas do Vale do Ribeira (Abobral margem esquerda, Galvão, São Pedro, Pedro Cubas, Pilões e Maria Rosa) para utilização como ferramenta de um estudo de subestruturação populacional realizado por Cotrim (2003).

Foram genotipados, também, marcadores presentes em genes candidatos que podem estar associados à predisposição do fenótipo de obesidade: *PLIN1* (rs2289487), *INSIG2* (rs7566605), *LEP* (rs2167270), *LEPR* (rs1137101), *ADRB2* (rs1042713), *PPARG* (rs1801282) e *RETN* (rs1862513) (Angeli, 2008; Angeli et al., 2011).

Foram estudadas, ainda, as mutações S e C do gene da hemoglobina (*HBB*) nos isolados quilombolas do Vale do Ribeira, uma vez que a mutação S é considerada a hemoglobinopatia mais comum no Brasil, estando presente em cerca de 4% dessas populações quilombolas (Auricchio et al., 2007).

3.3.2. Extração de DNA genômico

As amostras de DNA foram obtidas pelo grupo da Dra. Mingronni-Netto a partir de amostras do sangue periférico (aproximadamente 5 mL), as quais foram processadas e o DNA foi extraído de duas formas: segundo o protocolo de rotina do Laboratório de Genética Humana, com fenol e clorofórmio (usado nas primeiras amostras coletadas); ou utilizando-se o aparelho *Autopure LS* da Gentra Systems (empregado nas amostras mais recentemente coletadas).

Cada amostra de DNA foi quantificada com o auxílio de um espectrofotômetro (Nanodrop ND-1000, da Thermo Scientific) e uma alíquota de cada uma foi diluída em água *milli-Q* autoclavada com a finalidade de obterem-se amostras

de DNA com concentração entre 100 e 200 ng/ μ L para a genotipagem dos locos citados no item 3.3.1.; e amostras de DNA com concentração de 10 ng/ μ L para a genotipagem dos locos de microssatélites.

3.3.3. Genotipagem de microssatélites genômicos

Constituiu a parte prática de bancada do presente trabalho a obtenção de dados de genotipagem de 16 locos de microssatélites (sequências repetitivas de DNA que possuem trechos de dois a seis nucleotídeos) de aproximadamente 300 amostras de indivíduos quilombolas. Os *primers* para amplificar os microssatélites foram desenhados por meio do *software Primer3* (Rozen e Scaletsky, 2000) e as sequências *forward* foram marcadas com duas fluorescências distintas. A Tabela 3.2 apresenta as principais características dos locos de microssatélites utilizados.

A amplificação dos fragmentos contendo os locos de microssatélites foi realizada por meio de reações em cadeia da polimerase (PCR), em quatro sistemas multiplex distintos (no caso, reações que amplificam ao mesmo tempo quatro microssatélites de interesse). Os multiplex I a IV utilizados no presente trabalho foram organizados respectivamente da seguinte forma: D16S539, D10S1426, D7S821, D9S922 (I); D1S551, D4S3248, D7S3061, D5S816 (II); D18S535, D9S301, D20S482, D21S1437 (III); e D13S317, D6S1040, D19S559, D8S2324 (IV).

Tabela 3.2: Localização cromossômica, sequências dos *primers* dos 16 microssatélites analisados no presente estudo e respectivas marcações.

Locus	Cromossomo	Primer F 5'- 3'	Primer R 5'- 3'	Marcação
D1S551	1	TGGTGATCTGCCCCTATTCTA	TGGGAGTGTGCTCATTTTAAAC	FAM
D4S3248	4	CACACAGACAGAAAGCGTTACA	AATGCAGTGGGCCTATGTATCTA	FAM
D5S816	5	GAGCTATTGCCACTGAAAATCA	CTACTTGGCATCCCTGATGG	FAM
D6S1040	6	ATTGGATGAGGCTGGTGAGA	GGAATGGCCAGAAAATCAG	FAM
D7S821	7	TTTAAGATGGTGTGTGAAGCAGTAG	GGGGCAATAGGTAGGGAACATAA	HEX
D7S3061	7	CCTGGCCTACTATAGGATTTTATCA	GGAAGAGTGGGTGAGGAAAGTA	FAM
D8S2324	8	GCAGGTGTTCTGTCCATAATC	TGACGGAATGAGACTCCATCTAA	FAM
D9S922	9	GAATTCACCTCACGGAGCATAACA	TCACAGCCACACAAGGACATA	HEX
D9S301	9	TTCAAGACAGACAGGCAGACA	GGAAGGTGTGCAAGGATGTT	HEX
D10S1426	10	TTTGCTTGGCACCAACTATTTC	GTTGAAAACAGGGGCCTACAC	HEX
D13S317	13	GAAGTCTGGGATGTGGAGGA	TCCTTCAACTTGGGTTGAGC	FAM
D16S539	16	CAAGCTCTTCCTCTCCCTAGAT	GTGTGTGCATCTGTAAGCATGTAT	HEX
D18S535	18	GACAAAAGCCACACCATAACT	GCAGTTCCTTTCTGGGATAAT	HEX
D19S559	19	ACCAGCCTGACCAACATAGTG	GGAGGTCGATTTGGGACATA	FAM
D20S482	20	ATCAGAGGACAGCCTCCATATC	CAGAGACACCGAACCAATAAGA	HEX
D21S1437	21	GGTTGATTCCATGTCTTTGCT	TGAGGTGCTCCCAAACCTCTT	HEX

A reação de PCR foi realizada sob as seguintes condições: volume final de 25 µL; cerca de 5 ng de amostra de DNA; 1,5 pmol de cada *primer*; 4 nmol de cada dNTP (*Invitrogen*); 10 % do volume final de tampão (*Invitrogen*); 5% do volume final de DMSO; 0,15 µmol de MgCl₂; e 1,75 U de *Platinum Taq DNA Polimerase* (*Invitrogen*).

Para a amplificação foi utilizada uma sequência de ciclos na qual é feita uma desnaturação inicial a 94° C por 11 minutos, seguida por 30 ciclos de desnaturação a 94° C por 1 minuto, hibridação a 60° C por 1 minuto e extensão a 70° C por 1 minuto; seguindo-se de um ciclo de extensão final a 60° C por 60 minutos.

Após a reação de PCR, 1 µL do produto amplificado diluído (10x) foi misturado a 8,925 µL de uma solução de formamida deionizada *Hi-Di^{TM}* e 0,075 µL do peso molecular *GeneScan^{TM} -500 Rox^{TM} Size Standard* (ambos da *Applied Biosystems*).

Para a realização da eletroforese capilar, os produtos de PCR dos multiplex I e II (e, analogamente, dos multiplex III e IV), foram combinados. Essa estratégia possibilitou a análise simultânea de oito microssatélites por amostra, ao invés de analisar apenas quatro marcadores por amostra, a cada corrida.

Os fragmentos de DNA amplificados foram submetidos à eletroforese capilar no analisador automático de DNA *ABI 3730 DNA analyser* e analisados com o software *Free Peak Scanner^{TM} v1.0* (ambos da *Applied Biosystems*).

3.3.4. Análise computacional dos resultados obtidos a partir dos marcadores moleculares

Os dados provenientes da genotipagem dos marcadores moleculares foram organizados em planilhas e transferidos para arquivos-texto de dados para serem analisados por programas computacionais originais desenvolvidos em dialeto BASIC (*Liberty BASIC v4.04*, © Shoptalk Systems). Esses programas utilizam métodos de estimativas de parâmetros genético-populacionais e testes de hipóteses genéticas clássicos encontrados em diversos livros-texto de genética

de populações, como os de Li, 1955; Crow e Kimura, 1970; Spiess, 1977; Weir, 1999; Otto 2008; Hartl e Clark, 2010.

No Anexo 3 apresentamos a listagem do programa-fonte em BASIC (Liberty Basic) utilizado na obtenção das estimativas dos parâmetros genéticos de interesse. O programa determina, a partir da contagem dos genótipos dos marcadores moleculares dos indivíduos, as frequências alélicas, o valor do coeficiente médio de endocruzamento e os respectivos erros-padrão e variâncias de cada estimativa; na sequência o programa realiza testes de significância por meio do teste do qui-quadrado e determina o valor **P** de probabilidade associado ao valor obtido do teste.

Após todos os cálculos realizados loco a loco, o programa obtém o valor médio do coeficiente **F** para cada isolado quilombola. Para se conseguir essa única estimativa de **F** em relação a todos os locos autossômicos codominantes analisados em cada uma das populações, os diversos valores **F_i** estimados para cada um dos locos são ponderados pelos recíprocos de suas variâncias segundo

$$\mathbf{F} = \sum[\mathbf{F}_i/\mathbf{Var}(\mathbf{F}_i)]/\sum[1/\mathbf{Var}(\mathbf{F}_i)] .$$

Como os valores esperados dos números observados **N₁**, **N₂** e **N₃** respectivamente de indivíduos **AA**, **Aa** e **aa** em uma amostra qualquer de tamanho **N = N₁ + N₂ + N₃** e coeficiente de endocruzamento médio **F ≠ 0** são **N_e(AA) = N(p²+pqF)**
= N[p²+p(1-p)F] = Np[p+(1-p)F], **N_e(Aa) = 2Npq(1-F) = 2Np(1**

$-p)(1-F)$ e $N_e(aa) = N(q^2 + pqF) = N[(1-p)^2 + p(1-p)F] = N(1-p)[(1-p) + pF]$, a função de verossimilhança é dada pela expressão

$$P = K \cdot [p^2 + p(1-p)F]^{N_1} \cdot [2Np(1-p)(1-F)]^{N_2} \cdot [(1-p)^2 + p(1-p)F]^{N_3} ,$$

em que $K = \text{constante} = N!/(N_1!N_2!N_3!)$.

As estimativas de máxima verossimilhança são obtidas do sistema de equações $\{\partial P/\partial p = 0, \partial P/\partial F = 0\}$, ou, mais convenientemente, do sistema $\{\partial L/\partial p = 0, \partial L/\partial F = 0\}$, em que

$$L = \log(P) = N_1 \cdot \log[p^2 + p(1-p)F] + N_2 \cdot \log[2p(1-p)(1-F)] + N_3 \cdot \log [(1-p)^2 + p(1-p)F] .$$

As derivadas parciais de L em função dos argumentos p e F são respectivamente

$$\partial L/\partial p = (N_1 + N_2)/p - (N_2 + N_3)/q + N_1(1-F)/(p+qF) - N_3(1-F)/(q+pF)$$

e

$$\partial L/\partial F = N_1q/(p+qF) - N_2/(1-F) + N_3p/(q+pF) .$$

Não é difícil determinar-se que as soluções do sistema $\{\partial L/\partial p = 0, \partial L/\partial F = 0\}$ são idênticas às estimativas de p e F obtidas intuitivamente através da aplicação de métodos diretos: $p = (2N_1 + N_2)/2N$ e $F = 1 - (N_2/N)/[2p(1-p)]$.

A determinação dos valores exatos das variâncias de F e de p é problemática, uma vez que é praticamente impossível, mesmo para o caso relativamente simples de dois alelos autossômicos, inverter a matriz de informação dos

recíprocos das derivadas segundas puras e mistas para se obter a matriz de variâncias-covariâncias [argumentação rigorosa a respeito é fornecida por Weir (1996) e Fyfe e Bailey (1951)]. Estes últimos autores conseguiram, no entanto, obter expressões aproximadas para o valor das variâncias de p e de F : $\text{var}(p) = (1+F)p(1-p)/2N$ e $\text{var}(F) = (1-F)^2(1-2F)/N + F(1-F)(2-F)/[2Np(1-p)]$, em que $p = (2N_1+N_2)/2N$ e $F = 1 - (N_2/N)/[2p(1-p)]$.

A finalidade da aplicação do método de máxima verossimilhança acima é a obtenção de um valor aproximado da variância da estimativa F , o que pode ser realizado tratando p e $q = 1-p$ como constantes. A variância de F é então obtida de

$$\text{var}(F) = -1/[(d^2L/dF^2)_{p,F}] ,$$

ou seja, o recíproco do valor da derivada segunda de L em função de F , com sinal trocado, calculada no ponto correspondente às estimativas $p=(2N_1+N_2)/2$ e $F=(N_2/N)/[2p(1-p)]$.

A derivada segunda de L em função apenas do argumento F tem a forma

$$d^2L/dF^2 = - N_1[q/(p+qF)]^2 - N_2/(1-F)^2 - N_3[p/(q+pF)]^2 ,$$

de modo que para os valores das estimativas p , $q = 1-p$ e F mostrados mais acima a variância de F toma exatamente o valor

$$\begin{aligned} \text{var}(F) &= [(Npq)^2(1/N_1 + 4/N_2 + 1/N_3)]^{-1} \\ &= N_1 \cdot N_2 \cdot N_3 / [(Npq)^2 \cdot (N_2 \cdot N_3 + 4 \cdot N_1 \cdot N_3 + N_1 \cdot N_2)] . \end{aligned}$$

Como mostraremos posteriormente, o valor assim obtido de $\text{var}(\mathbf{F})$ não difere significativamente do valor aproximado proposto por Fyfe e Bailey (1951). Além disso, utilizando-se diretamente a fórmula $\text{var}(\mathbf{F}) = -(\mathbf{d}^2\mathbf{L}/\mathbf{dF}^2)^{-1}$, é fácil generalizá-la para o caso genérico de qualquer número de alelos segregando num loco autossômico, uma vez que $-(\mathbf{d}^2\mathbf{L}/\mathbf{dF}^2)$ reduz-se a

$$\sum N(\mathbf{a}_i\mathbf{a}_i) \cdot \left\{ (1-p_i) / [p_i + (1-p_i)\mathbf{F}] \right\}^2 + \sum N(\mathbf{a}_i\mathbf{a}_j) / (1-\mathbf{F})^2$$

em que $\mathbf{a}_i\mathbf{a}_i$ indica o homozigoto quanto ao alelo $\mathbf{a}_i$ e $\mathbf{a}_i\mathbf{a}_j$ o heterozigoto quanto aos alelos $\mathbf{a}_i$ e $\mathbf{a}_j$.

4. RESULTADOS

4.1. Estimativas do coeficiente médio de endocruzamento por meio de análise de genealogias

Foram obtidos os coeficientes médios de endocruzamento das genealogias de todas as comunidades. As genealogias correspondentes, embora contivessem um número enorme de informações devido à grande quantidade de indivíduos nelas representados, possuíam diversas lacunas provocadas pela falta de informações confiáveis.

Nenhuma das comunidades analisadas agrupava todos os indivíduos da população em apenas uma genealogia; para cada comunidade foi analisado um grupo de heredogramas (Tabela 4.1) com complexidade e tamanho diversos.

Tabela 4.1: Número de heredogramas analisados para as comunidades de Abobral margem esquerda e direita (AB/DR), André Lopes (AN), Galvão e São Pedro (GA/SP), Ivaporanduva (IV), Maria Rosa e Pilões (MR/PS), Nhunguara (NH), Pedro Cubas (PC), Sapatu (TU), Poça (PA) e Reginaldo (RE).

	AB/DR	AN	GA/SP	IV	MR/PS	NH	PC	TU	PA	RE
Nº de heredogramas	33	14	3	5	6	11	26	9	4	6

Em todos os heredogramas analisados, o número de indivíduos foi computado levando-se em conta apenas aqueles vivos nascidos ou inseridos nas comunidades, excluindo-se, portanto, indivíduos falecidos e ancestrais de indivíduos imigrantes. Devido a esse critério de exclusão, alguns

heredogramas das comunidades reduziram-se a apenas um indivíduo.

Grande parte desses heredogramas possui indivíduos contidos nas genealogias de outras comunidades. Esses indivíduos foram identificados por meio de análise comparativa dos heredogramas e incluídos, em geral, apenas no núcleo mais complexo no qual o indivíduo estava inserido.

Apresentamos na Figura 4.1 a representação de todos os diferentes tipos de casamentos consanguíneos encontrados nas populações quilombolas.

Figura 4.1: Heredogramas simplificados de casamentos consanguíneos encontrados nas genealogias das populações quilombolas (**a**: tio(a)-sobrinho(a); **b**: primos em primeiro grau; **c**: primos em segundo grau; **d**: primos em terceiro grau; **e**: primos em quarto grau; **f**: primos em quinto grau; **g**: meio-primos em primeiro grau).

A Tabela 4.2 reúne os dados referentes aos heredogramas da Figura 4.1, definindo, para cada uma das representações o valor do coeficiente de endocruzamento

(F_i) do indivíduo endocruzado e o número total de endocruzados provenientes de cada tipo de casamento em cada comunidade.

Tabela 4.2: Coeficiente de endocruzamento individual e número de indivíduos endocruzados em cada população, de acordo com os heredogramas representados na Figura 4.1.

	Heredograma						
	a.	b.	c.	d.	e.	f.	g.
F_i	1/8	1/16	1/32	1/64	1/128	1/256	1/32
AB/DR	-	36	5	-	-	-	4
AN	6	7	6	-	-	3	-
GA/SP	-	-	-	8	-	-	-
IV	-	3	-	-	-	-	-
MR/PS	-	4	-	5	-	-	-
NH	-	2	18	2	4	-	-
PC	-	-	-	-	-	-	-
TU	-	-	-	-	-	-	-
Total	6	52	29	15	4	3	4
PA	-	-	-	-	-	-	-
RE	-	1	-	-	-	-	-

Partindo-se dos dados apresentados acima, obtivemos os valores do coeficiente médio de endocruzamento para cada população, (Tabela 4.3).

Tabela 4.3: Número de indivíduos, coeficiente médio de endocruzamento e porcentagem de indivíduos provenientes de casamentos consanguíneos de cada população e em relação à população total.

Comunidade	Nº de indivíduos	F	Porcentagem endocruzada
Abobral	773	0,00327	5,82 %
André Lopes	567	0,00089	1,94 %
Galvão e São Pedro	446	0,00028	1,79 %
Ivaporanduva	575	0,00033	0,52 %
Maria Rosa e Pilões	324	0,00101	2,78 %
Nhunguara	434	0,00178	5,99 %
Pedro Cubas	368	0	-
Sapatu	472	0	-
Total	3959	0,00134	2,85 %
Poça	197	0	-
Reginaldo	173	0,00036	0,58 %

4.1.1. Teste da metodologia de análise

Previamente à aplicação da metodologia de análise de genealogias aqui proposta, realizamos dois testes. O primeiro dos testes foi realizado no isolado indígena Arara do Iriri (população formada por indivíduos que se reproduzem exclusivamente por casamentos endogâmicos) com a finalidade de selecionar, dentre os métodos descritos no item 3.1.2., o mais eficiente a ser aplicado em nossas análises.

Posteriormente, realizamos um segundo teste com o intuito de compararmos os valores populacionais de **F** obtidos por nossa metodologia aos de outros trabalhos da literatura. Aplicamos o método proposto ao heredograma

representativo da população de Valongo, estudada por Souza e Culpi (1992).

4.1.1.1. Isolado indígena Arara do Iriri

Estimamos o coeficiente médio de endocruzamento da tribo indígena Arara do Iriri pela média ponderada dos valores individuais (F_i) obtidos utilizando-se tanto o método dos passos genéticos (descrito em diversos livros básicos como o de Beiguelman, 2005), quanto o método que possibilita a listagem dos possíveis caminhos percorridos por cada alelo (descrito em livros-texto como o de Crow e Kimura, 1970). Para a aplicação de ambos os métodos, desmembramos a genealogia do isolado, apresentada no Anexo 1 (gentilmente cedida pelos Profs. Drs. João Farias Guerreiro, Sidney Emanuel Batista dos Santos, Andrea Ribeiro dos Santos e Eduardo José Melo dos Santos, todos do Dpto. de Genética da Universidade Federal do Pará), em heredogramas menores (Figura 4.2), contendo os indivíduos de interesse.

Figura 4.2: Heredogramas resultantes do desmembramento da genealogia complexa apresentada no Anexo 1.

Os símbolos destacados nos heredogramas acima representam, respectivamente, 7, 9, 19, 1 e 2 indivíduos, cujos valores individuais F_i obtidos são 1/4, 3/8, 3/8, 15/32 e 7/16.

A ponderação dos valores de F_i resulta na estimativa do coeficiente médio de endocruzamento dessa população (F_{AI}), que toma o valor

$$F_{AI} = (2 \times 0 + 7 \times 1/4 + 9 \times 3/8 + 19 \times 3/8 + 1 \times 15/32 + 2 \times 7/16) / 40$$

$$= 87/256 = 0,3398.$$

É interessante notar que, dentre os 40 indivíduos da população analisada, apenas os dois fundadores não são endocruzados ($F_i = 0$) e todos os demais casamentos ocorridos são exclusivamente endogâmicos. Isso explica o valor insolitamente alto do coeficiente médio de endocruzamento desse isolado indígena.

4.1.1.2. Isolado quilombola de Valongo (SC)

Com a finalidade de verificarmos a confiabilidade da estimativa de F obtida pelo método proposto em nosso trabalho, analisamos a genealogia da comunidade quilombola de Valongo (localizada no estado de Santa Catarina) e comparamos a nossa estimativa com a obtida por Souza e Culpi (1992). Nesse trabalho, o valor médio de F foi obtido por formulação originalmente desenvolvida por Wright (1931), na qual $F = 2(N_r - 1) / [2N_e - (2N_e - 1)(1 - m_e)^2]$, onde N_r é o

tamanho da população que se cruza, N_e é o tamanho efetivo da população e m_e número efetivo de migrantes.

Os valores de N_r e m_e foram obtidos por contagem, enquanto N_e foi obtido por meio da expressão $N_e = 2(N_r - 1) / (k - 1 + \sigma_k^2/k)$, na qual k é o número médio de filhos por casal na população.

A genealogia completa da comunidade de Valongo (Anexo 2) contém 162 indivíduos, dos quais 82 estão vivos, residentes na comunidade. Para a obtenção do valor médio de F , desmembramos a genealogia em heredogramas contendo apenas os indivíduos de interesse (Figura 4.3).

Figura 4.3: Heredogramas resultantes do desmembramento da genealogia de Valongo (Anexo 2).

Os símbolos destacados nos heredogramas acima representam, respectivamente, 33, 2, 7, 4, 3, 6 e 1

indivíduos, cujos valores de F_i obtidos a partir de nossa análise resultaram nas estimativas 1/16, 5/64, 1/64, 3/32, 9/128, 1/8 e 11/128.

A partir desses resultados temos

$$F_v = (33.1/16 + 2.5/64 + 7.1/64 + 4.3/32 + 3.9/128 + 6.1/8 + 1.11/128 + 26.0)/82 = 15/328 \sim 0,04573 .$$

O valor obtido pelo método que aplicamos mostrou-se próximo ao valor 0,04774 obtido por Souza e Culpi (1992). A diferença entre esses valores se deve ao fato de que enquanto a metodologia proposta em nosso trabalho pondera os valores estimados exatos de todos os indivíduos da população, a metodologia aplicada por Souza e Culpi (1992) leva em conta os valores médios do número de crianças nascidas por casal.

Os resultados acima indicam que o nosso método fornece valores de F virtualmente semelhantes aos obtidos por meio de outros métodos já consagrados. Por não utilizar valores médios, nosso método deve ser necessariamente mais preciso.

4.2. Correção do método de estimativa do valor do coeficiente médio de endocruzamento por meio da análise da frequência de pessoas e casais com o mesmo sobrenome

O método original descrito por Crow e Mange (1965) calcula o coeficiente de endocruzamento médio (F) da população na hipótese de que as frequências dos pares

irmão-irmão, irmão-irmã, irmã-irmão e irmã-irmã (Figura 4.4), genitores de um casal qualquer de primos em primeiro grau, sejam iguais (25% para cada par), como se espera em modelos baseados em casamentos aleatórios.

Figura 4.4: Em destaque, as combinações possíveis de genitores de um casal qualquer de primos em primeiro grau.

Em quase todas as populações, entretanto, as frequências de cada um dos pares de irmãos (genitores dos primos em primeiro grau) diferem entre si (Morton, 1955; Freire-Maia e Freire-Maia, 1961).

Realizamos uma correção no método original de Crow e Mange (1965), levando em conta essa diferença. Seja k a frequência na população de pares irmão-irmão genitores dos casais consanguíneos. Nessas condições o valor do coeficiente F (F_{IT} na notação de Wright) toma o valor $F = P.k$ onde P é a frequência de casais com mesmo sobrenome na população.

Fazendo-se F_r (F_{ST} na notação de Wright) igual a $\sum p_i q_i / 4$, obtemos a expressão para F_n (F_{IS} na notação de

Wright) por meio da relação $F = F_r + F_n - F_r F_n$ ou, explicitamente, $F_n = (F - F_r) / (1 - F_r)$, chegando ao valor

$$F_n = (Pk - \sum p_i q_i / 4) / (1 - \sum p_i q_i / 4)$$

$$= (4Pk - \sum p_i q_i) / (4 - \sum p_i q_i) .$$

4.3. Análise dos marcadores moleculares

4.3.1. Obtenção dos coeficientes médios de endocruzamento

Foram obtidos os coeficientes médios de endocruzamento dos isolados em relação aos 31 locos genômicos propostos no estudo (inserção de *Alu*, SNPs associados à obesidade e hipertensão e locos de microssatélites).

A Tabela 4.4 apresenta as estimativas do coeficiente de endocruzamento de 30 marcadores, destacando os valores significativamente diferentes de zero. A tabela contém, também, os valores médios do coeficiente F por população obtidos pela média ponderada dos valores médios de F de cada loco pelo recíproco de suas variâncias. Não apresentamos o valor médio de F do loco *AGT* para a comunidade de André Lopes, uma vez que esse loco revelou-se monomórfico.

Além disso, calculamos os valores médios de F por loco considerando as comunidades com mesma origem ancestral como uma única população.

Tabela 4.4: Valores do coeficiente médio de endocruzamento **F** de cada loco por população. *: valores de **P** < 0,05 sob hipótese de **F** = 0; **: valores de **P** < 0,01 sob hipótese de **F** = 0.

	AB	AN	IV	NH	PC	PS/MR	SP/GA	TU	Total	PA
ACE (rs1799752)	0.2028*	-0.0012	0.0753	0.0297	0.0304	-0.2054	0.0833	0.1151	0.0971*	-0.0286
NOS3 (rs1799983)	0.4987**	-0.0808	0.1069	-0.0806	-0.2222	0.1346	0.0330	0.0699	0.0462	-
GNB3 (rs5443)	0.0351	-0.0528	0.0088	0.1152	-0.1578	-0.0263	-0.0138	0.1619	0.0428	-
GNB3 (rs5441)	0.1069	0.0896	0.0159	0.0251	-0.0526	0.4631**	0.2167*	0.2235	0.1655**	0.2381
AGT (rs669)	0.0333	-	-0.0053	0.1765	-0.0097	-0.1657	-0.0848	-0.1151	0.0330	-
ADD2 (rs3755351)	0.1104	0.0103	0.0028	-0.0005	0.0308	0.0374	0.0498	0.0752	0.0587	-
GRK4 (rs1801058)	-0.0191	0.0026	-0.0979	-0.0715	-0.2072	0.1431	-0.0232	0.1692	-0.0145	-0.1758
PLIN1 (rs2289487)	0.1357	-0.0171	0.1364	0.0447	-0.0145	0.2646*	0.0508	0.4451**	0.1389**	-
INSIG2 (rs7566605)	-0.0260	0.0137	0.1808*	-0.1271	0.0330	-0.0208	0.2118*	-0.2061	0.0265	-
LEP (rs2167270)	0.0542	-0.0142	-0.0933	0.0538	0.0994	0.0578	-0.0556	-0.0374	0.0277	-
LEPR (rs1137101)	-0.0682	-0.0909	-0.0724	-0.0370	-0.1005	0.2998*	0.0260	-0.1441	0.0007	-
ADRB2 (rs1042713)	-0.0114	0.1363	-0.1665	-0.2858**	-0.0286	-0.1255	-0.0141	-0.0528	-0.0440	-
PPARG (rs1801282)	-0.0941	0.0440	-0.0943	-0.1585	-0.0636	0.3021*	-0.0447	-0.0323	0.0597	-
RETN (rs1862513)	0.0451	-0.0145	-0.1986*	-0.0705	-0.0407	-0.0577	0.0742	0.1958	-0.0043	-
D5S816	-0.1087	-0.0553	-0.1305	-0.0268	-0.0672	-0.0459	-0.0636	0.0292	-0.0500	-
D1S551	-0.0880	-0.0756	0.1479	0.0495	-0.0069	0.1904	-0.1707	0.0728	0.0471	-
D7S3061	0.0457	-0.1175	0.0230	0.0517	0.0033	0.0133	0.0850	0.0651	0.0443	-
D4S3248	-0.0093	0.0366	0.0119	-0.0752	0.0467	0.0606	0.1253	-0.0107	0.0365	-
D16S539	0.0882	-0.0328	-0.0968	-0.0112	-0.0314	-0.0821	-0.1023	-0.0541	-0.0294	-
D9S922	0.0090	0.0599	0.0248	-0.0675	0.0346	0.0751	-0.0519	-0.0068	0.0294	-
D10S1426	0.0645	-0.1859	0.1430	-0.1298	-0.1933	0.0184	-0.1775	-0.0747	-0.0169	-
D7S821	-0.0802	0.0151	-0.0276	0.0113	0.0824	-0.0622	-0.1766	-0.0963	-0.0208	-
D13S317	-0.0433	0.0601	0.0196	-0.0529	-0.0436	0.0662	0.0113	-0.0034	0.0238	-
D8S2324	0.0024	0.1955	0.0164	-0.0846	0.1319	0.0411	0.0397	-0.0254	0.0476	-
D19S559	-0.0237	0.3642	-0.0248	-0.1338	-0.0423	0.0079	-0.0983	0.0144	0.0075	-
D6S1040	-0.1615	-0.1860	-0.1544	-0.0354	-0.0638	-0.1100	-0.0073	0.0449	-0.0680	-
D20S482	0.1325	0.1280	0.1155	0.0879	0.1566	0.0999	0.0533	-0.1547	0.1038	-
D21S1437	0.3552	0.1336	0.1372	0.1619	-0.0304	0.2953	0.1582	0.3082	0.2093**	-
D9S301	0.1306	0.0679	-0.0076	-0.0308	-0.0230	-0.0535	-0.1790	0.0686	0.0182	-
D18S535	-0.0174	-0.1165	-0.0742	0.0662	-0.0293	0.0946	-0.0467	0.1611	0.0141	-
Fm	0.0492	-0.0007	-0.0139	-0.0546	-0.0449	0.0490	-0.0242	0.0183	0.0341	0.0116

Os dados referentes à comunidade de Poça mostrados na Tabela 4.4 são os únicos gerados até o momento para essa comunidade, já que ela foi incluída apenas mais recentemente aos estudos do grupo da Dra. Mingroni-Netto.

Para efeito comparativo, estimamos os valores médios de F (F_m) por população considerando dois grupos diferentes de marcadores moleculares (microsatélites e demais locos), já que os marcadores de microsatélites são multialélicos enquanto os demais são marcadores polimórficos bialélicos (Tabela 4.5).

Tabela 4.5: Estimativas de F_m em cada uma das populações para os conjuntos de locos de microsatélites e de todos os demais locos.

Comunidade	Microsatélites	Demais locos
Abobral	-0,0087	0,0812
André Lopes	-0,0298	-0,0144
Ivaporanduva	-0,0375	-0,0060
Nhunguara	-0,0571	-0,0572
Pedro Cubas	-0,0351	-0,0252
Pilões/Maria Rosa	-0,0254	0,0835
São Pedro/Galvão	-0,1244	0,0127
Sapatu	-0,0186	0,0064
Total	0,0047	0,0430

Em nossas análises excluimos os dados relativos ao loco *HBB* das estimativas apresentadas nas Tabelas 4.4 e 4.5, uma vez que o loco sofre seleção natural. Apresentamos separadamente (Tabela 4.6) os parâmetros genético-

populacionais obtidos a partir da análise dos dados referentes a esse loco.

Tabela 4.6: Dados de frequências genotípicas e alélicas, coeficiente médio de endocruzamento (com respectiva variância) em cada uma das populações, referentes ao loco *HBB*.

	P(AA)	P(AS)	P(SS)	P(AC)	P(A)	P(S)	P(C)	F	var(F)
AB	0,8558	0,1298	0,0144	-	0,9207	0,0793	-	0,1113	0,0002
AN	0,8947	0,1053	-	-	0,9474	0,0526	-	-0,0556	0,0002
IV	0,9306	0,0486	-	0,0208	0,9653	0,0243	0,0104	-0,0282	0,1038
NH	0,8593	0,1406	-	-	0,9297	0,0703	-	-0,0756	0,0003
PC	0,9790	0,0210	-	-	0,9895	0,0105	-	-0,0106	0,0000
PS/MR	0,9868	-	-	0,0132	0,9934	-	0,0066	-0,0066	0,0000
SP/GA	0,9762	0,0238	-	-	0,9881	0,0119	-	-0,0120	0,0000
TU	0,9216	0,0784	-	-	0,9608	0,0392	-	-0,0408	0,0002
Total	0,9180	0,0754	0,0028	0,0038	0,9575	0,0405	0,0019	0,0272	0,0014

4.3.2. Obtenção da variância de *F*

A variância de **F** foi obtida neste trabalho por meio de um método original, que permite conseguir uma estimativa aproximada mesmo para locos multialélicos (como é o caso dos marcadores de microssatélites).

Na Figura 4.5 abaixo, apresentamos um gráfico comparativo, correlacionando as estimativas de variância de **F** obtidas por meio da metodologia proposta neste trabalho (eixo X) e de uma outra (eixo Y) proposta por Fyfe e Bailey (1951), a qual obtém o valor aproximado da variância de **F** em locos bialélicos por meio da expressão $\text{var}(\mathbf{F}) = (1 - \mathbf{F})^2 \cdot (1 - 2\mathbf{F}) / N + \mathbf{F}(1 - \mathbf{F})(2 - \mathbf{F}) / [2Np(1 - p)]$. Para que fosse possível essa comparação, consideramos os marcadores de

microssatélites como sendo, também, dialélicos com o alelo **A** correspondendo ao alelo mais frequente na população e o alelo **a**, ao total dos alelos restantes.

Figura 4.5: Gráfico correlacionando as estimativas de variância pelo método proposto neste trabalho (eixo X) e pelo proposto na literatura (eixo Y).

4.3.3. Análise de subestruturação populacional

Estimamos os valores dos índices de fixação F_{IT} , F_{ST} e F_{IS} para cada um dos 30 locos considerados e, com a finalidade de testar a significância dessas estimativas, obtivemos os mesmos índices de fixação por meio de simulações. No Anexo 4 apresentamos a listagem do programa-fonte em BASIC (Liberty Basic), desenvolvido em colaboração com o orientador, utilizado na obtenção das

Tabela 4.7: Estimativas de F_{IT} , F_{ST} e F_{IS} obtidos pela aplicação direta das fórmulas e dos valores de suas medianas e respectivos intervalos de confiança a 95% obtidos por simulação.

	F_{IT}	F_{IT} simulado	F_{ST}	F_{ST} simulado	F_{IS}	F_{IS} simulado
ACE (rs1799752)	0,0956	0,095 (0,022, 0,164)	0,0442	0,048 (0,029, 0,075)	0,0538	0,048 (-0,029, 0,119)
NOS3 (rs1799983)	0,0459	0,042 (-0,039, 0,138)	0,0211	0,026 (0,012, 0,046)	0,0253	0,015 (-0,070, 0,112)
GNB3 (rs5443)	0,0415	0,040 (-0,038, 0,119)	0,0360	0,042 (0,022, 0,064)	0,0057	-0,002 (-0,078, 0,076)
GNB3 (rs5441)	0,1655	0,163 (0,079, 0,246)	0,0366	0,042 (0,023, 0,070)	0,1339	0,122 (0,040, 0,207)
AGT (rs669)	-0,0068	-0,007 (-0,100, 0,083)	0,0171	0,022 (0,009, 0,047)	-0,0243	-0,033 (-0,123, 0,061)
ADD2 (rs3755351)	0,0439	0,040 (-0,044, 0,120)	0,0183	0,024 (0,009, 0,044)	0,0260	0,015 (-0,067, 0,098)
GRK4 (rs1801058)	-0,0116	-0,013 (-0,090, 0,068)	0,0134	0,018 (0,008, 0,033)	-0,0253	-0,032 (-0,111, 0,048)
PLIN1 (rs2289487)	0,1404	0,136 (0,069, 0,207)	0,0356	0,040 (0,022, 0,062)	0,1087	0,099 (0,031, 0,173)
INSIG2 (rs7566605)	0,0263	0,021 (-0,045, 0,099)	0,0192	0,023 (0,011, 0,040)	0,0072	-0,004 (-0,070, 0,079)
LEP (rs2167270)	0,0354	0,032 (-0,033, 0,104)	0,0258	0,030 (0,016, 0,047)	0,0099	0,003 (-0,061, 0,074)
LEPR (rs1137101)	0,0020	0,002 (-0,062, 0,068)	0,0325	0,036 (0,020, 0,057)	-0,0315	-0,036 (-0,098, 0,028)
ADRB2 (rs1042713)	-0,0526	-0,054 (-0,128, 0,013)	0,0208	0,025 (0,012, 0,041)	-0,0749	-0,082 (-0,157, -0,012)
PPARG (rs1801282)	0,0596	0,046 (-0,040, 0,123)	0,0611	0,063 (0,037, 0,094)	-0,0017	-0,015 (-0,084, 0,060)
RETN (rs1862513)	-0,0030	-0,004 (-0,076, 0,065)	0,0152	0,019 (0,009, 0,032)	-0,0184	-0,024 (-0,097, 0,046)
D5S816	-0,1217	-0,127 (-0,219, -0,029)	0,0014	0,011 (0,003, 0,028)	-0,1232	-0,141 (-0,231, -0,041)
D1S551	0,0973	0,095 (-0,014, 0,207)	0,0242	0,036 (0,014, 0,068)	0,0749	0,062 (-0,049, 0,174)
D7S3061	0,0921	0,088 (-0,030, 0,209)	0,0068	0,018 (0,005, 0,045)	0,0859	0,070 (-0,046, 0,190)
D4S3248	0,0667	0,063 (-0,056, 0,186)	0,0118	0,023 (0,007, 0,049)	0,0555	0,039 (-0,081, 0,160)
D16S539	-0,0154	-0,021 (-0,122, 0,098)	0,0106	0,022 (0,006, 0,047)	-0,0264	-0,044 (-0,149, 0,073)
D9S922	-0,0624	-0,071 (-0,182, 0,045)	0,0179	0,028 (0,010, 0,057)	-0,0817	-0,100 (-0,215, 0,013)
D10S1426	-0,0046	-0,011 (-0,128, 0,111)	0,0473	0,059 (0,028, 0,099)	-0,0544	-0,075 (-0,189, 0,042)
D7S821	-0,0873	-0,092 (-0,195, 0,023)	0,0107	0,020 (0,006, 0,046)	-0,0990	-0,115 (-0,220, -0,009)
D13S317	0,0170	0,012 (-0,089, 0,131)	0,0327	0,042 (0,021, 0,078)	-0,0163	-0,032 (-0,140, 0,089)
D8S2324	0,1515	0,147 (0,028, 0,272)	0,0195	0,031 (0,011, 0,067)	0,1347	0,119 (-0,001, 0,249)
D19S559	0,0254	0,020 (-0,083, 0,131)	0,0167	0,028 (0,009, 0,057)	0,0088	-0,009 (-0,117, 0,100)
D6S1040	-0,0818	-0,083 (-0,201, 0,023)	0,0061	0,015 (0,004, 0,038)	-0,0884	-0,103 (-0,217, 0,006)
D20S482	0,1190	0,115 (0,000, 0,236)	0,0193	0,031 (0,010, 0,069)	0,1017	0,086 (-0,030, 0,207)
D21S1437	0,2849	0,279 (0,152, 0,418)	0,0220	0,040 (0,014, 0,090)	0,2688	0,247 (0,113, 0,387)
D9S301	-0,0235	-0,027 (-0,139, 0,080)	0,0352	0,046 (0,021, 0,081)	-0,0609	-0,078 (-0,188, 0,035)
D18S535	-0,0209	-0,026 (-0,140, 0,092)	0,0066	0,017 (0,005, 0,038)	-0,0276	-0,045 (-0,158, 0,072)

estimativas dos índices de fixação F_{IT} , F_{ST} e F_{IS} para a população quilombola. A tabela 4.7 apresenta os valores calculados, juntamente com as medianas desses índices e seus respectivos intervalos de confiança de 95%.

As estimativas de F_{IT} , F_{IS} e F_{ST} foram obtidas por meio de combinações de 1000 reamostragens de cada uma das populações, em relação a cada loco, por meio de simulações em computador utilizando-se o método *bootstrap* (com reposição). Em relação ao loco *PPARG* esses valores foram obtidos apenas quando o número de simulações foi reduzido a 100.

Em negrito estão apresentados os poucos valores dos índices F_{IT} e F_{IS} cujos intervalos de confiança a 95% não contêm o zero e que por isso são considerados diferentes de zero a um nível de rejeição de hipótese nula $P=0,05$ ou menos. Quanto aos valores de F_{ST} , a maioria tem valores modestos, com o limite inferior do intervalo de confiança a 95% muito próximo de zero. Devem ser significativos apenas os poucos valores altos (geralmente da ordem de 4% ou mais) correspondentes a valores de F_{IT} significativamente diferentes de zero. Quando o índice F_{IS} não difere de zero, F_{IT} deve ser igual a F_{ST} . Portanto, nessas condições, sempre que F_{IT} não diferir de zero, F_{ST} também não deverá ser significativamente diferente de zero. Os casos de F_{ST} não incluídos nessa situação, ou seja, aqueles que devem ser diferentes de zero, estão também representados em negrito.

Os quatro casos em que o valor de F_{ST} está representado em **negrito itálico** também devem ser significativamente diferentes de zero, porém a contribuição principal para o F_{IT} é dada pelo F_{IS} (endogamia dentro da população) e não pelo F_{ST} (atribuível à deriva).

4.4. Evolução de F em sistemas regulares de endocruzamento

No Anexo 5 apresentamos uma seção de cunho teórico sobre a evolução da taxa de homozigose por origem comum em sistemas regulares de cruzamentos (casamentos exclusivos entre irmãos, meio-irmãos, primos duplos em primeiro grau e casamentos circulares entre meio-irmãos), etapa esta que desenvolvemos paralelamente à execução do presente trabalho.

5. DISCUSSÃO

5.1. Estimativa do coeficiente médio de endocruzamento por meio da análise de genealogias

Observando as estimativas de **F** obtidas pela análise de genealogias (Tabela 4.3), verificamos que o valor médio do parâmetro é $F = 0,00134$, o qual é ligeiramente mais alto do que o de uma população formada exclusivamente por filhos de primos em sétimo grau ($F_i = 1/1024 \sim 0,00098$). O valor por nós obtido é cerca de 1,5 vezes mais elevado do que a estimativa apresentada para a população total brasileira ($F = 0,00088$) e cerca de duas vezes maior que a obtida para o estado de São Paulo (Freire-Maia, 1957; 1990).

A população quilombola que apresentou o maior valor de **F** foi a comunidade de Abobral ($F = 0,00327$), estimativa comparável ao que seria esperado numa população em que todos os habitantes são filhos de casais de primos em quinto grau ($F_i = 1/256 \sim 0,00391$).

Os valores estimados para as populações quilombolas do Vale do Ribeira são muito menores do que o calculado para o isolado quilombola localizado no sertão de Valongo (SC) descrito por Souza e Culpi (1992): o valor do coeficiente **F** é 0,04774, que é aproximadamente 34 vezes maior do que o valor médio obtido para as comunidades do Vale do Ribeira. Esse valor muito alto está em harmonia com o fato de que a

população foi gerada quase em sua totalidade por apenas quatro casais fundadores.

A Tabela 5.1 apresenta estimativas de **F** obtidas a partir do estudo de diversos isolados e semi-isolados. O isolamento dessas populações não ocorre simplesmente pelo fato de se apresentarem geograficamente isoladas, como é o caso, por exemplo, da população da Índia, na qual os casamentos estão restritos às castas em que estão inseridos (Bittles, 2002).

Tabela 5.1: Estimativas do coeficiente médio de endocruzamento e da frequência de casamentos consanguíneos (P cons.) em diversos isolados e semi-isolados populacionais encontrados na literatura.

População	F	P cons.	Pesquisador(es)
Judeus de Curitiba	0,0013	4,0%	Freire-Maia e Krieger (1963)
Amish do condado de Adams	0,0195	66,5%	Jackson <i>et al.</i> (1968)
Törbel (Suiça)	0,0058	-	Ellis e Starmer (1978)
Quilombo de Valongo	0,0477	85,0%	Souza e Culpi (1992)
Old Order Amish	0,0166	-	Dorsten <i>et al.</i> (1999)
Hutterites	0,0340	-	Abney <i>et al.</i> (2000)
Índia	0,0075	11,9%	Bittles (2002)
Sul da Índia	0,0212	31,0%	Bittles (2002)
Amman (Jordânia)	0,0142	28,4%	Hamamy <i>et al.</i> (2005)
Tribo Arara do Iriri	0,3398	91,7%	Presente estudo
Quilombos do Vale do Ribeira	0,0013	2,8%	Presente estudo
Quilombo Abobral	0,0033	5,8%	Presente estudo

Dos valores mostrados na tabela acima, destaca-se o da tribo indígena de Arara do Iriri (v. item 4.1.1.1.); devido à presença de casamentos exclusivamente endogâmicos, ao

número extremamente reduzido de indivíduos fundadores e à ausência total de migração.

Os quilombos do Vale do Ribeira apresentaram um valor médio de F muito próximo ao da população judaica descrita no Paraná por Freire-Maia e Krieger (1963). Constata-se um número menor de indivíduos endocruzados nas populações quilombolas, o que parece contraditório dada a forma com a qual essas populações se formaram (isolamento geográfico aliado a um pequeno número de fundadores).

A obtenção das estimativas de F nos isolados quilombolas foi extremamente dificultada pela falta de informações genealógicas mais completas, com diversas lacunas presentes entre núcleos notoriamente aparentados. Devido a isso, os valores do coeficiente F devem constituir na realidade subestimativas dos valores reais. Uma evidência disso é dada pelas populações de Pedro Cubas, Sapatu e Poça, cujos valores de F se mostraram nulos. Foram exatamente essas as comunidades das quais conseguimos o menor número de informações genealógicas confiáveis.

Os casamentos consanguíneos mais frequentemente observados foram os entre primos em primeiro grau (aproximadamente 46%), seguindo-se os de primos em segundo grau (aproximadamente 16%), como era de se esperar.

A análise das genealogias dos quilombos do Vale do Ribeira indica a presença de um padrão aparentemente contraditório de informações, posto que essas genealogias

apresentam, em média, um coeficiente de endocruzamento relativamente alto. Os índices médios de endocruzamento constituem, no entanto, estimativas baixas quando comparadas à maioria dos isolados estudados na literatura.

Parte dos problemas discutidos acima se deve, provavelmente, às dificuldades óbvias na construção de estruturas genealógicas confiáveis a partir de informações obtidas de agregados populacionais com alto grau de analfabetismo e desaculturação, como é o caso dos isolados quilombolas, originados por escravos fugidos, abandonados ou alforriados que se mantiveram isolados do restante da população por várias gerações.

As genealogias, além disso, foram levantadas em ocasiões distintas por pesquisadores diferentes. Isso, aliado à falta de memória cultural das populações e à ausência de uma transmissão coerente dos sobrenomes das famílias, resultou, por exemplo, na inclusão de vários indivíduos em mais de uma genealogia ou mesmo população.

Tentamos corrigir essas incongruências durante a análise das genealogias das comunidades, porém sem muito êxito.

5.2. Obtenção do coeficiente de endocruzamento por meio da frequência de casamentos de pessoas com o mesmo sobrenome

A metodologia de obtenção do coeficiente **F** por meio da análise de sobrenomes nas populações quilombolas foi proposta como uma maneira alternativa de obter a estimativa média de **F** quando as informações da genealogia são incompletas, como já apontado por Crow e Mange (1965). Esses autores apontaram também que inconsistências no padrão de transmissão dos sobrenomes é um dos fatores que podem dificultar a interpretação das estimativas obtidas.

Na tentativa de aplicarmos essa metodologia, procuramos verificar a existência de padrões na transmissão dos sobrenomes nas comunidades quilombolas. A partir das genealogias e das fichas de registro, notamos que a transmissão de sobrenomes mostrou-se totalmente aleatória, não respeitando nenhuma regra fixa de transmissão, sendo os sobrenomes dos indivíduos provenientes dos genitores masculino ou feminino, ou até mesmo, gerados ao acaso. Com isso, em expedições de campo realizadas às comunidades buscamos entrevistar moradores da região, com a finalidade de verificar possíveis padrões de transmissão dos sobrenomes.

Questionamos 30 indivíduos de diferentes comunidades, dos quais 28 não souberam explicar o motivo do padrão de transmissão não ser convencional. O primeiro dos indivíduos

a apresentar justificativa relatou que era comum a alteração proposital de sobrenomes já na vida adulta, com a intenção de contornar o impedimento imposto pela Igreja, que não autorizava casamentos entre pessoas aparentadas até primos em primeiro grau. Há cerca de 50 anos os cartórios não possuíam registros confiáveis de todos os indivíduos e, deste modo, não eram capazes de impedir essas alterações. A outra justificativa apresentada possui fundamento histórico e foi descrita por um dos representantes da comunidade de Ivaporanduva, o qual relata que na época da abolição da escravidão houve a alteração intencional da transmissão de determinados sobrenomes para evitar estigmatização. Por exemplo, os portadores do sobrenome "Meira", o mais comum na região entre os negros da época, tinham dificuldade em encontrar parceiros, uma vez que sofriam grande preconceito imposto pela comunidade branca. Procurando evitar o preconceito, muitos dos que portavam esse sobrenome deixaram de transmiti-lo à prole propositalmente.

Não foi possível, por conseguinte, a aplicação da metodologia proposta nas populações quilombolas do Vale do Ribeira. Pretendemos aplicar o método oportunamente em isolados populacionais do nordeste brasileiro com alto grau de endogamia, em colaboração com o grupo da Dra. Silvana Santos da Universidade Federal da Paraíba.

5.3. Obtenção do coeficiente de endocruzamento por meio da análise dos marcadores moleculares

Apresentamos na Tabela 4.4 (v. item 4.3.1.) os valores médios de **F** calculados em cada um dos 239 locos de todas as populações. Apenas 12 são significativamente diferentes de zero, considerando valores de $P < 0,05$ obtidos por meio da aplicação de testes estatísticos convencionais. Portanto, do total de locos genotipados, praticamente 5% desviam das proporções de Hardy-Weinberg, de acordo com o que se espera ao acaso.

Se considerarmos valores de $P < 0,01$, seria esperado que, também de forma aleatória, em cerca de 1% dos casos os valores de **F** se mostrem estatisticamente diferentes de zero. Dos resultados obtidos, observamos que em apenas quatro casos (de um total de 239) o valor de **F** foi considerado diferente de zero àquele nível de significância.

Esses resultados sugerem que os isolados quilombolas não apresentam individualmente coeficientes de endocruzamento significativamente diferentes de zero.

Entre os valores de **F** obtidos para os 30 locos no conjunto das comunidades, quatro (13,33%) mostraram-se significativamente diferentes de zero ao nível de 5% e três (10%), ao nível de 1%.

A taxa de homozigose não se apresenta distribuída de forma uniforme no genoma dos indivíduos, uma vez que

algumas regiões devem apresentar desvios das proporções de Hardy-Weinberg, devido à presença de locos contíguos em desequilíbrio de ligação (Weir, 2004). No trabalho de Purfield *et al.* (2012), os autores descrevem uma relação linear entre regiões de homozigose e o coeficiente de endocruzamento estimado por meio de análise de genealogias.

Essas regiões genômicas de homozigose encontradas em indivíduos endocruzados têm sido foco de estudo de alguns trabalhos, os quais envolvem grande número de marcadores de microssatélites e densos painéis de marcadores de SNPs (Leutenegger *et al.*, 2003; Weir, 2004; Kearney *et al.*; 2011; Purfield *et al.*, 2012).

A presença de um excesso de valores de **F** significativamente diferentes de zero, sugere que os indivíduos quilombolas possam apresentar regiões do genoma em excesso de homozigose, o que se espera em isolados populacionais com excesso de casamentos endogâmicos. Isso estaria de acordo com os resultados obtidos por meio da análise genealógica descrita mais anteriormente. A existência dessas regiões poderá ser eventualmente verificada por meio da análise de um número maior de marcadores genéticos.

5.4. Análise comparativa dos valores de F

A tabela 5.2 mostra os valores do coeficiente médio de endocruzamento obtidos por meio da análise genealógica (F_G) e por meio da análise de marcadores moleculares (F_A), em relação a cada uma das comunidades e em relação ao total das comunidades.

Tabela 5.2: Valores de F_m obtidos por análise genealógica (F_G) e pela análise de marcadores moleculares (F_A).

	F_G	F_A
Abobral	0,00327	0,0492
André Lopes	0,00089	-0,0007
Ivaporanduva	0,00033	-0,0139
Nhunguara	0,00178	-0,0546
Pedro Cubas	0	-0,0449
Pilões/Maria Rosa	0,00101	-0,0490
São Pedro/Galvão	0,00028	-0,0242
Sapatu	0	0,0183
Total	0,00134	0,0341

No caso baseado na análise genealógica, o F_G total foi obtido como no caso das comunidades isoladas, ou seja, ponderando-se todos os valores de F_i de todos os indivíduos de todas as comunidades.

No caso baseado na análise molecular, repetiu-se a análise de cada loco somando-se os resultados de genotipagem de todas as populações, estimando-se um valor de F para cada um dos locos e ponderando-se, finalmente, todos os valores de F pelo inverso de suas variâncias.

Tanto em relação às comunidades consideradas isoladamente como em conjunto, verifica-se que os valores de F_G são de uma ordem de grandeza inferior quando comparadas aos valores correspondentes de F_A .

Isso demonstra claramente o já apontado na discussão dos itens anteriores, ou seja, de que as informações genealógicas disponíveis dos isolados quilombolas devem fornecer subestimativas dos verdadeiros valores de F .

É digno de nota o fato de que o valor médio de F_A , obtido para o conjunto dos isolados, é comparável aos obtidos em isolados da literatura com níveis relativamente altos de endogamia, como os descritos entre os huteritas por Abney *et al.*, 2000 ($F = 0,034$) e no quilombo de Valongo por Souza e Culpi em 1992 ($F = 0,048$).

Trabalhos pioneiros sobre amostragens realizadas em populações de tamanho pequeno (revisão em Cannings e Edwards, 1969) mostraram que nelas deve ocorrer um excesso de heterozigotos. Isso explica o número grande de valores de F_A negativos observados nas comunidades analisadas isoladamente e o valor médio positivo de F_A quando se analisa o conjunto de todas as populações.

5.5. Análise de subestruturação populacional

Nas análises feitas, consideramos as comunidades quilombolas Abobral (margens esquerda e direita), André Lopes, São Pedro/Galvão, Ivaporanduva, Pilões/Maria Rosa,

Nhunguara, Pedro Cubas e Sapatu tanto separadamente, como de forma conjunta.

De acordo com o Instituto Socioambiental (em parceria com o Instituto de Terras de São Paulo), um dos fundadores de São Pedro seria Bernardo Furquim, um escravo fugido da região de Campinas, que havia se casado com pelo menos duas mulheres, das quais teve 24 filhos, os quais se estabeleceram na região, migrando oportunamente para as outras comunidades. A partir desse momento, iniciou-se um processo de povoamento da região onde se localizam as comunidades de Galvão e São Pedro (Santo e Tatto, 2008). Bernardo Furquim teria chegado por volta de 1830 às terras de Pedro Cubas e Ivaporunduva. A ocupação das terras de Nhunguara está relacionada à expansão de grupos ao entorno, como Ivaporunduva e São Pedro, populações com importante papel na consolidação da ocupação negra no Vale do Ribeira (Oliveira Junior *et al.*, 2000). A ocupação de André Lopes se deu a partir de expansões de grupos negros estabelecidos no entorno de Ivaporunduva, São Pedro e Nhunguara. Sapatu, por sua vez, foi formada por negros que fugiram do recrutamento forçado na Guerra do Paraguai e pelo estabelecimento de indivíduos de outras comunidades, como é o caso de Júlio Furquim (neto de Bernardo Furquim) (Santos e Tatto, 2008).

Por terem um estoque ancestral único, portanto, são constantemente observadas relações genealógicas em todas as

comunidades até hoje, o que explica o fluxo gênico entre as populações. Isso significa que a utilização de todas as populações como um único grupo é totalmente plausível dada a estrutura populacional encontrada.

Para testar essa hipótese, estimamos os valores dos índices de fixação F_{IT} , F_{ST} e F_{IS} de cada um dos 30 locos considerados (Tabela 4.7) para o isolado quilombola como um todo.

Os valores estimados de F_{ST} são invariavelmente muito baixos, o que sugere a existência de níveis apreciáveis de fluxo gênico entre as populações. Esse resultado foi obtido também por Kimura *et al.* (2013), que estimaram o nível de subestruturação populacional das comunidades quilombolas do Vale do Ribeira, mostrando que essas comunidades são mais similares geneticamente umas com as outras do que outro conjunto de populações quilombolas amazônico. Dentre os valores de F_{ST} , o único com valores acima de 5% ocorreu em relação ao loco *PPARG*, cujas amostras continham apenas duas classes genotípicas (CC e CG) em quatro populações (Abobral, Nhunguara, Pedro Cubas e Sapatu) das oito consideradas. No item 4.3.3. já apresentamos uma discussão aprofundada sobre interpretação dos resultados obtidos em relação a esse índice.

Excluindo-se os dados em relação ao loco *PPARG*, todos os outros estão em conformidade com o que foi encontrado por Kimura *et al.* (2013) e com as inferências históricas e

genealógicas favoráveis à existência de fluxo gênico entre populações.

Com relação aos valores obtidos do índice F_{IT} (coeficiente total de endocruzamento), verificou-se que em relação a seis locos eles são significativamente diferentes de zero (cinco com valores positivos e um negativo). Dos cinco locos cujos valores são positivos, quatro (*ACE*, *GNB3* (rs5441), *PLIN1* e D21S1437) revelaram ser diferentes de zero considerando-se a população total (Tabela 4.4). Os valores de intervalo de confiança obtidos por *bootstrap* de 1000 simulações estão, portanto, condizentes com os cálculos anteriormente detalhados.

6. CONCLUSÃO

Durante nosso trabalho, a obtenção das estimativas médias de **F** das comunidades quilombolas do Vale do Ribeira por meio da análise genealógica resultou em valores de **F** maiores dos que os encontrados para as populações brasileira e paulista, indicando um aumento médio na quantidade de casamentos endogâmicos. Os valores dessas estimativas são considerados altos mesmo na ausência de informações genealógicas confiáveis, que resultou na subestimativa dos valores de **F**.

Não foi possível a obtenção do coeficiente de endocruzamento por meio da frequência estimada de casamentos de pessoas com o mesmo sobrenome, dada a ausência completa, nas comunidades, de um padrão lógico de transmissão desses sobrenomes, os quais são transmitidos por qualquer um dos genitores ou gerados aleatoriamente.

A estimativa média de **F** considerando-se os dados de todos os locos de todas as populações é comparável às estimativas obtidas nas análises genealógicas de outros isolados da literatura, indicando que os valores obtidos nas análises das genealogias são, de fato, subestimados.

Apresentamos um método alternativo, inédito na literatura, para se calcular o valor da variância da estimativa **F** baseado na análise de marcadores moleculares. Ao contrário do método existente na literatura, apresenta a

vantagem de ser facilmente aplicável ao caso de alelos múltiplos.

7. RESUMO

A porção paulista do Vale do Ribeira concentra a maior quantidade de comunidades remanescentes de quilombos do estado de São Paulo, abrangendo uma área de cerca de 10% de seu território.

Por meio das análises de marcadores moleculares, de frequências de casais com mesmo sobrenome e de dados genealógicos, procuramos obter parâmetros globais de caracterização das comunidades: sistema de cruzamentos e medidas de subestruturação populacional.

Utilizamos dados genealógicos de cerca de 2000 indivíduos e moleculares de cerca de 1000 indivíduos das comunidades de Maria Rosa, Pilões, Galvão, São Pedro, Pedro Cubas, Ivaporanduva, Sapatu, André Lopes, Nhunguara, Abobral (margens esquerda e direita), Poça e Reginaldo.

A estimativa média de F obtida pela análise de genealogias apresentou valor 0,00134, o qual, embora subestimado devido à falta de informações genealógicas, é cerca de 1,5 vezes mais elevado do que a estimativa apresentada para a população total brasileira e duas vezes maior que a obtida para o estado de São Paulo, comparando-se a valores apresentados em outros isolados da literatura.

A partir das análises de locos genômicos obtivemos, para as comunidades separadamente, os valores médios de F relativos aos 239 locos de todas as comunidades, dentre os

quais 12 (5%) mostraram-se estatisticamente diferentes de zero ao nível de $P \leq 0,05$, frequência esperada de desvios ocorrendo ao acaso. Quando analisada de forma conjunta, a população apresentou quatro dos 30 locos (13,33%) com desvios significativos de pan-mixia, valor acima do esperado ao acaso, o que indica um excesso de homozigose no isolado total.

Obtivemos o valor médio total de F pela ponderação dos F de cada um dos locos pelos recíprocos de suas variâncias, estas calculadas por meio de uma metodologia inédita proposta neste trabalho, a qual é aplicável a casos de marcadores contendo mais de dois alelos. O valor médio de F que obtivemos é comparável aos obtidos de outros isolados da literatura.

Os valores do índice F_{ST} obtidos em uma análise de subestruturação populacional tiveram valores modestos geralmente bem menores que 5%, indicando a presença de níveis de subestruturação muito modestos.

8. ABSTRACT

Vale do Ribeira is a region located at the southern part of the state of São Paulo, corresponding to about 10% of its territory. Most of the *quilombo* remnants of the state are placed inside this region.

Using both molecular markers and genealogical data analyses, we estimated population genetic parameters from the communities (breeding system and substructure organization).

Genealogical and molecular data (collected from 2000 and 1000 individuals respectively) were obtained from 13 *quilombo* communities: Maria Rosa, Pilões, Galvão, São Pedro, Pedro Cubas, Ivaporanduva, Sapatu, André Lopes, Nhunguara, Abobral (both left and right edges), Poça e Reginaldo.

Genealogical analysis enabled us to obtain a mean **F** value of 0.00134, that represents an underestimate of the true value due to lack of reliable genealogical information. Even so, this value is almost 1.5 times higher than the value estimated for the total Brazilian population and almost twice as high than the same parameter estimated for the state of São Paulo.

By means of genomic *loci* data analysis, we obtained mean **F** value for the *quilombo* communities separately. Twelve (5%) out of a total of 239 *loci* from eight

communities were in $p^2:2pq:q^2$ ratios, as expected by chance; and for the set of all *quilombo* communities, four (13.33%) out of 30 *loci* deviated significantly from Hardy-Weinberg ratios, indicating an excess of homozygosis.

We also estimated the weighted mean value of F for the whole population by averaging the F values obtained from each *locus* by the reciprocal of their corresponding variances. For calculating the variance of estimated F values we developed a novel method that can be easily generalized to the case of any number of alleles segregating at an autosomal *locus*.

No significant levels of population substructure were detected since the estimated F_{ST} values among populations were in general quite modest.

We present also, as attachment to this work, the listings of the main computer program codes we used in our calculations and a section on the evolution of the fixation index F under different systems of regular endogamy.

9. REFERÊNCIAS BIBLIOGRÁFICAS

Abney M, McPeck MS, Ober C. Estimation of variance components of quantitative traits in inbred populations. **Am. J. Hum. Genet.** 66: 629-650, 2000.

Angeli CB. **Variabilidade genética em populações afro-brasileiras: estudo dos locos DXS548, FRAXAC1 e D1S80.** Dissertação de Mestrado, Universidade de São Paulo, 2003.

Angeli CB. **Susceptibilidade genética e outros fatores de risco associados ao sobrepeso e à obesidade em populações afro-descendentes do Vale do Ribeira-SP.** Tese de Doutorado, Universidade de São Paulo, 2008.

Angeli CB, Capelli LP, Auricchio MTBM, Vianna-Morgante AM, Mingronni-Netto RC, Leal-Mesquita ER, Ribeiro-dos-Santos AKC, Ferrari I, Oliveira SF, Klautau-Guimarães MN. AGG interspersed patterns in the CGG repeat of the FMR1 gene and linked DXS548/FRAXAC1 haplotypes in Brazilian populations. **Am. J. Med. Genet.** 132A: 210-214, 2005.

Angeli CB, Kimura L, Auricchio MTBM, Vicente JP, Mattevi VS, Zembruski VM, Hutz MH, Pereira AC, Pereira TV, Mingronni-Netto RC. Multilocus analyses of seven candidate genes suggest interacting pathways for obesity-related traits in Brazilian populations. **Obesity.** 19(6): 1244-1251, 2011.

Auricchio MTBM, Vicente JP, Meyer D, Mingronni-Netto RC. Frequency and origins of hemoglobin S mutation in

African-derived Brazilian populations. **Hum. Biol.** 79: 667-678, 2007.

Arcos-Burgos M, Muenke M. Genetics of population isolates. **Clin. Genet.** 61: 233-247, 2002.

Beiguelman B. **Genética de populações humanas.** Campinas, 2005. Disponível gratuitamente como e-book (formato .pdf) em <http://www.desvirtual.com/bbeiguel/ebook02.htm>.

Bittles AH. Endogamy, consanguinity and community genetics. **J. Genet.** 81(3): 91-98, 2002.

Brown JL, Eklund A. Kin recognition and the major histocompatibility complex: an integrative review. **Am. Nat.** 143: 435-461, 1994.

Cannings C, Edwards AW. Expected genotypic frequencies in a small sample: deviation from Hardy-Weinberg equilibrium. **Am. J. Hum. Genet.** 21(3): 245-247, 1969.

Cotrim NH. **Variabilidade genética das inserções de Alu em remanescentes de quilombos.** Dissertação de Mestrado, Universidade de São Paulo, 2003.

Cotrim NH, Auricchio MT, Vicente JP, Otto PA, Mingroni-Netto RC. Polymorphic Alu insertion in six brazilian african-derived populations. **Am. J. Hum. Biol.** 16: 264-277, 2004.

Crevelaro MA. **Transição nutricional em populações quilombolas da mata atlântica.** Dissertação de Mestrado, Universidade de São Paulo, 2009.

Crow JF, Felsenstein J. The effect of assortative mating on the genetic composition of a population. **Eugen. Quart.** 15(2): 85-97, 1968.

Crow JF, Kimura M. **An introduction population genetics theory.** Alpha Editions, Madison, 1970.

Dorsten LE, Hotchkiss L, King TM. The effect of inbreeding on early childhood mortality: twelve generations of an Amish settlement. **Demography.** 36(2): 263-271, 1999.

Ellis WS, Starmer WT. Inbreeding as measured by isonymy, pedigrees, and population size in Törbel, Switzerland. **Am. J. Hum. Genet.** 30: 366-376, 1978.

Freire-Maia N. Inbreeding in Brazil. **Am. J. Hum. Genet.** 9: 284-298, 1957.

Freire-Maia N. Genetic effects in Brazilian populations due to consanguineous marriages. **Am. J. Med. Genet.** 35: 115-117, 1990.

Freire-Maia N, Freire-Maia A. The structure of consanguineous marriages and its genetic implications. **Ann. Hum. Genet.** 25: 29-39, 1961.

Freire-Maia N, Krieger H. A Jewish isolate in southern Brazil. **Ann. Hum. Genet. Lond.** 27(31): 31-39, 1963.

Fyfe JL, Bailey NT. Plant breeding studies in leguminous forage crops I. Natural cross-breeding in winter beans. **J. Agric. Sci.** 41(04): 371-378, 1951.

Hamamy H, Jamhawi L, Al-Darawsheh J, Ajlouni K. Consanguineous marriages in Jordan: why is the rate changing with time? **Clin. Genet.** 67: 511-516, 2005.

Hardy GH. Mendelian proportions in a mixed population. **Science.** 28: 49-50, 1908.

Hartl DL, Clark AG. **Princípios de genética de populações.** Artmed, Porto Alegre, 4ª Ed., 2010.

Jackson CE, Symon WE, Pruden EL, Kaehr IM, Mann JD. Consanguinity and Blood Group Distribution in an Amish Isolate. **Am. J. Hum. Genet.** 20(6): 522-527, 1968.

Jorde LB, Watkins WS, Kere J, Nyman D, Eriksson AW. Gene mapping in isolated populations: new roles for old friends? **Hum. Hered.** 50: 57-65, 2000.

Kearney HM, Kearney JB, Conlin LK. Diagnostic implications of excessive homozygosity detected by SNP-based microarrays: consanguinity, uniparental disomy, and recessive single-gene mutations. **Clin. Lab. Med.** 31: 595-613, 2011.

Kimura L. **Fatores genéticos associados à hipertensão essencial em populações remanescentes de quilombos do Vale do Ribeira - São Paulo.** Tese de Doutorado, Universidade de São Paulo, 2010.

Kimura L, Angeli CB, Auricchio MTBM, Fernandes GR, Pereira AC, Vicente JP, Pereira TV, Mingroni-Netto RC. Multilocus family-based association analysis of seven candidate polymorphisms with essential hypertension in an

african-derived semi-isolated brazilian population. **Am. J. Hypertens.** 2012: 859219, 2012.

Kimura L, Ribeiro-Rodrigues EM, Auricchio MTBM, Vicente JP, Santos SEB, Mingroni-Netto RC. Genomic ancestry of rural african-derived populations from southeastern Brazil. **Am. J. Hum. Genet.** 25: 35-41, 2013.

Lautenegger AL, Prum B, Génin E, Verny C, Lemaingue A, Clerget-Darpoux F, Thompson EA. Estimation of the inbreeding coefficient through use of genomic data. **Am. J. Hum. Genet.** 73: 516-523, 2003.

Lewontin R, Kirk D, Crow JF. Selective mating, assortative mating, and inbreeding: definitions and implications. **Eugen. Quart.** 15: 141-143, 1968.

Li CC. **First course in population genetics.** Boxwood Press, Chicago, 1955.

Li CC, Sacks L. The derivation of joint distribution and correlation between relatives by the use of stochastic matrices. **Biometrics** 10: 347-360, 1954.

Macedo de Souza LI. **Variabilidade molecular do cromossomo Y em remanescentes de quilombos do Vale do Ribeira.** Dissertação de Mestrado, Universidade de São Paulo, 2003.

Malécot G. **Les mathématiques de l'hérédité.** Masson et Cie, Paris, 1948.

Mingroni-Netto RC, Angeli CB, Auricchio MT, Leal-Mesquita ER, Ribeiro-dos-Santos AK, Ferrari I, Hutz MH,

Salzano FM, Hill K, Hurtado AM, Vianna-Morgante AM. Distribution of CGG repeats and FRAXAC1/DXS548 alleles in South American populations. **Am. J. Med. Genet.** 111A: 243-252, 2002.

Mingroni-Netto RC, Angeli CB, Kimura L, Auricchio MTBM, Vicente JP. Doenças modernas nos antigos quilombos: a obesidade e a hipertensão no Vale do Ribeira. In: Volochko A, Batista LE. **Saúde nos quilombos**. Instituto da Saúde, São Paulo, 2009a, pp. 179-191.

Mingroni-Netto RC, Auricchio MTBM, Vicente JP. Importância da pesquisa do traço e da anemia falciforme nos remanescentes de quilombos do Vale do Ribeira-SP. In: Volochko A, Batista LE. **Saúde nos quilombos**. Instituto da Saúde, São Paulo, 2009b, pp. 169-177.

Morton NE. Non-randomness in consanguineous marriage. **Ann. Hum. Genet.** 20(2): 116-124, 1955.

Morton NE, Crow JF, Muller HJ. An estimate of the mutational damage in man from data on consanguineous marriages. **Proc. N. A. S.** 42: 855-863, 1956.

Ober C, Weitkamp LR, Cox N, Dytch H, Kostyu D, Elias S. HLA and mate choice in humans. **Am. J. Hum. Genet.** 61: 497-504, 1997.

Oliveira Junior AN, Stucchi D, Chagas MF, Brasileiro SS. Comunidades negras de Ivaporunduva, São Pedro, Pedro Cubas, Sapatu, Nhunguara, André Lópes, Maria Rosa e Pilões. In: Andrade T, Pereira CAC, Andrade MRO. (Eds.). **Negros do**

Ribeira: reconhecimento étnico e conquista do território.
Páginas & Letras - Editora Gráfica, São Paulo, 2000. pp.
39-192.

Otto PA. **Genética de poblaciones humanas.** EdUNaM
(Editorial Universitaria de la Universidad Nacional de
Misiones, Argentina), Posadas, 2008, 214 pp. ISBN 978-950-
579-113-2. Disponível gratuitamente como e-book (formato
.pdf) em http://www.ib.usp.br/~otto/pop_genetics.htm ou em
<http://www.lacygh.com.ar/abajo.htm>.

Pasinato R, Rettl KI. Desenvolvimento local
sustentável: a contribuição das comunidades quilombolas do
Vale do Ribeira. In: Volochko A, Batista LE. (Orgs.). **Saúde
nos quilombos.** Instituto da Saúde, São Paulo, 2009, pp. 43-
56.

Peltonen L, Pekkarinen P, Aaltonen J. Messages from an
isolate: lessons from the finnish gene pool. **Biol. Chem.
Hoppe-Seyler.** 376: 697-704, 1995.

Penn D, Potts W. MHC-disassortative mating preferences
reversed by cross-fostering. **Proc. R. Soc. Lond. B.** 265:
1299-1306, 1998.

Purfield DC, Berry DP, McParland S, Bradley DG. Runs
of homozygosity and population history in cattle. **BMC
Genet.** 13(70): 1471-2156, 2012.

Rincon D. **Estudos de DNA mitocondrial em populações
remanescentes de quilombos do Vale do Ribeira - São Paulo.**
Dissertação de Mestrado, Universidade de São Paulo, 2009.

Roberts SC, Gosling LM, Carter V, Petrie M. MHC-correlated odour preferences in humans and the use of oral contraceptives. **Proc. R. Soc. B.** 275: 2715-2722, 2008.

Rozen S, Skaletsky, HJ. Primer3 on the WWW for general users and for biologist programmers. *In*: Krawetz S, Misener S (Eds.). **Bioinformatics Methods and Protocols: Methods in Molecular Biology**. Humana Press, Totowa, NJ, 2000, pp. 365-386.

Santos KMP, Tatto N (Eds.). **Agenda socioambiental de comunidades quilombolas do Vale do Ribeira**. Ipsis Gráfica e Editora, São Paulo, 2008.

Souza IR, Culpi L. Valongo, an isolated Brazilian Black community. I. Structure of the population. **Brazil. J. Genet.** 15: 439-447, 1992.

Spiess EB. **Genes in populations**. John Wiley & Sons, Chicago, 1977.

Stulp G, Buunk AP, Pollet TV, Nettle D, Verhulst S. Are human matings preferences with respect to height reflected in actual pairings? **PLOS ONE.** 8(1): e54186, 2013.

Weir BS. **Genetic data analysis II**. Sinauer Associates, Sunderland, 1996.

Weir BS, Hill WG, Cardon LR. Allelic Association Patterns for a Dense SNP Map. **Genet. Epidemiol.** 27: 442-450, 2004.

Wright S. *Systems of mating. I. The biometric relations between parent and offspring.* **Genetics. 6: 111-123, 1921a.**

Wright S. *Systems of mating. II. The effects of inbreeding on the genetic composition of a population.* **Genetics. 6: 124-143, 1921b.**

Wright S. *Systems of mating. III. Assortative mating based on somatic resemblance.* **Genetics. 6: 144-161, 1921c.**

Wright S. *Systems of mating. IV. The effects of selection.* **Genetics. 6: 162-166, 1921d.**

Wright S. *Systems of mating. V. General considerations.* **Genetics. 6: 167-178, 1921e.**

Wright S. *Coefficients of inbreeding and relationship.* **Am. Nat. 56: 330-338, 1922.**

Wright S. *Evolution in Mendelian populations.* **Genetics. 16: 97-159, 1931.**

Yeh E, Kimura L, Errera FI, Angeli CB, Mingroni-Netto RC, Silva ME, Canani LH, Passos-Bueno MR. Association of polymorphisms at the ADIPOR1 regulatory region with type 2 diabetes and body mass index in a Brazilian population with European or African ancestry. **Braz. J. Med. Biol. Res. 41: 468-472, 2008.**

10. ANEXOS

10.1. Anexo 1

10.2. Anexo 2

10.3. Anexo 3

```
' program filename f_and_vf_est_c3.bas
' F and var(F) estimation
' codominant autosomal loci
' any number of alleles (n < 13)

OPEN "c:/temp/total.dat" FOR INPUT AS #1
INPUT #1, local$, indiv
DIM indiv$(indiv)
FOR i = 1 TO indiv
INPUT #1, a$
indiv$(i) = a$
NEXT i
CLOSE #1

DATA 31
READ locus
DIM locus$(indiv,locus), all$(locus,12), locusname$(locus)
DIM
naa(locus),nab(locus),nac(locus),nad(locus),nae(locus),naf(locus),nag(locus),nah(locus)
DIM
nai(locus),naj(locus),nak(locus),nal(locus),nbb(locus),nbc(locus),nbd(locus),nbe(locus)
DIM
nbf(locus),nbg(locus),nbh(locus),nbi(locus),nbj(locus),nbk(locus),nbl(locus),ncc(locus)
DIM
ncd(locus),nce(locus),ncf(locus),ncg(locus),nch(locus),nci(locus),ncj(locus),nck(locus)
DIM
ncl(locus),ndd(locus),nde(locus),ndf(locus),ndg(locus),ndh(locus),ndi(locus),ndj(locus)
DIM
ndk(locus),ndl(locus),nee(locus),nef(locus),neg(locus),neh(locus),nei(locus),nej(locus)
DIM
nek(locus),nel(locus),nff(locus),nfg(locus),nfh(locus),nfi(locus),nfj(locus),nfk(locus)
DIM
nfl(locus),ngg(locus),ngh(locus),ngi(locus),ngj(locus),ngk(locus),ngl(locus),nhh(locus)
DIM
nhi(locus),nhj(locus),nhk(locus),nhl(locus),nii(locus),nij(locus),nik(locus),nil(locus)
DIM njj(locus),njk(locus),njl(locus),nkk(locus),nkl(locus),nll(locus)
DIM pa(locus), pb(locus), pc(locus),pd(locus), pe(locus), pf(locus),pg(locus),
ph(locus), pi(locus)
DIM pj(locus), pk(locus), pl(locus)
DIM F(locus), n(locus)

all$(1,1) = "D"      : all$(1,2) = "I"
all$(2,1) = "G"      : all$(2,2) = "T"
all$(3,1) = "C"      : all$(3,2) = "T"
all$(4,1) = "A"      : all$(4,2) = "G"
all$(5,1) = "C"      : all$(5,2) = "T"
all$(6,1) = "A"      : all$(6,2) = "C"
all$(7,1) = "C"      : all$(7,2) = "T"
all$(8,1) = "A"      : all$(8,2) = "G"
all$(9,1) = "C"      : all$(9,2) = "G"
all$(10,1) = "A"     : all$(10,2) = "G"
all$(11,1) = "1"     : all$(11,2) = "2"
all$(12,1) = "A"     : all$(12,2) = "G"
all$(13,1) = "C"     : all$(13,2) = "G"
all$(14,1) = "C"     : all$(14,2) = "G"
all$(15,1) = "A"     : all$(15,2) = "S"      : all$(15,3) = "C"
all$(16,1) = "122"   : all$(16,2) = "130"   : all$(16,3) = "134"   : all$(16,4) = "138" :
all$(16,5) = "142"
all$(16,6) = "146"   : all$(16,7) = "148"   : all$(16,8) = "154"   : all$(16,9) = "158"
all$(17,1) = "163"   : all$(17,2) = "167"   : all$(17,3) = "171"   : all$(17,4) = "175" :
all$(17,5) = "179"
all$(17,6) = "183"   : all$(17,7) = "187"
all$(18,1) = "202"   : all$(18,2) = "204"   : all$(18,3) = "216"   : all$(18,4) = "220" :
all$(18,5) = "224"
all$(18,6) = "228"   : all$(18,7) = "232"   : all$(18,8) = "236"   : all$(18,9) = "240" :
all$(18,10) = "244"
all$(18,11) = "248"  : all$(18,12) = "252"
all$(19,1) = "268"   : all$(19,2) = "272"   : all$(19,3) = "276"   : all$(19,4) = "280" :
all$(19,5) = "284"
all$(19,6) = "288"   : all$(19,7) = "292"   : all$(19,8) = "296"
```

```

all$(20,1) = "138"      : all$(20,2) = "142"      : all$(20,3) = "146"      : all$(20,4) = "150"      :
all$(20,5) = "154"
all$(20,6) = "158"      : all$(20,7) = "162"      : all$(20,8) = "166"
all$(21,1) = "176"      : all$(21,2) = "180"      : all$(21,3) = "184"      : all$(21,4) = "188"      :
all$(21,5) = "192"
all$(21,6) = "196"
all$(22,1) = "214"      : all$(22,2) = "222"      : all$(22,3) = "226"      : all$(22,4) = "230"      :
all$(22,5) = "234"
all$(22,6) = "238"      : all$(22,7) = "242"      : all$(22,8) = "246"
all$(23,1) = "265"      : all$(23,2) = "269"      : all$(23,3) = "273"      : all$(23,4) = "277"      :
all$(23,5) = "281"
all$(23,6) = "285"      : all$(23,7) = "289"      : all$(23,8) = "293"      : all$(23,9) = "297"
all$(24,1) = "120"      : all$(24,2) = "124"      : all$(24,3) = "128"      : all$(24,4) = "132"      :
all$(24,5) = "136"
all$(24,6) = "140"      : all$(24,7) = "144"      : all$(24,8) = "148"
all$(25,1) = "192"      : all$(25,2) = "196"      : all$(25,3) = "200"      : all$(25,4) = "204"      :
all$(25,5) = "208"
all$(25,6) = "212"      : all$(25,7) = "216"      : all$(25,8) = "220"      : all$(25,9) = "224"
all$(26,1) = "228"      : all$(26,2) = "232"      : all$(26,3) = "236"      : all$(26,4) = "240"      :
all$(26,5) = "244"
all$(26,6) = "248"
all$(27,1) = "277"      : all$(27,2) = "285"      : all$(27,3) = "289"      : all$(27,4) = "293"      :
all$(27,5) = "297"
all$(27,6) = "301"      : all$(27,7) = "305"      : all$(27,8) = "309"
all$(28,1) = "106"      : all$(28,2) = "114"      : all$(28,3) = "118"      : all$(28,4) = "122"      :
all$(28,5) = "126"
all$(28,6) = "130"      : all$(28,7) = "134"
all$(29,1) = "147"      : all$(29,2) = "151"      : all$(29,3) = "155"      : all$(29,4) = "159"      :
all$(29,5) = "163"
all$(29,6) = "167"      : all$(29,7) = "171"      : all$(29,8) = "175"
all$(30,1) = "198"      : all$(30,2) = "202"      : all$(30,3) = "206"      : all$(30,4) = "210"      :
all$(30,5) = "214"
all$(30,6) = "218"      : all$(30,7) = "222"      : all$(30,8) = "226"      : all$(30,9) = "230"      :
all$(30,10) = "234"
all$(31,1) = "243"      : all$(31,2) = "247"      : all$(31,3) = "251"      : all$(31,4) = "255"      :
all$(31,5) = "259"
all$(31,6) = "263"      : all$(31,7) = "267"      : all$(31,8) = "271"

```

```

locusname$(1) = "ACE - rs1799752"      : locusname$(2) = "NOS3 - rs1799983"      :
locusname$(3) = "GNB3 - rs5443"
locusname$(4) = "GNB3 - rs5441"      : locusname$(5) = "AGT - rs669"      :
locusname$(6) = "ADD2 - rs3755351"
locusname$(7) = "GRK4 - rs1801058"      : locusname$(8) = "PLIN1 - rs2289487"      :
locusname$(9) = "INSIG2 - rs7566605"
locusname$(10) = "LEP - rs2167270"      : locusname$(11) = "LEPR - rs1137101"      :
locusname$(12) = "ADRB2 - rs1042713"
locusname$(13) = "PPARG - rs1801282"      : locusname$(14) = "RETN - rs1862513"      :
locusname$(15) = "HBB"
locusname$(16) = "D5S816"      : locusname$(17) = "D1S551"      :
locusname$(18) = "D7S3061"
locusname$(19) = "D4S3248"      : locusname$(20) = "D16S539"      :
locusname$(21) = "D9S922"
locusname$(22) = "D10S1426"      : locusname$(23) = "D7S821"      :
locusname$(24) = "D13S317"
locusname$(25) = "D8S2324"      : locusname$(26) = "D19S559"      :
locusname$(27) = "D6D9S301"
locusname$(28) = "D20S482"      : locusname$(29) = "D21S1437"      :
locusname$(30) = "D9S301"
locusname$(31) = "D18S535"

```

```

FOR i = 1 TO indiv
FOR j = 1 TO locus
locus$(i, j) = MID$(indiv$(i), 2 * j - 1, 2)
NEXT j: NEXT i
FOR i = 1 TO indiv
FOR j = 1 TO locus
IF locus$(i, j) = "aa" THEN naa(j) = naa(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ab" THEN nab(j) = nab(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ac" THEN nac(j) = nac(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ad" THEN nad(j) = nad(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ae" THEN nae(j) = nae(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "af" THEN naf(j) = naf(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ag" THEN nag(j) = nag(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ah" THEN nah(j) = nah(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ai" THEN nai(j) = nai(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "aj" THEN naj(j) = naj(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ak" THEN nak(j) = nak(j) + 1 : n(j) = n(j) + 1

```



```

IF locus$(i, j) = "al" THEN nal(j) = nal(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "bb" THEN nbb(j) = nbb(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "bc" THEN nbc(j) = nbc(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "bd" THEN nbd(j) = nbd(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "be" THEN nbe(j) = nbe(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "bf" THEN nbf(j) = nbf(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "bg" THEN nbg(j) = nbg(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "bh" THEN nbh(j) = nbh(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "bi" THEN nbi(j) = nbi(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "bj" THEN nbj(j) = nbj(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "bk" THEN nbk(j) = nbk(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "bl" THEN nbl(j) = nbl(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "cc" THEN ncc(j) = ncc(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "cd" THEN ncd(j) = ncd(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ce" THEN nce(j) = nce(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "cf" THEN ncf(j) = ncf(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "cg" THEN ncg(j) = ncg(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ch" THEN nch(j) = nch(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ci" THEN nci(j) = nci(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "cj" THEN ncj(j) = ncj(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ck" THEN nck(j) = nck(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "cl" THEN ncl(j) = ncl(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "dd" THEN ndd(j) = ndd(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "de" THEN nde(j) = nde(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "df" THEN ndf(j) = ndf(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "dg" THEN ndg(j) = ndg(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "dh" THEN ndh(j) = ndh(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "di" THEN ndi(j) = ndi(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "dj" THEN ndj(j) = ndj(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "dk" THEN ndk(j) = ndk(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "dl" THEN ndl(j) = ndl(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ee" THEN nee(j) = nee(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ef" THEN nef(j) = nef(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "eg" THEN neg(j) = neg(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "eh" THEN neh(j) = neh(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ei" THEN nei(j) = nei(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ej" THEN nej(j) = nej(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ek" THEN nek(j) = nek(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "el" THEN nel(j) = nel(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ff" THEN nff(j) = nff(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "fg" THEN nfg(j) = nfg(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "fh" THEN nfh(j) = nfh(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "fi" THEN nfi(j) = nfi(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "fj" THEN nfj(j) = nfj(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "fk" THEN nfk(j) = nfk(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "fl" THEN nfl(j) = nfl(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "gg" THEN ngg(j) = ngg(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "gh" THEN ngh(j) = ngh(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "gi" THEN ngi(j) = ngi(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "gj" THEN ngj(j) = ngj(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "gk" THEN ngk(j) = ngk(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "gl" THEN ngl(j) = ngl(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "hh" THEN nhh(j) = nhh(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "hi" THEN nhi(j) = nhi(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "hj" THEN nhj(j) = nhj(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "hk" THEN nhk(j) = nhk(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "hl" THEN nhl(j) = nhl(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ii" THEN nii(j) = nii(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ij" THEN nij(j) = nij(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ik" THEN nik(j) = nik(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "il" THEN nil(j) = nil(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "jj" THEN njj(j) = njj(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "jk" THEN njk(j) = njk(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "jl" THEN njl(j) = njl(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "kk" THEN nkk(j) = nkk(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "kl" THEN nkl(j) = nkl(j) + 1 : n(j) = n(j) + 1
IF locus$(i, j) = "ll" THEN nll(j) = nll(j) + 1 : n(j) = n(j) + 1
NEXT j: NEXT i
PRINT " "; local$ : PRINT
FOR j = 1 TO locus
IF n(j) = 0 THEN PRINT "'no data for locus # "; j; " ("; locusname$(j) ;") " : GOTO
[getout1]
pa(j) = (2 * naa(j) + nab(j) + nac(j) + nad(j) + nae(j) + naf(j) + nag(j) + nah(j) +
nai(j) + naj(j) + nak(j) + nal(j)) / (2 * n(j))
pb(j) = (2 * nbb(j) + nab(j) + nbc(j) + nbd(j) + nbe(j) + nbf(j) + nbg(j) + nbh(j) +
nbi(j) + nbj(j) + nbk(j) + nbl(j)) / (2 * n(j))

```

```

pc(j) = (2 * ncc(j) + nac(j) + nbc(j) + ncd(j) + nce(j) + ncf(j) + ncg(j) + nch(j) +
nci(j) + ncj(j) + nck(j) + ncl(j)) / (2 * n(j))
pd(j) = (2 * ndd(j) + nad(j) + nbd(j) + ncd(j) + nde(j) + ndf(j) + ndg(j) + ndh(j) +
ndi(j) + ndj(j) + ndk(j) + ndl(j)) / (2 * n(j))
pe(j) = (2 * nee(j) + nae(j) + nbe(j) + nce(j) + nde(j) + nef(j) + neg(j) + neh(j) +
nei(j) + nej(j) + nek(j) + nel(j)) / (2 * n(j))
pf(j) = (2 * nff(j) + naf(j) + nbf(j) + ncf(j) + ndf(j) + nef(j) + nfg(j) + nfh(j) +
nfi(j) + nfj(j) + nfk(j) + nfl(j)) / (2 * n(j))
pg(j) = (2 * ngg(j) + nag(j) + nbj(j) + ncg(j) + ndg(j) + neg(j) + nfg(j) + ngh(j) +
ngi(j) + ngj(j) + ngk(j) + ngl(j)) / (2 * n(j))
ph(j) = (2 * nhh(j) + nah(j) + nbh(j) + nch(j) + ndh(j) + neh(j) + nfh(j) + ngh(j) +
nhi(j) + nhj(j) + nhk(j) + nhl(j)) / (2 * n(j))
pi(j) = (2 * nii(j) + nai(j) + nbi(j) + nci(j) + ndi(j) + nei(j) + nfi(j) + ngi(j) +
nhi(j) + nij(j) + nik(j) + nil(j)) / (2 * n(j))
pj(j) = (2 * njj(j) + naj(j) + nbj(j) + ncj(j) + ndj(j) + nej(j) + nfj(j) + ngj(j) +
nhj(j) + nij(j) + njk(j) + njl(j)) / (2 * n(j))
pk(j) = (2 * nkk(j) + nak(j) + nbk(j) + nck(j) + ndk(j) + nek(j) + nfk(j) + ngk(j) +
nhk(j) + nik(j) + njk(j) + nkl(j)) / (2 * n(j))
pl(j) = (2 * nll(j) + nal(j) + nbl(j) + ncl(j) + ndl(j) + nel(j) + nfl(j) + ngl(j) +
nhl(j) + nil(j) + njl(j) + nkl(j)) / (2 * n(j))
K = 0
IF pa(j) <> 0 THEN K = K + 1
IF pb(j) <> 0 THEN K = K + 1
IF pc(j) <> 0 THEN K = K + 1
IF pd(j) <> 0 THEN K = K + 1
IF pe(j) <> 0 THEN K = K + 1
IF pf(j) <> 0 THEN K = K + 1
IF pg(j) <> 0 THEN K = K + 1
IF ph(j) <> 0 THEN K = K + 1
IF pi(j) <> 0 THEN K = K + 1
IF pj(j) <> 0 THEN K = K + 1
IF pk(j) <> 0 THEN K = K + 1
IF pl(j) <> 0 THEN K = K + 1
PRINT "'locus # "; j; " ("'; locusname$(j) ;") "
IF naa(j) <> 0 THEN PRINT "'n(';all$(j,1);"/";all$(j,1);") = "'; PRINT USING ("#####",
naa(j))
IF nab(j) <> 0 THEN PRINT "'n(';all$(j,1);"/";all$(j,2);") = "'; PRINT USING ("#####",
nab(j))
IF nac(j) <> 0 THEN PRINT "'n(';all$(j,1);"/";all$(j,3);") = "'; PRINT USING ("#####",
nac(j))
IF nad(j) <> 0 THEN PRINT "'n(';all$(j,1);"/";all$(j,4);") = "'; PRINT USING ("#####",
nad(j))
IF nae(j) <> 0 THEN PRINT "'n(';all$(j,1);"/";all$(j,5);") = "'; PRINT USING ("#####",
nae(j))
IF naf(j) <> 0 THEN PRINT "'n(';all$(j,1);"/";all$(j,6);") = "'; PRINT USING ("#####",
naf(j))
IF nag(j) <> 0 THEN PRINT "'n(';all$(j,1);"/";all$(j,7);") = "'; PRINT USING ("#####",
nag(j))
IF nah(j) <> 0 THEN PRINT "'n(';all$(j,1);"/";all$(j,8);") = "'; PRINT USING ("#####",
nah(j))
IF nai(j) <> 0 THEN PRINT "'n(';all$(j,1);"/";all$(j,9);") = "'; PRINT USING ("#####",
nai(j))
IF naj(j) <> 0 THEN PRINT "'n(';all$(j,1);"/";all$(j,10);") = "'; PRINT USING ("#####",
naj(j))
IF nak(j) <> 0 THEN PRINT "'n(';all$(j,1);"/";all$(j,11);") = "'; PRINT USING ("#####",
nak(j))
IF nal(j) <> 0 THEN PRINT "'n(';all$(j,1);"/";all$(j,12);") = "'; PRINT USING ("#####",
nal(j))
IF nbb(j) <> 0 THEN PRINT "'n(';all$(j,2);"/";all$(j,2);") = "'; PRINT USING ("#####",
nbb(j))
IF nbc(j) <> 0 THEN PRINT "'n(';all$(j,2);"/";all$(j,3);") = "'; PRINT USING ("#####",
nbc(j))
IF nbd(j) <> 0 THEN PRINT "'n(';all$(j,2);"/";all$(j,4);") = "'; PRINT USING ("#####",
nbd(j))
IF nbe(j) <> 0 THEN PRINT "'n(';all$(j,2);"/";all$(j,5);") = "'; PRINT USING ("#####",
nbe(j))
IF nbf(j) <> 0 THEN PRINT "'n(';all$(j,2);"/";all$(j,6);") = "'; PRINT USING ("#####",
nbf(j))
IF nbj(j) <> 0 THEN PRINT "'n(';all$(j,2);"/";all$(j,7);") = "'; PRINT USING ("#####",
nbj(j))
IF nbh(j) <> 0 THEN PRINT "'n(';all$(j,2);"/";all$(j,8);") = "'; PRINT USING ("#####",
nbh(j))
IF nbi(j) <> 0 THEN PRINT "'n(';all$(j,2);"/";all$(j,9);") = "'; PRINT USING ("#####",
nbi(j))
IF nbj(j) <> 0 THEN PRINT "'n(';all$(j,2);"/";all$(j,10);") = "'; PRINT USING ("#####",
nbj(j))

```

```

IF nbk(j) <> 0 THEN PRINT "n(";all$(j,2);"/";all$(j,11);") = ";; PRINT USING ("####",
nbk(j))
IF nbl(j) <> 0 THEN PRINT "n(";all$(j,2);"/";all$(j,12);") = ";; PRINT USING ("####",
nbl(j))
IF ncc(j) <> 0 THEN PRINT "n(";all$(j,3);"/";all$(j,3);") = ";; PRINT USING ("####",
ncc(j))
IF ncd(j) <> 0 THEN PRINT "n(";all$(j,3);"/";all$(j,4);") = ";; PRINT USING ("####",
ncd(j))
IF nce(j) <> 0 THEN PRINT "n(";all$(j,3);"/";all$(j,5);") = ";; PRINT USING ("####",
nce(j))
IF ncf(j) <> 0 THEN PRINT "n(";all$(j,3);"/";all$(j,6);") = ";; PRINT USING ("####",
ncf(j))
IF ncg(j) <> 0 THEN PRINT "n(";all$(j,3);"/";all$(j,7);") = ";; PRINT USING ("####",
ncg(j))
IF nch(j) <> 0 THEN PRINT "n(";all$(j,3);"/";all$(j,8);") = ";; PRINT USING ("####",
nch(j))
IF nci(j) <> 0 THEN PRINT "n(";all$(j,3);"/";all$(j,9);") = ";; PRINT USING ("####",
nci(j))
IF ncj(j) <> 0 THEN PRINT "n(";all$(j,3);"/";all$(j,10);") = ";; PRINT USING ("####",
ncj(j))
IF nck(j) <> 0 THEN PRINT "n(";all$(j,3);"/";all$(j,11);") = ";; PRINT USING ("####",
nck(j))
IF ncl(j) <> 0 THEN PRINT "n(";all$(j,3);"/";all$(j,12);") = ";; PRINT USING ("####",
ncl(j))
IF ndd(j) <> 0 THEN PRINT "n(";all$(j,4);"/";all$(j,4);") = ";; PRINT USING ("####",
ndd(j))
IF nde(j) <> 0 THEN PRINT "n(";all$(j,4);"/";all$(j,5);") = ";; PRINT USING ("####",
nde(j))
IF ndf(j) <> 0 THEN PRINT "n(";all$(j,4);"/";all$(j,6);") = ";; PRINT USING ("####",
ndf(j))
IF ndg(j) <> 0 THEN PRINT "n(";all$(j,4);"/";all$(j,7);") = ";; PRINT USING ("####",
ndg(j))
IF ndh(j) <> 0 THEN PRINT "n(";all$(j,4);"/";all$(j,8);") = ";; PRINT USING ("####",
ndh(j))
IF ndi(j) <> 0 THEN PRINT "n(";all$(j,4);"/";all$(j,9);") = ";; PRINT USING ("####",
ndi(j))
IF ndj(j) <> 0 THEN PRINT "n(";all$(j,4);"/";all$(j,10);") = ";; PRINT USING ("####",
ndj(j))
IF ndk(j) <> 0 THEN PRINT "n(";all$(j,4);"/";all$(j,11);") = ";; PRINT USING ("####",
ndk(j))
IF ndl(j) <> 0 THEN PRINT "n(";all$(j,4);"/";all$(j,12);") = ";; PRINT USING ("####",
ndl(j))
IF nee(j) <> 0 THEN PRINT "n(";all$(j,5);"/";all$(j,5);") = ";; PRINT USING ("####",
nee(j))
IF nef(j) <> 0 THEN PRINT "n(";all$(j,5);"/";all$(j,6);") = ";; PRINT USING ("####",
nef(j))
IF neg(j) <> 0 THEN PRINT "n(";all$(j,5);"/";all$(j,7);") = ";; PRINT USING ("####",
neg(j))
IF neh(j) <> 0 THEN PRINT "n(";all$(j,5);"/";all$(j,8);") = ";; PRINT USING ("####",
neh(j))
IF nei(j) <> 0 THEN PRINT "n(";all$(j,5);"/";all$(j,9);") = ";; PRINT USING ("####",
nei(j))
IF nej(j) <> 0 THEN PRINT "n(";all$(j,5);"/";all$(j,10);") = ";; PRINT USING ("####",
nej(j))
IF nek(j) <> 0 THEN PRINT "n(";all$(j,5);"/";all$(j,11);") = ";; PRINT USING ("####",
nek(j))
IF nel(j) <> 0 THEN PRINT "n(";all$(j,5);"/";all$(j,12);") = ";; PRINT USING ("####",
nel(j))
IF nff(j) <> 0 THEN PRINT "n(";all$(j,6);"/";all$(j,6);") = ";; PRINT USING ("####",
nff(j))
IF nfg(j) <> 0 THEN PRINT "n(";all$(j,6);"/";all$(j,7);") = ";; PRINT USING ("####",
nfg(j))
IF nfh(j) <> 0 THEN PRINT "n(";all$(j,6);"/";all$(j,8);") = ";; PRINT USING ("####",
nfh(j))
IF nfi(j) <> 0 THEN PRINT "n(";all$(j,6);"/";all$(j,9);") = ";; PRINT USING ("####",
nfi(j))
IF nfj(j) <> 0 THEN PRINT "n(";all$(j,6);"/";all$(j,10);") = ";; PRINT USING ("####",
nfj(j))
IF nfk(j) <> 0 THEN PRINT "n(";all$(j,6);"/";all$(j,11);") = ";; PRINT USING ("####",
nfk(j))
IF nfl(j) <> 0 THEN PRINT "n(";all$(j,6);"/";all$(j,12);") = ";; PRINT USING ("####",
nfl(j))
IF ngg(j) <> 0 THEN PRINT "n(";all$(j,7);"/";all$(j,7);") = ";; PRINT USING ("####",
ngg(j))
IF ngh(j) <> 0 THEN PRINT "n(";all$(j,7);"/";all$(j,8);") = ";; PRINT USING ("####",
ngh(j))

```

```

IF ngi(j) <> 0 THEN PRINT "n(";all$(j,7);"/";all$(j,9);") = ";; PRINT USING ("####",
ngi(j))
IF ngj(j) <> 0 THEN PRINT "n(";all$(j,7);"/";all$(j,10);") = ";; PRINT USING ("####",
ngj(j))
IF ngk(j) <> 0 THEN PRINT "n(";all$(j,7);"/";all$(j,11);") = ";; PRINT USING ("####",
ngk(j))
IF ngl(j) <> 0 THEN PRINT "n(";all$(j,7);"/";all$(j,12);") = ";; PRINT USING ("####",
ngl(j))
IF nhh(j) <> 0 THEN PRINT "n(";all$(j,8);"/";all$(j,8);") = ";; PRINT USING ("####",
nhh(j))
IF nhi(j) <> 0 THEN PRINT "n(";all$(j,8);"/";all$(j,9);") = ";; PRINT USING ("####",
nhi(j))
IF nhj(j) <> 0 THEN PRINT "n(";all$(j,8);"/";all$(j,10);") = ";; PRINT USING ("####",
nhj(j))
IF nhk(j) <> 0 THEN PRINT "n(";all$(j,8);"/";all$(j,11);") = ";; PRINT USING ("####",
nhk(j))
IF nhl(j) <> 0 THEN PRINT "n(";all$(j,8);"/";all$(j,12);") = ";; PRINT USING ("####",
nhl(j))
IF nii(j) <> 0 THEN PRINT "n(";all$(j,9);"/";all$(j,9);") = ";; PRINT USING ("####",
nii(j))
IF nij(j) <> 0 THEN PRINT "n(";all$(j,9);"/";all$(j,10);") = ";; PRINT USING ("####",
nij(j))
IF nik(j) <> 0 THEN PRINT "n(";all$(j,9);"/";all$(j,11);") = ";; PRINT USING ("####",
nik(j))
IF nil(j) <> 0 THEN PRINT "n(";all$(j,9);"/";all$(j,12);") = ";; PRINT USING ("####",
nil(j))
IF njj(j) <> 0 THEN PRINT "n(";all$(j,10);"/";all$(j,10);") = ";; PRINT USING ("####",
njj(j))
IF njk(j) <> 0 THEN PRINT "n(";all$(j,10);"/";all$(j,11);") = ";; PRINT USING ("####",
njk(j))
IF njl(j) <> 0 THEN PRINT "n(";all$(j,10);"/";all$(j,12);") = ";; PRINT USING ("####",
njl(j))
IF nkk(j) <> 0 THEN PRINT "n(";all$(j,11);"/";all$(j,11);") = ";; PRINT USING ("####",
nkk(j))
IF nkl(j) <> 0 THEN PRINT "n(";all$(j,11);"/";all$(j,12);") = ";; PRINT USING ("####",
nkl(j))
IF nll(j) <> 0 THEN PRINT "n(";all$(j,12);"/";all$(j,12);") = ";; PRINT USING ("####",
nll(j))
PRINT "n = ";; PRINT USING ("####", n(j))
IF pa(j) <> 0 THEN
PRINT "P(";all$(j,1);") = ";; PRINT USING ("###.####", pa(j))
PRINT "var[P(";all$(j,1);")] = ";; PRINT USING ("###.####", pa(j)*(1-pa(j))/(2*n(j)))
PRINT "s.e.[P(";all$(j,1);")] = ";; PRINT USING ("###.####", SQR(pa(j)*(1-
pa(j))/(2*n(j))))
END IF
IF pb(j) <> 0 THEN
PRINT "P(";all$(j,2);") = ";; PRINT USING ("###.####", pb(j))
PRINT "var[P(";all$(j,2);")] = ";; PRINT USING ("###.####", pb(j)*(1-pb(j))/(2*n(j)))
PRINT "s.e.[P(";all$(j,2);")] = ";; PRINT USING ("###.####", SQR(pb(j)*(1-
pb(j))/(2*n(j))))
END IF
IF pc(j) <> 0 THEN
PRINT "P(";all$(j,3);") = ";; PRINT USING ("###.####", pc(j))
PRINT "var[P(";all$(j,3);")] = ";; PRINT USING ("###.####", pc(j)*(1-pc(j))/(2*n(j)))
PRINT "s.e.[P(";all$(j,3);")] = ";; PRINT USING ("###.####", SQR(pc(j)*(1-
pc(j))/(2*n(j))))
END IF
IF pd(j) <> 0 THEN
PRINT "P(";all$(j,4);") = ";; PRINT USING ("###.####", pd(j))
PRINT "var[P(";all$(j,4);")] = ";; PRINT USING ("###.####", pd(j)*(1-pd(j))/(2*n(j)))
PRINT "s.e.[P(";all$(j,4);")] = ";; PRINT USING ("###.####", SQR(pd(j)*(1-
pd(j))/(2*n(j))))
END IF
IF pe(j) <> 0 THEN
PRINT "P(";all$(j,5);") = ";; PRINT USING ("###.####", pe(j))
PRINT "var[P(";all$(j,5);")] = ";; PRINT USING ("###.####", pe(j)*(1-pe(j))/(2*n(j)))
PRINT "s.e.[P(";all$(j,5);")] = ";; PRINT USING ("###.####", SQR(pe(j)*(1-
pe(j))/(2*n(j))))
END IF
IF pf(j) <> 0 THEN
PRINT "P(";all$(j,6);") = ";; PRINT USING ("###.####", pf(j))
PRINT "var[P(";all$(j,6);")] = ";; PRINT USING ("###.####", pf(j)*(1-pf(j))/(2*n(j)))
PRINT "s.e.[P(";all$(j,6);")] = ";; PRINT USING ("###.####", SQR(pf(j)*(1-
pf(j))/(2*n(j))))
END IF
IF pg(j) <> 0 THEN
PRINT "P(";all$(j,7);") = ";; PRINT USING ("###.####", pg(j))

```

```

PRINT "var[P(";all$(j,7);")]" = ";: PRINT USING ("###.####", pg(j)*(1-pg(j))/(2*n(j)))
PRINT "s.e.[P(";all$(j,7);")]" = ";: PRINT USING ("###.####", SQR(pg(j)*(1-
pg(j))/(2*n(j))))
END IF
IF ph(j) <> 0 THEN
PRINT "P(";all$(j,8);")" = ";: PRINT USING ("###.####", ph(j))
PRINT "var[P(";all$(j,8);")]" = ";: PRINT USING ("###.####", ph(j)*(1-ph(j))/(2*n(j)))
PRINT "s.e.[P(";all$(j,8);")]" = ";: PRINT USING ("###.####", SQR(ph(j)*(1-
ph(j))/(2*n(j))))
END IF
IF pi(j) <> 0 THEN
PRINT "P(";all$(j,9);")" = ";: PRINT USING ("###.####", pi(j))
PRINT "var[P(";all$(j,9);")]" = ";: PRINT USING ("###.####", pi(j)*(1-pi(j))/(2*n(j)))
PRINT "s.e.[P(";all$(j,9);")]" = ";: PRINT USING ("###.####", SQR(pi(j)*(1-
pi(j))/(2*n(j))))
END IF
IF pj(j) <> 0 THEN
PRINT "P(";all$(j,10);")" = ";: PRINT USING ("###.####", pj(j))
PRINT "var[P(";all$(j,10);")]" = ";: PRINT USING ("###.####", pj(j)*(1-
pj(j))/(2*n(j)))
PRINT "s.e.[P(";all$(j,10);")]" = ";: PRINT USING ("###.####", SQR(pj(j)*(1-
pj(j))/(2*n(j))))
END IF
IF pk(j) <> 0 THEN
PRINT "P(";all$(j,11);")" = ";: PRINT USING ("###.####", pk(j))
PRINT "var[P(";all$(j,11);")]" = ";: PRINT USING ("###.####", pk(j)*(1-
pk(j))/(2*n(j)))
PRINT "s.e.[P(";all$(j,11);")]" = ";: PRINT USING ("###.####", SQR(pk(j)*(1-
pk(j))/(2*n(j))))
END IF
IF pl(j) <> 0 THEN
PRINT "P(";all$(j,12);")" = ";: PRINT USING ("###.####", pl(j))
PRINT "var[P(";all$(j,12);")]" = ";: PRINT USING ("###.####", pl(j)*(1-
pl(j))/(2*n(j)))
PRINT "s.e.[P(";all$(j,12);")]" = ";: PRINT USING ("###.####", SQR(pl(j)*(1-
pl(j))/(2*n(j))))
END IF
obshet = nab(j) + nac(j) + nad(j) + nae(j) + naf(j) + nag(j) + nah(j) + nai(j)
+ naj(j) + nak(j) + nal(j)
obshet = obshet + nbc(j) + nbd(j) + nbe(j) + nbf(j) + nbj(j) + nbh(j) + nbi(j) + nbj(j)
+ nbk(j) + nbl(j)
obshet = obshet + ncd(j) + nce(j) + ncf(j) + ncg(j) + nch(j) + nci(j) + ncj(j) + nck(j)
+ ncl(j)
obshet = obshet + nde(j) + ndf(j) + ndg(j) + ndh(j) + ndi(j) + ndj(j) + ndk(j) + ndl(j)
obshet = obshet + nef(j) + neg(j) + neh(j) + nei(j) + nej(j) + nek(j) + nel(j)
obshet = obshet + nfg(j) + nfh(j) + nfi(j) + nfj(j) + nfk(j) + nfl(j)
obshet = obshet + ngh(j) + ngi(j) + ngj(j) + ngk(j) + ngl(j)
obshet = obshet + nhi(j) + nhj(j) + nhk(j) + nhl(j)
obshet = obshet + nij(j) + nik(j) + nil(j)
obshet = obshet + njk(j) + njl(j)
obshet = obshet + nkl(j)
exphet = 2 * pa(j) * (pb(j) + pc(j) + pd(j) + pe(j) + pf(j) + pg(j) + ph(j) +
pi(j) + pj(j) + pk(j) + pl(j))
exphet = exphet + 2 * pb(j) * (pc(j) + pd(j) + pe(j) + pf(j) + pg(j) + ph(j) + pi(j) +
pj(j) + pk(j) + pl(j))
exphet = exphet + 2 * pc(j) * (pd(j) + pe(j) + pf(j) + pg(j) + ph(j) + pi(j) + pj(j) +
pk(j) + pl(j))
exphet = exphet + 2 * pd(j) * (pe(j) + pf(j) + pg(j) + ph(j) + pi(j) + pj(j) + pk(j) +
pl(j))
exphet = exphet + 2 * pe(j) * (pf(j) + pg(j) + ph(j) + pi(j) + pj(j) + pk(j) + pl(j))
exphet = exphet + 2 * pf(j) * (pg(j) + ph(j) + pi(j) + pj(j) + pk(j) + pl(j))
exphet = exphet + 2 * pg(j) * (ph(j) + pi(j) + pj(j) + pk(j) + pl(j))
exphet = exphet + 2 * ph(j) * (pi(j) + pj(j) + pk(j) + pl(j))
exphet = exphet + 2 * pi(j) * (pj(j) + pk(j) + pl(j))
exphet = exphet + 2 * pj(j) * (pk(j) + pl(j))
exphet = exphet + 2 * pk(j) * pl(j)
IF exphet = 0 THEN GOTO [getout1]
F(j) = 1 - (obshet/n(j))/exphet
PRINT "F = ";: PRINT USING ("###.####", F(j))
IF F(j) = 0 THEN GOTO [getout1]
IF pa(j)*pb(j) = 0 THEN GOTO [getout2]
IF F(j) = 1 - (nab(j)/n(j))/(2*pa(j)*pb(j)) THEN VARF = (pa(j)*pb(j))/(2*n(j)): GOTO
[getout3]
[getout2]
VARF = naa(j)*((1-pa(j))/(pa(j)+(1-pa(j))*F(j)))^2
VARF = VARF + nbb(j)*((1-pb(j))/(pb(j)+(1-pb(j))*F(j)))^2
VARF = VARF + ncc(j)*((1-pc(j))/(pc(j)+(1-pc(j))*F(j)))^2

```

```

VARF = VARF + ndd(j)*((1-pd(j))/(pd(j)+(1-pd(j))*F(j)))^2
VARF = VARF + nee(j)*((1-pe(j))/(pe(j)+(1-pe(j))*F(j)))^2
VARF = VARF + nff(j)*((1-pf(j))/(pf(j)+(1-pf(j))*F(j)))^2
VARF = VARF + ngg(j)*((1-pg(j))/(pg(j)+(1-pg(j))*F(j)))^2
VARF = VARF + nhh(j)*((1-ph(j))/(ph(j)+(1-ph(j))*F(j)))^2
VARF = VARF + nii(j)*((1-pi(j))/(pi(j)+(1-pi(j))*F(j)))^2
VARF = VARF + njj(j)*((1-pj(j))/(pj(j)+(1-pj(j))*F(j)))^2
VARF = VARF + nkk(j)*((1-pk(j))/(pk(j)+(1-pk(j))*F(j)))^2
VARF = VARF + nll(j)*((1-pl(j))/(pl(j)+(1-pl(j))*F(j)))^2
VARF = VARF + obshet/(1-F(j))^2
VARF = 1/VARF
[getout3]
FV = FV + F(j)/VARF
IV = IV + 1/VARF
PRINT "var(F) = "; : PRINT USING ("###.####", VARF)
PRINT "s.e.(F) = "; : PRINT USING ("###.####", SQR(VARF))
CS = n(j)*F(j)^2*(K-1)
DF = K*(K-1)/2
PRINT "chi-sq.{F=0} , d.f. = "; : PRINT USING ("##", DF); : PRINT "] = " ; : PRINT
USING ("####.####", CS)
IF CS < 0.0000000000000000001 AND INT(DF/2)<>DF/2 THEN PRINT "'P = 1.0000" : GOTO
[getout1]
GOSUB [CHISQDIS]
PRINT "'P = "; : PRINT USING ("#.####", 1 - J9 * K9 * L9)
[getout1]
PRINT
NEXT j
PRINT "'Fm = "; : PRINT USING ("###.####", FV/IV)
END
[CHISQDIS]
W = CS : V = DF
R9 = 1: FOR I9 = V TO 2 STEP -2: R9 = R9 * I9: NEXT I9
K9 = W ^ (INT((V + 1) / 2)) / EXP(W / 2) / R9
IF INT(V / 2) = V / 2 THEN 2400
J9 = SQR(2 / W / 3.141592653599)
GOTO 2500
2400 J9 = 1
2500 L9 = 1: M9 = 1
2700 V = V + 2: M9 = M9 * W / V
IF M9 < .0000001 THEN 3300
L9 = L9 + M9: GOTO 2700
3300
RETURN

```

10.4. Anexo 4

```

REM PROGRAM FILENAME FSTATISTICS_locus_01.BAS
REM Maria_Rosa + Piloes & Galvao + Sao_Pedro
INPUT "Number of simulations to be performed (n < 10000) = ";T
K = 8
DIM P9(T)
DIM D(K), H(K), R(K), N(K), P(K), X(K), FILENAME$(K + 3), H88(25)
NAME$(1) = "FIT" : NAME$(2) = "FST" : NAME$(3) = "FIS" : "
data 033, 038, 025
data 023, 043, 020
data 016, 035, 026
data 014, 041, 034
data 010, 034, 033
data 019, 031, 005
data 042, 042, 015
data 024, 023, 009
K1 = K + 1 : K2 = K + 2 : K3 = K + 3
FILENAME$ = "c:\temp\results_locus_01.txt"
OPEN FILENAME$ FOR OUTPUT AS #99
PRINT #99, "SUBSTRUCTURE ANALYSIS : ALL QUILOMBOLA POPULATIONS"
PRINT #99, ""
PRINT #99, "LOCUS # 01 : ACE - rs1799752, alleles D & I"
PRINT #99, ""
PRINT #99, "ESTIMATES BASED ON SAMPLE NUMBERS"
PRINT #99, ""
PRINT #99, "SUBPOP.    N(DD)    N(DI)    N(II)    N"
PRINT #99, "-----"
FOR I = 1 TO K
  READ D9, H9, R9
  D(I) = D9: H(I) = H9: R(I) = R9
  N(I) = D(I) + H(I) + R(I)
  N = N + N(I): D = D + D(I): H = H + H(I): R = R + R(I)
  PRINT #99, USING ("####", I);
  PRINT #99, USING ("#####", D(I)); : PRINT #99, USING ("#####", H(I));
  PRINT #99, USING ("#####", R(I)); : PRINT #99, USING ("#####", N(I))
NEXT I
PRINT #99, "-----"
PRINT #99, " total ";
PRINT #99, USING ("####", D);
PRINT #99, USING ("#####", H);
PRINT #99, USING ("#####", R);
PRINT #99, USING ("#####", N)
PRINT #99, ""
FOR I = 1 TO K
  P(I) = (2 * D(I) + H(I)) / (2 * N(I))
  F(I) = 1 - (H(I) / N(I)) / (2 * P(I) * (1 - P(I)))
  PRINT #99, "p("; : PRINT #99, USING ("##", I);
  PRINT #99, ") = "; : PRINT #99, USING ("##.####", P(I))
  PRINT #99, "F("; : PRINT #99, USING ("##", I);
  PRINT #99, ") = "; : PRINT #99, USING ("##.####", F(I))
  X(I) = N(I) / N: P = P + X(I) * P(I)
  VP = VP + X(I) * P(I) * P(I)
NEXT I
VP = VP - P * P: FIT = 1 - (H / N) / (2 * P * (1 - P))
FST = VP / (P * (1 - P)): FIS = (FIT - FST) / (1 - FST)
PRINT #99, "p      = "; : PRINT #99, USING ("##.####", P)
PRINT #99, "var(p) = "; : PRINT #99, USING ("##.####", VP)
PRINT #99, "FIT      = "; : PRINT #99, USING ("##.####", FIT)
PRINT #99, "FST      = "; : PRINT #99, USING ("##.####", FST)
PRINT #99, "FIS      = "; : PRINT #99, USING ("##.####", FIS)
PRINT #99, ""
PRINT #99, "AVERAGE ESTIMATES BASED ON ";
PRINT #99, USING ("####", T); : PRINT #99, " SIMULATIONS"
PRINT #99, ""
FOR I=1 TO K
  FILENAME$(I) = "c:\temp\POPULAT"+STR$(I)+".DAT"
NEXT I
FILENAME$(K1) = "c:\temp\DATA_Fit.DAT"
FILENAME$(K2) = "c:\temp\DATA_Fst.DAT"
FILENAME$(K3) = "c:\temp\DATA_Fis.DAT"
FOR I9 = 1 TO K
  OPEN FILENAME$(I9) FOR OUTPUT AS #I9

```

```

SSP = 0: FOR I = 1 TO 3: SP(I) = 0: NEXT I
P(1) = D(I9) / N(I9): P(2) = H(I9) / N(I9): P(3) = R(I9) / N(I9)
FOR I = 1 TO 3
    SSP = SSP + P(I): SP(I) = SP(I) + SSP
NEXT I
FOR I = 1 TO T
    D1 = 0: H1 = 0: R1 = 0
    FOR INDIV = 1 TO N(I9)
        A = RND(1)
        IF A <= SP(1) THEN D1 = D1 + 1
        IF A > SP(1) AND A <= SP(2) THEN H1 = H1 + 1
        IF A > SP(2) AND A <= SP(3) THEN R1 = R1 + 1
    NEXT INDIV
    P1 = 2 * D1 + H1: Q1 = H1 + 2 * R1
    F1 = 1 - H1 / (P1 * Q1 / (P1 + Q1))
    PRINT #I9, D1 : PRINT #I9, H1 : PRINT #I9, R1 : PRINT #I9, P1
    SUMF1 = SUMF1 + F1: SUMP1 = SUMP1 + P1 / (2 * N(I9))
NEXT I
CLOSE #I9
PRINT #99,"P("; : PRINT #99, USING ("##", I9);
PRINT #99,") = "; : PRINT #99, USING ("##.####", SUMP1 / T)
PRINT #99,"F("; : PRINT #99, USING ("##", I9);
PRINT #99,") = "; : PRINT #99, USING ("##.####", SUMF1 / T)
SUMF1 = 0: SUMP1 = 0
NEXT I9
OPEN FILENAME$(1) FOR INPUT AS #1
OPEN FILENAME$(2) FOR INPUT AS #2
OPEN FILENAME$(3) FOR INPUT AS #3
OPEN FILENAME$(4) FOR INPUT AS #4
OPEN FILENAME$(5) FOR INPUT AS #5
OPEN FILENAME$(6) FOR INPUT AS #6
OPEN FILENAME$(7) FOR INPUT AS #7
OPEN FILENAME$(8) FOR INPUT AS #8
OPEN FILENAME$(9) FOR OUTPUT AS #9
OPEN FILENAME$(10) FOR OUTPUT AS #10
OPEN FILENAME$(11) FOR OUTPUT AS #11
FOR I = 1 TO T
    FOR J = 1 TO K
        IF J=1 THEN INPUT #1, D, H, R, P
        IF J=2 THEN INPUT #2, D, H, R, P
        IF J=3 THEN INPUT #3, D, H, R, P
        IF J=4 THEN INPUT #4, D, H, R, P
        IF J=5 THEN INPUT #5, D, H, R, P
        IF J=6 THEN INPUT #6, D, H, R, P
        IF J=7 THEN INPUT #7, D, H, R, P
        IF J=8 THEN INPUT #8, D, H, R, P
        HT = HT + H: P = P / (2 * N(J))
        PT = PT + X(J) * P: VARP = VARP + X(J) * P * P
    NEXT J
    VARP = VARP - PT * PT: FIT = 1 - (HT / N) / (2 * PT * (1 - PT))
    FST = VARP / (PT * (1 - PT)): FIS = (FIT - FST) / (1 - FST)
    HT = 0: PT = 0: VARP = 0
    PRINT #9, FIT : PRINT #10, FST : PRINT #11, FIS
    SUMFIT = SUMFIT + FIT: SUMFST = SUMFST + FST: SUMFIS = SUMFIS + FIS
NEXT I
CLOSE #1
CLOSE #2
CLOSE #3
CLOSE #4
CLOSE #5
CLOSE #6
CLOSE #7
CLOSE #8
CLOSE #9
CLOSE #10
CLOSE #11
PRINT #99, "FIT      = "; : PRINT #99, USING ("##.####", SUMFIT / T)
PRINT #99, "FST      = "; : PRINT #99, USING ("##.####", SUMFST / T)
PRINT #99, "FIS      = "; : PRINT #99, USING ("##.####", SUMFIS / T)
PRINT #99, ""
PRINT #99, "MEDIAN AND 95% BOOTSTRAP CONFIDENCE INTERVALS"
PRINT #99, ""
FOR I=1 TO 11
    OPEN FILENAME$(I) FOR INPUT AS #I
    FOR J = 1 TO T
        IF I > K THEN INPUT #I, F ELSE INPUT #I, D, H, R, P
        IF I <= K THEN F = 1 - H / (P * (1 - P / (2 * N(I))))

```



```

    P9(J) = F
NEXT J
CLOSE #I
GOSUB [SORTVALUES]
T1 = T/2 : T2 = T/2 + 1
T3 = T/40 + 1 : T4 = T-T/40
IF I <= K THEN
    PRINT #99, "F("; : PRINT #99, USING ("##", I); : PRINT #99,") : ";
ELSE
    PRINT #99, NAME$(I - 8);
END IF
PRINT #99, USING ("##.###", (P9(T1) + P9(T2)) / 2); : PRINT #99, " {";
PRINT #99, USING ("##.###", P9(T3)); : PRINT #99, " ,";
PRINT #99, USING ("##.###}", P9(T4)); : PRINT #99, " }"
NEXT I
CLOSE #99
END
[ SORTVALUES ]
D88 = 10: K88 = 1: L88 = 1: R88 = T
560 REM EXCHANGE SORT IF BLOCKSIZE <= D88
    IF R88 - L88 + 1 > D88 THEN 590
    GOSUB 900: GOTO 780
590 I88 = L88: J88 = R88: W88 = RND(1) * (R88 - L88) + .5
    W88 = INT(W88 + L88): F88 = P9(W88)
    REM PARTITION PHASE
620 IF P9(I88) >= F88 THEN 640
    I88 = I88 + 1: GOTO 620
640 IF P9(J88) <= F88 THEN 660
    J88 = J88 - 1: GOTO 640
660 IF I88 > J88 THEN 680
    W88 = P9(I88): P9(I88) = P9(J88): P9(J88) = W88: I88 = I88 + 1: J88 = J88 - 1
680 IF I88 <= J88 THEN 620
    REM BOOKKEEPING PHASE
    IF J88 - L88 >= R88 - I88 THEN 740
    IF I88 >= R88 THEN 730
    H88(K88) = I88: K88 = K88 + 1: H88(K88) = R88: K88 = K88 + 1
730 R88 = J88: GOTO 770
740 IF L88 >= J88 THEN 760
    H88(K88) = L88: K88 = K88 + 1: H88(K88) = J88: K88 = K88 + 1
760 L88 = I88
770 IF L88 < R88 THEN 590
780 IF K88 <= 1 THEN 800
    K88 = K88 - 1: R88 = H88(K88): K88 = K88 - 1: L88 = H88(K88): GOTO 560
800 RETURN
900 REM EXCHANGE SORT
    FOR I88 = L88 TO R88 - 1: K77 = I88
    FOR J88 = I88 + 1 TO R88: IF P9(J88) >= P9(K77) THEN 940
    K77 = J88
940 NEXT J88
    IF K77 = I88 THEN 970
    W88 = P9(I88): P9(I88) = P9(K77): P9(K77) = W88
970 NEXT I88
RETURN

```

10.5. Anexo 5

Os livros-texto que utilizamos como referência (Crow & Kimura, 1970; Li, 1955; Spiess, 1977) apresentam o problema praticamente como foi introduzido originalmente por Wright na década de 20 (Wright, 1921a, 1921b, 1921c, 1921d, 1921e, 1922). Outros trabalhos contemporâneos ou mais modernos (por exemplo, Kimura & Crow, 1963; e Weir, 1996) reestudam o problema apenas de uma maneira mais formal sob o ponto de vista matemático. Os resultados que apresentamos a seguir foram desenvolvidos em co-autoria com o orientador sob a justificativa da necessidade de um aprimoramento nas técnicas de matemática aplicada e de computação usuais em genética de populações.

Apesar de todos os resultados finais já serem conhecidos da literatura, a seção inteira foi desenvolvida de uma maneira totalmente independente, com vários detalhes (indicados oportunamente) ainda inéditos na literatura especializada.

10.5.1. Autofecundação

Em regime de autofecundação a taxa de heterozigose decai à metade a cada geração e isso pode ser observado por meio da fórmula

$$h_{t+1} = 2pq(1-F_{t+1}) = h_t/2 = (1/2) \cdot 2pq(1-F_t) ;$$

considerando-se apenas a equação

$$2pq(1-F_{t+1}) = (1/2) \cdot 2pq(1-F_t) ,$$

obtemos imediatamente, de

$$1-F_{t+1} = (1/2) \cdot (1-F_t) , \quad (1)$$

a solução geral

$$1-F_t = (1/2)^t \cdot (1-F_0) ,$$

de modo que

$$F_t = 1 - (1/2)^t \cdot (1-F_0) .$$

Podemos, alternativamente, obter o valor de F em uma geração em função da anterior; considerando-se apenas a equação (1)

$$1-F_{t+1} = (1/2) \cdot (1-F_t) ,$$

obtemos a expressão

$$F_{t+1} = F_t/2 - 1/2 + 1 ,$$

de forma que

$$F_{t+1} = F_t/2 + 1/2 .$$

A partir das equações obtidas podemos notar que os valores de F_t tendem rapidamente a um (1) conforme o valor de t tende a infinito.

10.5.2. Casamentos exclusivos entre irmãos

A Figura 10.1 mostra claramente a complexificação crescente, com o correr das gerações, de um sistema regular de casamentos exclusivos entre irmãos. O diagrama deixa claro que o sistema equivale a um sistema de cruzamentos ao acaso que ocorrem exclusivamente dentro de unidades subpopulacionais de tamanho dois ($N = 2$).

Figura 10.1: Sistema de casamentos exclusivos entre pares de irmãos.

Os diversos valores de F_t são obtidos aplicando-se recursivamente a regra dos caminhos possíveis pelos quais um alelo pode ser transmitido a um indivíduo, a partir dos ancestrais comuns (representados pelas letras sublinhadas nas fórmulas abaixo).

$$F_0 = 0 = 0/1$$

$$F_1 = 0 = 0/2$$

$$\begin{aligned} F_2 &= P(\underline{C}\underline{A}\underline{D}) \cdot (1+F_A) + P(\underline{C}\underline{B}\underline{D}) \cdot (1+F_B) \\ &= (1/2)^3 + (1/2)^3 \\ &= 1/4 \end{aligned}$$

$$\begin{aligned} F_3 &= P(\underline{E}\underline{C}\underline{A}\underline{D}\underline{F}) \cdot (1+F_A) + P(\underline{E}\underline{D}\underline{A}\underline{C}\underline{F}) \cdot (1+F_A) + P(\underline{E}\underline{C}\underline{B}\underline{D}\underline{F}) \cdot (1+F_B) \\ &\quad + P(\underline{E}\underline{D}\underline{B}\underline{C}\underline{F}) \cdot (1+F_B) + P(\underline{E}\underline{C}\underline{F}) \cdot (1+F_C) + P(\underline{E}\underline{D}\underline{F}) \cdot (1+F_D) \\ &= (1/2)^5 + (1/2)^5 + (1/2)^5 + (1/2)^5 + (1/2)^3 + (1/2)^3 \\ &= (1/2)^3 + (1/2)^2 \\ &= 3/8 \end{aligned}$$

$$F_4 = P(\underline{G}\underline{E}\underline{C}\underline{A}\underline{D}\underline{F}\underline{H}) \cdot (1+F_A) + P(\underline{G}\underline{E}\underline{D}\underline{A}\underline{C}\underline{F}\underline{H}) \cdot (1+F_A)$$

$$\begin{aligned}
& + P(\text{HEC}\underline{\text{AD}}\text{FG}).(1+F_A) + P(\text{HED}\underline{\text{AC}}\text{FG}).(1+F_A) \\
& + P(\text{GEC}\underline{\text{BD}}\text{FH}).(1+F_B) + P(\text{GED}\underline{\text{BC}}\text{FH}).(1+F_B) \\
& + P(\text{HEC}\underline{\text{BD}}\text{FG}).(1+F_B) + P(\text{HED}\underline{\text{BC}}\text{FG}).(1+F_B) \\
& + P(\text{GEC}\underline{\text{F}}\text{H}).(1+F_C) + P(\text{HEC}\underline{\text{F}}\text{G}).(1+F_C) + P(\text{GED}\underline{\text{F}}\text{H}).(1+F_D) \\
& + P(\text{HED}\underline{\text{F}}\text{G}).(1+F_D) + P(\text{GE}\underline{\text{H}}).(1+F_E) + P(\text{GF}\underline{\text{H}}).(1+F_F) \\
= & (1/2)^7 + (1/2)^7 + (1/2)^7 + (1/2)^7 + (1/2)^7 + (1/2)^7 \\
& + (1/2)^7 + (1/2)^7 + (1/2)^5 + (1/2)^5 + (1/2)^5 + (1/2)^5 \\
& + (1/2)^3.(1+1/4) + (1/2)^3.(1+1/4) = (1/2)^4 + (1/2)^3 \\
& + (1/2)^2.(5/4) \\
= & 1/16 + 1/8 + 5/16 \\
= & 1/2 = 8/16
\end{aligned}$$

...

Os termos acima obedecem obviamente à aplicação da fórmula

$$F_t = \sum_{\{k=2,t\}} \left\{ (1/2)^k . [1+F_{t-k}] \right\} , \quad t \geq k ,$$

com condições iniciais $F_0 = 0$ e $F_1 = 0$.

Partindo-se de

$$F_0 = 0.00000000 = 0 \quad = \quad 0/1$$

$$F_1 = 0.00000000 = 0 \quad = \quad 0/2$$

e aplicando-se recursivamente a fórmula acima, obtemos sucessivamente

$$F_2 = 0.25000000 = \quad \quad \quad 1/4$$

$$F_3 = 0.37500000 = \quad \quad \quad 3/8$$

$$F_4 = 0.50000000 = 1/2 \quad = \quad 8/16$$

$$F_5 = 0.59375000 = \quad \quad \quad 19/32$$

$$F_6 = 0.67187500 = \quad \quad \quad 43/64$$

$$F_7 = 0.73437500 = 47/64 = 94/128$$

$$F_8 = 0.78515625 = 201/256$$

O programa BASIC (*Liberty Basic*) não compilado
(programa-fonte) mostrado abaixo executa todos os cálculos
mostrados acima:

```
REM PROGRAM FILENAME INBREEDING_COEFF_SIB_MATING_03.BAS
REM F(i) = S(k=2,i){(1/2)^k.[1+F(i-k)]}, i >= 2
dim F(20)
F(0) = 0 : F(1) = 0
for i = 2 to 20
  sumfi = 0
  for j = 2 to i
    sumfi = sumfi + (1/2)^j*(1+F(i-j))
  next j
  F(i) = sumfi
next i
print " '          t          F(t)"
print "-----";
print "-----"
for i = 0 to 20
  print " ' ";
  print using ("#####", i);
  print using ("#####.#####", F(i));
  print " = ";
  print using ("#####", F(i)*2^(i));
  print "/";
  print using ("#####", 2^(i))
next i
print "-----";
print "-----"
```

'	t	F(t)		

'	0	0.0000000000000000 =	0/	1
'	1	0.0000000000000000 =	0/	2
'	2	0.2500000000000000 =	1/	4
'	3	0.3750000000000000 =	3/	8
'	4	0.5000000000000000 =	8/	16
'	5	0.5937500000000000 =	19/	32
'	6	0.6718750000000000 =	43/	64
'	7	0.7343750000000000 =	94/	128
'	8	0.7851562500000000 =	201/	256
'	9	0.8261718750000000 =	423/	512
'	10	0.8593750000000000 =	880/	1024
'	11	0.8862304687500000 =	1815/	2048
'	12	0.9079589843750000 =	3719/	4096
'	13	0.9255371093750000 =	7582/	8192
'	14	0.9397583007812500 =	15397/	16384
'	15	0.9512634277343750 =	31171/	32768
'	16	0.9605712890625000 =	62952/	65536
'	17	0.9681015014648444 =	126891/	131072
'	18	0.9741935729980477 =	255379/	262144
'	19	0.9791221618652344 =	513342/	524288
'	20	0.9831094741821299 =	1030865/	1048576

Os termos acima, que evidentemente formam uma série recorrente de primeiro grau e segunda ordem, não obedecem, no entanto, à equação homogênea

$$F_{t+2} = aF_{t+1} + bF_t ,$$

indicando que provavelmente a equação de diferença é do tipo

$$F_{t+2} = aF_{t+1} + bF_t + k ;$$

Utilizando-se os valores numéricos dos cinco primeiros valores de F_t , obtemos sucessivamente

$$1/4 = a.0/2 + b.0/1 + k$$

$$3/8 = a.1/4 + b.0/2 + k$$

$$1/2 = a.3/8 + b.1/4 + k ;$$

a solução desse sistema de equações (em a , b , k) fornece os resultados $a = 1/2$, $b = 1/4$ e $k = 1/4$, de modo que a equação de diferença completa (não homogênea) é

$$F_{t+2} = aF_{t+1} + bF_t + k = (1/2).F_{t+1} + (1/4).F_t + 1/4 .$$

Se diminuirmos da unidade ambos os lados dessa equação, no entanto, obtemos

$$\begin{aligned} 1-F_{t+2} &= 3/4 - (1/2).F_{t+1} - (1/4).F_t \\ &= (1/2).(1-F_{t+1}) + (1/4).(1-F_t) , \end{aligned}$$

que é agora uma equação de diferença homogênea de primeiro grau e segunda ordem, cuja solução geral é obtida facilmente, tendo a forma

$$1-F_t = C_1 r_1^t + C_2 r_2^t , \text{ em que}$$

$$r_1 = (1+\sqrt{5})/4$$

$$r_2 = (1-\sqrt{5})/4$$

$$C_1 = (1-r_2)/(r_1-r_2)$$

$$C_2 = (r_1-1)/(r_1-r_2) .$$

Evidentemente, a solução para F_t é

$$F_t = 1 - (1-F_t) = 1 - (C_1 r_1^t + C_2 r_2^t) ;$$

O programa BASIC (dialecto *Liberty Basic*) abaixo calcula os diversos valores de F_t aplicando diretamente a fórmula

$$F_t = 1 - (C_1 r_1^t + C_2 r_2^t) :$$

```
REM PROGRAM FILENAME INBREEDING_COEFF_SIB_MATING_05.BAS
dim F(20)
r1 = (1+sqr(5))/4 : r2 = (1-sqr(5))/4
C1 = (1-r2)/(r1-r2) : C2 = (r1-1)/(r1-r2)
print " '          t          F(t)"
print "-----";
print "-----"
for i = 0 to 20
  F(i) = 1 - (C1*r1^i + C2*r2^i)
  print " ' ";
  print using ("#####", i);
  print using ("#####.#####", F(i));
  print " = ";
  print using ("#####", F(i)*2^(i));
  print "/";
  print using ("#####", 2^(i))
next i
print "-----";
print "-----"
```

'	t	F(t)		

'	0	0.0000000000000000	=	0/ 1
'	1	0.0000000000000000	=	0/ 2
'	2	0.2500000000000000	=	1/ 4
'	3	0.3750000000000000	=	3/ 8
'	4	0.5000000000000000	=	8/ 16
'	5	0.5937500000000000	=	19/ 32
'	6	0.6718750000000000	=	43/ 64
'	7	0.7343750000000000	=	94/ 128
'	8	0.7851562500000000	=	201/ 256
'	9	0.8261718750000000	=	423/ 512
'	10	0.8593750000000000	=	880/ 1024
'	11	0.8862304687500000	=	1815/ 2048
'	12	0.9079589843750000	=	3719/ 4096
'	13	0.9255371093750000	=	7582/ 8192
'	14	0.9397583007812500	=	15397/ 16384
'	15	0.9512634277343750	=	31171/ 32768
'	16	0.9605712890625000	=	62952/ 65536
'	17	0.9681015014648440	=	126891/ 131072
'	18	0.9741935729980470	=	255379/ 262144
'	19	0.9791221618652340	=	513342/ 524288
'	20	0.9831094741821290	=	1030865/ 1048576

Um método alternativo (contribuição original deste trabalho) para se obter a equação de diferença aplicável ao caso de cruzamentos exclusivos entre irmãos é obtido enumerando-se todas as alternativas possíveis de transmissão do alelo em cada geração. Assim, no caso de utilizarmos as três primeiras gerações (sendo $F_0 = 0$ e $F_1 = 0$) o valor de F_2 é obtido sem maiores dificuldades por meio da Figura 10.2 ($F_2 = 1/4$).

Figura 10.2: Representação dos caminhos possíveis de transmissão de alelo para a geração dois.

Podemos notar que o termo $(1/2)^4$ é multiplicado por quatro devido aos indivíduos da geração zero (ancestrais comuns aos genitores dos indivíduos da geração dois) possuírem somados quatro alelos no mesmo locus.

Na geração três, temos $F_3 = 1/4 + 1/8 = 3/8$ (Figura 10.3).

Figura 10.3: Representação dos caminhos possíveis de transmissão de alelo para a geração três.

O valor de F na geração seguinte é $F_4 = 1/2$, como mostra a Figura 10.4.

Figura 10.4: Representação dos caminhos possíveis de transmissão de alelo para a geração quatro.

Obtemos, assim como calculado anteriormente, os valores do coeficiente F e podemos organizá-los da seguinte maneira:

$$F_0 = 0 ;$$

$$F_1 = 0 ;$$

$$F_2 = 1/4 ;$$

$$F_3 = 3/8 = 1/4 + 1/8 ;$$

$$F_4 = 1/2 = 1/4 + 1/8 + 1/16 + 1/16 .$$

O último termo de F_4 corresponde à complicação adicional de que um dos ascendentes comuns dos genitores do indivíduo da última geração é homozigoto por origem comum com probabilidade $F = 1/4$. Como para qualquer geração acima da quarta o valor de F é a soma de $1/4 + 1/8 + 1/16 + 1/32 + \dots$, basta verificarmos, para o caso do cálculo de F na geração seguinte, quais são as situações possíveis correspondentes a essa complicação acima referida por meio da Figura 10.5.

Figura 10.5: Representação dos caminhos possíveis de transmissão de alelo para a geração cinco, nos quais um ancestral comum a ambos os genitores da geração cinco também é endocruzado.

Somando-se as probabilidades correspondentes a cada um desses cinco eventos possíveis obtemos

$$1/16 + 2.4.(1/2)^4.(1/2)^4 + 2.4.(1/2)^6.(1/2)^2$$

$$= 1/16 + 1/16 = 1/8,$$

de maneira que

$$F_5 = 1/4 + 1/8 + 1/16 + 1/32 + 1/8 = 15/32 + 1/8 = 19/32.$$

Esses elementos já nos permitem obter a equação de diferença correspondente ao caso de cruzamentos exclusivos entre irmãos, já que subtraindo F da unidade obtemos o conjunto $\{1, 1, 3/4, 5/8, 1/2, 13/32, \dots\}$.

Assim, do sistema

$$3/4 = a + b$$

$$5/8 = 3a/4 + b$$

obtemos, facilmente, $a = 1/2$ e $b = 1/4$, de modo que

$$h_{t+2} = h_{t+1}/2 + h_t/4$$

$$2pq(1-F_{t+2}) = 2pq(1-F_{t+1})/2 + 2pq(1-F_t)/4$$

$$1-F_{t+2} = (1-F_{t+1})/2 + (1-F_t)/4$$

$$1-(1-F_{t+2}) = 1 - (1-F_{t+1})/2 - (1-F_t)/4$$

e, finalmente,

$$F_{t+2} = F_{t+1}/2 + F_t/4 + 1/4 .$$

Além dessas duas maneiras alternativas, o modo de evolução dos valores de F num sistema de cruzamentos exclusivos entre irmãos pode ser obtido por meio da utilização do conceito de coeficiente de consanguinidade (g).

O coeficiente de consanguinidade de dois indivíduos é definido pela probabilidade de que dois alelos escolhidos ao acaso, um de cada indivíduo, sejam idênticos por origem comum (descendência). Portanto, o coeficiente de endocruzamento de um indivíduo possui o mesmo valor do

coeficiente de consanguinidade de seus genitores (Malécot, 1948).

Figura 10.6: Sistema de casamentos exclusivos entre pares de irmãos.

A partir da Figura 10.6 obtemos o sistema de equações

$$F_{t+2} = g_{t+1}$$

$$g_{t+2} = [(1+F_{t+1})/2]/2 + g_{t+1}/2$$

e, deste, a equação geral

$$F_{t+3} = 1/4 + F_{t+1}/4 + F_{t+2}/2 ,$$

que pode ser colocada sob a forma

$$F_{t+2} = F_{t+1}/2 + F_t/4 + 1/4 .$$

Partindo da equação de recorrência acima e subtraindo ambos os lados de 1, obtemos

$$1-F_{t+2} = 1 - (F_{t+1}/2 + F_t/4 + 1/4)$$

$$(1-F_{t+2}) = (1-F_{t+1})/2 + (1-F_t)/4$$

$$2pq(1-F_{t+2}) = 2pq[(1-F_{t+1})/2 + (1-F_t)/4]$$

$$2pq(1-F_{t+2}) = [2pq(1-F_{t+1})]/2 + [2pq(1-F_t)]/4$$

e, finalmente,

$$h_{t+2} = h_{t+1}/2 + h_t/4 .$$

Definindo

$$r = \lim_{t \rightarrow \infty} h_{t+1}/h_t ,$$

como sendo o limite da razão entre dois valores sucessivos da taxa de heterozigose quando t tende a infinito, obtemos a equação característica

$$4r^2 - 2r - 1 = 0 ,$$

cujos autovalores (raízes da equação) são $r_1 \sim 0,809$ e $r_2 \sim -0,309$.

Quando t é grande ou tende a infinito, a taxa de heterozigose decai de acordo com a razão de maior valor em módulo (raiz dominante), sendo, portanto, a taxa de heterozigose da população em uma determinada geração cerca de 80% da observada na geração anterior.

O programa a seguir, em BASIC, calcula os valores de F_t e a taxa com a qual decai a heterozigose a cada geração.

```
' program filename Sib_Mat.bas
' System of exclusive matings between full sibs
DIM h(20)
h(0) = 1 : h(1) = 1
FOR t = 2 TO 20 : h(t) = h(t-1)/2 + h(t-2)/4 : NEXT t
PRINT "   G       "; "F(t)"
PRINT "1-F(t)" ; "h(t)/h(t-1)"
PRINT "-----";
PRINT "-----"
FOR t = 0 TO 20
template1$ = "#.####"
template2$ = "###"
template3$ = "#####"
PRINT " ";
PRINT USING (template2$, t); " ";
PRINT USING (template1$, 1-h(t)); " = ";
PRINT USING (template3$, (1-h(t))*2^t); " / ";
PRINT USING (template3$, 2^t); " ";
PRINT USING (template1$, h(t)); " ";
IF t <> 0 THEN PRINT USING (template1$, h(t)/h(t-1)) ELSE PRINT "-----"
NEXT t
PRINT "-----";
PRINT "-----"
```

G	F(t)			1-F(t)	h(t)/h(t-1)
0	0.0000	=	0 /	1	1.0000
1	0.0000	=	0 /	2	1.0000
2	0.2500	=	1 /	4	0.7500
3	0.3750	=	3 /	8	0.6250
4	0.5000	=	8 /	16	0.5000
5	0.5938	=	19 /	32	0.4062
6	0.6719	=	43 /	64	0.3281
7	0.7344	=	94 /	128	0.2656
8	0.7852	=	201 /	256	0.2148
9	0.8262	=	423 /	512	0.1738
10	0.8594	=	880 /	1024	0.1406
11	0.8862	=	1815 /	2048	0.1138
12	0.9080	=	3719 /	4096	0.0920
13	0.9255	=	7582 /	8192	0.0745
14	0.9398	=	15397 /	16384	0.0602
15	0.9513	=	31171 /	32768	0.0487
16	0.9606	=	62952 /	65536	0.0394
17	0.9681	=	126891 /	131072	0.0319
18	0.9742	=	255379 /	262144	0.0258
19	0.9791	=	513342 /	524288	0.0209
20	0.9831	=	1030865 /	1048576	0.0169

Outra contribuição original consiste na aplicação recursiva, em programa de computação eletrônica, das propriedades fundamentais

(a) frequências dos genótipos AA, Aa e aa numa população qualquer com $F > 0$:

$$P(AA) = p^2 + Fpq$$

$$P(Aa) = 2pq - 2pqF$$

$$P(aa) = q^2 + Fpq;$$

(b) frequências de pares de irmãos AA.AA, AA.Aa, ..., aa.aa, tiradas diretamente de

$$\begin{aligned}
P(AA.AA) &= P(AAxAA) \cdot P(AA.AA | AxAx) \\
&\quad + P(AAxAa) \cdot P(AA.AA | AxAa) \\
&\quad + P(AaxAa) \cdot P(AA.AA | AaxAa)
\end{aligned}$$

...

e

$$(c) \quad P_{t+1}(AA) = P_t(AA \times AA) \cdot 1 + P_t(AA \times Aa) \cdot 1/2 + P_t(Aa \times Aa) \cdot 1/4$$

...

Para obter os valores numéricos de interesse, basta iterar, em programa de computação, os três conjuntos (a), (b) e (c) de equações acima. Esse método pode ser aplicado a qualquer situação de endogamia sistemática e regular, uma vez que as frequências de pares possíveis de consanguíneos (item b) podem ser obtidas sem grandes dificuldades de matrizes de probabilidade populacionais de pares de consanguíneos (Li & Sacks, 1954) ou da aplicação direta de princípios básicos de genética de populações.

Um exemplo é o código Lyberty Basic abaixo,

```
rem program filename inb_syst_01.bas
rem full sib matings
F = 0 : p = 1/2 : q = 1-p
P1 = p^2+p*q*F : P2 = 2*p*q*(1-F) : P3 = q^2+p*q*F
for i = 1 to 20
  if i = 1 then
    P11 = P1^2 : P12 = 2*P1*P2 : P13 = 2*P1*P3
    P22 = P2^2 : P23 = 2*P2*P3 : P33 = P3^2
  else
    P11 = P1111 : P12 = P1112 : P13 = P1122
    P22 = P1212 : P23 = P1222 : P33 = P2222
  end if
  P1111 = P11 * 1 + P12 * 1/4 + P13 * 0 + P22 * 1/16 + P23 * 0 + P33 * 0
  P1112 = P11 * 0 + P12 * 1/2 + P13 * 0 + P22 * 1/4 + P23 * 0 + P33 * 0
  P1122 = P11 * 0 + P12 * 0 + P13 * 0 + P22 * 1/8 + P23 * 0 + P33 * 0
  P1212 = P11 * 0 + P12 * 1/4 + P13 * 1 + P22 * 1/4 + P23 * 1/4 + P33 * 0
  P1222 = P11 * 0 + P12 * 0 + P13 * 0 + P22 * 1/4 + P23 * 1/2 + P33 * 0
  P2222 = P11 * 0 + P12 * 0 + P13 * 0 + P22 * 1/16 + P23 * 1/4 + P33 * 1
  P1 = P1111 + P1112/2 + P1212/4
  P3 = P2222 + P1222/2 + P1212/4
  P2 = 1-P1-P3
  F = 1-P2/(2*p*q)
  print using ("#.#####",1-F)
next i
```

que fornece os seguintes resultados:

0.750000000
 0.625000000
 0.500000000
 0.406250000
 0.328125000
 0.265625000
 0.214843750
 0.173828125
 0.140625000
 0.113769531
 0.092041016
 0.074462891
 0.060241699
 0.048736572
 0.039428711
 0.031898499
 0.025806427
 0.020877838
 0.016890526
 0.013664722

10.5.3. Casamentos exclusivos entre meio-irmãos

Os casamentos exclusivos entre meio-irmãos (Figura 10.7) representam a melhor forma de se evitar o endocruzamento em subunidades populacionais de tamanho três ($N = 3$).

Figura 10.7: Sistema de casamentos exclusivos entre meio-irmãos.

Nesse tipo de casamento existem, por geração, dois tipos diferentes de coeficientes de consanguinidade, um se

referindo aos pares de meio-irmãos (g) e o outro aos pares de irmãos (j).

Por meio do heredograma representado na Figura 10.7 obtemos o sistema de equações

$$F_{t+3} = g_{t+2} \quad (2)$$

$$g_{t+3} = (2/8)[(1+F_{t+2})/2] + (4/8)g_{t+2} + (2/8)j_{t+2} \quad (3)$$

$$j_{t+3} = (2/4)[(1+F_{t+2})/2] + (2/4)g_{t+2} . \quad (4)$$

Substituindo (2) em (4) e resolvendo a equação resultante, obtemos

$$j_{t+3} = 1/4 + F_{t+2}/4 + F_{t+3}/4 . \quad (5)$$

Substituindo (2) e (5) em (3) obtemos, finalmente,

$$F_{t+3} = F_{t+2}/2 + F_{t+1}/4 + F_t/16 + 3/16 .$$

Diminuindo ambos os lados dessa equação de um (1) e multiplicando-os por 2pq obtemos

$$h_{t+3} = h_{t+2}/2 + h_{t+1}/4 + h_t/16 . \quad (6)$$

O programa a seguir, em BASIC, calcula os valores de F_t e a taxa de heterozigose retida a cada geração.

```
' program filename HSib_Mat.bas
' System of exclusive matings between half-sibs
DIM h(20)
h(0) = 1 : h(1) = 1 : h(2) = 7/8
FOR t = 3 TO 20 : h(t) = h(t-1)/2 + h(t-2)/4 + h(t-3)/16 : NEXT t
PRINT "    G    "; "F(t)";
PRINT "1-F(t) "; "h(t)/h(t-1)"
PRINT "-----";
PRINT "-----"
FOR t = 0 TO 20
template1$ = "#.####"
template2$ = "##"
template3$ = "#####"
IF t <> 0 THEN D = 2^(2*t-1) ELSE D = 1
PRINT " ";
PRINT USING (template2$, t); " ";
PRINT USING (template1$, 1-h(t)); " = ";
PRINT USING (template3$, (1-h(t))*D); " / ";
PRINT USING (template3$, D); " ";
PRINT USING (template1$, h(t)); " ";
```

```

IF t <> 0 THEN PRINT USING (templatel$, h(t)/h(t-1)) ELSE PRINT "-----"
NEXT t
PRINT "-----";
PRINT "-----"

```

G	F(t)			1-F(t)	h(t)/h(t-1)
0	0.0000	=	0 /	1	1.0000
1	0.0000	=	0 /	2	1.0000
2	0.1250	=	1 /	8	0.8750
3	0.2500	=	8 /	32	0.7500
4	0.3438	=	44 /	128	0.6563
5	0.4297	=	220 /	512	0.5703
6	0.5039	=	1032 /	2048	0.4961
7	0.5684	=	4656 /	8192	0.4316
8	0.6245	=	20464 /	32768	0.3755
9	0.6733	=	88256 /	131072	0.3267
10	0.7158	=	375296 /	524288	0.2842
11	0.7528	=	1578688 /	2097152	0.2472
12	0.7849	=	6584448 /	8388608	0.2151
13	0.8129	=	27276288 /	33554432	0.1871
14	0.8372	=	112370944 /	134217728	0.1628
15	0.8584	=	460848128 /	536870912	0.1416
16	0.8768	=	1882938368 /	2147483648	0.1232
17	0.8928	=	7669365760 /	8589934592	0.1072
18	0.9068	=	31156328448 /	34359738368	0.0932
19	0.9189	=	126291677184 /	137438953472	0.0811
20	0.9294	=	510965346304 /	549755813888	0.0706

Quando t tende a infinito, o valor de F_t tende a 1 e o valor de h_t tende a zero. Podemos, portanto, obter a equação característica da equação de diferença dividindo ambos os lados da equação (6) por h_t , uma vez que r é o limite de h_{t+1}/h_t quando t tende a infinito:

$$16r^3 - 8r^2 - 4r - 1 = 0.$$

Isso equivale, também, à substituição $h_t = Cr^t$: a equação (6) toma então a forma

$$Cr^{t+3} = Cr^{t+2}/2 + Cr^{t+1}/4 + Cr^t/16;$$

dividindo-se todos os membros por Cr^t obtemos imediatamente a equação característica acima.

O auto-valor dominante dessa equação ($r \sim 0,870$) mede a taxa de heterozigose retida pela população após certo número de gerações de aplicação do sistema.

10.5.4. Casamentos exclusivos entre primos duplos em primeiro grau

O sistema de casamentos exclusivos entre primos duplos em primeiro grau (Figura 10.8) equivale a um sistema de cruzamentos ao acaso ocorrendo exclusivamente dentro de subunidades populacionais de tamanho quatro ($N = 4$).

Figura 10.8: Sistema de casamentos exclusivos entre primos duplos em primeiro grau.

Nesse tipo de casamento também existem, por geração, dois tipos diferentes de coeficiente de consanguinidade. Um deles se referindo aos pares de primos (g) e o outro se referindo aos pares de irmãos (j).

Por meio do heredograma representado na Figura 10.8 obtemos o sistema de equações

$$F_{t+3} = g_{t+2} , \quad (7)$$

$$g_{t+3} = (2/4)j_{t+2} + (2/4)g_{t+2} \quad (8)$$

$$j_{t+3} = (4/8)[(1+F_{t+2})/2] + (4/8)g_{t+2} . \quad (9)$$

Substituindo (7) em (9) e resolvendo a equação, obtemos

$$j_{t+3} = F_{t+2}/4 + F_{t+3}/2 + 1/4 ; \quad (10)$$

Substituindo (7) e (10) em (8) obtemos finalmente

$$F_{t+3} = F_{t+2}/2 + F_{t+1}/4 + F_t/8 + 1/8 .$$

Ao diminuirmos ambos os lados dessa equação da unidade obtemos, após multiplicarmos todos os membros por 2pq,

$$h_{t+3} = h_{t+2}/2 + h_{t+1}/4 + h_t/8 . \quad (11)$$

Procedendo como anteriormente, obtemos a equação característica da equação de diferença:

$$8r^3 - 4r^2 - 2r - 1 = 0 .$$

A maior raiz da equação cúbica acima (auto-valor dominante) tem valor $r \sim 0,920$ e mede a taxa de heterozigose retida na população por geração.

O programa a seguir, em BASIC, calcula os valores de F_t e a taxa de heterozigose retida a cada geração.

```
' program filename DFC_Mat.bas
' System of exclusive matings between double first cousins
DIM h(20)
h(0) = 1 : h(1) = 1 : h(2) = 1
FOR t = 3 TO 20 : h(t) = h(t-1)/2 + h(t-2)/4 + h(t-3)/8 : NEXT t
PRINT "   G       "; "F(t)"
PRINT "1-F(t)   "; "h(t)/h(t-1)"
PRINT "-----";
PRINT "-----"
FOR t = 0 TO 20
D = 2^(2*t-1)
template1$ = "#.####"
template2$ = "##"
template3$ = "#####"
PRINT "   ";
PRINT USING (template2$, t); "   ";
PRINT USING (template1$, 1-h(t)); " = ";
PRINT USING (template3$, (1-h(t))*D); " / ";
PRINT USING (template3$, D); "   ";
PRINT USING (template1$, h(t)); "   ";
IF t <> 0 THEN PRINT USING (template1$, h(t)/h(t-1)) ELSE PRINT "-----"
NEXT t
PRINT "-----";
PRINT "-----"
```

G	F(t)			1-F(t)	h(t)/h(t-1)
0	0.0000	=	0 /	1	1.0000
1	0.0000	=	0 /	2	1.0000
2	0.0000	=	0 /	8	1.0000
3	0.1250	=	4 /	32	0.8750
4	0.1875	=	24 /	128	0.8125
5	0.2500	=	128 /	512	0.7500
6	0.3125	=	640 /	2048	0.6875
7	0.3672	=	3008 /	8192	0.6328
8	0.4180	=	13696 /	32768	0.5820
9	0.4648	=	60928 /	131072	0.5352
10	0.5078	=	266240 /	524288	0.4922
11	0.5474	=	1147904 /	2097152	0.4526
12	0.5837	=	4896768 /	8388608	0.4163
13	0.6172	=	20709376 /	33554432	0.3828
14	0.6479	=	86966272 /	134217728	0.3521
15	0.6762	=	363053056 /	536870912	0.3238
16	0.7023	=	1508081664 /	2147483648	0.2977
17	0.7262	=	6237847552 /	8589934592	0.2738
18	0.7482	=	25707413504 /	34359738368	0.2518
19	0.7684	=	105610739712 /	137438953472	0.2316
20	0.7870	=	432673390592 /	549755813888	0.2130

10.5.5. Casamentos circulares entre meio-irmãos

Os casamentos circulares entre meio-irmãos (Figura 10.9) constituem uma forma menos vantajosa, ao menos nas primeiras gerações, de se evitar ao máximo o endocruzamento em subunidades populacionais de tamanho quatro ($N = 4$).

Figura 10.9: Sistema de casamentos circulares entre meio-irmãos.

Neste sistema de cruzamentos consideramos dois tipos de coeficiente de consanguinidade por geração a cada geração, um relativo aos pares de meio-irmãos (g) e outro relativo aos pares de primos em primeiro grau.

Podemos obter, a partir do heredograma representado na Figura 10.9, o sistema de equações

$$F_{t+3} = g_{t+2}$$

$$g_{t+3} = (1/4)[(1+F_{t+2})/2] + (2/4)g_{t+2} + (1/4)j_{t+2}$$

$$j_{t+3} = (2/4)j_{t+2} + (2/4)g_{t+2} .$$

Fazendo

$$h = 1-F ,$$

$$m = 1-g$$

e

$$n = 1-j ,$$

obtemos o sistema alternativo de equações

$$h_{t+3} = m_{t+2}$$

$$m_{t+3} = (1/8)h_{t+2} + (1/2)m_{t+2} + (1/4)n_{t+2}$$

$$n_{t+3} = (1/2)m_{t+2} + (1/2)n_{t+2} ;$$

Representando esse sistema sob forma matricial, temos

$$\begin{pmatrix} h_{t+3} \\ m_{t+3} \\ n_{t+3} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1/8 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{pmatrix} \cdot \begin{pmatrix} h_{t+2} \\ m_{t+2} \\ n_{t+2} \end{pmatrix}$$

A partir da matriz

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1/8 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{pmatrix} ,$$

e da equação $\det(\mathbf{A}-r\mathbf{I}) = 0$ obtemos a equação característica

$$r^3 - r^2 + 1/16 = 0 . \quad (12)$$

O autovalor dominante da equação (12) é $r \sim 0,927$.
Levando-se em conta a razão r (limite de h_{t+1}/h_t quando t tende a infinito) obtemos, finalmente, a equação de recorrência

$$h_{t+3} = h_{t+2} - h_t/16$$

a qual descreve o comportamento da taxa de heterozigose retida pela população ao longo das gerações. Dividindo ambos os lados dessa equação por $2pq$ e subtraindo-os de 1 obtemos, finalmente, a equação de recorrência

$$F_{t+3} = F_{t+2} - F_t/16 + 1/16 .$$

O programa-fonte (em BASIC) a seguir calcula os valores de F_t e a taxa de heterozigose retida a cada geração nesse sistema de cruzamentos.

```
' program filename HSCirc_Mat.bas
' System of exclusive matings between double first cousins
DIM h(20)
h(0) = 1 : h(1) = 1 : h(2) = 7/8
FOR t = 3 TO 20 : h(t) = h(t-1) - h(t-3)/16 : NEXT t
PRINT "   G       "; "F(t)"
PRINT "1-F(t)" ; "h(t)/h(t-1)"
PRINT "-----";
PRINT "-----"
FOR t = 0 TO 20
  template1$ = "#.####"
  template2$ = "###"
  template3$ = "#####"
  PRINT " ";
  PRINT USING (template2$, t); " ";
  PRINT USING (template1$, 1-h(t)); " = ";
  PRINT USING (template3$, (1-h(t))*2^(2*t-1)); " / ";
  PRINT USING (template3$, 2^(2*t-1)); " ";
  PRINT USING (template1$, h(t)); " ";
  IF t <> 0 THEN PRINT USING (template1$, h(t)/h(t-1)) ELSE PRINT "-----"
NEXT t
PRINT "-----";
PRINT "-----"
```


G	F(t)			1-F(t)	h(t)/h(t-1)
0	0.0000	=	0 /	1	1.0000
1	0.0000	=	0 /	2	1.0000
2	0.1250	=	1 /	8	0.8750
3	0.1875	=	6 /	32	0.8125
4	0.2500	=	32 /	128	0.7500
5	0.3047	=	156 /	512	0.6953
6	0.3555	=	728 /	2048	0.6445
7	0.4023	=	3296 /	8192	0.5977
8	0.4458	=	14608 /	32768	0.5542
9	0.4861	=	63712 /	131072	0.5139
10	0.5234	=	274432 /	524288	0.4766
11	0.5581	=	1170368 /	2097152	0.4419
12	0.5902	=	4950912 /	8388608	0.4098
13	0.6200	=	20803072 /	33554432	0.3800
14	0.6476	=	86919424 /	134217728	0.3524
15	0.6732	=	361428480 /	536870912	0.3268
16	0.6970	=	1496719360 /	2147483648	0.3030
17	0.7190	=	6176070656 /	8589934592	0.2810
18	0.7394	=	25406052352 /	34359738368	0.2606
19	0.7584	=	104227266560 /	137438953472	0.2416
20	0.7759	=	426564521984 /	549755813888	0.2241

10.5.6. Análise comparativa dos sistemas de casamentos exclusivos

Durante a descrição da evolução dos valores de F nos sistemas regulares de endocruzamento apresentados, pode-se notar que em todos os sistemas, conforme t aumenta, os valores de F tendem a 1 enquanto os de h tendem a zero, o que é de se esperar pois F e h são inversamente proporcionais, variando de acordo com a expressão $F = 1 - h/2pq$, em que 2pq é um fator constante em populações submetidas a sistemas de casamentos endogâmicos exclusivos.

As taxas segundo as quais a homozigose cresce e a heterozigose decresce variam proporcionalmente ao grau de proximidade biológica dos indivíduos cruzados.

O programa abaixo criado usando o pacote de rotinas matemáticas Mathematica v8.0.4.0 (© Wolfram Research, Inc.)

mostra graficamente a evolução da taxa segundo a qual os valores do coeficiente F aumentam em cada sistema.

```
(*Program Filename SysMat_alt _ 02.nb*)
S[1,0]=0; S[1,t_]:=S[1,t]= S[1,t-1]/2 + 1/2
S[2,0]=0; S[2,1]=0; S[2,t_]:=S[2,t]=S[2,t-1]/2+S[2,t-2]/4 + 1/4
S[3,0]=0; S[3,1]=0; S[3,2]= 1/8;S[3,t_]:=S[3,t]= S[3,t-1]/2+S[3,t-2]/4+S[3,t-3]/16 + 3/16
S[4,0]=0; S[4,1]=0; S[4,2]=1/8; S[4,t_]:=S[4,t]=S[4,t-1]-S[4,t-3]/16+1/16
S[5,0]=0; S[5,1]=0; S[5,2]=0;S[5,t_]:=S[5,t]=S[5,t-1]/2+S[5,t-2]/4+S[5,t-3]/8+1/8
IdLb[1] = "Selfing system (exclusive)";
IdLb[2] = "System of exclusive matings between full sibs";
IdLb[3] = "System of exclusive matings between half-sibs";
IdLb[4] = "System of exclusive circular matings between half-sibs";
IdLb[5] = "System of exclusive matings between double first cousins";
For[j=1, j<= 5, j++,
  Print[""]; Print[IdLb[j]];Print[""];
  Print[Table[S[j,i],{i,0,20}]];
  Print[PaddedForm[N[Table[S[j,i],{i,0,20}]], {6, 6}]];
  Print[ListLinePlot[Table[{i,S[j,i]},{i, 0, 30}], Frame->True,
    FrameTicks->{{True, False}, {True, False}}, PlotRange->{{0, 30}, {0, 1}}]]
];

Selfing system (exclusive)

{0,  $\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}, \frac{31}{32}, \frac{63}{64}, \frac{127}{128}, \frac{255}{256}, \frac{511}{512}, \frac{1023}{1024}, \frac{2047}{2048}, \frac{4095}{4096}, \frac{8191}{8192}, \frac{16383}{16384}, \frac{32767}{32768}, \frac{65535}{65536}, \frac{131071}{131072},$ 
 $\frac{262143}{262144}, \frac{524287}{524288}, \frac{1048575}{1048576}$ }
{ 0.000000, 0.500000, 0.750000, 0.875000, 0.937500, 0.968750, 0.984375,
0.992188, 0.996094, 0.998047, 0.999023, 0.999512, 0.999756, 0.999878,
0.999939, 0.999969, 0.999985, 0.999992, 0.999996, 0.999998, 0.999999}
```

System of exclusive matings between full sibs

```
{0, 1 3 1 19 43 47 201 423 55 1815 3719 3791 15397 31171 7869 126891
  4'8'2'32'64'64'256'512'64'2048'4096'4096'16384'32768'8192'131072'
  255379 256671 1030865
  262144'262144'1048576}
{ 0.000000, 0.000000, 0.250000, 0.375000, 0.500000, 0.593750, 0.671875,
  0.734375, 0.785156, 0.826172, 0.859375, 0.886230, 0.907959, 0.925537,
  0.939758, 0.951263, 0.960571, 0.968102, 0.974194, 0.979122, 0.983109}
```

System of exclusive matings between half-sibs

```
{0, 1 1 11 55 129 291 1279 1379 733 24667 51441 26637 438949 450047 1838807 7489615
  8'4'32'128'256'512'2048'2048'1024'32768'65536'32768'524288'524288'2097152'8388608'
  15213051 30832929 124747399
  16777216'33554432'134217728}
{ 0.000000, 0.000000, 0.125000, 0.250000, 0.343750, 0.429688, 0.503906,
  0.568359, 0.624512, 0.673340, 0.715820, 0.752777, 0.784927, 0.812897,
  0.837229, 0.858397, 0.876812, 0.892832, 0.906769, 0.918893, 0.929441}
```

System of exclusive circular matings between half-sibs

{0,0, $\frac{1}{8}, \frac{3}{16}, \frac{1}{4}, \frac{39}{128}, \frac{91}{256}, \frac{103}{256}, \frac{913}{2048}, \frac{1991}{4096}, \frac{67}{128}, \frac{18287}{32768}, \frac{38679}{65536}, \frac{40631}{65536}, \frac{339529}{524288}, \frac{705915}{1048576}, \frac{182705}{262144},$
 $\frac{6031319}{8388608}, \frac{12405299}{16777216}, \frac{12723055}{16777216}, \frac{104141729}{134217728}$ }

{ 0.000000, 0.000000, 0.125000, 0.187500, 0.250000, 0.304688, 0.355469,
 0.402344, 0.445801, 0.486084, 0.523438, 0.558075, 0.590195, 0.619980,
 0.647600, 0.673213, 0.696964, 0.718989, 0.739413, 0.758353, 0.775916 }

System of exclusive matings between double first cousins

{0,0,0, $\frac{1}{8}, \frac{3}{16}, \frac{1}{4}, \frac{5}{16}, \frac{47}{128}, \frac{107}{256}, \frac{119}{256}, \frac{65}{128}, \frac{1121}{2048}, \frac{2391}{4096}, \frac{79}{128}, \frac{1327}{2048}, \frac{22159}{32768}, \frac{46023}{65536}, \frac{47591}{65536}, \frac{49033}{65536},$
 $\frac{402873}{524288}, \frac{825259}{1048576}$ }

{ 0.000000, 0.000000, 0.000000, 0.125000, 0.187500, 0.250000, 0.312500,
 0.367188, 0.417969, 0.464844, 0.507813, 0.547363, 0.583740, 0.617188,
 0.647949, 0.676239, 0.702255, 0.726181, 0.748184, 0.768419, 0.787028 }

Para a melhor apreciação da evolução dos valores de F nos vários sistemas de endocruzamento unimos todos os gráficos por meio do código-fonte apresentado a seguir, também criado com o pacote de rotinas matemáticas Mathematica.

```
(*Program Filename AllSys.nb*)
S[1,0]=0; S[1,t_]:=S[1,t]= S[1,t-1]/2 + 1/2
S[2,0]=0; S[2,1]=0; S[2,t_]:=S[2,t]=S[2,t-1]/2+S[2,t-2]/4 + 1/4
S[3,0]=0; S[3,1]=0; S[3,2]= 1/8;S[3,t_]:=S[3,t]= S[3,t-1]/2+S[3,t-2]/4+S[3,t-3]/16 + 3/16
S[4,0]=0; S[4,1]=0; S[4,2]=1/8; S[4,t_]:=S[4,t]=S[4,t-1]-S[4,t-3]/16+1/16
S[5,0]=0; S[5,1]=0; S[5,2]=0;S[5,t_]:=S[5,t]=S[5,t-1]/2+S[5,t-2]/4+S[5,t-3]/8+1/8
For[j=1, j<= 5, j++,
  IbdGraph[j]= Table[{i,S[j,i]},{i, 0, 30}]
];
Lbl = "All Systems"
ListLinePlot[{IbdGraph [1], IbdGraph [2], IbdGraph [3], IbdGraph [4],IbdGraph
[5]}, Frame->True,
FrameTicks->{{True, False}, {True, False}}, PlotRange->{{0, 30}, {0, 1}}]
```