

**UNIVERSIDADE FEDERAL DO PAMPA
CAMPUS SÃO GABRIEL
CURSO DE ENGENHARIA FLORESTAL**

**CONSERVAÇÃO DA BIODIVERSIDADE
FLORESTAL: USO DE SIMULAÇÕES NO ESTUDO DE
DIVERSIDADE GENÉTICA**

TRABALHO DE CONCLUSÃO DE CURSO

Leonardo Severo da Costa

**São Gabriel, RS, Brasil
2011**

LEONARDO SEVERO DA COSTA

**CONSERVAÇÃO DA BIODIVERSIDADE FLORESTAL: USO DE
SIMULAÇÕES NO ESTUDO DE DIVERSIDADE GENÉTICA**

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia Florestal, da Universidade Federal do Pampa (UNIPAMPA, RS), como requisito parcial para obtenção do grau de Graduação em Engenharia Florestal.

Orientador: Prof. Dr. Valdir Marcos Stefenon

**São Gabriel
2011**

LEONARDO SEVERO DA COSTA

**CONSERVAÇÃO DA BIODIVERSIDADE FLORESTAL: USO DE
SIMULAÇÕES NO ESTUDO DE DIVERSIDADE GENÉTICA**

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia Florestal, da Universidade Federal do Pampa (UNIPAMPA, RS), como requisito parcial para obtenção do grau de Graduação em Engenharia Florestal.

Trabalho de Conclusão de Curso defendido e aprovado em: 29 de novembro de 2011.
Banca examinadora:

Prof. Dr. Valdir Marcos Stefenon
Orientador
(Unipampa)

Prof. Dr. Andres Delgado Cañedo
(Unipampa)

Prof. Dr. Rubem Samuel de Avila Jr.
(Unipampa)

AGRADECIMENTOS

Aos meus pais e minha irmã pelo constante apoio e incentivo nos momentos mais difíceis.

Aos meus amigos pela consideração e compreensão.

Ao Professor Doutor Valdir Marcos Stefenon, orientador e amigo, pelos ensinamentos transmitidos, pois o conhecimento sempre se constrói em cooperação.

A Deus, pela vida.

RESUMO

A pesquisa tratou do desenvolvimento e aplicação de simulações computacionais no estudo de parâmetros genéticos como a diversidade genética, riqueza alélica e coeficiente de endogamia em populações florestais, de modo a possibilitar o desenvolvimento de novas estratégias de manejo e conservação dos recursos genéticos florestais. O estudo utilizou como programa padrão para elaborar as simulações o software EASYPOP versão 1.0 e para análise dos dados o software FSTAT versão 2.9.3. Além dos dados simulados, foram utilizados no estudo, como parâmetro comparativo, dados amostrados a campo de três espécies do gênero *Araucária*: *A. angustifolia*, *A. nemorosa* and *A. columnaris*. Na primeira parte do trabalho foi analisada a interferência do tamanho amostral na estimativa de parâmetros genéticos em espécies florestais. Nas populações com alta diversidade genética, o índice mais influenciado pelo tamanho da amostra foi à riqueza alélica, sendo que as menores estimativas foram para os menores tamanho de amostra (50 e 100 indivíduos), o sistema de cruzamento também interferiu este índice, com valores menores para maiores taxas de endogamia, nas amostras de 50 e 100 indivíduos. O coeficiente de endogamia não foi significativamente alterado pelo tamanho da amostra quando comparado à população total. As comparações dos resultados simulados com os dados obtidos a campo também demonstraram uma tendência de aumento da riqueza alélica com o aumento do tamanho amostral. Na segunda parte do estudo, foi avaliado o comportamento da riqueza alélica e coeficiente de endogamia ao longo de sucessivas gerações em populações florestais fragmentadas. Os resultados revelaram uma susceptibilidade da riqueza alélica em todas as simulações e a perda de alelos ocorreu em todas as gerações sendo fortemente influenciada pelo tamanho da amostra. As comparações dos resultados simulados com os dados obtidos a campo corroboraram a hipótese de que populações isoladas tendem a sofrer mais drasticamente com a perda de alelos e aumento do coeficiente de endogamia. Em geral, os resultados obtidos destacam a importância da manutenção da diversidade genética bem como a recuperação de populações florestais fragmentadas através da implantação de corredores ecológicos. Os dados obtidos neste estudo podem ser usados no planejamento de programas de recuperação/enriquecimento de áreas degradadas, assim como programas de melhoramento genético baseados em populações base para coleta de sementes.

Palavras chave: diversidade genética, riqueza alélica, gerações, amostragem.

ABSTRACT

The research addressed the development and application of computer simulations in the study of genetic parameters such as genetic diversity, allelic richness and inbreeding coefficient in forest populations, to enable the development of new strategies for management and conservation of forest genetic resources. The study used as the default program to develop simulations EASYPOP version 1.0 software and data analysis software FSTAT version 2.9.3. Besides the simulated data were used in the study as comparator, the field data sampled three species of the genus *Araucaria*: *A. angustifolia*, *A. nemorosa* and *A. columnaris*. In the first part of the work was analyzed the influence of the sample size in the estimation of genetic parameters on forest species. In populations with high genetic diversity, the index more influenced by the sample size was allelic richness, and the estimates were smaller for the smaller sample size (50 and 100 individuals), the mating system also interfered with this index, with values lower for higher rates of inbreeding, in samples of 50 and 100 individuals. The inbreeding coefficient was not significantly altered by the sample size when compared to the total population. Comparisons of simulated results with field data obtained also showed a trend of increasing allelic richness with increasing sample size. In the second part of the study, we examined the behavior of allelic richness and coefficient of inbreeding over successive generations in fragmented forest populations. The results revealed a susceptibility allele of wealth in all simulations and the loss of alleles occurred in every generation is strongly influenced by sample size. Comparisons of simulated results with field data obtained corroborate the hypothesis that isolated populations tend to suffer more dramatically with the loss of alleles and increased inbreeding coefficient. In general, the results highlight the importance of maintaining genetic diversity as well as the recovery of fragmented forest populations through the establishment of ecological corridors. The data from this study can be used in planning programs for the recovery / enrichment of degraded areas, as well as breeding programs based on base populations for seed collection.

Keywords: genetic diversity, allelic richness, generations, sampling.

LISTA DE FIGURAS

Figura 1: Fronteira entre área de cultivo agrícola e área de floresta nativa no município de Rosário do Sul, Rio Grande do Sul.	14
Figura 2: Análise comparativa temporal de fragmentos florestais em uma mesma área entre 2003 (A) e 2009 (B).	15
Figura 3: Conseqüências genéticas de uma população grande e única (UG) versus vários fragmentos de populações pequenas (VP).....	18
A simulation study on the behavior of allelic richness and inbreeding coefficient over generations in fragmented populations of tree species	
Figure 1: Estimations of genetic diversity measures for four different populations size (50, 100, 500 and 1000 individuals) over 0, 5, 45, 100, 500 and 1000 generations of inbred mating. A) Estimations of the number of alleles. B) Estimations of the inbreeding coefficient.....	46

LISTA DE TABELAS

The correlation of sample size with the estimation of allelic richness and inbreeding coefficient in forest tree species

Table 1: Summary of the estimated mean values of gene diversity (H), number of alleles (A) and inbreeding coefficient (f) over 10 loci for the total population (200000 individuals) and the mean values and standard deviation for each sub-sample (mean values over 100 repetitions)..... 33

Table 2: Summary of the estimated mean values of gene diversity (H), number of alleles (A) and inbreeding coefficient (f) over 5 microsatellite loci for the bulked sample of 384 individuals and the mean values and standard deviation for each sub-sample (mean values over 10 repetitions), for field data from natural populations of *Araucaria angustifolia* (Brazilian Pine)..... 34

A simulation study on the behavior of allelic richness and inbreeding coefficient over generations in fragmented populations of tree species

Table 1: Statistical significance of pair-wise comparisons of the number of alleles (A) and coefficient of endogamy (G_{IS}) for population-pairs of different sizes over 5, 45, 100, 500 and 1000 generations, based on an ANOVA test..... 48

Table 2: Statistical significance of pair-wise comparison of the number of alleles (A) and coefficient of endogamy (G_{IS}) across generations of inbred mating over populations of different sizes, based on an ANOVA test..... 49

SUMÁRIO

1. INTRODUÇÃO GERAL	11
1.1 Conservação de populações florestais nativas.....	11
1.2 Conservação de recursos genéticos florestais.....	12
1.3 Regeneração da diversidade genética em populações florestais	13
1.4 Fragmentação florestal	14
1.5 Conseqüências genéticas da redução do tamanho efetivo populacional	16
1.6 Conseqüências genéticas da fragmentação florestal.....	17
1.7 Simulação de dados genéticos através de aplicativos computacionais.....	19
1.8 Metodologia e objetivos	21
2. REFERÊNCIAS	22
3. THE CORRELATION OF SAMPLE SIZE WITH THE ESTIMATION OF ALLELIC RICHNESS AND INBREEDING COEFFICIENT IN FOREST TREE SPECIES	27
3.1 Abstract.....	27
3.2 Introduction	28
3.3 Methods	29
3.4 Results and Discussion	31
3.5 Conclusion	35
3.6 Acknowledgements	36
3.7 References	36
4. A SIMULATION STUDY ON THE BEHAVIOR OF ALLELIC RICHNESS AND INBREEDING COEFFICIENT OVER GENERATIONS IN FRAGMENTED POPULATIONS OF TREE SPECIES	39
4.1 Abstract.....	39
4.2 Introduction	40
4.3 Methods	41
4.4 Results	43
4.5 Discussion.....	48
4.6 Conclusion.....	50
4.7 Acknowledgements	50
4.8 References	51
5. DISCUSSÃO	53

6. CONCLUSÕES FINAIS E PERSPECTIVAS FUTURAS.....	55
APÊNDICES	56
APÊNDICE A – Entrada de dados no software EASYPOP para simulação de população com cruzamentos endogâmicos.....	56
APÊNDICE B – Entrada de dados no software EASYPOP para simulação de população com cruzamentos endogâmicos, tendo coeficiente de endogamia igual a 0,5.	57
APÊNDICE C – Entrada de dados no software EASYPOP para simulação de população com cruzamentos endogâmicos, tendo coeficiente de endogamia igual a 0,8.	58
APÊNDICE D – Entrada de dados no software EASYPOP para simulação de população com cruzamentos completamente aleatórios.	59
APÊNDICE E – Entrada de dados no software EASYPOP para simulação de população com cruzamentos completamente aleatórios.	60

1. INTRODUÇÃO GERAL

1.1 Conservação de populações florestais nativas

Um dos maiores desafios deste final de século é a conservação da biodiversidade, em função do elevado nível de perturbações antrópicas dos ecossistemas naturais, sendo que uma das principais conseqüências dessas perturbações é a fragmentação de ecossistemas naturais (VIANA, 1998).

Nos países em desenvolvimento e nas regiões tropicais, muitas espécies e ecossistemas florestais estão seriamente ameaçados devido ao desmatamento descontrolado.

A exploração extrativista de florestas, muito antes de se estudar e conhecer a sua biologia vem reduzindo rapidamente as populações de muitas espécies potencialmente valiosas. Esse empobrecimento genético tem sido causado principalmente pelo avanço da fronteira agrícola, ocupações humanas inadequadas e mal planejadas e exploração florestal para fins agropecuários e/ou energéticos, o que dificulta a regeneração natural da floresta (SILVA, 1986).

Como exemplo, temos a Floresta Atlântica, que hoje apresenta menos de 5% da sua cobertura florestal original, sendo que a maioria dos seus remanescentes apresenta-se em estado degradado ou em fragmentos. A mesma situação, em menor escala, ocorre com a vegetação do Cerrado, reduzida a 30% da original. Essa situação ocorre principalmente devido à expansão das cidades, plantações e pastagens.

É fundamental encontrar o equilíbrio entre a conservação das florestas, tendo em vista a sua futura utilização, bem como a demanda atual pelos seus bens e serviços. Se a disponibilidade dessas áreas não fosse adequada, a conseqüência desse desequilíbrio seria a redução de opções de uso de recursos à sociedade humana (GRIFFITH, 1987).

1.2 Conservação de recursos genéticos florestais

Os ecossistemas tropicais em especial vêm sofrendo grandes pressões que acarretam mudanças no uso do solo, e conseqüentemente, perdas irreparáveis da diversidade biológica, a qual é normalmente dividida em três níveis hierárquicos: diversidade de ecossistemas, diversidade de espécies e diversidade genética (RODRIGUEZ, 2002).

O primeiro nível refere-se ao tipo e número de espécies, bem como as funções e processos ecológicos que ocorrem em diferentes cenários físicos. A diversidade de espécies é a variedade de organismos vivos encontrados em um determinado lugar. A diversidade genética diz respeito à variedade de informação genética encontrada entre os representantes individuais de uma espécie (ou de espécies diferentes) (STEFENON et al., 2007).

O principal dilema dos projetos de uso e conservação de recursos naturais, bem como dos recursos genéticos envolvidos, é conciliar uso econômico com a manutenção de habitats e ecossistemas viáveis à dinâmica evolutiva do conjunto de espécies que os formam (VIVAM, 2008).

Os ecossistemas florestais representam um extraordinário recurso que se comporta como reserva de diversidade genética, estabelecendo um imenso potencial para pesquisas.

No entanto, grande parte dos trabalhos em pesquisa sobre biodiversidade não abordam a importância dos recursos genéticos florestais, tampouco das potenciais implicações que a perda desses recursos pode acarretar.

A conservação dos recursos genéticos é um trabalho contínuo, a longo prazo, que exige investimentos importantes em tempo, pessoal, instalações e operacionalidade, justificáveis em função das necessidades e não do desejo ou conveniência de conservar o material (SANTOS 2001).

Os motivos para conservar, bem como a escolha das espécies que devem ser conservadas, devem ser definidos com base em critérios lógicos, científicos, econômicos e sociais.

A conservação dos recursos genéticos pode ocorrer dentro (*in situ*) e fora (*ex situ*) dos seus ambientes de origem, dependendo do que se pretenda fazer com eles (MARTINS, 1987).

Os modelos clássicos de conservação genética, *in-situ* e *ex-situ*, são amplamente empregados no Brasil, no entanto verifica-se cada vez mais, a necessidade do uso de estratégias complementares que promovam, de maneira definitiva, a conservação dos recursos genéticos florestais brasileiros. Espera-se que a conservação e uso dos recursos genéticos florestais possam, através de programas de melhoramento genético, contribuir para a diminuição das pressões predatórias, principalmente, sobre as florestas nativas.

1.3 Regeneração da diversidade genética em populações florestais

Segundo Frankham (2008) a diversidade genética é a variedade de alelos e genótipos presentes no grupo sob estudo (populações, espécies ou grupos de espécies).

A perda da diversidade genética reduz o potencial evolutivo e está também associada com a redução do sucesso reprodutivo, essa redução de variabilidade genética é conhecida como erosão genética.

A erosão genética, além de reduzir significativamente a variabilidade genética, aumenta a uniformidade de populações e as deixa mais vulneráveis a fatores adversos, o dificulta a prática de ações de conservação.

De acordo com Kageyama (1998), o papel de conservação pode ser potencializado através de: i) enriquecimento genético de matas secundárias, com a inclusão de espécies localmente extintas ou muito erodidas; ii) troca artificial de sementes ou propágulos entre fragmentos não distantes, para populações sabidamente com erosão genética; iii) aumento do tamanho de fragmentos com plantio misto de espécies nativas, a partir de sementes coletadas adequadamente quanto ao N_e ; e iv) implantar corredores de fluxo gênico, possibilitando a conectividade gênica entre fragmentos com populações pequenas.

Uma estratégia de manejo efetiva para a recuperação de populações pequenas, endocruzadas e com baixa diversidade genética é a introdução de indivíduos de outras populações, o que permite melhorar a eficiência reprodutiva e restaurar a diversidade genética.

Os indivíduos selecionados para serem inseridos em populações endocruzadas para recuperação da capacidade reprodutiva e da diversidade genética devem ser: não endocruzados (se disponíveis) ou endocruzados, mas geneticamente diferenciados em relação à população na qual estejam sendo introduzidos.

1.4 Fragmentação florestal

A fragmentação de habitats florestais é cada vez mais intensificada devido ao contínuo crescimento da população humana, acarretando um aumento na demanda por recursos naturais. Com esse avanço da civilização, grandes florestas são devastadas de modo a possibilitar o uso da terra com outras finalidades (Figura 1).

A fragmentação é decorrente do desmatamento de extratos de florestas contínuas, que isola pedaços de floresta de diferentes tamanhos, em meio a áreas perturbadas. A partir do isolamento, a periferia dessas ilhas de vegetação fica mais exposta à insolação e à modificação do regime de ventos (DELAMÔNICA, 2001).

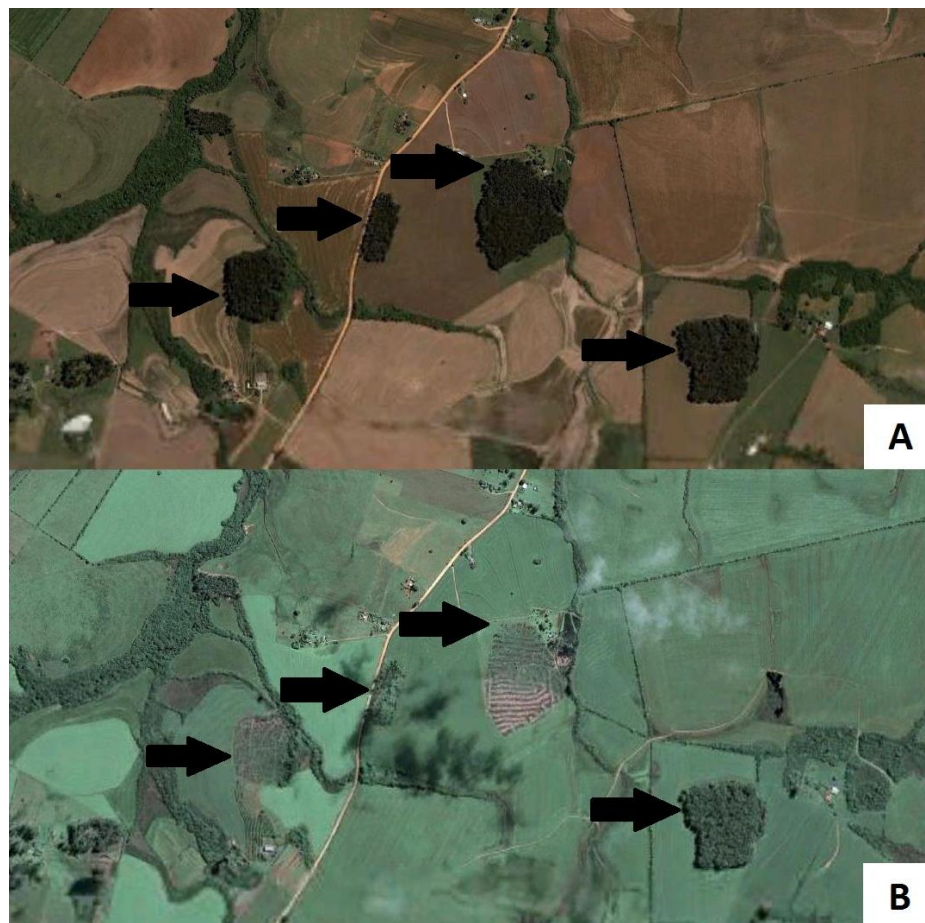
O processo de fragmentação apresenta efeito negativo sobre o ecossistema, que adquire aspectos ecológicos diferentes em relação ao ambiente original. A descontinuidade da distribuição da vegetação original reduz o habitat disponível às plantas (DAROS, 2006).

De acordo com Viana (1992), os principais fatores que comprometem a dinâmica de fragmentos florestais são: tamanho, forma, grau de isolamento, tipo de vizinhança (cidades, plantações agrícolas, espécies exóticas) e histórico de perturbações, sendo que estes apresentam ligações com fenômenos biológicos que afetam a natalidade e a mortalidade de plantas como, por exemplo, o efeito de borda, a deriva genética (perda aleatória de alelos) e as interações entre plantas e animais.

A Figura 2 apresenta um comparativo de uma mesma área rural do município de São Gabriel no ano de 2003 (A) e 2009 (B), evidenciando o processo de fragmentação florestal com a finalidade de expandir a área agrícola.



FIGURA 1 – Fronteira entre área de cultivo agrícola e área de floresta nativa no município de Rosário do Sul, Rio Grande do Sul.



Fonte: Google Earth

FIGURA 2 – Análise temporal comparativa de fragmentos florestais em uma mesma área entre 2003 (A) e 2009 (B).

A exploração predatória de espécies arbóreas, nas florestas tropicais, pode promover uma série de alterações, tais como: redução no tamanho efetivo populacional, mudanças no padrão de distribuição espacial, no comportamento de polinizadores e dispersores de sementes, na taxa de cruzamento (RAPOSO, 2007).

1.5 Conseqüências genéticas da redução do tamanho efetivo populacional

O tamanho efetivo populacional de uma população real é o número de indivíduos que, em uma população ideal, sofreria a mesma magnitude de deriva genética aleatória que a população real (HARTL, 2010).

O tamanho efetivo é sempre definido em relação a uma população de referência ideal. A população ideal consiste de uma população-base de tamanho infinito, subdividida em infinitas subpopulações de cruzamentos aleatórios, cada uma com número (N) constante de indivíduos reprodutivos por geração (SEBBENN, 2005).

Em cada subpopulação, os indivíduos produzem um infinito número de gametas masculinos e femininos dentro de um grande conjunto, do qual apenas $2N$ gametas são amostrados e unidos para produzir os N zigotos da seguinte geração.

O impacto do tamanho populacional reduzido ocorre em cada geração e as perdas acumulam-se com o tempo. Todas as conseqüências genéticas adversas do tamanho populacional pequeno dependem do tamanho efetivo populacional, ao invés do número absoluto de indivíduos, tendo em vista que o tamanho efetivo pode diferir muito do tamanho senso de uma população, sendo usualmente muito menor (WRIGHT, 1938).

Quando as populações são pequenas são mais propensas à extinção, sendo que a redução do número de indivíduos de uma população deixa a mesma vulnerável a distúrbios posteriores e uma das principais conseqüências pode ser a mudança na proporção de machos para fêmeas, o que poderia acarretar numa diminuição brusca nas taxas reprodutivas (ROCHA, 2007).

A redução do tamanho de uma população aumenta o número de indivíduos homocigotos e reduz os indivíduos heterocigotos (WRIGHT, 1921), existindo uma oscilação genética, o que acarreta a perda de alelos e variabilidade genética (causada pelo aumento de cruzamentos consanguíneos), além de reduzir o potencial adaptativo da população a distúrbios externos (PRIMACK, 2002).

1.6 Conseqüências genéticas da fragmentação florestal

A grande maioria dos estudos sobre conservação e recuperação de fragmentos fundamenta-se em conhecimentos ecológicos e demográficos, não abordando aspectos genéticos, fundamentais para a elaboração de um plano de manejo bem sucedido, tendo em vista, que um dos processos mais relevantes para a conservação de florestas são os efeitos da fragmentação sobre a diversidade genética das espécies (SEOANE, 2007).

Os impactos da fragmentação de populações sobre a diversidade genética, diferenciação, endogamia e risco de extinção, dependem diretamente do nível do fluxo gênico dentro e entre os fragmentos, sendo que este depende do(a) (FRANKHAM, 2004):

- i. Número de fragmentos populacionais;
- ii. Distribuição dos tamanhos populacionais nos fragmentos;
- iii. Distribuição geográfica das populações;
- iv. Distância entre os fragmentos;
- v. Habilidade de dispersão das espécies;
- vi. Ambiente da matriz entre os fragmentos e seu impacto sobre a dispersão;
- vii. Tempo desde a fragmentação;
- viii. Taxas de extinção e recolonização nos fragmentos.

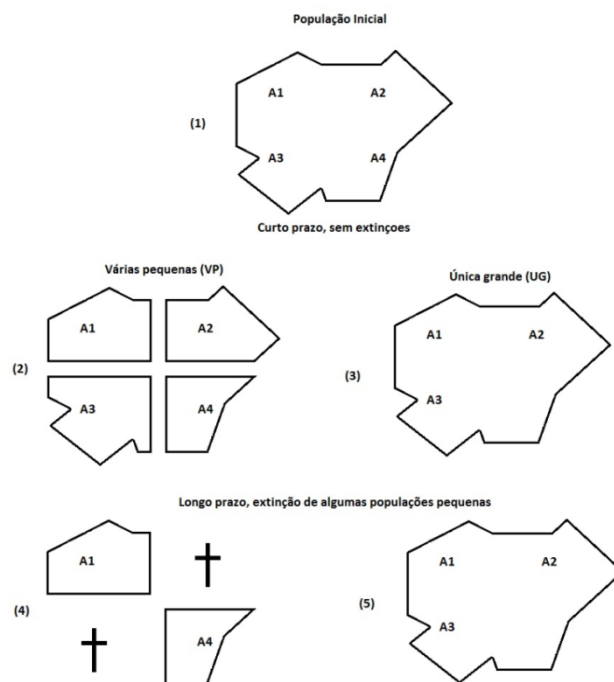
A fragmentação florestal provoca à diminuição do número de indivíduos de uma população, favorecendo a perda de variabilidade genética, com isso a população remanescente passa a ter um tamanho menor que o mínimo adequado para que a mesma possa ter sua normal continuidade e evolução (KAGEYAMA, 1998).

Juntamente com a redução do número de indivíduos de uma população, ocorre o isolamento reprodutivo dos indivíduos que contêm apenas uma pequena amostra do conjunto

gênico da população original (gargalo genético), o que acarretará contínua perda de alelos devido à deriva genética, caso a população remanescente permaneça isolada por várias gerações (SOUZA, 1997).

De acordo com Kageyama (1998), o processo de fragmentação florestal pode acarretar, em curto prazo, deriva genética, o que significa ter as frequências de seus genes afastadas daquelas da população original, inclusive chegando a perder alelos. Em longo prazo, ainda pode haver um aumento da endogamia, decorrente da maior probabilidade de autofecundação e acasalamento entre indivíduos aparentados.

Os efeitos da fragmentação são ilustrados na Figura 3 onde quatro fragmentos pequenos e isolados (VP) são comparados com uma grande população única (UG) de mesmo tamanho total inicial.



Fonte: Frankham, 2008.

FIGURA 3 - Conseqüências genéticas de uma população grande e única (UG) versus vários fragmentos de populações pequenas (VP).

A população grande (1) possui quatro alelos, A1, A2, A3 e A4. Em curto prazo as quatro populações VP (2) tornam-se rapidamente homozigotas e perdem valor adaptativo através da depressão endogâmica (redução do valor fenotípico médio de uma população devido aos acasalamentos consangüíneos). A perda de diversidade genética é mais lenta na população grande e única (3) – ela perde somente o alelo A4. No entanto, a fixação da população VP ocorre ao acaso, de modo que todos os quatro alelos são conservados, enquanto que a população UG perdeu o alelo A4. Com isso, desde que não ocorra extinção de nenhuma população VP, elas mantêm maior diversidade genética do que a população UG.

Porém, em longo prazo, quando extinções de populações pequenas ocorrem, o total das pequenas populações sobreviventes (4) reterá menos diversidade genética do que a população grande e única (5).

Logo, a caracterização dos graus de variabilidade, estrutura genética da população e o conhecimento da movimentação de alelos proporcionam as bases necessárias da execução de estratégias, para maximizar a eficiência dos programas de manejo e conservação genética.

Os parâmetros mais utilizados para estimar a diversidade genética em populações naturais e as alterações decorrentes da ação antrópica, empregando-se marcadores genéticos altamente polimórficos como os microssatélites ou SSR (“simple sequence repeats”), são: número médio de alelos por loco (A), heterozigosidade esperada ou diversidade gênica (H_e); heterozigosidade observada (H_o).

1.7 Simulação de dados genéticos através de aplicativos computacionais

Estudos genéticos em geral envolvem grande conjunto de dados, compostos por variáveis (características fenotípicas, multicategóricas, marcadores moleculares e/ou geográficas) e unidades taxonômicas (espécies, populações, cultivares, famílias, indivíduos) (FERREIRA, 2007).

Contudo, um estudo genético torna-se complexo quando são necessárias informações adicionais sobre a espécie em questão, cabendo ao pesquisador determinar qual a metodologia mais adequada para obter tais informações.

No entanto, muitas vezes esses dados apresentam-se inacessíveis, seja por uma determinada localização geográfica que impede o acesso ou por uma quantidade insuficiente de indivíduos a serem coletados, resultando uma amostragem inconsistente do ponto de vista estatístico.

Além disso, devem-se levar em consideração dois aspectos fundamentais: tempo disponível para a realização da pesquisa e sua viabilidade econômica.

Com base nesse cenário, o uso da informática passou a ser uma ferramenta indispensável na criação, processamento, armazenamento e gerenciamento de dados.

As simulações de dados genéticos nos permitem explorar situações muito complexas de serem analisadas analiticamente, permite validar inferências de estudos empíricos, executar simulações com diferentes parâmetros estimados e compará-los com dados originais além de possibilitar a construção de hipóteses.

1.8 Metodologia e objetivos

Todas as simulações foram realizadas através do software EASYPOP versão 1.0 (BALLOUX, 2001). Para a estimativa dos parâmetros de diversidade genética dos dados gerados foi utilizado o software FSTAT versão 2.9.3 (GOUDET, 1995). As análises estatísticas, análise de variância (ANOVA) e Teste-T foram utilizadas para verificar a significância da diferença das medidas de diversidade genética sobre todas as gerações e populações, usando o software StatPlus 2009.

No desenvolvimento deste trabalho foram realizadas inúmeras simulações, com diferentes parâmetros de entrada (*input*), gerando uma grande quantidade de dados a serem analisados, sendo que a partir do processamento destes os parâmetros genéticos foram estimados e interpretados.

A descrição detalhada de cada procedimento é apresentada nos itens 3.4 e 4.4, sendo que os parâmetros utilizados (*input*) encontram-se em anexo.

Objetivo geral

O objetivo central deste estudo foi avaliar o uso de simulações computacionais no estudo da diversidade genética, riqueza alélica e coeficiente de endogamia em populações florestais, visando o planejamento de estratégias de conservação da biodiversidade florestal.

Objetivos específicos

- Avaliar como o tamanho da amostra afeta o desempenho da inferência da diversidade genética de espécies florestais;
- Determinar o tamanho amostral capaz de representar o mais próximo possível a população total;
- Analisar as implicações de sucessivas gerações de cruzamentos endogâmicos em espécies florestais;
- Estimar o número de gerações necessárias para populações florestais isoladas e com elevada taxa de endogamia atingirem o equilíbrio genético.

2. REFERÊNCIAS

BALLOUX, F. EASYPOP (version 1.7): a computer program for population genetics simulations. **The Journal of Heredity**. 92: 301-302. 2001.

DAROS, L. T. **Sistema reprodutivo e estrutura genética de uma população de Imbuia (*Ocotea porosa* (Ness e Mart.) Barroso – Lauraceae)**. 2006. 78 f. Dissertação (Mestrado em Taxonomia e biologia de fanerógamas, pteridófitas e líquens) - Universidade Federal do Paraná. Curitiba. 2006.

DELAMÔNICA, P.; LAURANCE, W. F.; LAURANCE, S. G. A fragmentação da paisagem. In: OLIVEIRA, A. A.; DALY, D. C. *As florestas do Rio Negro*. Cia. das Letras/UNIP, São Paulo: Cia. das Letras, 2001. p. 285-301.

FERREIRA, F. M. **Diversidade em populações simuladas com base em locos multialélicos**. 2007. 192 f. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa. 2007.

FRANKHAM, R.; BALLOU, J. D.; BRISCOE, D. A. Conseqüências genéticas do tamanho populacional pequeno. In: **Fundamentos da Genética da Conservação**. Ribeirão Preto, SP: SBG (Sociedade Brasileira de Genética), 2008. p. 51-76.

GOOGLE MAPS. Disponível em: <<http://maps.google.com.br/maps?hl=pt-BR&tab=wl>>. Acessado em: 20 jul. 2011.

GOUDET, J. Fstat version 1.2: a computer program to calculate Fstatistics. **Journal of Heredity**. 86(6): 485-486. 1995.

GRIFFITH, J. J. Economia da conservação *in situ* de recursos genéticos florestais. **Instituto de Pesquisas e Estudos Florestais - IPEF**. n. 35, p. 85-92, abr. 1987.

HARTL, D. L.; CLARK, A. G. Deriva genética aleatória. In: **Princípios de genética de populações**. 4. ed. Porto Alegre: Artmed, 2010. p. 109-164.

KAGEYAMA, P. Y.; GANDARA, F. B.; SOUZA, L. M. I. Conseqüências genéticas da fragmentação sobre populações de espécies arbóreas. **Série Técnica IPEF**. v. 12, n. 32, p. 65-70, dez. 1998.

MARTINS, P. S. Estrutura populacional, fluxo gênico e conservação “in situ”. **Instituto de Pesquisas e Estudos Florestais - IPEF**. n. 35, p. 71-78, abr. 1987.

MORALES, E. A. V.; VALOIS, A. C. C. Princípios genéticos para recursos hídricos. In: PUIGNAU, J. P. **Conservacion de germoplasma vegetal**. Montevideo: IICA – PROCISUR, 1996. p. 35-45.

PRIMACK, B. R.; RODRIGUES, E. **Biologia da Conservação**. Londrina, Paraná: Editora Planta, 2002.

RAPOSO, A.; MARTINS, K.; CIAMPI, A. Y.; WADT, L. H. O. VEASEY, E. A. Diversidade genética de populações de andiroba no Baixo Acre. **Pesquisa Agropecuária Brasileira**. Brasília, v. 12, n. 9, p. 1291-1298, set. 2007.

ROCHA, M.; GASCA, J. Ecología molecular de la conservación. In: EGUIARTE, L. E.; SOUZA, V.; AGUIRRE, X. **Ecología molecular**. México: Instituto Nacional de Ecología, 2007. p. 251-278.

RODRIGUEZ, L. C. E. (Coord.). Ciência e Tecnologia no Setor Florestal Brasileiro: Diagnóstico, Prioridades e Modelo de Financiamento. **Instituto de Pesquisas e Estudos Florestais - IPEF**. Piracicaba, SP, 2002.

SANTOS, E.; BETTENCOURT, E. Manual de apoio à formação e treino em Conservação ex situ de Recursos Fitogenéticos. **Instituto Nacional de Investigação Agrária (INIA)**, Lisboa, Portugal e **Instituto Internacional para os Recursos Fitogenéticos (IPGRI-SSA)**, Nairobi, Quênia, 2001.

SEBBENN, A. M.; SEOANE, C. E. S. Estimativa do tamanho efetivo de endogamia por marcadores genéticos. **Revista Árvore**. v.29, n.1, p.1-7, 2005.

STEFENON VM, GAILING O AND FINKELDEY R. Genetic structure of populations in Brazil: implications for the *Araucaria angustifolia* (Araucariaceae) in situ conservation of genetic resources. **Plant Biology**. v. 9, p.516–525, jun. 2007.

SOUZA, L. M. I. Estrutura genética de populações naturais de *Chorisia speciosa* St. Hill. (Bombacaceae) em fragmentos florestais na região de Bauru (SP). **Scientia Forestalis**. n. 65, p. 70-79, jun. 2004.

SEOANE, C. E. S. Efeitos da fragmentação Florestal sobre a Genética de Populações de Guarantã. **Embrapa Florestas**. Documento 159. Out. 2007.

SEOANE, C. E. S.; KAGEYAMA, P. Y.; SEBBENN, A. M. Efeitos da fragmentação florestal na estrutura genética de populações de *Esenbeckia leiocarpa* Engl. (Guarantã). **Scientia Forestalis**. n. 57, p.123-139, jun. 2000.

SILVA, J. A.; LEITE, E. J.; GRIPP, A. Estratégia para conservação genética de espécies florestais prioritárias na Amazônia. **Acta Amazonica**. 16/17, n. único, p. 535-548, 1986/87.

VIANA, V. M.; PINHEIRO, L. A. F. V. Conservação da biodiversidade em fragmentos florestais. **Série Técnica IPEF**. v. 12, n. 32, p. 25-42, dez. 1998.

VIANA, V. M.; TABANEZ, A. A. J.; MARTINS, J. L. A. Restauração e manejo de fragmentos florestais. In: **CONGRESSO NACIONAL SOBRE ESSÊNCIAS NATIVAS**, 2, São Paulo, 1992. Anais. São Paulo: Instituto Florestal de São Paulo. p. 400-407, abr. 1992.

VIVAM, J. L. **Análise da tomada de decisão para o uso e conservação de recursos genéticos vegetais em florestas manejadas e sistemas agroflorestais**. 2008. 292 f. Tese (Doutorado em Recursos Genéticos Florestais) - Universidade Federal de Santa Catarina. Florianópolis. 2008.

WRIGHT, S. Size of population and breeding structure in relation to evolution. **Science**. v. 87, n. 2263, p. 430-2264, 1938.

WRIGHT, S. Systems of Mating II. The effects of inbreeding on the genetic composition of a population. **Genetics**, v. 6, p. 124-143, 1921.

The correlation of sample size with the estimation of allelic richness and inbreeding coefficient in forest tree species

Submetido para publicação em *Silvae Genetica*

Resumo

A aquisição de estimativas confiáveis de diversidade genética em populações naturais de espécies de árvores florestais é um fator chave no planejamento de conservação dos recursos genéticos. A fim de examinar o efeito do uso de pequenos tamanhos amostrais no desempenho da inferência de diversidade genética, estimativas de riqueza alélica e coeficiente de endogamia foram obtidas a partir de amostras de diferentes tamanhos coletadas em grandes populações simuladas e dentro de um grupo de genótipos de populações naturais de pinheiro brasileiro, utilizando dados de microssatélites. Com base nos resultados obtidos, este estudo sugere que o desenho amostral dos inventários genéticos relacionados com a conservação de recursos genéticos de espécies florestais deve levar em consideração fatores como o tamanho do censo das populações, o nível de heterozigidade e da capacidade de dispersão de pólen e sementes das espécies. Quanto à estimativa de riqueza alélica, devido à ampla distribuição dos alelos nas áreas de ocorrência, as populações com maior heterozigidade e maior dispersão de pólen e sementes terão um menor tamanho amostral para estimativas precisas, em comparação com populações com baixa heterozigidade e dispersão limitada de pólen e sementes. Com relação ao coeficiente de endogamia, pequenos tamanhos amostrais parecem gerar estimativas seguras, independentemente dos níveis de heterozigidade da população ou a capacidade de dispersão de genes.

Palavras chave: estratégia de amostragem, heterozigidade, riqueza alélica, coeficiente de endogamia, conservação de recursos genéticos.

3. THE CORRELATION OF SAMPLE SIZE WITH THE ESTIMATION OF ALLELIC RICHNESS AND INBREEDING COEFFICIENT IN FOREST TREE SPECIES

Leonardo Severo da Costa, Nathana da Silva Corneleo, Valdir Marcos Stefenon¹
Universidade Federal do Pampa, Campus São Gabriel, Av. Antonio Trilha 1847, 97300-000,
São Gabriel, RS, Brasil

3.1 Abstract

Acquiring reliable estimations of genetic diversity from natural populations of forest tree species is a key factor for planning genetic resources conservation. In order to examine the effect of using small sample sizes in the performance of genetic diversity inference, estimations of allelic richness and inbreeding coefficient were obtained from samples of different sizes collected within large simulated populations and within a bulk of genotypes from natural populations of Brazilian Pine, using microsatellite data. Based on the obtained results, this study suggests that the sampling design of genetic inventories related to genetic resources conservation of forest tree species has to take in consideration factors such as the census size of the populations, the level of heterozygosity and the capacity of pollen and seed dispersal of the species. Concerning the estimation of allelic richness, due to the widespread distribution of the alleles in the occurrence areas, populations with higher heterozygosity and larger pollen and seed dispersal will need a smaller sample size for accurate estimations, in comparison to populations with low heterozygosity and limited pollen and seed dispersal. Regarding the inbreeding coefficient, small sample sizes seems to generate secure estimations, independent of levels of heterozygosity of the population or the capacity of gene dispersal.

Key words: Sampling strategy, heterozygosity, allelic richness, inbreeding coefficient, genetic resources conservation.

¹ LSC – Realizou as simulações, analisou os dados e escreveu o artigo. NSC – auxiliou nas simulações. VMS – Planejou o experimento, auxiliou na interpretação dos dados e traduziu o artigo.

3.2 Introduction

Acquiring secure estimations of genetic diversity from natural populations of forest tree species is an essential factor for planning genetic resources conservation. However, since researchers usually work with samples of large populations, the effect of sample size in such estimations becomes a matter.

Genetic diversity is the fraction of the biodiversity related to the variety of genetic information found among individual representatives of a species (or of different species) and has been widely used as an approach for the determination of operational taxonomic units of real importance for nature conservation (Stefenon et al., 2007), for the elucidation of species evolution (Stefenon et al., 2008) and for the progress of breeding programs (Staub et al., 1996; Vendramin and Hansen 2005).

Despite the extensive use of genetic diversity data obtained from molecular marker analysis, the confidence of such information is directly correlated with an efficient study design. A biased design can generate unfair results, leading to mistakes and wrong actions when taking decisions about ecosystem conservation or landscape restoration, among others. An efficient study design includes planning the number and geographical distribution of the analyzed populations, the category and number of molecular markers to be employed and the spatial distribution and number of individuals to be sampled within each population.

The number and geographical distribution of the populations is directly linked to the general distribution of the studied species and funds allocated to develop the research. This step is usually easy to plan and do not offers major challenges. The kind of molecular marker to be employed depends mainly on the financial support owed to the research, and on the main objective of the study (Gillet et al., 1999). Usually, using a large number of highly polymorphic markers is desired. Similarly, sampling a high number of individuals is preferred in studies of populations genetic, expecting to acquire estimations which are as close to the total population as possible.

While the selection of a large number of loci is routine in population genetic studies, the sampling and analysis of an outsized number of individuals belonging to a large natural population is much complex, expensive and laborious. Therefore, determining the minimal sample size needed for reliable estimations of genetic diversity measures in large natural

populations of tree species may significantly help to plan genetic diversity inventories. While estimations of heterozygosity are less influenced by the sample size (Nei and Roychoudhury, 1974), allelic richness and inbreeding coefficient are more dependent on the number of analyzed individuals. In this study, estimations of allelic richness and inbreeding coefficient were obtained from samples of different sizes collected within large simulated populations and within a bulk of genotypes from natural populations of Brazilian Pine. Intending to contribute with the development of projects aiming genetic resource conservation, we delineated the main factors to be considered when planning sampling strategies for genetic studies in forest tree species, by comparing the estimations obtained from replicates of different sample sizes with the known values of allelic richness and inbreeding coefficient of the total populations.

3.3 Methods

Simulation of genetic data

All stochastic individual-based simulations were performed using the software EASYPOP version 1.0 (Balloux, 2001). Simulations involved populations consisting of a fixed number of 200000 individuals displaying information on diploid genotypes at 10 microsatellite loci, each reaching a maximal of 40 alleles. Simulated populations displayed different mating systems (out-crossing or selfing) and different levels of genetic diversity (high or low diversity, determined by its mean heterozygosity over the 10 microsatellite loci). The out-crossing population comprised random mating, with an effective population size $N_e = 100000$ male and $N_e = 100000$ female individuals. Selfing populations were simulated with an effective population size $N_e = 200000$ hermaphrodite individuals and three different levels of inbreeding: 10% of self pollination (low inbreeding), 50% of self pollination (medium inbreeding) and 80% of self pollination (high inbreeding). In the population with high diversity, non-linked alleles were randomly assigned and all individuals boasted the same probability of having any of the 40 alleles. For the low diversity population, all individuals started with the same allele, with 10% of recombination between adjacent loci. All simulations were run for 20 generations before data collection. In this study, the number of generations is not relevant, since the object of study is the actual level of diversity of the total

population and not the evolutionary events changing the genetic structure of the population. The stepwise mutation model (SMM; Kimura and Ohta, 1975) was chosen for the development of the genotypes, since it seems to be a good approximation of the mutation system occurring in the nature for these genetic loci (Weber and Wong, 1993). This mutation model relies on the assumption that mutations change the state of an allele by one step forward or backward with equal probability.

Genetic data of Brazilian pine

Genetic data scored at five microsatellite loci for 320 adult trees belonging to five natural populations (64 individuals from each population) of out-crossing conifer species Brazilian pine [*Araucaria angustifolia* (Bert.) O. Kuntze] were bulked in a single group. These populations formed a wide continuous forest in the recent past (about 200 years ago), before the exploitation and fragmentation of the Araucaria forest in Brazil. Considering the long life span of this species, the adult trees genotyped for this study should closely represent a real large population of Brazilian pine. Details about the geographic distribution of the populations and the microsatellite analysis were previously described by Stefenon et al. (2007).

Data analysis

Measures of allelic richness (A) and inbreeding coefficient (f) were estimated for the microsatellite genotypes using the software FSTAT version 2.9.3 (Goudet, 1995). For the simulated data set, both measures were estimated for the total population and for sub-sets of 50, 100, 500 and 1000 individuals randomly sampled among the 200000 individuals of the total populations. Each sub-set was re-sampled 99 times with replacement of individuals, totalizing 100 samplings. Mean values over the 100 estimations of each genetic diversity measure were computed for each sample size and compared with the values of the total populations. The standard deviation was used to examine the variation among samplings of the same size.

For the genetic data of Brazilian pine, the genetic diversity estimators were computed for the bulked group (320 individuals) and for sub-sets of 50, 100 and 200 individuals randomly sampled from the bulk (50 repetitions with replacement of individuals). The standard deviation was computed to determine the variation among samplings of the same size.

3.4 Results and Discussion

High diversity population

For the populations with high genetic diversity ($H = 0.97$), the allelic richness was highly influenced by the sample size, with significant lower estimations for the samples of 50 and 100 individuals, in comparison to the total populations (Table 1). For these two sample sizes, the estimations are very variable through the different samplings, with higher standard deviation for higher inbreeding rates. The mating system also influenced this measure, with lower values for higher inbreeding rates in the samples of 50 and 100 individuals (Table 1). No loss of alleles was revealed by the samplings of 500 and 1000 individuals. The estimations of the inbreeding coefficient were just slightly dependent on the sample size, while the mating system do not influenced this inference (Table 1). The standard deviation suggests low variation of the inbreeding coefficient estimations through the different samplings.

Low diversity population

In the populations with low genetic diversity ($H = 0.11$), the allelic richness of the sub-samples was smaller than in the total population for all sample sizes, being smaller for the smaller sample sizes (Table 1). For populations with low diversity, the mating system did not influence this measure, with similar values for the same sample size over different levels of inbreeding. Like for the population with high diversity, the estimations of the inbreeding coefficient were strongly influenced by the mating system, but just slightly dependent on the sample size, with low standard deviation (Table 1).

Table 1: Summary of the estimated mean values of gene diversity (H), number of alleles (A) and inbreeding coefficient (f) over 10 loci for the total population (200000 individuals) and the mean values and standard deviation for each sub-sample (mean values over 100 repetitions).

Sample size	Gene diversity measure	Population with high diversity				Population with low diversity			
		Random	Inbred0.1	Inbred0.5	Inbred0.8	Random	Inbred0.1	Inbred0.5	Inbred0.8
Total pop	H	0.975	0.975	0.975	0.975	0.11	0.11	0.11	0.11
	A	40	40	40	40	5.09	5.01	5.22	5.03
	f	0.000	0.0002	0.327	0.592	0.001	0.000	0.288	0.399
50	H	0.975 (± 0.002)	0.975 (± 0.002)	0.975 (± 0.003)	0.975 (± 0.004)	0.109 (± 0.012)	0.109 (± 0.015)	0.109 (± 0.015)	0.110 (± 0.015)
	A	36.9 (± 1.49)	36.8 (± 1.49)	35.2 (± 1.67)	33.3 (± 1.89)	3.0 (± 0.11)	2.9 (± 0.16)	2.9 (± 0.14)	2.9 (± 0.15)
	f	0.0014 (± 0.023)	0.0013 (± 0.023)	0.319 (± 0.047)	0.585 (± 0.051)	0.002 (± 0.028)	-0.004 (± 0.032)	0.225 (± 0.083)	0.362 (± 0.089)
100	H	0.975 (± 0.001)	0.975 (± 0.001)	0.975 (± 0.001)	0.975 (± 0.002)	0.110 (± 0.010)	0.109 (± 0.009)	0.111 (± 0.010)	0.108 (± 0.011)
	A	39.7 (± 0.44)	39.7 (± 0.46)	39.4 (± 0.67)	38.8 (± 1.03)	3.1 (± 0.11)	3.1 (± 0.11)	3.1 (± 0.116)	3.1 (± 0.13)
	f	0.000 (± 0.015)	0.0005 (± 0.015)	0.328 (± 0.034)	0.599 (± 0.036)	0.002 (± 0.022)	-0.003 (± 0.021)	0.280 (± 0.059)	0.365 (± 0.073)
500	H	0.975 (± 0.000)	0.975 (± 0.000)	0.975 (± 0.000)	0.975 (± 0.000)	0.109 (± 0.004)	0.109 (± 0.003)	0.111 (± 0.005)	0.110 (± 0.005)
	A	40 (± 0.00)	40 (± 0.00)	40 (± 0.00)	40 (± 0.00)	3.7 (± 0.22)	3.6 (± 0.19)	3.6 (± 0.195)	3.6 (± 0.19)
	f	0.000 (± 0.006)	0.000 (± 0.007)	0.327 (± 0.014)	0.595 (± 0.016)	0.0007 (± 0.011)	-0.0001 (± 0.010)	0.288 (± 0.027)	0.395 (± 0.029)
1000	H	0.975 (± 0.000)	0.975 (± 0.000)	0.975 (± 0.000)	0.975 (± 0.000)	0.110 (± 0.003)	0.109 (± 0.002)	0.110 (± 0.003)	0.109 (± 0.003)
	A	40 (± 0.00)	40 (± 0.00)	40 (± 0.00)	40 (± 0.00)	4.1 (± 0.21)	4.1 (± 0.22)	4.0 (± 0.203)	4.0 (± 0.21)
	f	0.000 (± 0.005)	0.000 (± 0.005)	0.327 (± 0.010)	0.593 (± 0.011)	0.0001 (± 0.008)	0.0005 (± 0.008)	0.284 (± 0.017)	0.396 (± 0.019)

Field data

For the bulked data (320 individuals, $H = 0.69$) based on five microsatellite markers the mean allelic richness equaled $A = 13.8$ and the inbreeding coefficient was $f = 0.137$ (Table 2). The values estimated for the sub-sets of 50, 100 and 200 samples are described in Table 2. The mean allelic richness increased according to the sample size, ranging from $A = 10.49$ (± 0.96) in the smaller sub-set (50 samples) to $A = 12.79$ (± 0.42) in the larger sub-set (200 samples). Similarly, the inbreeding coefficient increased from 0.127 (± 0.033) in the sub-set of 50 individuals, to 0.141 (± 0.011) in the subset of 200 individuals.

Table 2: Summary of the estimated mean values of gene diversity (H), number of alleles (A) and inbreeding coefficient (f) over 5 microsatellite loci for the bulked sample of 384 individuals and the mean values and standard deviation for each sub-sample (mean values over 10 repetitions), for field data from natural populations of *Araucaria angustifolia* (Brazilian Pine).

Gene diversity measure	Sample size				
	Bulked sample	50	100	150	200
H	0.72	0.69 (± 0.029)	0.70 (± 0.036)	0.70 (± 0.044)	0.71 (± 0.034)
A	14.6	9.88 (± 0.52)	11.32 (± 0.60)	12.34 (± 0.53)	12.86 (± 0.67)
f	0.194	0.148 (± 0.035)	0.170 (± 0.048)	0.168 (± 0.043)	0.179 (± 0.033)

Discussion

Determining the size of samples to correctly estimate genetic diversity in natural populations is a very important issue to a range of studies addressing key topics on conservation genetics such as biodiversity management, delimitation of zones for seed collection, characterization of recent evolutionary events and selection of areas of primary importance for conservation. Genetic diversity is one of the three fundamental levels of biodiversity and is very important for its conservation, since genetic variability is essential in

the preservation of species and ecosystem. Preservation of the genetic variability is imperative to the overall health of populations because decreased genetic variability tends to increase the levels of inbreeding, reducing population fitness.

As expected, this study shows that inbreeding coefficient is strongly influenced by the mating system of the species. However, the sample size seems to have really low influence in the estimation of this genetic index. On the other hand, estimating allelic richness is primarily determined by the analyzed sample size, but also strongly influenced by the heterozygosity level of the population.

According to the simulated data, populations with low diversity ($H = 0.11$ in our simulation) tend to need larger sample sizes to better estimate the real allelic richness. This is related with the higher amount of homozygote individuals which will be sampled, representing just a fraction of the alleles. Populations with higher diversity ($H = 0.97$ in our simulation) will present higher amount of heterozygote individuals, which better represent the alleles present in the population, allowing to reach an accurate estimation with a relatively smaller sample size.

Also for the field data from Brazilian pine populations ($H = 0.69$), the estimations obtained for allelic richness increased with the increase of the sample size tending to reach the value of the bulked population. The estimations of inbreeding coefficient also increased, reaching, in the sub-samples of 100 and 200 individuals, values higher than the value of the bulked population. Given that the standard deviation of the estimations overlapped the real values of the bulked population, this behavior can be resulted from Wahlund effect. The five populations of Brazilian pine bulked in this study express significant genetic differentiation, in a pattern consistent with the isolation-by-distance model (Stefenon et al., 2007), besides to spatial genetic structure within populations, as effect of limited gene flow (Stefenon et al., 2009). Given that the existence of spatial genetic structure in a population can affect the estimations of allelic richness (Bashalkhanov et al., 2009), the results obtained in the present study suggests the need of particular care in analysis of very widespread populations of species with limited gene flow, which can reveal genetically differentiated sub-populations in continuous large forests. In such cases, the sub-populations should be analyzed individually, determining the number of individuals to be collected based on the numerical census of the population, covering homogeneously the area of occurrence of the species in a more extensive sample effort.

In conservation of crop genetic resources the focus is to save alleles that may be of value for future use in breeding. Concerning the conservation of forest tree genetic resources, the main objective is to save the potential for adaptation (Eriksson, 2005). Following this reasoning, some rare alleles could be neglected, considering that their low frequency is caused by poor adaptation to the present environment (leading them to be eradicated by selection) or because they recently emerged in the population as mutants that are in threat of loss by drift (Gillet et al., 2007). So, the loss of some of such alleles can be considered irrelevant. This was the case of the estimations of the number of alleles in the population with low diversity, which presented alleles with very low frequency ($p < 10^{-4}$, data not shown) after the 20 generations simulated. Capturing such alleles in a large population would require sample sizes close to the entire number of individuals in the population (Bashalkhanov et al., 2009).

Conservation of genetic diversity is crucial for the maintenance of forest ecosystems and a reliable estimation of the genetic diversity parameters is the source of information needed to choose the individuals, populations or communities of particular importance. The number of sampled individuals should be carefully considered when planning genetic studies, in order to estimate the adaptive potential of a larger number of populations, based on a minimal sample within each one, capable to guarantee the confidence of the estimations. Differences in sample sizes through the studied populations will not introduce bias in the comparisons of inbreeding coefficient, while the use of the rarefaction strategy may solve this problem for comparisons of the number of alleles (Petit et al., 1998).

3.5 Conclusion

As a general conclusion, this study suggests that the sampling design of genetic inventories related to genetic resources conservation of forest tree species has to take in consideration factors such as the census size of the populations, the level of heterozygosity and the capacity of pollen and seed dispersal of the species. Concerning the estimation of allelic richness, due to the widespread distribution of the alleles in the occurrence areas, populations with higher heterozygosity and larger pollen and seed dispersal will need a smaller sample size for accurate estimations, in comparison to populations with low heterozygosity and limited pollen and seed dispersal. Regarding the inbreeding coefficient, small sample sizes seems to generate reliable estimations, considering that the standard

deviation of the estimations overlapped the real values for all sample sizes analyzed, for the simulated and the bulked populations.

3.6 Acknowledgements

This research was made possible by the financial support from FAPERGS (Process n. 0902170). LSC and NSC were sponsored by UNIPAMPA (PBDA scholarships) and VMS receives a grant from CNPq (Process n. 314664/2009-2).

3.7 References

- Balloux, F. (2001) EASYPOP (version 1.7): a computer program for population genetics simulations. *The Journal of Heredity* 92:301-302.
- Bashalkhanov, S.; Pandey, M. and Rojora, O. (2009) A simple method for estimating genetic diversity in large populations from finite sample sizes. *BMC Genetics* 10:84 doi:10.1186/1471-2156-10-84
- Eriksson, G. (2005) Evolution and evolutionary factors, adaptation and adaptability. In: Geburek, Th. and Turok, J. (eds.) *Conservation and Management of Forest Genetic Resources in Europe*. Arbora Publishers, Zvolen. pp. 199-211.
- Goudet, J. (1995). Fstat version 1.2: a computer program to calculate Fstatistics. *Journal of Heredity*. 86(6): 485-486.
- Gillet, E.M. (1999) Which DNA Marker for Which Purpose? Final Compendium of the Research Project Development, optimization and validation of molecular tools for assessment of biodiversity in forest trees in the European Union DGXII Biotechnology FW IV Research Program Molecular Tools for Biodiversity. URL <http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm>

Kimura, M. and T. Ohta (1975) Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. *Proceedings of the National Academy of Science of the USA* 72:2761-2764.

Nei, M. and A.K. Roychoudhury (1974) Sampling variances of heterozygosity and genetic distance. *Genetics* 76: 379-390.

Petit, R.J., A. El Mousadik and O. Pons (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* 12:844–855.

Staub, J.E., Serquen, F.C. and Gupta, M. (1996) Genetic markers, map construction, and their application in plant breeding. *HortScience* 31:729-741.

Stefenon, V.M., O. Gailing and R. Finkeldey (2007) Genetic structure of *Araucaria angustifolia* (Araucariaceae) in Brazil: Implications for the *in situ* conservation of genetic resources. *Plant Biology* 9:516-525.

Stefenon, V.M., H. Behling, O. Gailing and R. Finkeldey (2008) Evidences of delayed size recovery in *Araucaria angustifolia* populations after post-glacial colonization of highlands in Southeastern Brazil. *Anais da Academia Brasileira de Ciências* 80:433-443.

Vendramin, G.G. and Hansen, D.B. (2005) Molecular markers for characterizing diversity in forest trees. In: Geburek, Th. and Turok, J. (eds.) *Conservation and Management of Forest Genetic Resources in Europe*. Arbora Publishers, Zvolen. pp. 337-368.

Weber, J. and C. Wong (1993) Mutation of human short tandem repeats. *Human Molecular Genetics* 2:1123–1128.

A simulation study on the behavior of allelic richness and inbreeding coefficient over generations in fragmented populations of tree species

Aceito para publicação em *Annals of Forest Research*

Resumo

As simulações computacionais foram empregadas nesse estudo visando compreender os efeitos de sucessivos ciclos de cruzamentos endogâmicos em populações isoladas de espécies arbóreas com diferentes tamanhos efetivos até 1000 gerações. Os resultados revelaram uma susceptibilidade de riqueza alélica tanto para o tamanho da população quanto para o número de gerações de cruzamentos e um baixo crescimento do coeficiente de endogamia ao longo de gerações em populações com 50 e 100 plantas, com exceção das populações com 500 e 1000 indivíduos. A perda de alelos ocorreu por todas as gerações e foi fortemente influenciada pelo tamanho da população. O aumento mais notável no coeficiente de endogamia ocorreu a partir da geração inicial para a 5ª geração, independente do tamanho populacional. A comparação dos resultados simulados com os dados obtidos a campo corroboraram a hipótese de que populações isoladas tendem a sofrer mais drasticamente com a perda de alelos e aumento do coeficiente de endogamia, enquanto as florestas contínuas com produção efetiva de sementes férteis e regeneração de mudas estão inclinados a preservar comparativamente maior riqueza alélica e menor coeficiente de consanguinidade ao longo de gerações. Em geral, os resultados obtidos destacam a importância de um cuidado especial na seleção e preservando populações isoladas, para a conservação de recursos genéticos florestais e preservação adaptatedness.

Palavras chave: adaptabilidade, riqueza alélica, G_{IS} , heterozigosidade, cruzamento endogâmico, coeficiente de endogamia.

4. A SIMULATION STUDY ON THE BEHAVIOR OF ALLELIC RICHNESS AND INBREEDING COEFFICIENT OVER GENERATIONS IN FRAGMENTED POPULATIONS OF TREE SPECIES

Leonardo Severo da Costa, Valdir Marcos Stefenon²

Universidade Federal do Pampa, Campus São Gabriel, Av. Antonio Trilha 1847, 97300-000, São Gabriel, RS, Brasil

4.1 Abstract

Computer simulations were employed in this study aiming to understand the effects of repeated cycles of inbred mating in isolated populations of tree species with different effective sizes and over up to 1000 generations. The results revealed a susceptibility of allelic richness to both, population size and repeated generations under inbred mating and a low but significant increase of the inbreeding coefficient over generations in populations with 50 and 100 plants, but not in populations with 500 and 1000 individuals. The loss of alleles occurred throughout all generations and was largely influenced by the population size. The most outstanding increase in the inbreeding coefficient occurred from the initial generation to the 5th generation, independent of the population size. The comparison of simulated results with data obtained from a field studies corroborated the hypothesis that isolated populations tend to more drastically suffer with loss of alleles and increase of inbreeding coefficient, while continuous forests with effective production of fertile seeds and regeneration of seedlings are inclined to preserve comparatively higher allelic richness and lower inbreeding coefficient over generations. In general, the results obtained highlight the importance of special care in selecting ESUs and preserving isolated populations, towards the conservation of forest genetic resources and adaptedness preservation.

Key words: adaptability, allelic richness, G_{IS} , heterozygosity, inbred mating, inbreeding coefficient.

² LSC – Planejou o experimento, realizou as simulações, analisou os dados e escreveu o artigo. VMS – Traduziu o artigo e auxiliou na interpretação dos dados.

4.2 Introduction

The long life span and the sessile nature of tree species imply that they need to survive under different environmental conditions during their lifetime. Consequently, adaptation of trees to the environment is ubiquitous and required for survival (Finkeldey and Hattermer 2007). Adaptation is directly related to the species adaptedness – the degree to which an organism is able to live and reproduce in a specific environment (Erikson 2005) – promptly responding to environmental inconsistencies. In times of emergent climatic instability, such adaptedness is even more important and maintenance of genetic diversity is essential to uphold such competence, towards conservation of species genetic resources.

As result of forest fragmentation, large continuous forests have been converted to isolated fragments of different effective population sizes, becoming a dilemma to conservation geneticists. Conservation of genetic resources usually has been centered in determining evolutionary significant units (ESUs), which may be defined as populations reproductively separate from other populations and having unique or different adaptations (Waples, 1991) or as populations that show significant divergence of allele frequencies at nuclear loci in comparison to other populations (Moritz, 1994). Operationally, efforts to determine ESUs have emphasized reproductive isolation of the populations (Crandal et al., 2000). Usually, isolated populations tend to present high genetic divergence of evolutionary lineages by local adaptation, in addition to high levels of inbreeding and lower allelic richness (Prakash et al. 1969; Newmann and Talmonn 2001; Stefenon et al. 2007).

Concerning conservation of forest tree genetic resources, the effects of such isolation in the levels of inbreeding and allelic richness over generations of such long-lived species is still a query.

Studies on population genetics of forest tree species usually report data from a single generation and are the only source of information towards planning management, conservation and restoration of forest environments. Given the long life span of tree species, a field survey aiming to understand the effect of repeated generations of inbred mating inside completely isolated populations is impracticable since many hundred years will be necessary in order to study the population during tens of generations. In this study, we report the results of a simulation-based survey employed to explore the levels of allelic richness and inbreeding coefficient in completely isolated populations of different sizes over several generations of inbred mating. These simulations were employed to test the hypothesis that isolated populations tend to more drastically suffer with loss of alleles and increase of inbreeding coefficient, while continuous forests with effective production of fertile seeds and regeneration of seedlings are inclined to preserve comparatively higher allelic richness and lower inbreeding coefficient over generations. In order to test to which extent the simulated data fits the events occurring in real populations, we evaluated the correspondence of the simulated data with microsatellite records reported from two generations of three different conifer species.

4.3 Methods

This study relied on individual-based simulations of microsatellite genotypes of mature hermaphrodite diploid individuals, belonging to populations of four different effective sizes: 50, 100, 500 and 1000 individuals. These populations were simulated so that they replicate the behavior of completely isolated populations, lacking any level of gene influx. Data on number of alleles (A) and inbreeding coefficient (G_{IS}) were recorded for each population after 5, 45, 100, 500 and 1000 generations of mating among individuals, with a

proportion of 50% of self pollination and 50% out-crossing. A preliminary round of five generations was run in order to obtain the initial data on A and G_{IS} (generation 0). In order to avoid stochastic variations of the simulation process, all simulations were repeated 1000 times and the mean value over these simulations are reported. The most common assumption employed in genetic models was assumed for these simulations: same mutation rate over all individuals and generations, non-overlapping generations, same mutation model and constant population size over generations. Simulations were performed using the software EasyPop version 1.0 (Balloux 2001). Individuals were simulated in order to display information on diploid genotypes at 20 microsatellite loci, each reaching a maximal of 80 alleles. The stepwise mutation model (SMM; Kimura and Ohta 1975) was chosen for the evolutionary development of the genotypes. The mutation rate (mutations/locus/generation) was set to $\mu = 0.003$. Given that tree species are usually characterized by high levels of within population genetic diversity and relatively low levels of differentiation among populations (Loveless and Hamrick 1984, Finkeldey and Hattermer 2007), all simulations started from a population with maximal diversity (all individuals having the same probability of display any of the 80 alleles). Estimations of number of alleles (direct count) and inbreeding coefficient ($G_{IS} = 1 - H_o / H_S$, where H_o is the observed heterozygosity and H_S is the expected heterozygosity under Hardy-Weinberg equilibrium) for each simulated population were obtained using the software Fstat (Goudet 1995) version 2.9.3. The statistical significance of the difference of genetic diversity measures over generations and over populations were tested through an analysis of variance (ANOVA), using the software StatPlus 2009.

The matching of simulated and empirical data was obtained by comparing the computer-generated data with number of alleles and inbreeding coefficient recorded for microsatellite markers in two generations (adults and juveniles) of three species of the genus *Araucaria*: *A. angustifolia* (Sant'Anna, 2011), *A. nemorosa* and *A. columnaris* (Kettle et al.,

2007). The statistical significance of the differences obtained across generations for number of alleles and inbreeding coefficient were determined using an uni-modal *t*-test.

4.4 Results

The recorded number of alleles was lower in smaller populations and reduced significantly over generations of inbred mating (Figure 1a). The reduction in number of alleles from the initial to the 1000th generation was about 5-fold for the populations with 50 and with 100 individuals, 3-fold for the populations with 500 individuals and 2-fold for the populations with 1000 individuals, confirming that the effective population size is very important for the maintenance of the allelic richness in isolated populations, which will develop under repeated generations of inbred mating. In all pair-wise comparisons of populations with different size for the same number of generations (Table 1), the differences were statistically significant at 1% level, according to the ANOVA test ($p < 0.001$). In the pair-wise comparisons of generations for each population size (Table 2), just the pair 500 vs 1000 generations revealed a non-significant difference ($p = 0.145$) for the population with 50 individuals. All the other pairs revealed a significant difference in number of alleles ($p < 0.001$).

The inbreeding coefficient (G_{IS}), measured as the loss of heterozygote individuals in relation to expected under Hardy-Weinberg equilibrium, was slightly higher in larger populations and increased over generations. The enhancement over generations was proportional in all populations, independent of the effective population size, but with an outstanding increase in the five initial generations and a relative constancy over the following generations (Figure 1b). Despite the overall small increase, the pair-wise comparisons of populations with different size for the same number of generations (Table 1) revealed a

statistically significant difference for all pair of populations ($p < 0.014$), except for the pair 100 vs 500 individuals after 100 generations ($p = 0.154$) and for the pair 500 vs 1000 individuals ($p > 0.459$), independent of the number of generations. Concerning the pair-wise comparisons of generations for each population size (Table 2), all populations differed ($p < 0.001$) from the generation 0 (0 vs 5, 0 vs 45, 0 vs 100, 0 vs 500, 0 vs 1000). The pairs 5 vs 45 and 5 vs 100 generations for populations with 1000 individuals also revealed a significant difference ($p < 0.05$) in the measure of inbreeding coefficient. The non-significant difference in the measure of inbreeding coefficient in all other pair-wise comparisons suggests a relative stability of the levels of inbreeding, mainly after ten generations, and high influence of the effective population size over this index.

The data from two out of the three field studies matched the behavior revealed by our simulation study. The records from *A. angustifolia* (Sant'Anna, 2011) revealed an increase in the inbreeding coefficient from the first to the second generation ($G_{IS} = 0.096$ and $G_{IS} = 0.163$, respectively; uni-modal t-test $p = 0.11$) and a decrease in the mean number of alleles over generations ($A = 9.33$ for the first generation and $A = 8.67$ for the second generation; uni-modal t-test $p = 0.05$).

Similarly, the data from *A. nemorosa* expressed the same patterns for inbreeding coefficient ($G_{IS} = 0.108$ for the first generation and $G_{IS} = 0.192$ for the second generation; uni-modal t-test $p = 0.008$) and for the mean number of alleles ($A = 7.45$ for the first generation and $A = 6.98$ for the second generation; uni-modal t-test $p = 0.02$). The data from *A. columnaris*, on the other hand, expressed a decrease level of inbreeding coefficient over generations ($G_{IS} = 0.123$ for the first generation and $G_{IS} = 0.093$ for the second generation; uni-modal t-test $p = 0.45$), while the mean number of alleles did not differ ($A = 6.0$ for both generations; uni-modal t-test $p = 0.011$).

Figure 1: Estimations of genetic diversity measures for four different populations size (50, 100, 500 and 1000 individuals) over 0, 5, 45, 100, 500 and 1000 generations of inbred mating.

A) Estimations of the number of alleles. B) Estimations of the inbreeding coefficient.

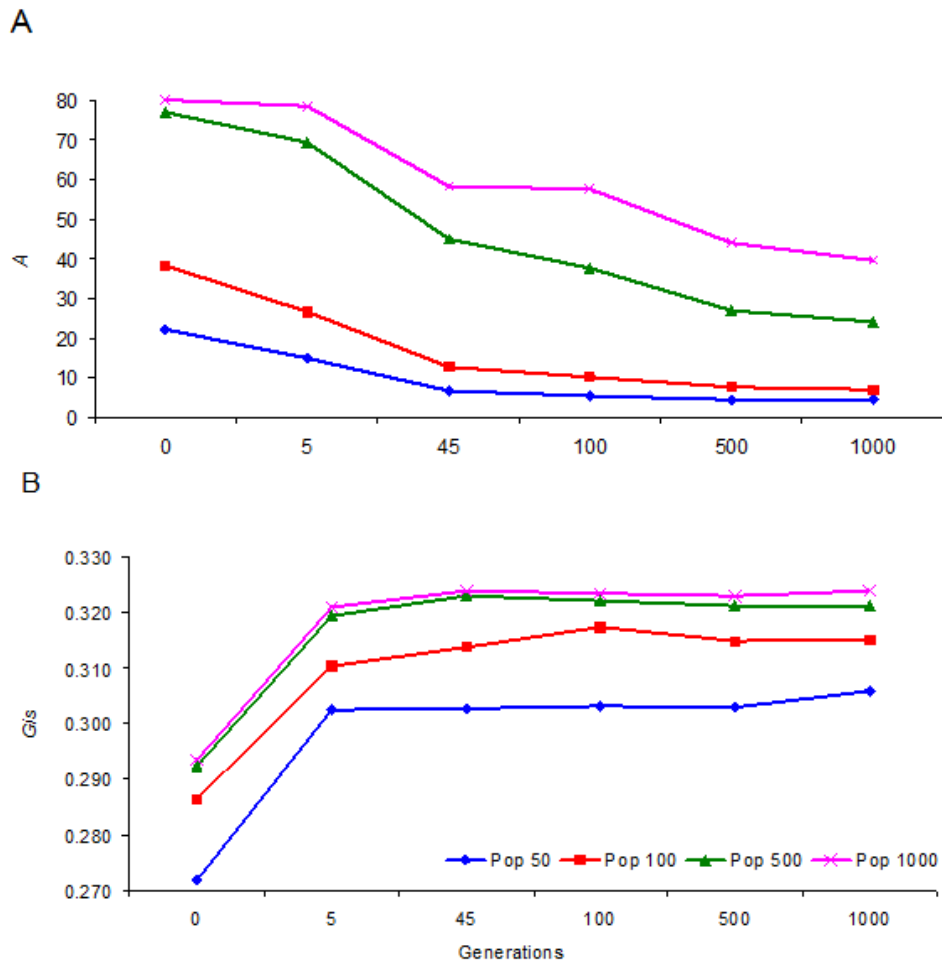


Table 1: Statistical significance of pair-wise comparisons of the number of alleles (A) and coefficient of endogamy (G_{IS}) for population-pairs of different sizes over 5, 45, 100, 500 and 1000 generations, based on an ANOVA test.

Population Size	A (generations)					G_{IS} (generations)				
	5	45	100	500	1000	5	45	100	500	1000
50 vs 100	***	***	***	***	***	***	**	***	***	***
50 vs 500	***	***	***	***	***	***	***	***	***	***
50 vs 1000	***	***	***	***	***	***	***	***	***	***
100 vs 500	***	***	***	***	***	**	**	**	*	**
100 vs 1000	***	***	***	***	***	**	***	***	***	***
500 vs 1000	***	***	***	***	***	n.s.	n.s.	n.s.	n.s.	n.s.

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, n.s.: not significant

Table 2: Statistical significance of pair-wise comparison of the number of alleles (A) and coefficient of endogamy (G_{IS}) across generations of inbred mating over populations of different sizes, based on an ANOVA test.

Number of Generations	A (population size)				G_{IS} (population size)			
	50	100	500	1000	50	100	500	1000
0 vs 5	***	***	***	***	***	***	***	***
0 vs 45	***	***	***	***	***	***	***	***
0 vs 100	***	***	***	***	***	***	***	***
0 vs 500	***	***	***	***	***	***	***	***
0 vs 1000	***	***	***	***	***	***	***	***
5 vs 45	***	***	***	***	n.s.	n.s.	n.s.	**
5 vs 100	***	***	***	***	n.s.	n.s.	n.s.	*
5 vs 500	***	***	***	***	n.s.	n.s.	n.s.	n.s.
5 vs 1000	***	***	***	***	n.s.	n.s.	n.s.	n.s.
45 vs 100	***	***	***	***	n.s.	n.s.	n.s.	n.s.
45 vs 500	***	***	***	***	n.s.	n.s.	n.s.	n.s.
45 vs 1000	***	***	***	***	n.s.	n.s.	n.s.	n.s.
100 vs 500	***	***	***	***	n.s.	n.s.	n.s.	n.s.
100 vs 1000	***	***	***	***	n.s.	n.s.	n.s.	n.s.
500 vs 1000	n.s.	***	***	***	n.s.	n.s.	n.s.	n.s.

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, n.s.: not significant

4.5 Discussion

Even if the simulations employed in this study fail to cover all the possible scenarios concerning mating systems and ecological issues, they bring out important considerations related to the behavior of allelic richness and inbreeding coefficient over generations in isolated populations of tree species.

The larger decrease in the number of alleles over generations for smaller populations (50 and 100 individuals) under inbred mating observed in this study is not surprising, but exposes the need of special attention for small forests, which usually display low number of alleles. Low number of alleles can retain high heterozygosity without maintaining allelic richness. However, allelic richness is essential for safeguarding populations' adaptedness, that is vital for forest tree populations which are constantly exposed to different evolutionary factors, changing frequencies of alleles and/or genotypes (Erikson 2005).

Due to their sessile nature, the production of inbred offspring in plant populations is likely. While some plant species presents the possibility of self pollination, out-crossing species may be pollinated by genetically related plants growing in the near vicinity. Both source of inbreeding are usually more intense in small populations and for species with limited gene flow (e.g. Stefenon et al. 2008b). An important aspect of our study on this subject is the higher increase of the inbreeding coefficient occurred from the generation 0 to the 5th generation, that is, in the initial period after the forest isolation (around 50 to 100 years, considering species starting to reproduce with 10 to 20 years), followed by a scenario of stabilization through the further generations.

The comparatively lower inbreeding coefficient observed in the smaller populations (50 and 100 individuals) in our simulation may be effect of the very high levels of heterozygosity ($H = 0.98$), emphasizing the importance of the maintenance of high diversity. This is observed when comparing the level of increase in the inbreeding coefficient over populations. The smaller population (50 individuals) revealed the higher loss of alleles over generations, with a decrease of 5-fold in number of alleles from the generation 0 to the 1000th generation. This population revealed a small but comparatively higher (uni-modal t-test $p < 0.001$) increase in the inbreeding coefficient in the same time interval, in comparison to the

larger population, likely as effect of the lower allelic richness in the smaller population. This result highlights the importance of maintaining allelic richness towards the conservation of forest genetic resources.

The contrasting result concerning inbreeding coefficient and mean number of alleles for the three species of *Araucaria* evaluated in this study reflects the importance of considering ecological and reproductive/regenerative attributes of each species. Present-day populations of *A. angustifolia* and *A. nemorosa* are highly fragmented and present relatively low seedling regeneration (Stefenon et al., 2007, 2008a; Bittencourt and Sebbenn 2008; Kettle et al., 2007), whereas populations of *A. columnaris* comprise many thousands of individuals which produce abundant fertile seeds and show plentiful seedling regeneration (Kettle et al., 2007).

In addition to the ecological aspects, the kind of genetic marker used in the study and the pressure of selection should also be considered. For instance, a natural population of *A. angustifolia* studied with isozyme markers (Klöch, 2011), which may suffer adaptive pressure, revealed decrease in allelic richness from adults (first generation) to plantlets ($A = 1.9$ and 1.6 , respectively), but an increased number of alleles in the seeds of the first generation ($A = 2.2$), suggesting that many of the seeds will not further develop and alleles will be lost.

Considering that microsatellite markers are usually neutral, are highly polymorphic and presents elevated number of alleles, they may better represent the general behavior of allelic richness and inbreeding coefficient over generations, but do not necessarily assess selective patterns.

Considering that heterozygosity, allelic richness and inbreeding coefficient are typically estimated in population genetic surveys, these measures should be the start point for planning forest genetic resources conservation and management. However, care is needed when interpreting data concerning genetic diversity and conservation. Some studies based on computation of H and F -statistics consider critical populations for conservation based on the estimations of F_{ST} and similar measures. If estimated F_{ST} is low, the subpopulations are usually considered to be not highly differentiated and therefore, the maintenance of different subpopulations of the species is considered not important (Jost 2008). However, F_{ST} and similar metrics are not measures of genetic differentiation and may lead to equivocated

decisions concerning genetic resources conservation (for a broad discussion on this topic, see Jost 2008).

4.6 Conclusion

Although the levels of heterozygosity are an important index for the selection of important ESUs, it should not be the main feature evaluated since high levels of heterozygosity can exist also for loci containing just two alleles. Special attention should be given to the number of alleles of the selected populations. Reduction of the allelic richness and increase of the inbreeding coefficient over generations are expected for isolated populations of forest tree species, which usually display relatively small effective population size. Thus, allelic richness and levels of inbreeding should also be considered when selecting ESUs aiming the conservation of forest species genetic resources.

Moreover, sustainable management of planted forests and forest remnants, as well as forestation/reforestation enterprises should care on population genetic structure and levels of gene flow of the target species, towards maintenance of allelic richness (Stefenon et al. 2008a). Since ecological, demographic and reproductive features of the species are important for the decisions about genetic resource conservation, further studies covering a larger range of options have to be considered, in order to add more information in this issue.

4.7 Acknowledgements

This research was made possible by the financial support from FAPERGS (Process n. 0902170). The Authors are sponsored by a PIBIC scholarship (LSC) and a PQ grant (VMS) from CNPq (Process n. 314664/2009-2).

4.8 References

- Balloux F., 2001. EASYPOP (version 1.7): a computer program for population genetics simulations. *The Journal of Heredity* 92: 301-302.
- Bittencourt J.V.M., Sebbenn A.M., 2007. Patterns of pollen and seed dispersal in a small, fragmented population of the wind-pollinated tree *Araucaria angustifolia* in southern Brazil. *Heredity* 99: 580-591.
- Erikson G., 2005. Evolution and evolutionary factors, adaptations and adaptability. In Geburek Th., Turok J. (eds.) *Conservation and Management of Forest Genetic Resources in Europe*. Arbora Publishers, Zvolen, pp. 199-211.
- Finkeldey R., Hattmer H.H., 2007. *Tropical Forest Genetics*. Springer, Berlin, Heidelberg, New York.
- Gauli A., Gailing O., Stefenon V.M., Finkeldey R., 2009. Genetic similarity of natural populations and plantations of Sarg. in Nepal. *Annals of Forest Science* 66: 703.
- Goudet J., 1995. *Fstat* version 1.2: a computer program to calculate F-statistics. *The Journal of Heredity* 86: 485-486.
- Jost L., 2008. G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* 17: 4015-4026.
- Kettle C.J., Hollingsworth P.M., Jaffré T., Moran B., Ennos A., 2007 Identifying the early genetic consequences of habitat degradation in a highly threatened tropical conifer, *Araucaria nemorosa* Laubenfels. *Molecular Ecology* 16: 3581–3591.
- Kimura M., Ohta T., 1975. Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. *Proceedings of the National Academy of Sciences of USA* 72: 2761-2764.
- Sant'Anna C., 2011. Diversidade genética, estrutura genética espacial e dispersão realizada de pólen e sementes em uma população contínua de *Araucaria angustifolia* (Bertol.) Kuntze no planalto norte de Santa Catarina. [Genetic diversity, spatial genetic structure and effective dispersion of pollen and seeds in a continuous population of *Araucaria angustifolia* (Bertol.) Kuntze in the north Highlands of Santa Catarina] Msc Dissertation. Department of Plant Sciences, Federal University of Santa Catarina, Florianópolis, 89 p.

Stefenon V.M., Gailing O., Finkeldey R., 2007. Genetic Structure of *Araucaria angustifolia* (Araucariaceae) Populations in Brazil: Implications for the *in situ* Conservation of Genetic Resources. *Plant Biology* 9: 516-525.

Stefenon V.M., Gailing O., Finkeldey R., 2008a. Genetic structure of plantations and the conservation of genetic resources of Brazilian pine (*Araucaria angustifolia*). *Forest Ecology and Management* 255: 2718–2725.

Stefenon V.M., Gailing O., Finkeldey R., 2008b. The role of gene flow in shaping genetic structures of the subtropical conifer species *Araucaria angustifolia*. *Plant Biology* 10: 356-364.

5. DISCUSSÃO

Para os parâmetros analisados, diversidade genética, riqueza alélica e coeficiente de endogamia, o uso de simulações computacionais foi capaz de gerar dados consistentes e condizentes com dados reais, obtidos a campo.

Com relação à influência do tamanho amostral na estimativa de índices de diversidade genética, o estudo verificou uma forte correlação entre o tamanho amostral e a estimativa da riqueza alélica, sendo influenciada também pela heterozigosidade da população, enquanto a estimativa do coeficiente de endogamia se mostrou quase independente do tamanho amostral e da diversidade da população.

No que diz respeito ao comportamento da riqueza alélica e coeficiente de endogamia ao longo de gerações em populações florestais isoladas, o estudo afirmou a hipótese de que populações pequenas (50 e 100 indivíduos) tendem a sofrer uma grande redução no número de alelos ao longo de gerações, ressaltando a necessidade de atenção especial na implantação de ações a fim de minimizar os efeitos resultantes de populações florestais pequenas e fragmentadas.

Tendo em vista a recuperação e manutenção da diversidade genética em populações florestais fragmentadas a implantação de corredores ecológicos é uma excelente alternativa para restabelecer uma conexão entre populações isoladas.

O manejo integrado de corredores ecológicos visa promover o fluxo de indivíduos e genes entre populações e subpopulações, aumentando a probabilidade de sua sobrevivência em longo prazo e garantindo a manutenção de processos ecológicos evolutivos.

A utilização de softwares genéticos no estudo de genética de populações é uma alternativa que permite uma enorme variedade de possibilidades, desde explorar situações muito complexas de serem analisadas analiticamente, como exemplo a simulação do comportamento de uma população com 1000 indivíduos submetidos a 1000 gerações, até a possibilidade de construir hipóteses de acordo com os dados obtidos, tendo em vista que são utilizadas inúmeras variáveis como parâmetros para uma simulação, permitindo assim a reprodução de fenômenos naturais evolutivos.

Conforme descrito por Balloux (2001), de uma forma geral, a única limitação que os softwares genéticos apresentam, diz respeito apenas a memória física do computador, que restringe a quantidade de dados a serem processados inviabilizando uma simulação muito extensa.

A disponibilidade dos mesmos tem a intuito de atender a uma demanda crescente de usuários das mais diversas instituições de ensino, que manipulam um grande volume de dados, os quais necessitam um processamento adequado, seja para fins didáticos ou pesquisas.

Os aplicativos desenvolvidos com a finalidade de realizar simulações computacionais bem como aplicativos de análise estatística são suficientes para o desenvolvimento de estudos na área de genética de populações, com enfoque em espécies arbóreas. Porém, na literatura especializada, encontram-se trabalhos que tratam de simulações referentes a espécies anuais de interesse econômico, como por exemplo, soja e arroz. No entanto esses estudos têm o objetivo de fornecer subsídios apenas ao melhoramento genético, visando um aumento de produtividade.

6. CONCLUSÕES FINAIS E PERSPECTIVAS FUTURAS

O uso de simulações computacionais no estudo da diversidade genética se mostrou bastante eficiente, revelando resultados compatíveis com dados reais e promissores para que seja dada continuidade a pesquisa.

O sucesso do estudo permitiu refletir sobre a possibilidade de realizar novas simulações com a finalidade de construir um banco de dados que sirva como base para elaborar um software capaz de inferir o comportamento de uma população após determinado número de gerações através de dados reais obtidos a campo.

Portanto, espera-se que os resultados encontrados aqui contribuam na fundamentação de ações de manejo, programas de recuperação e preservação de populações florestais fragmentadas, principalmente através da implantação de projetos de restauração ecológica, é fundamental o conhecimento dos conceitos de ecologia de paisagem. Dessa forma, incorporar na restauração os conceitos de fragmentação, permeabilidade da matriz, conectividade da paisagem, corredores biológicos, fluxo gênico e de organismos, faz avançar a visão de restauração e amplia os horizontes das nossas ações em áreas degradadas.

Com isso, a necessidade de reverter o quadro atual da degradação ambiental gera o desafio de se recuperar áreas desmatadas ou degradadas, tendo-se como apreensão ações para o restabelecimento das funções e da estrutura dos ecossistemas respeitando a diversidade de espécies, a sucessão ecológica e a representatividade genética entre populações.

APÊNDICES

APÊNDICE A – Entrada de dados no software EASYPOP para simulação de população com cruzamentos endogâmicos.

```

-----
EASYPOP (v. 2.0.1)
Author: F. Balloux
Available at: http://www.unil.ch/izea/software/easypop.html
-----

Ploidy level ? (0=haplo-diploid; 1=haploid; 2=diploid)
2
Two sexes?:y/n
n
Random mating?:y/n
Y
Number of populations ?
1
Number of individuals ?
200000
Number of loci ?
10
Free recombination between loci?:y/n
Y
Do all loci have the same mutation scheme?:y/n
Y
Mutation rate ?(between 0 and 1, e.g. 0.0001)
0.003
Mutation model
1= Kam, (same probability to mutate to any allelic state)
2= Ssm, (single step mutation model)
3= Mixed model of Ssm with a proportion of Kam mutation events)
4= Mixed model of Ssm with a proportion of double step mutation events)
2
Number of possible allelic states ?(below 1000)
40
Variability of the initial population
1= Maximal, (randomly assigned alleles)
2= Minimal, (all individuals start with the same allele)
1
Number of generations ?
20
Do you want the complete dataset in the '.dat' and '.gen' result files
?:y/n
Y
Do you want a file giving the complete pedigree of the simulation ?:y/n
(Please notice that this file can be very huge and will slow down
simulations)
n
Name of the file?
inbreed
Number of replicates? (between 1 and 999)
1

```


APÊNDICE B – Entrada de dados no software EASYPOP para simulação de população com cruzamentos endogâmicos, tendo coeficiente de endogamia igual a 0,5.

```

-----
EASYPOP (v. 1.8)
Author: F. Balloux
Available at: http://www.unil.ch/izea/software/easypop.html
-----

Ploidy level ? (0=haplo-diploid; 1=haploid; 2=diploid)
2
Two sexes?:y/n
n
Random mating?:y/n
n
Proportion of clonal reproduction (between 0 and 1)
0.5
Proportion of selfing (between 0 and 1)
of the individuals not born as clones
0.5
Number of populations ?
1
Number of individuals ?
200000
Number of loci ?
10
Free recombination between loci?:y/n
y
Do all loci have the same mutation scheme?:y/n
y
Mutation rate ?(between 0 and 1, e.g. 0.0001)
0.003
Mutation model
1= Kam, (same probability to mutate to any allelic state)
2= Ssm, (single step mutation model)
3= Mixed model of Ssm with a proportion of Kam mutation events)
4= Mixed model of Ssm with a proportion of double step mutation events)
2
Number of possible allelic states ?(below 1000)
40
Variability of the initial population
1= Maximal, (randomly assigned alleles)
2= Minimal, (all individuals start with the same allele)
1
Number of generations ?
20
Do you want the complete dataset in the '.dat' and '.gen' result files
?:y/n
y
Name of the file?
inbreed50
Number of replicates? (between 1 and 999)
1

```

APÊNDICE C – Entrada de dados no software EASYPOP para simulação de população com cruzamentos endogâmicos, tendo coeficiente de endogamia igual a 0,8.

```
-----
EASYPOP (v. 1.8)
Author: F. Balloux
Available at: http://www.unil.ch/izea/software/easypop.html
-----
```

```
Ploidy level ? (0=haplo-diploid; 1=haploid; 2=diploid)
2
Two sexes?:y/n
n
Random mating?:y/n
n
Proportion of clonal reproduction (between 0 and 1)
0.8
Proportion of selfing (between 0 and 1)
of the individuals not born as clones
0.8
Number of populations ?
1
Number of individuals ?
200000
Number of loci ?
10
Free recombination between loci?:y/n
y
Do all loci have the same mutation scheme?:y/n
y
Mutation rate ?(between 0 and 1, e.g. 0.0001)
0.003
Mutation model
1= Kam, (same probability to mutate to any allelic state)
2= Ssm, (single step mutation model)
3= Mixed model of Ssm with a proportion of Kam mutation events)
4= Mixed model of Ssm with a proportion of double step mutation events)
2
Number of possible allelic states ?(below 1000)
40
Variability of the initial population
1= Maximal, (randomly assigned alleles)
2= Minimal, (all individuals start with the same allele)
1
Number of generations ?
20
Do you want the complete dataset in the '.dat' and '.gen' result files
?:y/n
y
Name of the file?
high_inbreed10
Number of replicates? (between 1 and 999)
1
```

APÊNDICE D – Entrada de dados no software EASYPOP para simulação de população com cruzamentos completamente aleatórios.

```

-----
EASYPOP (v. 2.0.1)
Author: F. Balloux
Available at: http://www.unil.ch/izea/software/easypop.html
-----

Ploidy level ? (0=haplo-diploid; 1=haploid; 2=diploid)
2
Two sexes?:y/n
y
Mating system?
1= Random mating
2= Polygyny
3= Monogyny
1
Number of populations ?
1
Number of females in each population ?
100000
Number of males in each population?
100000
Number of loci ?
10
Free recombination between loci?:y/n
y
Do all loci have the same mutation scheme?:y/n
y
Mutation rate ?(between 0 and 1, e.g. 0.0001)
0.003
Mutation model
1= Kam, (same probability to mutate to any allelic state)
2= Ssm, (single step mutation model)
3= Mixed model of Ssm with a proportion of Kam mutation events)
4= Mixed model of Ssm with a proportion of double step mutation events)
2
Number of possible allelic states ?(below 1000)
40
Variability of the initial population
1= Maximal, (randomly assigned alleles)
2= Minimal, (all individuals start with the same allele)
1
Number of generations ?
20
Do you want the complete dataset in the '.dat' and '.gen' result files
?:y/n
y
Do you want a file giving the complete pedigree of the simulation ?:y/n
(Please notice that this file can be very huge and will slow down
simulations)
n
Name of the file?
random
Number of replicates? (between 1 and 999)
1

```

APÊNDICE E – Entrada de dados no software EASYPOP para simulação de população com cruzamentos completamente aleatórios.

Esse modelo foi utilizado para realizar as simulações de populações endogâmicas com diferentes tamanhos populacionais (50, 100, 500 e 1000 indivíduos) submetidos a diferentes números de gerações (5, 10, 50, 100, 500 e 1000).

```
-----
EASYPOP (v. 2.0.1)
Author: F. Balloux
Available at: http://www.unil.ch/izea/software/easypop.html
-----
```

```
Ploidy level ? (0=haplo-diploid; 1=haploid; 2=diploid)
2
Two sexes?:y/n
n
Random mating?:y/n
n
Proportion of clonal reproduction (between 0 and 1)
0.5
Proportion of selfing (between 0 and 1)
of the individuals not born as clones
0.5
Number of populations ?
100
Same number of individuals in each populations ?:y/n
y
Number of individuals ?
50
Same migration scheme over all simulation? (y/n)
n
migration model for the first part of the simulation?
migration model?
1 = 1-dimension stepping stone;
2 = 2-dimension stepping stone (only with a square
  number of populations, e.g. 9, 16,..144..);
3 = island model;
4 = hierarchical stepping stone ('contact zone');
5 = hierarchical island ;
6 = spatial models;
1
proportion of migration?(between 0 and 1)
0.005
migration model for the second part of the simulation?
migration model?
1 = 1-dimension stepping stone;
2 = 2-dimension stepping stone (only with a square
  number of populations, e.g. 9, 16,..144..);
3 = island model;
4 = hierarchical stepping stone ('contact zone');
5 = hierarchical island ;
6 = spatial models;
1
proportion of migration?(between 0 and 1)
0.005
```

Number of loci ?
20
Free recombination between loci?:y/n
y
Do all loci have the same mutation scheme?:y/n
y
Mutation rate ?(between 0 and 1, e.g. 0.0001)
0.005
Mutation model
1= Kam, (same probability to mutate to any allelic state)
2= Ssm, (single step mutation model)
3= Mixed model of Ssm with a proportion of Kam mutation events)
4= Mixed model of Ssm with a proportion of double step mutation events)
2
Number of possible allelic states ?(below 1000)
80
Variability of the initial population
1= Maximal, (randomly assigned alleles)
2= Minimal, (all individuals start with the same allele)
1
Number of generations ?
5
Number of generations before the migration scheme changes?
5
Do you want the complete dataset in the '.dat' and '.gen' result files
?:y/n
y
Do you want a file giving the complete pedigree of the simulation ?:y/n
(Please notice that this file can be very huge and will slow down
simulations)
n
Name of the file?
5g_50i
Number of replicates? (between 1 and 999)
10