UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE ECOLOGIA

VARIABILIDADE GENÉTICA EM *Caryocar brasiliense* Camb. (CARYOCARACEAE): ESTRUTURA GENÉTICA, SISTEMA DE CRUZAMENTO, FILOGEOGRAFIA E SUBSÍDIOS PARA A CONSERVAÇÃO

ROSANE GARCIA COLLEVATTI

Tese apresentada ao Departamento de Ecologia como requisito parcial para obtenção do grau de doutor em Ecologia

BRASÍLIA - DF
FEVEREIRO - 2000

Universidade de Brasília
Instituto de Ciências Biológicas
Departamento de Ecologia

VARIABILIDADE GENÉTICA EM *Caryocar brasiliense* Camb. (CARYOCARACEAE): ESTRUTURA GENÉTICA, SISTEMA DE CRUZAMENTO, FILOGEOGRAFIA E SUBSÍDIOS PARA A CONSERVAÇÃO

ROSANE GARCIA COLLEVATTI

Tese apresentada ao Departamento de Ecologia como requisito parcial para obtenção do grau de doutor em Ecologia

BRASÍLIA-DF
FEVEREIRO-2000

ROSANE GARCIA COLLEVATTI

VARIABILIDADE GENÉTICA EM *Caryocar brasiliense* Camb. (CARYOCARACEAE): ESTRUTURA GENÉTICA, SISTEMA DE CRUZAMENTO, FILOGEOGRAFIA E SUBSÍDIOS PARA A CONSERVAÇÃO

Data da defesa: 17 de Fevereiro de 2000

Dr. John DuVall Hay
Orientador

Dr. Dario Grattapaglia
Orientador

Dr. Aldicir Scariot

Dr. Rogério Gribel

Dr. Marcio Elias Ferreira

"...it takes all the running you can do, to keep in the same place.
If you want to get somewhere else, you must run at least twice as fast as that!
(L. Carroll, *Through the Looking-glass*)

Ao meu filho, Guilherme,

pelos momentos de ausência

pelo tempo que passou que eu espero recuperar

A minha mãe Cidinha,

pelo amor, carinho, compreensão

pelo apoio incondicional

Agradecimentos

Ao meu orientador, Prof. Dr. John Du Vall Hay, pela oportunidade de desenvolver este trabalho e pelo apoio constante. Ao Dr. Dario Grattapaglia, também meu orientador, sem o qual este trabalho não seria possível, por ter acreditado em minha capacidade e ter me dado oportunidade de desenvolver este trabalho, pelas discussões, sugestões e orientações e pelo apoio incondicional. A amiga e colega de trabalho, Rosana P.V. Brondani, pela ajuda constante no trabalho, principalmente no desenvolvimento dos microsatélites, primordial para alcançar os objetivos deste trabalho, e pelo apoio nos vários momentos difíceis. Ao Dr. Márcio Elias Ferreira pelo apoio e oportunidade de trabalhar no Laboratório de Genética de Plantas e pelas sugestões.

Ao Dr. Rogério Gribel, pelas discussões construtivas e informações sobre o pequizeiro. Ao Dr. Aldicir Scariot e Dr. Ernani Pilla, pelas sugestões. A Célia M. T. Cordeiro pelas discussões e assistência na análise estatística. Aos doutores Paul Lewis e Dmitri Zaykin pela autorização no uso do programa experimental GDA. A Aécio A. Santos, Antônio Dimas Costa Jr., Bruno Walter, Eugenia C. Amaral, Fabian Borghetti, Flavia Pinto, Frederico S. Lopes, Galdino R. Santos, João B. Pereira, Maura C. Almeida, Neilton R. Oliveira, Sebastião P.Cordovil, pela valorosíssima ajuda nas coletas. Ao IBAMA (Instituto Brasileiro do Meio Ambiente) IEF-SP (Instituto Estadual de Floresta, SP), FEMAGO (Instituto do Meio Ambiente de Goiás) e SEMATEC (Secretaria do Meio Ambiente e Tecnologia, DF) pela autorização para coletas no Parque Nacional de Brasília e de Grandes Sertões Veredas, Estação Ecológica de Itirapina, Parque Estadual da Serra de Caldas e na Estação Ecológica de Águas Emendadas, respectivamente.

Este trabalho foi financiado pela WWF (World Wildlife Fund) e pelo CNPq/PADCT (Ministério de Ciência e Tecnologia) através do financiamento competitivo 62.00059/97-4

coordenado pelo Dr. Dario Grattapaglia, além da bolsa de doutorado pelo CNPq, aos quais agradeço. Agradecimentos, também, a Universidade de Brasília, pela oportunidade em realizar o doutorado no Departamento de Ecologia, e à EMBRAPA – Recursos Genéticos e Biotecnologia, pela oportunidade em realizar o trabalho de laboratório.

Finalmente, aos amigos do Laboratório de Genética de Plantas e a todos que direta ou indiretamente possibilitaram a realização deste trabalho.

Índice

Resumo

Este trabalho teve como objetivo estudar a estrutura genética de populações e o sistema de acasalamento de *Caryocar brasiliense* (Camb., Caryocaraceae), gerando subsídios para a conservação e manejo desta espécie, além de estudar a filogeografia da espécie no Cerrado. Foram desenvolvidos dez locos microsatélites para *C. brasiliense* altamente informativos, com probabilidade de identidade genética da ordem de 10^{-17} e exclusão de paternidade combinada de 0,99999995. Os genótipos multilocos são únicos permitindo discriminar os indivíduos de *C. brasiliense* em estudos detalhados de vínculo genético em populações naturais. Os resultados da estrutura genética de populações indicaram que existe diferenciação genética significativa entre populações conforme o modelo de "isolamento por distância", com correlação positiva entre distância geográfica e F_{ST} entre pares de populações. As estimativas sugerem que a diferenciação deve ser ocasionada principalmente pelo cruzamento entre indivíduos aparentados. O estudo do sistema de acasalamento indicou que *C. brasiliense* é preferencialmente alógamo. Entretanto, a taxa de endogamia biparental indicou uma alta freqüência de acasalamento entre parentes. Nossos resultados indicaram que *C. brasiliense* apresenta uma estrutura em demes, com alta freqüência de acasalamentos entre parentes. Assim, a fragmentação pode limitar o fluxo gênico, isolando os dispersores de sementes e os polinizadores, aumentando a taxa de cruzamento entre parentes. Dessa forma, a conservação de populações não isoladas em grandes áreas pode ser necessária para garantir eventos de polinização cruzada entre indivíduos não aparentados e, assim, manter a viabilidade desta espécie no Cerrado. Os dados obtidos com cpDNA também indicam uma alta diferenciação das populações devido à restrição ao fluxo de semente. Para o genoma nuclear as populações tenderam a ser agrupadas de acordo com a distância geográfica, mas não para o genoma de cloroplasto. O padrão de quebra filogenética não coincide, entretanto,

com padrões geográficos. O padrão filogeográfico encontrado pode ser explicado por restrições ao fluxo de semente e por eventos de colonização posterior à ultima glaciação e efeito de fundador.

Introdução Geral

1. Aspectos Teóricos

A variação genética em populações naturais pode estar estruturada no tempo e no espaço. O desenvolvimento e manutenção dessa estrutura se deve a fatores como o "pool" gênico, a organização da variabilidade nos genótipos e a distribuição espacial destes genótipos, o sistema reprodutivo, o sistema de polinização e dispersão de sementes (fluxo gênico), eventos estocásticos como deriva gênica, e processos ecológicos como recrutamento, crescimento populacional, mortalidade e fecundidade (Wright, 1931, 1943; Gilpin & Soulé, 1986). A estruturação genética espacial pode ser favorecida por barreiras geográficas ou por uma distribuição espacial em manchas, que proporcionam um certo grau de isolamento entre populações, promovendo um acúmulo de diferenças genéticas. Quando uma espécie possui uma distribuição geográfica ampla, o tamanho da área ocupada comparada à capacidade de dispersão dos indivíduos pode impedir a formação de uma única unidade panmítica. Dessa forma, as restrições ao fluxo gênico podem levar as populações a diferenciarem geneticamente, ocasionando uma estruturação espacial da variabilidade genética. Este modelo de diferenciação genética é conhecido como "isolamento por distância", e foi proposto por Sewall Wright ("isolation by distance", Wright, 1931, 1943). Assim, a diferenciação genética entre populações pode ser o resultado da restrição ao fluxo gênico, e pode estar correlacionada com a distância geográfica.

O processo de "isolamento por distância" apresenta um padrão de mudanças genéticas semelhantes à endogamia, na medida que leva a um excesso de homozigotos. Dessa forma, os efeitos desse processo nas populações podem ser medidos em relação ao decréscimo na freqüência de heterozigotos pelas estatísticas de Wright (Wright, 1951). Nesse caso, uma

população subdividida teria três níveis de complexidade estrutural: (i) os indivíduos; (ii) as subpopulações; (iii) a população como um todo. Os efeitos da subdivisão são medidos, nestes três níveis, em relação ao desvio de uma população panmítica: a redução da heterozigosidade devido a acasalamentos não aleatórios dos indivíduos dentro das subpopulações (coeficiente de endogamia - F_{IS}); a redução na heterozigosidade da subpopulação devido a deriva gênica (índice de fixação - F_{ST}); redução na heterozigosidade em relação à população total, devido a acasalamentos não aleatórios dentro das subpopulações e deriva gênica (coeficiente de endogamia total - F_{IT}). Se todas as subpopulações estiverem em equilíbrio de Hardy-Weinberg, com mesma freqüência alélica, então Fst é igual a zero. Valores negativos de F_{IS} e F_{IT} indicam que a heterozigosidade observada nas subpopulações é maior que a esperada pelo equilíbrio de Hardy-Weinberg. A estrutura genética pode ser também avaliada pela correlação das freqüências alélicas em diferentes níveis de subdivisão da população (Cockerham, 1969). As correlações, ρ, estão relacionadas à componentes de variância (função do quadrado médio esperado) obtidos por uma análise de variância, considerando os níveis de subdivisão da população como fonte de variação. Assim, o componente de variância total, σ^2, pode ser decomposto em: $\sigma^2=\sigma^2_G+\sigma^2_I+\sigma^2_P$ onde: σ^2_G é o componente de variância para os alelos dentro dos indivíduos; σ^2_I – para os indivíduos dentro das populações; σ^2_P - componente de variância para populações. As correlações entre freqüências alélicas podem ser definidas como: θ – a correlação entre as freqüências alélicas de diferentes indivíduos em uma mesma população $\{\theta=\rho_P=(\sigma^2_P)/(\sigma^2_G+\sigma^2_I+\sigma^2_P)\}$; F – a correlação entre as freqüências alélicas dentro de indivíduos de diferentes populações $\{F=\rho_{PI}=(\sigma^2_P+\sigma^2_I)/(\sigma^2_G+\sigma^2_I+\sigma^2_P)\}$; f – a correlação entre as freqüências alélicas dentro dos indivíduos dentro das populações $\{f=\rho_I=(\sigma^2_I)/(\sigma^2_G+\sigma^2_I)\}$ (Cockerham, 1969; Weir, 1996). As correlações podem ser equivalentes às estatísticas de Wright: $\theta=F_{ST}$, o índice de fixação; $F=F_{IT}$, coeficiente total de endogamia; $f=F_{IS}$, coeficiente de endogamia (Wright, 1951, 1965; Cockerham, 1969).

Em populações subdivididas, a diferenciação genética deve refletir as diferenças locais em pressões seletivas, alta freqüência de endogamia, ou podem ser resultado de processos aleatórios como mutação e deriva gênica. Entretanto, o fluxo gênico pode representar um mecanismo homogeneizador, mantendo as populações conectadas e diminuindo a divergência genética (Slatkin, 1985).

O fluxo gênico pode ser interpretado como um processo de dispersão à longa distância, no qual todas as subpopulações têm a mesma probabilidade de enviar migrantes para qualquer outra subpopulação. Nesse caso, as subpopulações podem ser consideradas equivalentes, com uma taxa de migração constante e igual para todas as subpopulações. De acordo com este modelo de migração, conhecido por "modelo de ilhas", uma população grande é dividida em várias subpopulações dispersas geograficamente, como ilhas em um arquipélago (Wright, 1931). O número de migrantes por geração pode ser estimado pelo tamanho efetivo da população e pelo efeito da deriva gênica na heterozigosidade (F_{ST}). Wright (1968) demonstrou que se uma fração "m" da população for reposta por migrantes não haverá diferenciação genética entre as subpopulações se Nm>1, onde "N" é o tamanho efetivo da população. Com o aumento do número de migrantes, o índice de fixação (F_{ST}) diminui rapidamente. Com níveis muito baixos de migração há uma tendência à fixação de alelos nas subpopulações (F_{ST}~1,0).

Kimura (1953) propôs um modelo de migração, conhecido como "stepping stone". O modelo considera que as populações naturais estão subdivididas em "colônias" (subpopulações ou demes), e a migração dos indivíduos é restrita a "colônias" adjacentes. Os pares de subpopulações podem estar organizados espacialmente em uma (fluxo gênico entre uma subpopulação e dois vizinhos mais próximos), duas ou três dimensões, com a mesma fração de indivíduos movendo-se entre pares de subpopulações adjacentes. Este modelo representa o extremo em fluxo gênico a curta distância.

Apesar das populações estarem arranjadas de forma distinta no espaço, o modelo para duas dimensões apresenta as mesmas previsões que o "modelo de ilhas" (Kimura & Maruyama, 1971; Slatkin & Barton, 1989). A estimativa do fluxo gênico pelo método proposto por Wright (1931, 1943) é igualmente eficiente para ambos os modelos pois é sensível a mudanças nos valores de "Nm", relativamente insensível a diferenças nas taxas de mutação e seleção, ao número de demes, ou subpopulações no "modelo de ilhas", e prevê uma boa estimativa de Nm no modelo "stepping stone", para demes mais próximos (Slatkin & Barton, 1989). Entretanto, o número de migrantes por geração (Nm) estimado pelo modelo de ilhas considera que a população está em equilíbrio, com taxa de mutação desprezível e taxa de migração constante (Wright, 1943). Esta condição raramente é encontrada, principalmente quando existe um "isolamento por distância", com correlação positiva entre a diferenciação e distância geográfica (Whitlock & McCauley, 1999).

Muitas espécies com distribuição geográfica ampla podem ser compostas de populações geográficas cujos membros ocupam diferentes ramos de uma árvore filogenética intraespecífica ("pedigree"). O processo de subdivisão de populações pode levar, também, a uma distribuição geográfica de linhagens genealógicas, tanto a nível infragenérico como intraespecífico (Avise *et al.*, 1987, Avise, 1994). Espécies que não mostram uma estrutura populacional filogeográfica geralmente tiveram uma história de vida onde predominaram dispersão e fluxo gênico a longa distância. Por outro lado, espécies com grupos monofiléticos bastante distintos geralmente passaram por longos períodos de barreira ao fluxo gênico (Avise, 1994). Com o aumento do tempo desde o isolamento, espera-se que o grau de diferenciação filogeográfica entre grupos genealógicos aumente. A filogeografia (Avise *et al.*, 1987) é baseada na distribuição espacial de linhagens filogenéticas e permite a detecção de correlação entre a distribuição geográfica dos haplótipos e sua relação genealógica. Assim, a separação filogeográfica dentro das espécies pode estar correlacionada com os limites entre

províncias biogeográficas (Avise, 1992, 1994). Esses estudos baseiam-se na variabilidade no genoma de organelas que são mais conservados, com baixas taxas de mutação, quando comparado ao genoma nuclear, e sem recombinação, como o cloroplasto, para plantas (cpDNA) e mitocôndria, para animais (mtDNA) (Birky, 1988; Swofford & Olsen, 1999; Avise, 1994). Dessa forma, espera-se que a estrutura genética do genoma nuclear seja mais influenciado por fatores históricos como o fluxo gênico no passado, além de eventos geológicos como as glaciações e mudanças climáticas, que afetam a distribuição geográfica das espécies (Avise *et al.*, 1987; Avise, 1994; Schaal *et al.*, 1998). A análise de DNA de cloroplasto de plantas, que geralmente possui herança maternal em angiospermas, juntamente com o estudo do DNA genômico, com herança biparental, permite obter informações sobre a importância relativa do fluxo gênico via pólen (DNA nuclear) e semente (cpDNA) na estruturação das populações (McCauley, 1995; Schall *et al.*, 1998).

O sistema de acasalamento pode determinar o papel da endogamia na diferenciação genética entre populações (Wright, 1940). Nas plantas, o sistema de acasalamento pode ser determinado pela (1) característica do sistema reprodutivo como mecanismos de auto-incompatibilidade, protoginia, protandria, dioecia; (2) comportamento do polinizador; (3) aborto seletivo; (4) fenologia e densidade de indivíduos em floração (e.g. Shaanker *et al.*, 1988; Marshall & Folsom, 1991). As taxas de cruzamento geralmente são estimadas assumindo o modelo "sistema de acasalamento misto" (mixed mating model) (Ritland & Jain, 1981). Esse modelo assume que (1) cada acasalamento representa um evento aleatório de fertilização cruzada ou auto-fertilização, com probabilidades iguais a t e (1-t), respectivamente; (2) não ocorre mutação e seleção após a fertilização; (3) não há acasalamento preferencial (a probabilidade de fertilização cruzada é independente dos genótipos do pai ou da mãe) ou variação na freqüência de alelos no conjunto de pólen. Para estimar a taxa de cruzamento (probabilidade que o descendente de uma mãe seja derivado de

fertilização cruzada) vêm sendo utilizado este modelo, implementado pelo programa MLTR desenvolvido por Ritland (1996), que permite estimar vários parâmetros relacionados ao sistema de acasalamento.

Hipóteses anteriores previam que a maioria das espécies de plantas tropicais deveriam ser autógamas - altos níveis de endogamia e deriva gênica deveriam levar a uma rápida diferenciação entre populações e, eventualmente à especiação (Federov, 1966). Federov acreditava que a grande distância entre indivíduos co-específicos, somada à auto-compatibilidade levaria a uma alta taxa de endogamia, assim, haveria uma grande variabilidade entre populações. Por outro lado, Ashton (1969) sugeriu que a maioria das espécies tropicais são alógamas e que ocorreria especiação alopátrica, como em espécies animais.

Entretanto, os trabalhos sobre sistema reprodutivo de espécies tropicais (revisado por Bawa, 1992) têm demonstrado que, ao contrário do esperado, a maioria das plantas tropicais são auto-incompatíveis e há um alto nível de polinização cruzada. Além disso, estudos em florestas tropicais com espécies taxonomicamente não relacionadas, com forma de vida e hábito distintos, utilizando principalmente eletroforese de proteínas, têm mostrado não haver diferenciação genética significativa entre populações (estruturação genética espacial), além de terem sido encontrados altos níveis de fluxo gênico (Alvarez-Buylla & Garay, 1994, Chase *et al.*, 1995, Hamrick & Loveless, 1989, Loiselle *et al.*, 1995a,b), e sistema de acasalamento alógamo (e.g. O'Malley & Bawa, 1987; Murawski & Hamrick, 1991; Boshier *et al.*, 1995; James *et al.*, 1998; Loveless *et al.*, 1998). Entretanto, algumas espécies tropicais apresentam níveis significativos de auto-polinização (Murawski & Hamrick, 1992; Murawski *et al.*, 1994).

2. Genética e Conservação

Embora a perda de diversidade de espécies nas regiões tropicais, pela fragmentação e destruição de habitats, venha recebendo mais atenção nas últimas décadas (Wilson, 1994), pouco se sabe sobre a perda da diversidade genética. A grande maioria dos trabalhos sobre variabilidade genética e sistema de acasalamento foram desenvolvidos em regiões temperadas, e os trabalhos em florestas tropicais vêm sendo desenvolvidos, principalmente, na América Central, particularmente no Panamá e Costa Rica (e.g. Hamrick & Loveless, 1986, 1989; O'Malley & Bawa, 1987; Murawski & Hamrick, 1991, 1992; Boshier *et al.*, 1995; James *et al.*, 1998; Loveless *et al.*, 1998). Em ecossistemas como o Cerrado, considerado um dos ecossistemas mais ameaçados do Brasil, ao lado da Floresta Atlântica, principalmente pela expansão da fronteira agrícola, praticamente nada se sabe a respeito da diversidade genética das espécies e como ela está organizada espacialmente.

A subdivisão de uma população grande e amplamente distribuída em pequenas subpopulações isoladas pode aumentar a susceptibilidade da espécie à extinção local, e posteriormente à extinção global, pois cada subpopulação se torna mais vulnerável aos efeitos deletérios da imprevisibilidade demográfica, ambiental, genética e a catástrofes naturais (Terborgh & Winter 1980; Gilpin & Soulé 1986). As variações genéticas estocásticas resultam da deriva gênica, depressão endogâmica e efeito de fundador, resultado da perda de variabilidade, diminuição do tamanho e isolamento das populações decorrentes da fragmentação.

As populações remanescentes nos fragmentos podem ser produtos de um efeito de amostragem (sub-amostra dos alelos da população original). Parte da variabilidade genética original, portanto, pode ser perdida, inicialmente, somente por uma redução do tamanho da população, efeito conhecido como "genetic bottleneck" (Franklin, 1980, Ellstrand & Elam,

1993). A perda de variabilidade genética pode limitar a população a responder a mudanças a longo prazo no ambiente. Características neutras em condições atuais podem se tornar vantajosas em condições ambientais diferentes. Quando há a perda de alelos, a população tem poucas opções disponíveis e, portanto, tem menor probabilidade de sobreviver que populações maiores (Franklin, 1980).

A deriva gênica pode mudar a distribuição da variabilidade genética pela perda de variação dentro da população (perda de alelos) e pelo aumento da diferenciação entre populações por fixação de alelos. Por outro lado, pode provocar a fixação de alelos deletérios e provocar a diminuição do valor adaptativo dos indivíduos da população e, consequentemente, aumentar as chances de extinção (Franklin, 1980, Ellstrand & Elam, 1993).

A maior freqüência de acasalamentos entre parentes (endogamia), em populações pequenas e isoladas, resulta em uma perda da variabilidade genética por aumento da homozigose, e pode levar a uma depressão endogâmica. (Franklin, 1980; Ellstrand & Elam, 1993). A depressão endogâmica consiste na diminuição do valor adaptativo pela expressão de alelos deletérios em homozigose e depende de fatores como sistema reprodutivo e tamanho da população. Em populações tipicamente autógamas ou endogâmicas a freqüência de alelos deletérios recessivos declina com o aumento da homozigose. Estes alelos podem ser eliminados por seleção durante a história evolutiva da espécie (Franklin, 1980; Barrett & Kohn 1991; Ellstrand & Elam, 1993). Em populações tipicamente grandes, com pouca estruturação espacial, ou em populações que se tornaram pequenas por qualquer fator, a endogamia provoca depressão endogâmica com o aumento do nível de homozigose (Charlesworth & Charlesworth, 1987; Ellstrand & Elam, 1993; Hasting & Harrison, 1994).

O nível de fluxo gênico geralmente é alto o suficiente para contrabalançar os efeitos de níveis moderados de deriva gênica, endogamia e seleção. Mesmo em plantas autógamas, o

fluxo gênico a longas distâncias pode ocorrer a taxas significativas (Ellstrand & Elam, 1993). Entretanto, a taxa de fluxo gênico é fortemente dependente do tamanho da população e distância entre populações. A curtas distâncias, populações maiores recebem menor taxa de fluxo gênico que populações pequenas. Entretanto, com o aumento da distância, a tendência se reverte e as populações pequenas se tornam mais isoladas que as grandes. Populações maiores devem exportar maior quantidade de migrantes (pólen e semente para plantas) que populações pequenas, criando um fluxo gênico fortemente assimétrico (Slatkin, 1987; Ellstrand & Elam, 1993).

Uma vez que os efeitos da deriva gênica e depressão endogâmica são especialmente pronunciados em populações pequenas e isoladas, o planejamento de reservas para conservação de espécies e planos de manejo de espécies exploradas economicamente devem levar em conta estes riscos. O monitoramento periódico das populações restritas, considerando as mudanças na freqüência gênica, permite uma preservação a longo prazo, uma vez que a perda de variabilidade e depressão endogâmica podem ser detectados e amenizados com introdução de indivíduos de outras populações. O monitoramento pode ainda dar informações sobre a distribuição da variabilidade genética entre e dentro das populações. Se a maior quantidade de variabilidade estiver entre populações, então a preservação de um maior número de populações pode ser necessário, afim de preservar uma maior quantidade de diversidade genética (Franklin, 1980, Ellstrand & Elam, 1993).

3. O Cerrado e a Importância de sua Conservação

O Cerrado (*latu sensu*) ocupa uma área de aproximadamente 2 milhões de km^2, representando cerca de 23% do território brasileiro, distribuída principalmente no Planalto Central (Furley & Ratter, 1988). Este Bioma consiste de uma vegetação heterogênea, desde

floresta mesofítica até uma vegetação savânica, com arbustos e árvores de pequeno porte (cerrado *sensu stricto*) e campos, que podem ou não apresentar árvores e arbustos esparsos. Muitos fatores podem afetar a distribuição das espécies de plantas no Cerrado, como o clima, fertilidade e pH do solo, disponibilidade de água, geomorfologia e topografia, latitude, freqüência de fogo e fatores antrópicos, além da interação complexa entre estes fatores (Furley & Ratter, 1988). O Cerrado possui uma alta biodiversidade, com cerca de 160.000 espécies, incluindo plantas, animais e fungos. O número de arbustos e árvores no cerrado *sensu stricto* pode exceder a 800 espécies, das quais aproximadamente 40% são endêmicas (Ratter *et al.*, 1997).

Até a década de 60 essa região era considerada como marginal para a agricultura intensiva. A ocupação do cerrado foi motivada principalmente pela implantação de Brasília, provocando mudanças radicais na estrutura rodoviária, e pela criação do Programa de Desenvolvimento do Centro-Oeste (Polocentro), na década de 70, que levou a uma intensa migração para a região em busca de terras a custos mais baixos, em relação ao sul do país, e pelos incentivos fiscais para abertura de novas áreas agrícolas (Macedo,1995). Além disso, a perda de fertilidade e erosão nas regiões agrícolas tradicionais e a política agrícola de monoculturas que demandam grandes áreas para cultivo vêm provocando uma crescente pressão sobre a região do cerrado.

Atualmente, 47 milhões de hectares dos cerrados estão ocupados com áreas agrícolas, como pastagens cultivadas, culturas anuais e perenes, o que corresponde a 23% do território do cerrado. Estima-se uma área potencial de 89 milhões de hectares para uso agrícola futuro, que irá resultar em 136 milhões de hectares de ocupação agrícola do cerrado, ou 66% de todo o cerrado (Macedo, 1995), provocando uma maior fragmentação do ecossistema original.

O Cerrado constitui um ecossistema heterogêneo com alto nível de endemismo, cuja evolução foi influenciada pelo distúrbio natural do fogo e características do solo, criando uma

diversidade de fisionomias (Furley & Ratter, 1990). Em ecossistemas como o Cerrado, onde há alta freqüência de distúrbios, as populações podem estar estruturadas em um sistema de metapopulações, ou seja, distribuídas em manchas, ou subpopulações interdependentes, onde a manutenção da estabilidade e persistência da metapopulação é dependente do fluxo de indivíduos entre populações (Hanski, 1989). Nesse contexto, a fragmentação desse ecossistema, levando à diminuição do tamanho das populações e isolamento entre elas, pode comprometer a persistência a longo prazo de muitas espécies. Processos genéticos de diferenciação entre populações, como deriva gênica, efeito de fundador e seleção natural são grandemente influenciados pelo fluxo gênico (Wright, 1978; Hasting & Harrison, 1994). Dessa forma, o estudo da variabilidade genética presente nesse tipo de ambiente é de suma importância para avaliar a probabilidade de persistência das espécies nos remanescentes.

4. História Natural do *Caryocar brasiliense*

Caryocar brasiliense, conhecida popularmente como pequi, é a única espécie da família Caryocaraceae que ocorre no Cerrado. Essa espécie tem uma ampla distribuição geográfica, ocorrendo em pelo menos sete estados do Brasil, principalmente nas regiões Central e Sudeste (Prance & Silva, 1973). Regionalmente, apresenta distribuição agregada, em manchas bem definidas (Araújo, 1994), o que pode indicar a existência de uma organização regional em subpopulações. Em levantamentos fitossociológicos, geralmente está entre as 10 espécies com maior valor de importância (e.g. Felfili *et al*, 1994). Além disso, tem a segunda maior biomassa acima do solo, entre 35 espécies comuns do cerrado, em Brasília, DF, sendo, portanto, bastante importante na ciclagem de nutrientes (Silva, 1990).

A espécie é economicamente importante para populações locais do Centro Oeste e Sudeste, como no norte de Minas Gerais e Goiás (Araújo, 1994). Constituí fonte de material

para pequenas e médias industrias, que utilizam as flores, frutos, sementes, folhas e casca. O óleo é utilizado em indústria cosmética, para iluminação, lubrificação e alimentação. As folhas contem propriedades medicinais e o mesocarpo é bastante rico em vitaminas e sais minerais, sendo bastante utilizado na culinária do norte de Minas Gerais e Goiás (Araújo, 1994).

Além disso, constitui importante fonte de recurso para diversas espécies de animais do Cerrado. As flores são visitadas por nove espécies de morcegos, que são os principais polinizadores, além de diversas espécies de mariposas e abelhas, eventuais polinizadores (Gribel & Hay, 1993). Os frutos são utilizados por seis espécies de mamíferos, entre eles a cotia (*Dasyprocta* sp.), e diversas espécies de aves, como a gralha (*Cyanocorax cristatellus*), a arara canidé (*Ara ararauna*), e a ema (*Rhea americana*), que pode ser dispersora das sementes (Gribel, 1986). Outro fator importante é a dependência da planta dos polinizadores para a produção de frutos e sementes. A produção de sementes é mais freqüente com polinização cruzada (16,4%) que com geitonogamia (3,6%) ou autogamia (4,4%) aparentando ser auto-incompatível (Gribel & Hay, 1993). A floração ocorre entre agosto e setembro, mas pode variar de acordo com variações locais do início das chuvas, veranicos ou variações geográficas. A frutificação ocorre entre outubro e janeiro.

Apesar da grande abundância do pequizeiro no Cerrado, nos últimos anos tem sido observado um decréscimo no número de indivíduos em algumas regiões. Isso se deve, aparentemente, à exploração, pelas populações de algumas regiões, de virtualmente todos os frutos produzidos. Esse comportamento parece ser reforçado pela crença popular de que as sementes não germinam, não havendo, portanto, problemas na super-exploração dos frutos (Araújo, 1994). Nas regiões onde há degradação do Cerrado por atividades antrópicas, os indivíduos jovens e plântulas não sobrevivem à queimadas, comumente praticadas no Brasil. Somente os indivíduos adultos reprodutivos sobrevivem e são mantidos para aproveitamento

dos frutos. Além disso, são observados períodos de alta produção de frutos seguidos por períodos de baixa produção e altas taxas de predação por insetos e outros animais (Gribel & Hay, 1993; Araújo, 1994). Todos esses fatores contribuem para um baixo recrutamento, principalmente em regiões submetidas à distúrbios antrópicos. Araújo (1994) observou em Montes Claros, MG, a existência de poucas plântulas e indivíduos jovens. Mais de 70% da população estudada era constituída de indivíduos adultos, significando um baixo recrutamento. Nessa região, na época de frutificação do pequi, as atividades normais da população são desviadas para a coleta e comercialização de frutos, gerando emprego temporário no comércio e indústria.

Objetivos

Esse trabalho teve como objetivos estudar a estrutura genética, o sistema de acasalamento e a filogeografia do pequizeiro (*Caryocar brasiliense* Camb., Caryocaraceae) no Cerrado, gerando subsídios para a conservação e manejo desta espécie. As hipóteses de trabalho foram:

(1) As populações de *C. brasiliense* são diferenciadas por um processo de "isolamento por distância";

(2) A taxa de cruzamento é alta, evidenciada pela quiropterofilia, e pelo sistema reprodutivo preferencialmente alógamo;

(3) *C. brasiliense* apresenta uma estrutura filogeográfica, com linhagens filogenéticas bem definidas correlacionadas com barreiras geográficas.

(4) A subdivisão das populações de *C. brasiliense* deve ser maior para o genoma de cloroplasto que para o nuclear, assim, o fluxo gênico via pólen deve ser mais importante na homogeneização das populações que o fluxo via semente.

Para acessar a organização da variabilidade genética foram utilizados tecnologias de marcadores moleculares, baseados em DNA nuclear e de cloroplasto (cpDNA). A geração e análise de dados de marcadores moleculares neutros possui uma série de vantagens sobre análise morfométrica: (1) custo baixo, quando comparado a quantidade de informações geradas por unidade de tempo; (2) pode ser prontamente aplicada para qualquer espécie utilizando protocolos, reagentes e processos analíticos universais; (3) envolve amostragem não destrutiva do material biológico; (4) gera dados geneticamente interpretáveis sem influência ambiental; (5) permite um acúmulo rápido de dados (Ferreira & Grattapaglia,

1995). A análise direta a nível de DNA permite uma investigação refinada da variação genética em diferentes níveis hierárquicos de organização, i.e. entre espécies, entre populações da mesma espécie, dentro das populações. Além disso, a análise de DNA é flexível, permitindo trabalhar com o genoma nuclear e de organelas (cloroplasto e mitocôndria), que possuem características diferentes de herança e taxa de evolução molecular. Por um lado, o genoma nuclear é herdado de ambos os parentais, sofre recombinação na meiose e é submetido a uma taxa de mutação média relativamente mais alta, resultando em altos níveis de polimorfismo entre indivíduos de uma mesma população. Por outro lado o genoma de organelas como cloroplastos e mitocôndrias exibem diferentes padrões de diferenciação genética devido a herança geralmente uniparental e maternal, não sofrem recombinação e possuem taxas de mutação mais baixa, apresentando um padrão de diversidade genética muito mais conservador (Dowling *et al.*, 1990; Avise, 1994). A análise de DNA de cloroplasto (cpDNA) tem sido utilizada extensivamente para investigar a diferenciação interespecífica (e.g. White, 1990; Pennington, 1995; Smith & Doyle, 1995; McCauley *et al.*, 1996). Além disso, a variação intraespecífica entre populações tem permitido traçar linhagens maternais e entender processos como a extinção e colonização de populações de plantas.

O trabalho foi organizado em quatro capítulos, correspondendo aos quatro artigos que foram submetidos para publicação, os quais respondem aos diferentes objetivos e hipóteses apresentados acima. Antes da apresentação dos capítulos é apresentado um resumo do Material e Métodos correspondente aos quatro capítulos, são eles: (1) "Development and characterization of microsatellite markers for genetic analysis of a Brazilian endangered tree species *Caryocar brasiliense*", que trata do desenvolvimento, otimização e caracterização dos marcadores microsatélites para *C. brasiliense*, além da análise de herança e

transferibilidade destes marcadores para outras espécies do mesmo gênero; (2) "Population genetic structure of the endangered tropical tree species *Caryocar brasiliense*, based on variability at microsatellite loci", que trata da estrutura genética de dez populações de *C. brasiliense* situadas em áreas de preservação e em áreas fragmentadas, submetidas à distúrbios antrópicos, e discute o efeito da fragmentação na diversidade e estrutura genética; (3) "High resolution microsatellite based analysis of mating system allows the detection of significant biparental inbreeding in *Caryocar brasiliense,* endangered tropical tree species", onde é estudado o sistema de cruzamento de quatro populações de *C. brasiliense*; (4) "Phylogeography of the endangered Brazilian tree species C*aryocar brasiliense* based on variability at chloroplast microsatellite loci: genetic structure and a comparison with nuclear microsatellite loci" onde se compara a estrutura genética de populações baseada em cpDNA e DNA nuclear e discute-se a filogeografia de *C. brasiliense* em relação a restrições ao fluxo gênico via semente, e a influência das glaciações do Quaternário na distribuição de plantas do Cerrado. Finalmente, é apresentada uma discussão geral dos quatro capítulos.

Material e Métodos

1. Populações, amostragem e extração de DNA

Para análise da estrutura genética de populações de *C. brasiliense*, uma população por localidade (total de 10 populações) foram amostradas em toda a distribuição geográfica do pequizeiro. Estas populações apresentam diferentes histórias de fragmentação e distúrbio (Figura 1, Tabela 1): cinco áreas contínuas - AGE – Estação Ecológica de Águas Emendadas, Brasília, DF; CNV – Parque Estadual da Serra de Caldas, Caldas Novas, GO; GSV – Parque Nacional de Grandes Sertões Veredas, Arinos, MG; FAL – Parque Florestal de Água Limpa, Brasilia, DF; PNB – Parque Nacional de Brasília, Brasília, DF; e cinco áreas fragmentadas e isoladas, com altos níveis de distúrbios antrópicos – CGR - Campus da Universidade Federal de Mato Grosso do Sul, Campo Grande, MS; ITI – Reserva Ecológica de Itirapina, Itirapina, São Paulo; MTR - Rondonópolis, MT; TOC - Porto Nacional, TO; URU – área de influência indireta da Hidroelétrica de Serra da Mesa, Uruaçú, GO. Em cada população, todos os indivíduos (pelo menos 30) foram marcados, mapeados e folhas expandidas foram coletadas e estocadas em freezer a –80°C.

Para análise de sistema de cruzamento, foram amostradas quatro populações, CNV, FAL, PNB e URU. Nestas populações foram amostradas de 8 a 10 famílias de meios-irmãos de polinização aberta. Para cada família, em cada população, além da folha da árvore mãe, foram coletadas 16 sementes.

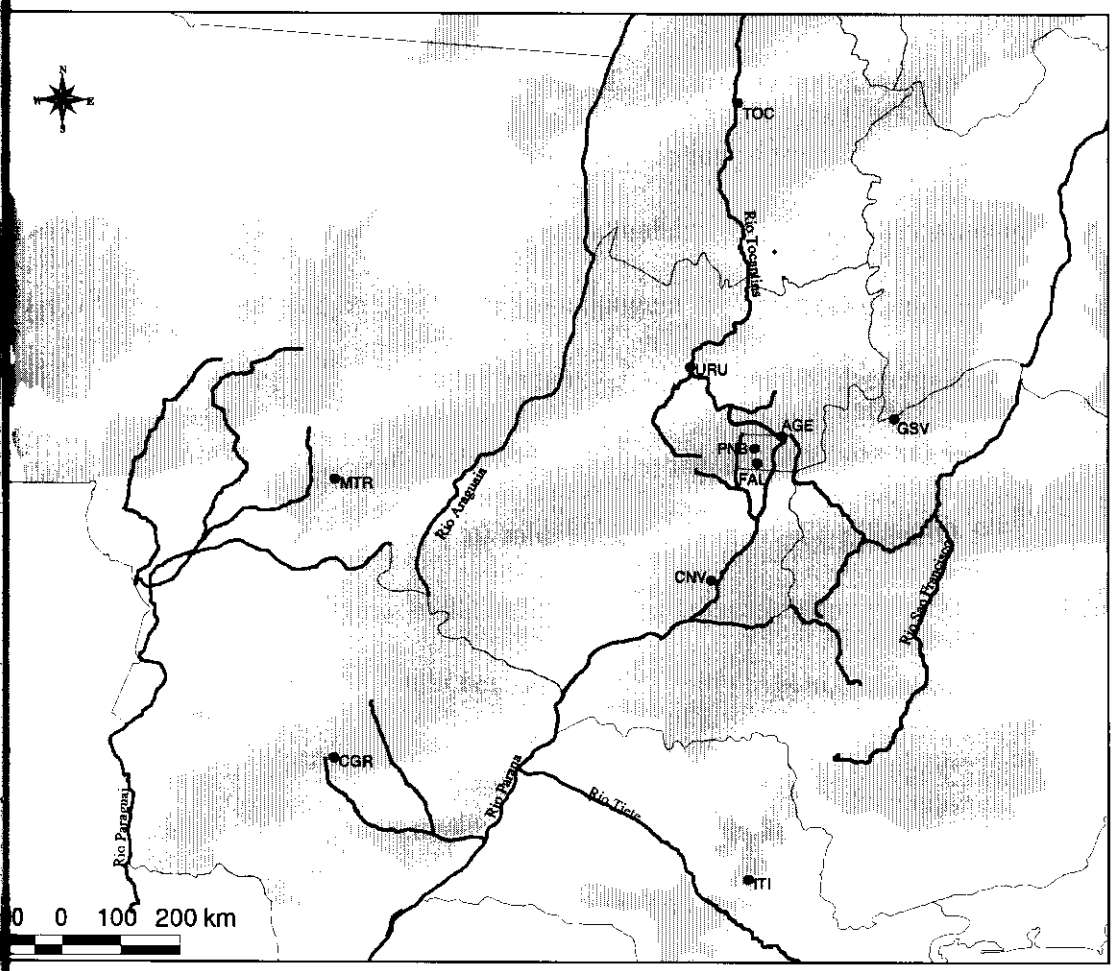
Para análise de filogeografia, foram utilizados 16 indivíduos de cada uma das dez populações descritas acima.

O DNA genômico de folhas foi extraído pelo método padrão de CTAB (Doyle & Doyle, 1987). Para análise da progênie no estudo de sistema de cruzamento, o DNA foi

extraído diretamente dos embriões, devido a baixa taxa de germinação das sementes. Para extração do DNA dos embriões foi utilizado um kit de extração de DNA - Fast DNA™ Kit H (BIO101, CA), e uma máquina de maceração - o FP120 FastPrep Cell Disruptor™ (BIO101/Savant Instruments Inc., CA), de acordo com as instruções do fabricante.

Tabela 1. Caracterização das dez localidades e populações de *C. brasiliense* amostradas. As áreas amostrais consistiam em retângulos. Ni – número de indivíduos amostrados em cada população.

Localidade	**Área (ha)**	**Característica da localidade**	**Área da amostra (m^2)**	**N_i**
AGE	10,547	Contínua	2,500	30
CGR	32	fragmentada/isolada/distúrbio	50,000	30
CNV	12,000	Contínua	6,000	30
FAL	4,000	contínua/distúrbio	2,500	41
GSV	84,000	Contínua	1,200	31
ITI	2,300	fragmentada/isolada/distúrbio/ regeneração	8,800	30
MTR	5	fragmentada/isolada/distúrbio/ pastagem	50,000	30
PNB	28,000	contínuo/distúrbio	8,000	32
TOC	10	fragmentada/isolada/distúrbio/ pastagem	100,000	30
URU	5	Fragmentada/isolada/distúrbio	2,500	30



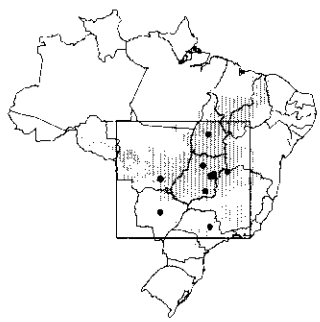


Figura 1. Localização das dez áreas nas quais as populações de *C. brasiliense* foram amostradas. Área cinza representa o bioma Cerrado. As linhas grossas representam os rios principais que passam no domínio do Cerrado e as linhas finas são as divisões dos Estados. Veja o texto para as legendas.

2. Análise genética

Dez locos microsatélites (cb1, cb3, cb5, cb6, cb9, cb11, cb12, cb13, cb20, cb23) previamente desenvolvidos e otimizados para *C. brasiliense* foram utilizados para genotipar todos os indivíduos coletados. Os locos SSR foram desenvolvidos de uma biblioteca genômica enriquecida para a sequência de dinucleotídeo AG/TC (Collevatti *et al*, 1999).

Para todos os experimentos de genotipagem as amplificações por PCR (reação da cadeia da polimerase) foram feitas em um volume total de 13 µl contendo 0.9 µM de cada primer, 1unidade de Taq DNA polimerase (Gibco, MD), 200 µM de cada dNTP, 1X de tampão (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), DMSO 50% e 10.0 ng de DNA. As amplificações foram feitas utilizando um termociclador PT-100 thermal controller (MJ Research) nas seguintes condições: 96°C por 2 min (1 ciclo), 94°C por 1 min, 54 a 56°C por 1 min (de acordo com cada loco), 72°C por 1 min (30 ciclos); e 72°C por 7 min (1 ciclo).

Para determinação dos genótipos, os produtos amplificados foram separados em géis denaturantes de poliacrilamida 4%, corados com prata (Bassam *et al.*, 1991). O tamanho dos alelos foi determinado por comparação com um DNA padrão de 10 pb (DNA ladder standard, Gibco, MD). O tamanho dos alelos foi estimado utilizando o programa SEQAID II (Rhoads & Roufa, 1990) levando em consideração o tamanho esperado dos alelos em pares de base, a partir do primer desenhado e o clone original de DNA do qual o loco SSR foi desenvolvido.

Para o estudo de filogeografia a partir de cpDNA, foram utilizados dez locos microsatélites de cloroplasto desenvolvidos por Weising & Gardner (1999) para angiospermas dicotiledôneas. Inicialmente, foi feita uma otimização, para cada par de primer, para o *C. brasiliense,* a fim de se obter uma amplificação robusta e clara das bandas de DNA. A amplificação por PCR dos microsatélites foi feita em um volume total de 13 µl contendo 0.9 µM de cada primer, 1unidade de Taq DNA polimerase (Gibco, MD), 200 µM de cada dNTP,

1X de tampão (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), DMSO 50% e 10.0 ng de DNA. As amplificações foram feitas, novamente, utilizando um termociclador PT-100 thermal controller (MJ Research) nas seguintes condições: 96°C por 2 min (1 ciclo), 94°C por 1 min, 56°C por 1 min, 72°C por 1 min (30 ciclo); e 72°C por 7 min (1 ciclo). A análise dos fragmentos amplificados foi feita em géis denaturantes de poliacrilamida 4%, corados com prata (Bassam *et al.*, 1991). O tamanho dos alelos foi determinado por comparação com um DNA padrão de 10 pb (DNA ladder standard, Gibco, MD).

Os locos microsatélites de cloroplasto que apresentaram amplificação clara e específica em vários genótipos de *C. brasiliense* foram marcados com pigmentos fluorescentes. Quatro primers foram marcados com fluorescência azul - 6-FAM (o) dye (ccmp1, ccpm2, ccpm5 e ccpm6), três com fluorescência verde - TET (&) dye (ccpm4, ccpm7 e ccpm8) e três com fluorescência amarela HEX (@) dye (ccpm3, ccpm9 e ccpm10). Para todos os experimentos de genotipagem com microsatélites fluorescentes, as reações de PCR foram feitas em um volume total de 10 µl contendo 10.0 µM de cada primer, 1unidade de Taq DNA polimerase (Gibco, MD), 200 µM de cada dNTP, 1X de tampão (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), BSA mg/ml e 10.0 ng de DNA. As amplificações foram feitas nas mesmas condições descritas acima, para cada loco separadamente. Posteriormente, as reações foram diluídas 1:5 em três multiplexes com três locos cada, marcados com diferentes fluorescências. Somente o loco ccmp2 foi diluído separadamente. Um microlitro de cada reação diluída 1:5 foi adicionado a 0,25 µL do marcador interno (GeneScan 500 internal lane standard, ROX, Perkin-Elmer, CA), 0,45 µL de tampão de corrida (25mM EDTA e 50 mg/ml Blue-Dextran) e 2,3 µL de formamida deionizada. As reações foram denaturadas a 95°C por 3 min, mantidas em gelo e submetidas a eletroforese em géis denaturantes de poliacrilamida 5% em um sequenciador automático de DNA (ABI Prism 377 automated

DNA, Perkin-Elmer, CA). Os produtos da PCR fluorescente foram interpretados automaticamente utilizando o programa Genescan (Perkin-Elmer, CA).

Devido a erros na determinação do tamanho dos fragmentos gerados pela amplificação pela Taq polimerase (Taq polymerase slippage) e pelo método de determinação de tamanho de fragmento comparativo, utilizando um marcador interno padrão (Haberl & Tautz, 1999), dois indivíduos da população CNV foram utilizados como controle em todas as amplificações e eletroforeses. Além disso, para todos os locos microsatélites, as amplificações e eletroforeses foram repetidas para quatro indivíduos de cada população, para verificar a variação na determinação do tamanho dos fragmentos. Adicionalmente, 24 indivíduos da população FAL foram genotipados, resultando em um total de 40 indivíduos para esta população, para verificar a existência de outros haplótipos.

3. Análise estatística

Os locos SSR foram caracterizados para o número de alelos por loco e heterozigosidade observada e esperada pelo equilíbrio de Hardy-Weinberg, em cada população (Nei, 1978). Uma vez que os marcadores microsatélites são multialélicos, foi utilizado o teste exato de Fisher para verificar se as heterozigosidades observadas diferiam das esperadas pelo equilíbrio de Hardy-Weinberg (Weir, 1996).

Para estudar a estrutura genética das populações foram estimadas as correlações na freqüência de alelos: θ – a correlação entre as frequências alélicas de diferentes indivíduos da mesma população; F – a correlação de alelos entre diferentes indivíduos de diferentes populações; f – a correlação de alelos dentro de indivíduos dentro de uma mesma população (Cockerham, 1969). Essas correlações podem ser equivalentes às estatísticas de Wright: $\theta=F_{ST}$, o índice de fixação, $F=F_{IT}$ o coeficiente total de endogamia e $f=F_{IS}$, o coeficiente de

endogamia (Wright, 1951, 1965; Cockerham, 1969). A estimativa foi realizada por uma análise de variância utilizando o programa GDA (Lewis & Zaykin, 1999). Os testes de significância para as correlações foram feitas por reamostragem (bootstrapping) com intervalo de confiança igual a 95% (Weir, 1996). Para testar a hipótese de isolamento por distância, foi obtida uma matriz de F_{ST} linearizado entre pares de população (Slatkin, 1991, 1995), pelo procedimento AMOVA do programa Arlequin (Schneider *et al.*, 1997), que foi correlacionada com uma matriz de distância geográfica pelo teste de Mantel (Mantel, 1967).

Uma vez que alelos com tamanhos próximos podem ser confundidos com bandas secundárias (stutter bands), ou podem não ser distinguidos, os heterozigotos podem ser genotipados como homozigotos, levando a uma superestimativa da endogamia. Por outro lado, bandas secundárias (stutter bands) podem ser designadas como diferentes alelos, levando a uma superestimativa do número de alelos por loco e, consequentemente, da heterozigosidade. Dessa forma, uma análise adicional mais conservadora foi realizada, estimando as correlações entre freqüências alélicas, nas qual alelos com baixa freqüência foram consolidados com os alelos mais próximos.

Para análise do sistema de cruzamento, as populações foram analisadas separadamente, utilizando o programa MLTR (Ritland, 1996). Neste programa, foram estimadas as taxas de fecundação cruzada "single locus" e "multilocus" (probabilidade de que cada descendente de uma árvore mãe seja produto de fecundação cruzada), utilizando o modelo de acasalamento misto (Ritland & Jain, 1981; Ritland, 1989).

As taxas de fecundação cruzada "single locus" e "multilocus" foram estimadas por máxima verossimilhança, ajustando as proporções observadas dos genótipos da progênie de um genótipo maternal conhecido, às proporções esperadas pelo modelo de acasalamento misto. O modelo de acasalamento misto assume que: (1) cada acasalamento é um evento aleatório de fecundação cruzada ou auto-fecundação, com probabilidades iguais a t e $(1-t)$,

respectivamente; (2) não há mutação ou seleção após a fertilização; (3) não há acasalamento preferencial (a probabilidade de fecundação cruzada é independente dos genótipos materno ou paterno) e não há variabilidade ou heterogeneidade na freqüência alélica do conjunto de pólen entre árvores mãe (Ritland & Jain, 1981).

Como o programa MLTR aceita somente oito alelos, quando o número de alelos excedeu este valor, o que ocorreu para todos os locos com exceção de cb1 e cb13, os alelos com freqüência abaixo de 0.05 foram consolidados em uma única classe. Este processo foi feito separadamente para cada população, uma vez que as distribuições de freqüência dos alelos foram diferentes entre populações (Collevatti *et al.*, 1999).

Como o genótipo maternal era conhecido, os seguintes parâmetros foram estimados para cada população: taxa de fecundação cruzada "single locus" (t_s) e "multilocus" (t_m); coeficiente de endogamia materna média (f); correlação da taxa de fecundação cruzada dentro das progênies (r_t); correlação de paternidade dentro das progênies (r_p). Além disso, foram estimadas as freqüências alélicas no conjunto de pólen e óvulo de cada população.

Para maximizar a equação de máxima verossimilhança foi utilizada o método de expectativa e maximização e 100 reamostragens (bootstrapping) foram realizadas para obter os erros padrão de cada estimativa.

Para verificar a heterogeneidade na freqüência alélica do conjunto de pólen entre árvores mãe, em cada população, foram construídas tabelas de contingência e realizado um teste exato de Fisher utilizando o método de "Markov Chain", implementado pelo programa GENEPOP (Raymond & Rousset, 1995).

A análise de estrutura genética para o genoma de cloroplasto foi realizada da mesma forma que para o genoma nuclear, assim como o teste de isolamento por distância. Como o cpDNA é não-recombinante e herdado maternalmente, os genótipos foram interpretados como haplótipos.

Para análise filogeográfica os dados de microsatélites de cloroplasto foram designados como caracteres multiestado, não ordenados com pesos iguais, e analisados utilizando o programa Phylogenetic Analysis Using Parsimony (PAUP, Swofford, 1993). Pelo critério de otimização de máxima parsimônia de Fitch, as árvores filogenéticas não enraizadas foram obtidas pelo procedimento de "Exhaustive Search". Posteriormente, a espécie *C. villosum* foi adicionada como um grupo externo (outgroup) e as árvores foram enraizadas, considerando o grupo externo como parafilético em relação ao grupo interno, monofilético. Foi realizada uma análise de reamostragem (bootstrapping), com 100 replicações para obtenção do nível de confidência de cada grupo (a freqüência de ocorrência de cada grupo ou valor de suporte do bootstrap). Neste caso, as populações foram mantidas constantes e os caracteres foram amostrados com reposição, para construir novos dados com o mesmo tamanho dos dados originais. Foi utilizado o método de busca "Heuristic search" para obtenção das árvores mais parsimoniosas e foi obtida uma árvore consenso pelo método do "50%majority-rule".

Para comparar os dados de microsatélite nuclear com a filogenia gerada pelos dados de microsatélites de cloroplasto, foram estimadas as distâncias genéticas de Cavalli-Sforza & Edwards (1967) para os dados de microsatélites nuclear e realizada uma análise de agrupamento pelo método de UPGMA.

Capítulo 1

Development and Characterization of Microsatellite Markers for Genetic Analysis of a Brazilian Endangered Tree Species *Caryocar brasiliense*

Artigo publicado na Heredity (Heredity 83:748-756, 1999)

Autores: Rosane Garcia Collevatti, Rosana Vianello Brondani & Dario Grattapaglia,

Resumo

Nesse trabalho apresentamos o desenvolvimento e caracterização de dez locos microsatélites para a espécie arbórea *Caryocar brasiliense*. A eficiência do desenvolvimento dos marcadores SSR, usando enriquecimento de biblioteca genômica, foi de 14,4%, do sequenciamento à obtenção de locos polimórficos com amplificação clara e robusta. A sequência dos "primers" para esses dez locos são disponibilizados, além das heterozigosidades esperadas, probabilidade de exclusão de paternidade e probabilidade de identidade. A herança Mendeliana e segregação foram confirmadas para os dez locos em uma família de meios-irmãos de polinização aberta, além da transferibilidade absoluta desses locos para cinco espécies do mesmo gênero. O número de alelos por locos variou de 10 a 22, com valor médio igual a 16 e heterozigosidade esperada variando de 0,84 a 0,94. A probabilidade de identidade genética combinada foi da ordem de 10^{-17} demonstrando que os genótipos multiloco são provavelmente únicos e capazes de discriminar prontamente os indivíduos de *C. brasiliense*. O alto valor de probabilidade de exclusão de paternidade combinada (0.99999995) indica que estes marcadores permitem um estudo detalhado de parentesco em populações naturais, mesmo em situações onde ambos os pais são desconhecidos. A bateria de marcadores microsatélites desenvolvida e caracterizada nesse trabalho abre uma nova perspectiva para a geração de dados sobre genética de populações fundamentais para desenvolver estratégias de conservação e manejo para o *C. brasiliense* e espécies relacionadas

redity 83 (1999) 000–000 Received 12 May 1999, accepted 24 August 1999

Development and characterization of microsatellite markers for genetic analysis of a Brazilian endangered tree species *Caryocar brasiliense*

ROSANE GARCIA COLLEVATTI†§, ROSANA VIANELLO BRONDANI‡§ & DÁRIO GRATTAPAGLIA*§

†Departamento de Ecologia, Universidade de Brasília, Brasília 70910-900, D.F. Brasil, ‡Departamento de Biologia Celular, Universidade de Brasília, Brasília 70910-900, D.F. Brasil, and §Laboratório de Genética de Plantas, Recursos Genéticos e Biotecnologia—EMBRAPA, C.P. 02372, Brasília, D.F. Brasil

In this work we report the development and characterization of 10 microsatellite loci for the endangered tree species *Caryocar brasiliense*. Using genomic library enrichment, the efficiency of SSR marker development was 14.4% from sequencing data to operationally useful loci. Primer sequences for this set of 10 loci are made available together with their estimates of expected heterozygosity, probability of paternity exclusion and probability of identity. Mendelian inheritance and segregation was confirmed for all 10 loci in open-pollinated half-sib families as well as the absolute transferability of these 10 loci to five other species of the same genus. Number of alleles per locus ranged from 10 to 22 with a mean value of 16 and expected heterozygosity varying from 0.84 to 0.94. The combined probability of genetic identity was on the order of 10^{-17} clearly demonstrating that SSR multilocus genotypes are likely to be unique and capable of readily discriminating individuals of *C. brasiliense*. The very high combined probability of paternity exclusion (0.99999995) also indicates that these markers will permit detailed parentage studies in natural populations even in situations where both maternity and paternity are unknown. The battery of microsatellite markers developed and characterized in this study opens a new perspective for the generation of fundamental population genetic data for devising sound collection and conservation procedures for *C. brasiliense* and related species of the genus.

Keywords: *Caryocar brasiliense*, Caryocaraceae, Cerrado, microsatellites, tropical tree.

oduction

ocar brasiliense Camb. (Caryocaraceae) is a wide-ad but endangered Brazilian Cerrado tree species, nated by small-sized nectarivorous bats. Seeds are ounded by a woody endocarp coated by a yellow y mesocarp rich in oil and vitamin A, which is eaten everal wild animals, such as greater rhea, macaws, ots, pampas deer and paca (Gribel & Hay, 1993). itionally; this species plays a significant role in the economy of the inhabitants of central Brazil, who he yellow mesocarp as an important source of oil ooking and for home-made recipes for candies, ice-n and liqueur.

e fragmentation of the 'Cerrado' vegetation in l and the higher frequency of fire caused by agricultural practices have been affecting recruitment and ultimately population size and dynamics of *C. brasiliense*, augmented by the intense commerce in its fruits. Habitat fragmentation may reduce genetic variability through genetic bottlenecks. Founder effect, genetic drift and restriction of gene flow, added to the enhancement of inbreeding, may increase population genetic isolation and divergence. Additionally, these genetic hazards may lead to fixation of deleterious alleles, endangering species persistence in habitat fragments (Gilpin & Soulé, 1986; Young *et al.*, 1996). It is therefore fully recognized that a number of genetic parameters such as inbreeding and outbreeding depression, genetic bottlenecks, loss of heterozygosity and adaptability have to be considered together with population demography in 'population vulnerability analysis' (PVA) and the estimation of 'minimum viable population' (MVP) (Gilpin & Soulé, 1986).

espondence. E-mail: dario@cenargen.embrapa.br

B 0 2 6 0 6 3 8 R
Dispatch: 5.11.99 Journal: Heredity
Author received: No. of pages: 9

lecular markers have been increasingly used as a ffective tool for the understanding of population structure, gene flow, parentage, population y and, ultimately, to quantify the effects of t fragmentation and to guide conservation strat- Young *et al.*, 1996; Parker *et al.*, 1998). DNA orphisms based on SSR (Simple Sequence Re- or microsatellites) are one of the most powerful lar markers to estimate genetic parameters of tions and understand detailed patterns of gene d parentage composition. Microsatellites display high content of genetic information, as they are inant and highly multiallelic, with expected het- osity values typically above 0.7. Furthermore they ndant, uniformly distributed in plant genomes, ically transportable among closely related spe- cause of genome sequence homology (Morgante eri, 1993). Additionally, it has been demonstra- t some classes of SSR constitute an important of quantitative genetic variation, coding for nal elements of protein and acting as regulatory s of transcription (Kashi *et al.*, 1997). In plants, otifs of many types are as frequently and widely ted as in human and other mammalian genomes & Renz, 1984; Weber, 1990; Morgante & 1993; Wang *et al.*, 1994). Despite the usefulness osatellite makers for the investigation of popu- enetics and conservation, reports on the devel- characterization and use of SSR loci in tropical cies are still scarce (but see Condit & Hubbell, ase *et al.*, 1996; White & Powell, 1997a; Aldrich 98). This is partly because of the tradition that polymorphisms still hold and the often not so le molecular technologies needed to develop a of polymorphic microsatellite markers.

re interested in understanding the population structure, patterns of gene flow and mating of *C. brasiliense*, in order to generate useful ion for conservation strategies. As part of this e report here the development, characterization ritance of a battery of highly polymorphic SSR *C. brasiliense*. Besides estimating the genetic ion content of this set of markers for the study c structure and parentage analysis, we investi- transferability of these loci to other species of genus.

ls and methods

aterial and DNA extraction

development total genomic DNA was extracted anded leaves of a single individual tree of *ense*, sampled at Água Limpa Forestry Park (15°57′12″S, 47°56′35″W), Brasília, Brazil. For SSR loci characterization, at least 30 individuals per population from four populations 200–1000 km apart (a total of 123 individuals) were used. These four populations were: (1) Campus of the Federal University of Mato Grosso do Sul (20°30′24″S, 54°36′53″W), Campo Grande; (2) Itirapina Ecological Reserve (22°13′13″S, 47°51′03″W), São Paulo: (3) Brasília National Park (15°44′26″S, 47°59′19″W), Brasília; (4) Grandes Sertões Veredas National Park (15°13′29″S, 45°49′12″W), Minas Gerais. Genomic DNA extraction from expanded leaves followed standard CTAB procedure (Doyle & Doyle, 1987) both for SSR development and genotyping experiments.

Construction of SSR-enriched genomic libraries

Protocols described by Rafalski *et al.* (1996) and optimized for tropical tree genomes at Embrapa — Genetic Resource and Biotechnology (Brondani *et al.*, 1998) were used. DNA from an individual of *C. brasiliense* was digested with three different restriction enzymes, *Mse*I, *Tsp*509 and *Sau*3A, according to manufacturer's instructions, in order to select one that would produce a larger amount of fractionated DNA in the range of 280–600 bp. Approximately 50 μg of genomic DNA was digested with *Mse*I (TTAA), and fragments between 280 and 600 bp were recovered by DEAE-cellulose NA-45 membrane (Schleicher and Schuell, NY) via electrophoresis on 2% agarose gel. Around 30 μg of DNA fragments were ligated to adaptors to the *Mse*I restriction site. Fragments containing SSR sequences were selected by hybridization with biotinylated oligonucleotides complementary to the repetitive sequence AG/CT, and recovered by magnetic beads linked to streptavidine. Fragments were amplified by PCR and cloned in the plasmid vector pGEM-T (Stratagene, CA) and then transformed by electroporation into *E. coli* strain XL1-Blue and grown on ampicillin and tetracycline containing agar plates. Transformants were picked, streaked on 132-mm plates (100 per plate) and regrown at 37°C for 12 h. Duplicate plates containing colonies from these transformants were stamped onto positively charged nylon membranes (Hybond N, Amersham Pharmacia), grown, lysed, denatured, neutralized and UV cross-linked.

Selection of recombinants for repeat sequences

Recombinant colonies having SSR were identified by hybridization with a poly (dA-dG) probe labelled with Digoxigenin-11-ddUTP using a DIG oligonucleotide 3′-end labelling Kit (Boehringer Mannheim) according to the manufacturer's instructions. The temperature

for prehybridization and hybridization was 65°C
e poly AG/TC oligonucleotide. Processed mem-
were exposed to X-ray film for 2–3 h at 37°C.

encing of positive clones rimer design

e clones were picked and grown overnight in ampicillin LB media. Plasmid DNA was extracted Wizard Minipreps (Promega Co., WI). DNA were sequenced on an Applied Biosystems 377 n-Elmer, CA) instrument using dye-terminator cent chemistry. Oligonucleotides complementary repeats were designed using the software 'PRIMER' In *et al.*, 1991). To reduce problems with spurious g patterns generated during amplification and to later development of single-reaction multiplex some stringent criteria in primer sequence design pplied: (i) primer T_m of 72°C; (ii) 3°C difference in ween primer pairs; (iii) GC content ranging from o 60%; and (iv) absence of complementarity n primers. Primers were synthesized by Operon logies Inc. (Alameda, CA).

r screening and PCR amplification

two individuals of *C. brasiliense*, randomly from the 123 individuals (eight individuals per ion), were used for primer screening. Microsat-mplification was performed in a 13 μL reaction taining 0.9 μM of each primer, 1 unit Taq DNA rase, 200 μM of each dNTP, 1% reaction buffer Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), 50% and 10.0 ng of template DNA. Amplifica-re performed using a PT-100 thermal controller search) with the following conditions: 96°C for 1 cycle), 94°C for 1 min, 54 or 56°C for 1 min ing to each primer pair annealing temperature), r 1 min (30 cycles); and 72°C for 7 min (1 cycle). rimer pair was initially screened for product rphism, and the annealing temperature was later ed to produce clear and robust DNA band ation in all loci. Analysis of amplified fragments d by these amplifications was carried out in Metaphor gels stained with ethidium bromide mL^{-1}) and sized by comparison to a 1 kb DNA tandard (Gibco, MD). For genotype determin-d precise estimates of allele sizes, the amplified s were separated on 4% denaturing polyacryl-gels stained with silver nitrate (Bassam *et al.*, nd sized by comparison to a 10-bp DNA ladder l (Gibco, MD) on a computer screen. Allele ere estimated using the software SEQAID II & Roufa, 1990) taking into consideration the expected allelic series in base pairs from the primer designed and the original DNA clone from which the SSR locus was developed.

Analysis of inheritance and transferability of SSR loci

To verify the inheritance of the microsatellites developed, we examined the segregation in an open-pollinated half-sib family. Sixteen seeds were collected from a mother tree and DNA was extracted directly from the embryo because of the low germination of dormant seeds. For DNA extraction from embryos we used the Fast DNA™ Kit H, and FP120 FastPrep Cell Disruptor™ (BIO101/Savant Instruments Inc., CA), according to manufacturer's instructions. PCR amplification and visualization of allele segregation followed the same protocols used for leaves.

To test the transferability of SSR loci we extracted DNA from leaves of eight individuals of five species of the same genus: *C. coriaceum*, from Riachão das Neves, Bahia, *C. edule*, Porto Seguro, Bahia, and *C. glabrum*, *C. pallidum and C. villosum*, from Manaus, Amazônia. PCR amplification followed the same protocol used for *C. brasiliense*. Transferability was visualized both in 3.5% agarose gels and on silver-stained 4% denaturing polyacrylamide gels.

SSR loci characterization

Ten selected SSR loci were characterized for number of alleles per locus, allelic frequency and observed and expected heterozygosities under Hardy–Weinberg (Nei, 1978), using the 123 individuals of *C. brasiliense*. Genetic analyses were carried out using the software Genetic Data Analysis (GDA: Lewis & Zaykin, 1998). Based on estimated allele frequencies, two parameters of genetic information content for parentage studies were estimated for each locus: (i) probability of genetic identity (I) (Paetkau *et al.*, 1995), which corresponds to the probability of two random individuals displaying the same genotype; and (ii) paternity exclusion probability (Q) (Weir, 1996), which corresponds to the power with which a locus excludes an individual tree of being the parent of an offspring. The combined probability of paternity exclusion, $QC = 1 - [\prod(1 - Q_i)]$ and the combined probability of genetic identity $IC = \prod I_i$ were also estimated for the combined battery of loci.

Results

SSR development

Digestion of the *C. brasiliense* genome with three different enzymes revealed that *Mse*I produced the most

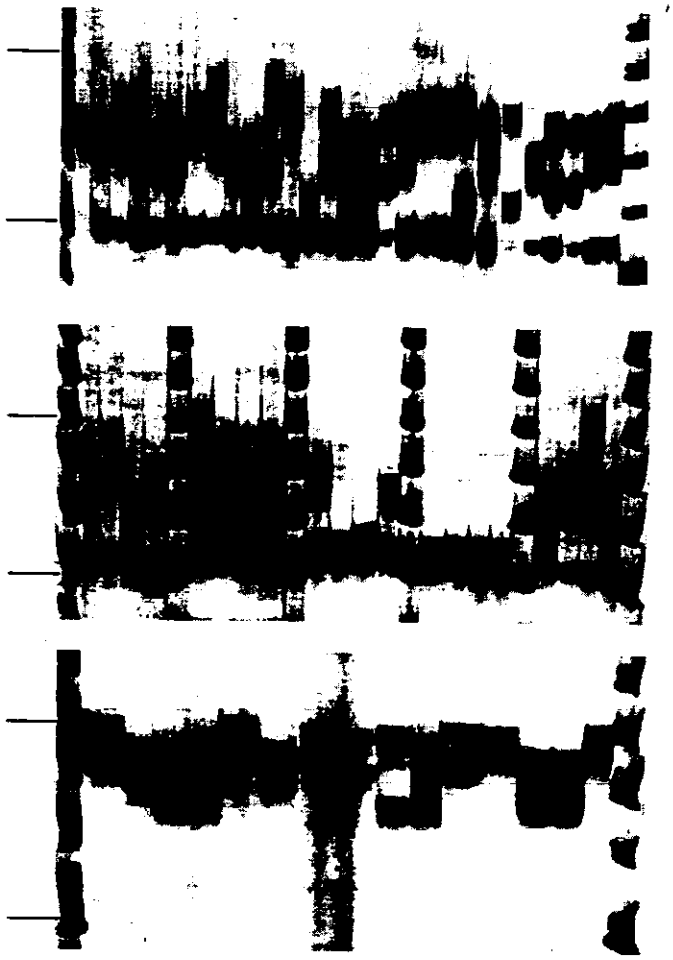


Fig. 1 Microsatellite polymorphisms in *Caryocar brasiliense* for locus *cb6* visualized in silver-stained denaturing polyacrylamide gels. First and last lanes are 10 bp ladder (Gibco). (a) DNA polymorphism for 28 unrelated individuals; (b) transferability of the locus to *C. coriaceum* (lanes 2–5), *C. edule* (lanes 7–10), *C. glabrum* (lanes 11–14), *C. pallidum* (lanes 16–19) and *C. villosum* (lanes 21–24); (c) inheritance and segregation in a open-pollinated half-sib family, lane 2 maternal tree followed by 14 progeny individuals.

e digestion profile for library development, with nts ranging from 200 to 800 bp. One library d for AG microsatellite repeats was constructed. he enrichment step, screening of 1000 recombiolonies with the AG/TC probe detected 195 clones (19.5%). From the 195 positive colonies ed, 28 were useful sequences (14.4%) from rimer pairs were designed. In 19 clones no SSRs und in the sequences, suggesting that positives were misidentified. In 78 clones the repeated calization was too close to the end of the insert t enough sequence for primer design. In 51 high quality DNA sequence could not be d for both regions flanking the SSR. Finally, e primer sequences could not be designed from A sequences of 19 clones.

nces of the DNA inserts containing microsatelwed three categories of repeats as classified by (1990). Twenty-two inserts contained single repeats (with no interruption in the repeat e) and four microsatellites were imperfect (with ion in repeat sequence), but none of them was a interpretable locus. Two microsatellites were compound repeats (with different repeats in tandem), one of them was monomorphic, and the other was polymorphic. Primers were named using the prefix cb (from *Caryocar brasiliense*).

Screening of SSR and allele size determination

From 28 primer pairs developed for *C. brasiliense* 15 (54%) were amplified using a single PCR protocol and generated clearly interpretable products. Of these, five (33%) were monomorphic in a sample of 32 individuals and the other 10 were polymorphic showing clear allele size variation (Fig. 1a). From the 10 loci characterized, only *cb12* was a compound microsatellite (Table 1). The other nine were perfect microsatellites, with repeat motif size ranging from 18 to 28 (Table 1).

Inheritance, transferability and characterization

Inheritance was verified for all 10 SSR loci by analysing a heterozygous mother tree for the locus and it

e 1 Primer sequences, repeat motifs, expected fragment size from sequencing data and observed size range in detected s, annealing temperature (T_a) and total number of alleles (A) for the 10 SSR loci developed for *Caryocar brasiliense*

s	Repeat motif	Primer pair sequences (5′–3′)	Fragment size and range (bp)	T_a (°C)	A
	$(AG)_{27}$	ggTgTgAgCTTAgAgCTgAA gTCCAgCTTAATgTCCgACT	189 150–195	54	18
	$(AG)_{25}$	CAgCCATggTTCACgTTAgT CgCACATggAAACgCTTA	158 130–175	56	19
	$(AG)_{19}$	gTCAgAATgAAggCAgCTTg ATAgAATCCAggCCACACCA	153 130–180	56	16
	$(AG)_{19}$	CTACCACAACTCggAgACAA gACACTCCTgCAACTCCATT	123 105–160	56	14
	$(AG)_{21}$	ATCgAgATgAgCCAACCgAC ggAAggTgTTgCAgCACTgA	92 55–95	56	13
	$(AG)_{28}$	ggTCgTTATTgCTgTggT gTgAACATgAgCATCggT	176 135–185	56	15
	$(AG)_{21}(AC)_9$	GACATgTggCAATAggCggT TTgTgTgTgAAggTgTgTTggTT	179 150–210	56	20
	$(AG)_{15}$	AgCAgTTAgATgATgAAgTg CTATTgCCATATgTCgTAgT	145 125–170	54	15
	$(AG)_{22}$	TgACACAACCATCACATTCT gCAACTgTCgCAATAAACAA	164 140–185	56	17
	$(AG)_{18}$	ATACCAgCTCTgACAgAA AAgCCTgAgAgTAgAgAA	153 110–185	56	23

pollinated half-sib family (Fig. 1b). All sibs dis- one of the maternal alleles, confirming Mende- heritance and suggesting no seed contamination. *cb6* and *cb12*, however, showed more than two in the profile suggesting locus duplication possi- cause of the ancient polyploid nature of the *car* genome (see below) or DNA or chromosome ation. At locus *cb12*, interpretation of polymor- was not a problem as the second locus was norphic. However at *cb6* interpretation of geno- for some individuals was impossible as alleles at ci co-migrated to the same position in the gel. All i were fully transferable to the five species d, displaying clear genotypes using the same ols for PCR amplification (Fig. 1c).

sidering the 10 SSR loci analysed in the present a total of 123 individuals genotyped and the -pollinated half-sib family, the least and the most e loci displayed 13 (*cb9*) and 23 alleles (*cb23*), ively (Table 1). Progeny individuals genotyped in eritance analysis displayed alleles that were not d in individuals genotyped in the characterization *cb1*, *cb6* and *cb23* — one allele, *cb9* — three, *cb13*). All loci presented three or four more frequent (Fig. 2), except the most polymorphic loci (*cb12*, *cb20*, *cb23*), with a more uniform frequency distribution. For *cb9*, the locus with the lowest number of alleles, only one allele (allele *60*) represented more than 30.0% of the total. Paternity exclusion probabilities ranged from 0.69 to 0.95 with a combined value (over all loci) of 0.99999995 (Table 2). As expected, the two loci with the lowest number of alleles, *cb9* with 10 alleles and *cb13* with 11, displayed the lowest values of paternity exclusion probability and the highest values of probability of genetic identity (Table 2). Probability of genetic identity ranged from 0.01 to 0.4 with a combined value (over all loci) of 3.1×10^{-17}.

Discussion

Our results show that AG sequence repeats in the *C. brasiliense* genome are relatively abundant and therefore amenable to isolation for the development of microsatellite markers. Primer pairs that amplified easily-interpretable markers were developed for 14.4% of the sequenced plasmid clones from an enriched library. An anchor-PCR screening prior to sequencing could have significantly improved the yield of useful sequences by eliminating false-positives and positive sequences wit repeated motifs positioned too close to the vector

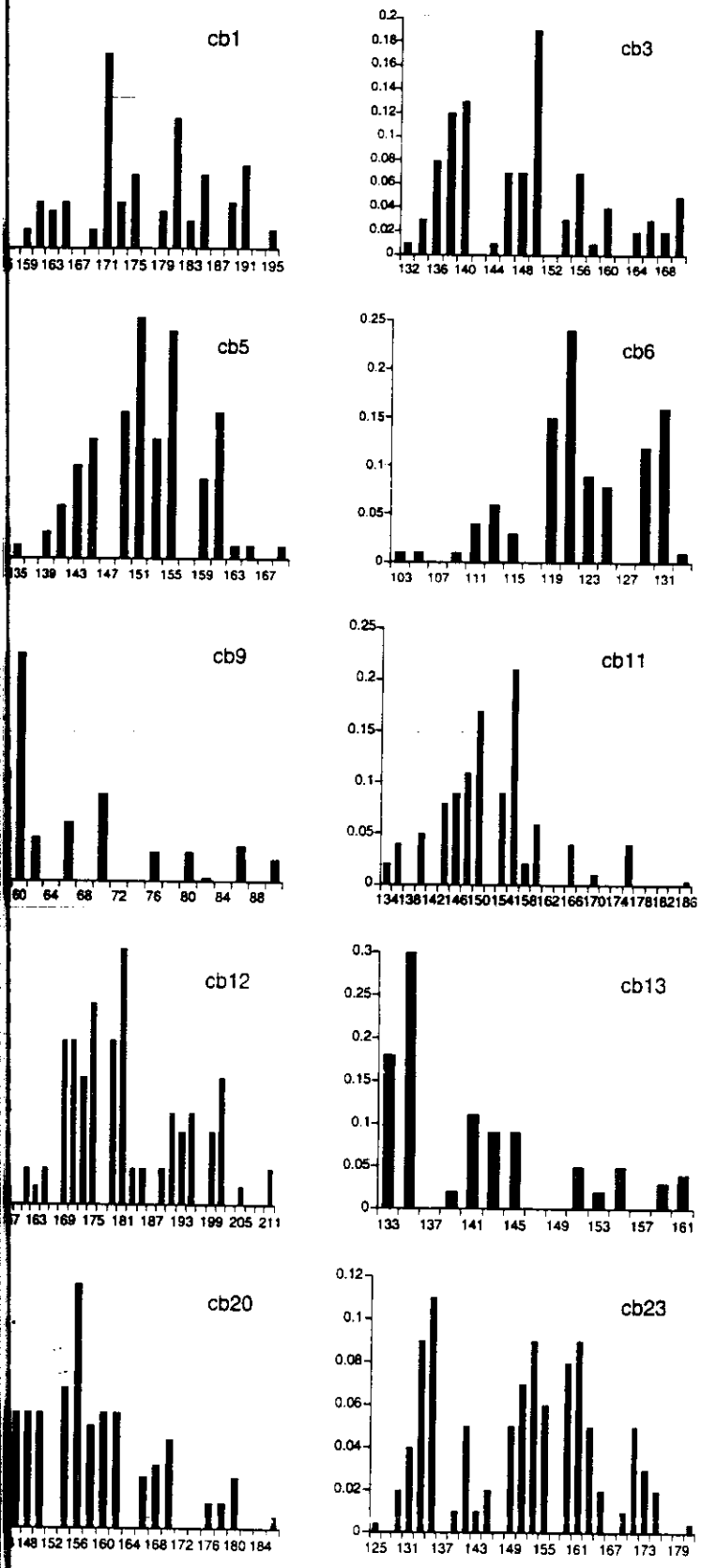


Fig. 2 Allele frequency distribution for the 10 SSR loci. *x*-axis, allele size in base pairs; *y*-axis, allele frequency.

le 2 Characterization of 10 SSR loci of *Caryocar iliense*, based on a sample of 123 unrelated individuals

cus	A	H_e	H_o	Q	I
	17	0.90	0.83	0.79	0.02
	19	0.91	0.95	0.79	0.02
	16	0.89	0.71	0.79	0.02
	13	0.86	0.81	0.74	0.04
	10	0.84	0.72	0.70	0.04
1	15	0.88	0.77	0.80	0.03
2	20	0.94	0.56	0.84	0.01
3	11	0.84	0.59	0.69	0.04
0	17	0.91	0.64	0.95	0.02
3	22	0.94	0.69	0.85	0.01
n	16.0	0.89	0.73	QC = 0.99999995	IC = 3.1×10^{-17}

umber of alleles; H_o, observed heterozygosity; H_e, expected ozygosity; Q, probability of paternity exclusion; QC, ined probability of paternity exclusion; I, probability of ic identity; IC, combined probability of genetic identity.

ndani *et al.*, 1998). However, given the likelihood of rving high expected heterozygosity values at the SSR and considering the research objectives contemplated this programme — population genetic struc- — only a relatively small number of loci was needed, at an intense screening of positive clones prior to ncing was deemed unnecessary.

crosatellite polymorphisms in *C. brasiliense* were ted by silver staining. Attempts were made to use escence labelling and semiautomated detection of loci. Both methods offer advantages and limitations. staining detection on polyacrylamide gels is an ible technology and allows the genetic analysis of s that show nonspecific fragment amplification se of low quality of DNA, as these products usually te off the allele size range. On the other hand, rapid generation by multiplexing of several loci in a single ne is limited. Fluorescence-based DNA detection the potential of multifluorescence multiplex yping ability and precise allele size determination hell *et al.*, 1997). On the other hand, fluorescence ion is not so accessible as silver staining because of igh cost of equipment for detection and the high y of DNA required. This last aspect was the limiting for employing fluorescence labelling and detection R loci in *C. brasiliense*. The DNA obtained from and seeds of *C. brasiliense* typically was minated with polysaccharides and polyphenols that hard to remove and seriously affected PCR. As al attempts were made to optimize conditions for escence detection and yielded unsatisfactory s, we chose to perform the genotyping work on -stained gels, because this method was found to yield robust data and to be significantly less influenced by DNA contamination. Moreover, adoption of this marker technology by other research groups in tropical countries should be significantly more straightforward if based on silver staining.

The most variable locus in *C. brasiliense* (*cb23* with 23 alleles) was the shortest one in number of repeat units, displaying only 18 AG repeats. Although the number of microsatellite loci surveyed in this study was limited, this observation does not support the view that the number of alleles per locus is positively correlated with the number of repeat motifs (Weber, 1990; Taramino & Tingey, 1996). In fact, this relationship is controversial because the size of a nonrepeat portion of the amplified fragment may be different among loci (Valdes *et al.*, 1993; Goldstein & Pollock, 1997).

Transferability of microsatellite loci between closely related species is a consequence of the homology of flanking regions of simple sequence repeats. Other studies in tropical trees have already demonstrated the high rate of transferability of SSR loci among taxonomically related tree species, such as in the Leguminosae (Dayanandan *et al.*, 1997), Meliaceae (White & Powell, 1997b) and among *Eucalyptus* species (Brondani *et al.*, 1998). The absolute transferability (100%) of the microsatellite loci developed for *C. brasiliense* to five other species of the genus (*C. coriaceum*, *C. edule*, *C. glabrum*, *C. pallidum* and *C. villosum*) indicates a high level of genome homology and will allow comparative studies of population genetic structure in all these species. *Caryocar brasiliense* and *C. villosum* have been described as having a high similarity in chromosome number and karyotype. Both are polyploid, as are most of the species of the order Theales, with $2n = 46$ (Ehrendorfer *et al.*, 1984). Despite polyploidy and therefore potential locus duplication, most of the microsatellite loci showed amplification from a unique site with the exception of loci *cb6* and *cb12*. This result suggests that the polyploidization event is a relatively ancient one and that sufficient time has passed to allow sequence divergence of the duplicated genomes. Alternatively, allopolyploidy might have occurred between species with disparate genomes such that only one set of homeologues contains a site that can be amplified.

Relatively high levels of multiallelism were observed at all 10 SSR loci developed. Mean number of alleles per locus (16.0, for 10 loci) and expected heterozygosity range (0.84–0.94) were higher than those found by White & Powell (1997a) for *Swietenia humilis*, an endangered tropical hardwood species in Central America (9.7 alleles per locus for 10 loci). The broad range observed in expected heterozygosity values results from the broad variation in number of alleles per locus, and allele frequency distribution within populations.

ith smaller numbers of alleles or with a skewed ncy distribution such as *cb9* and *cb13*, tend to lower heterozygosity values and consequently probability of paternity exclusion, and higher ility of genetic identity. The number of alleles per eported in this study is most likely a minimum Because of the widespread distribution of the in Brazil the number of alleles should increase ew populations are sampled. Indeed, new alleles sly undetected in the adults were seen when ng the progeny individuals.

combined probability of genetic identity, i.e. the ility that two individuals drawn at random from lation have identical multilocus genotypes at all i, was on the order of 10^{-17}. This clearly strates that SSR multilocus genotypes will be and capable of readily discriminating individuals *rasiliense*. This excellent power of discrimination ry useful tool to identify precisely clonality in populations. In *C. brasiliense* this battery of tellite markers should therefore allow the precise cation of clonal regeneration arising by root ng, a common event in some ecosystems and with nt implications for conservation strategies. The gh combined power of paternity exclusion 9995) also indicates that these markers will detailed parentage studies in natural populations, situations where both maternity and paternity are n *a priori*. Furthermore, the exact determination ntage of regenerants and seeds will allow a understanding of reproductive success of adults e dynamics of genetic structure of natural ions. In conclusion, the microsatellite markers ed and characterized in this study open a new tive for generating fundamental data to devise conservation procedures for *C. brasiliense* and species of the genus.

wledgements

dy is part of the Doctoral Thesis of the first and has been supported by the Universidade de WWF (World Wildlife Fund), CNPq/PADCT n Ministry of Science and Technology) competnt 62.00059/97-4 to D.G. and CAPES (National for Research and Technology) whose assistance fully acknowledge. We wish to thank John Hay nued support, Paul Lewis and Dmitri Zaykin for orization for the use of the GDA software as well nio Dimas Costa Jr, Fabian Borghetti, Flavia rederico S. Lopes, Gaidino R. Santos, Maura C. and Neilton R. Oliveira for their kind help in k. We also acknowledge the support of IBAMA n Ministry of Environment) and IEF-SP (State Forest Institute of São Paulo) for authorizing plant collections in Brasília and Grandes Sertões Veredas National Parks and in Itirapina Ecological Station.

References

ALDRICH, P., HAMRICK, J. L., CHAVARRIAGA, P. AND KOCHERT, G. 1998. Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree *Symphonia globulifera*. *Mol. Ecol.*, 7, 933–944.

BASSAM, B. J., CAETANO-ANOLLES, G. AND GRESSHOFF, P. M. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Analyt. Biochem.*, **196**, 80–83.

BRONDANI, R. P. V., BRONDANI, C., TARCHINI, R. AND GRATTAPAGLIA, D. 1998. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theor. Appl. Genet.*, **97**, 816–827.

CHASE, M., KESSELI, R. AND BAWA, K. 1996. Microsatellite markers for population and conservation genetics of tropical trees. *Am. J. Bot.*, **83**, 51–57.

CONDIT, R. AND HUBBELL, H. 1991. Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. *Genome*, **34**, 66–71.

DAYANANDAN, S., BAWA, K. S. AND KESSELI, R. 1997. Conservation of microsatellites among tropical trees (Leguminosae). *Am. J. Bot.*, **84**, 1658–1663.

DOYLE, J. J. AND DOYLE, J. L. 1987. Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13–15.

EHRENDORFER, F., MORAWETZ, W. AND DAWE, J. 1984. The neotropical angiosperm families Brunelliaceae and Caryocaraceae: first karyosystematical data and affinities. *Pl. Syst. Evol.*, **145**, 183–191.

GILPIN, M. E. AND SOULÉ, M. E. 1986. Minimum viable populations: process of species extinction. In: Soulé, M. E. (ed.) *Conservation Biology, the Science of Scarcity and Diversity*, pp. 19–34. Sinauer Associates, Sunderland, MA.

GOLDSTEIN, D. B. AND POLLOCK, D. D. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *J. Hered.*, **88**, 335–342.

GRIBEL, R. AND HAY, J. D. 1993. Pollination ecology of *Caryocar brasiliense* (Caryocaraceae) in Central Brazil cerrado vegetation. *J. Trop. Ecol.*, **9**, 199–211.

KASHI, Y., KING, D. AND SOLLER, M. 1997. Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.*, 13, 74–78.

LEWIS, P. AND ZAYKIN, D. 1998. *Genetic data analysis: computer program for the analysis of allelic data*. Version 1.0 (d12). Free program distributed by the authors over the internet from the GDA Home Page at http://chee.unm.edu/gda/.

LINCOLN, S. E., DALY, M. J. AND LANDER, E. S. 1991. *'Primer' software*. Available from E. Lander, Whitehead Institute, Cambridge, MA.

MITCHELL, S. E., KRESOVICH, S., JESTER, C. A., HERNANDEZ, C. J. AND SZEWC-McFADDEN, A. K. 1997. Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. *Crop Sci.*, **37**, 617–624.

NTE, M. AND OLIVIERI, A. M. 1993. PCR-amplified osatellites as markers in plant genetics. *Plant J.*, **3**, 182.

1978. Estimation of average heterozygosity and genetic nce from a small number of individual. *Genetics*, **89**, 590.

U, D., CALVERT, W., STIRLING, I. AND STROBECK, C. 1995. osatellite analysis of population structure in Canadian bears. *Mol. Ecol.*, **4**, 347–354.

, P. G., SNOW, A. A., SCHUG, M. D., BOOTON, G. C. AND ST, P. A. 1998. What molecules can tell us about lations: choosing and using a molecular marker. *gy*, **79**, 361–382.

KI, J. A., MORGANTE, M., POWELL, W., VOGEL, J. M. AND EY, S. V. 1996. Generating and using DNA markers in s. In: Birren, B. and Lai, E. (eds) *Analysis of Non-malian Genomes – A Practical Guide*, pp. 75–134. emic Press, New York.

, D. D. AND ROUFA, D. J. 1990. SEQAID II 3.80. Molecular tics Laboratory, Kansas State University, New York.

NO, G. AND TINGEY, S. 1996. Simple sequence repeats for plasm analysis and mapping in maize. *Genome*, **39**, 87.

TAUTZ, D. AND RENZ, M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucl. Acids Res.*, **12**, 4127–4138.

VALDES, A. M., SLATKIN, M. AND FREIMER, N. B. 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics*, **133**, 737–749.

WANG, Z., WEBER, J. L., ZHONG, G. AND TANKSLEY, S. D. 1994. Survey of plant short tandem repeats. *Theor. Appl. Genet.*, **88**, 1–6.

WEBER, J. L. 1990. Informativeness of human $(dC\text{-}dA)_n$ $(dG\text{-}dT)_n$ polymorphisms. *Genomics*, **7**, 524–530.

WEIR, B. S. 1996. *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.

WHITE, G. AND POWELL, W. 1997a. Isolation and characterization of microsatellite loci in *Swietenia humilis* (Meliaceae): an endangered tropical hardwood species. *Mol. Ecol.*, **6**, 851–860.

WHITE, G. AND POWELL, W. 1997b. Cross-species amplification of SSR loci in the Meliaceae family. *Mol. Ecol.*, **6**, 1195–1197.

YOUNG, A., BOYLE, T. AND BROWN, T. 1996. The population genetic consequences of habitat fragmentation for plants. *Trends Ecol. Evol.*, **11**, 413–418.

Capítulo 2

Population Genetic Structure of the Endangered Tropical Tree Species *Caryocar brasiliense*, Based on Variability at Microsatellite Loci

Artigo submetido a Molecular Ecology em Novembro de 1999

Autores: Rosane Garcia Collevatti, Dario Grattapaglia & John DuVall Hay,

Resumo

Nesse trabalho estudamos a estrutura genética de populações da espécie arbórea *Caryocar brasiliense*, baseado na variabilidade em dez locos microsatélites. Para um total de 314 indivíduos em dez populações o número de alelos por loco variou de 20 a 27 e a heterozigosidade esperada e observada variou entre 0,129 a 0,924 e 0,226 a 1,000, respectivamente. O coeficiente de endogamia (f) foi significativamente maior que zero. Valores significativos, porém baixos de θ e valores altos de F ($F>\theta$) sugerem que a diferenciação ocorre principalmente por acasalamento entre indivíduos aparentados. Conforme esperado pelo modelo de isolamento por distância, a matriz de F_{ST} entre pares de populações foi positivamente correlacionada com a matriz de distância geográfica. Esses dados sugerem que a restrição ao fluxo gênico contribui para o estabelecimento de pequenas unidades panmíticas em *C. brasiliense*. Como as sementes são dispersas principalmente por gravidade e animais terrestres, a fragmentação do Cerrado deve estar limitando significativamente o fluxo gênico mantendo os dispersores de sementes e as espécies de morcegos polinizadores, de pequeno tamanho corporal e territoriais, isolados dentro dos fragmentos, aumentando a taxa de acasalamento entre indivíduos aparentados. A manutenção de populações não isoladas em grandes áreas preservadas deve ser necessário para permitir o fluxo gênico por pólen e semente e, consequentemente manter a viabilidade da espécie no Cerrado.

Population genetic structure of the endangered tropical tree species *Caryocar brasiliense*, based on variability at microsatellite loci

OSANE G. COLLEVATTI[1,3], DARIO GRATTAPAGLIA[2] AND JOHN D. HAY[1]

Departamento de Ecologia, Universidade de Brasília, Brasília, D.F. Brasil, 0910-900. E-mail: rosanegc@hotmail.com

Laboratório de Genética de Plantas, Recursos Genéticos e Biotecnologia –EMBRAPA, C.P.02372, Brasília, D.F. Brasil. 70770-900

Pós-Graduação em Biotecnologia e Recursos Genéticos, Universidade Católica de Brasília, Brasília, DF. E-mail: rosanegc@pos.ucb.br

Key words: *Caryocar brasiliense,* Caryocaraceae, fragmentation, microsatellites, population genetic structure, tropical tree

Send proofs to: Rosane Garcia Collevatti, *Pós-Graduação em Biotecnologia e Recursos Genéticos, Universidade Católica de Brasília, SGAN 916, Mod. B, Asa Norte, Brasília, DF. 70790-160. E-mail:* rosanegc@pos.ucb.br Phone and FAX: +55613405550.

Running title: Population genetic structure of *C. brasiliense*

Abstract

In this work we report the population genetic structure of the endangered tree species *Caryocar brasiliense*, based on variability at ten microsatellite loci, and discuss the consequences for conservation. For a total of 314 individuals in 10 populations, number of alleles per locus ranged from 20 to 27 and expected and observed heterozygosity varied from 0.129 to 0.924 and 0.226 to 1.000, respectively. A significant amount of inbreeding was found ($f>0$). Significant but low values of θ and high values of F ($F>\theta$) suggest that differentiation occurs mainly due to mating between closely related individuals and not to genetic drift. As expected under the isolation by distance model, pairwise F_{ST} values were significantly and positively correlated with geographical distance. These data suggest that restriction in gene flow contribute for the establishment of small panmitic units in *C. brasiliense*. As seeds are dispersed mainly by gravity and terrestrial animals, cerrado fragmentation may thus be significantly limiting gene flow by keeping seed dispersers and territorial small sized bat pollinators isolated inside fragments, increasing the rate of mating between close relatives. No fragmentation effect could be detected in our study. We suggest that the lack of fragmentation effect on genetic structure is a result of the short time since Cerrado fragmentation, since it is a relatively recent event (~60 years) compared to the species life cycle. In addition, populations at fragmented area are composed mainly by adult individuals, prior to fragmentation. Nevertheless, population genetic structure indicated that the maintenance of non-isolated populations in many and large preserved areas should provide adequate opportunity for gene flow by pollen and seed and consequently maintain species viability in Cerrado biome.

Introduction

It is fully recognised that almost all organisms are patchily distributed. This may be caused by the discontinuity of habitat, with favourable areas surrounded by unfavourable ones, or by behaviour features, such as foraging and mating behaviours forming groups such as herds, flocks and colonies (Pickett & White 1985; Krebs & Davies 1997). This ecological spatial distribution may result in a spatial structure in allele frequency caused by patterns of gene flow or migration, differential selection among patches, genetic drift and mating system that determine inbreeding strength (Wright 1931, 1943). Even in species with continuous spatial distribution, "interbreeding is restricted to small distances by the occurrence of only short range means of dispersal" (Wright 1943). In this case, populations may become differentiated by a processes know as "isolation by distance" (Wright 1943). Hence, genetic differentiation among populations may be the outcome of gene flow restriction, and could be correlated to geographical distance.

Fragmentation of tropical communities, because of harvesting for wood products or agricultural expansion, has been changing the original landscape to a mosaic of fragments of remnant habitats surrounded by unfavourable ones (reviewed in Laurance & Bierregaard 1997). In this process, ancient larger populations are diminished and subdivided leading to the isolation of populations. Habitat fragmentation may reduce genetic variability through genetic bottlenecks. Subsequently, founder effect, genetic drift and gene flow restriction may increase population genetic isolation and divergence. These genetic hazards, added to the enhancement of endogamy, may lead to fixation of

deleterious alleles, endangering species persistence in fragments and jeopardizing their conservation (Gilpin & Soulé 1986).

Despite considerable efforts in conservation of tropical rain forests (e.g. Laurance & Bierregaard 1997, and references therein), the conservation of other threatened biomes has been neglected. Particularly in Brazil, the Cerrado biome is threatened to an even greater extent than the Amazonian forest, because of the rapid and intensive expansion of agricultural land in Central Brazil (Ratter *et al.* 1997).

The Brazilian Cerrado biome covers nearly 22% of the Brazilian territory (2 million km^2) consisting of a very heterogeneous vegetation with nearly 160,000 species, including plants, animals and fungi. The number of tree and shrub species in savannah-like vegetation (cerrado *sensu stricto*) exceeds 800, nearly 40% of them are endemic (Ratter *et al.* 1997). Despite this high biodiversity and endemism, this biome has been fragmented due to agricultural expansion resulting in continental islands of wild habitat surrounded by an "ocean" of crops, causing a fragmentation of the ecosystem and jeopardizing species viability.

Caryocar brasiliense Camb. (Caryocaraceae) is a widely distributed but endangered Brazilian Cerrado tree species, locally distributed in well delimited patches. This species plays an important role in economy of Central Brazil been a source of raw material for small and middle-sized industries (Araújo 1994). The seeds are surrounded by a woody endocarp coated with a yellow fleshy mesocarp rich in oil and vitamin A, and are eaten by several wild animals. *C. brasiliense* is completely outcrossed (Collevatti *et al.* 2000, in press), flowers are hermaphroditic and pollination is done mainly by small sized glossophagine bats (*Glossophaga soricina* and *Anoura geoffroyi*) (Gribel & Hay 1993). Despite its high abundance, ecological and economic importance, Cerrado

fragmentation and the higher frequency of fire due to agricultural practices have been affecting recruitment and ultimately population size and dynamics of this species, augmented by the intense commerce of fruits (Araújo 1994).

We are interested in understanding the population genetic structure, gene flow and mating system of *C. brasiliense*, in order to generate useful information for conservation strategies. In this study we report the population genetic structure of *C. brasiliense* in fragmented and continuous areas of Cerrado, and discuss the consequences for conservation. Our working hypothesis was that populations are significantly differentiated due to "isolation by distance" (Wright 1943). Genetic data was generated using microsatellite markers (Simple Sequence Repeats or SSR), one of the most powerful molecular markers to estimate genetic parameters and describe population genetic structure and gene flow (Morgante & Olivieri 1993).

Materials and Methods

Populations, sampling and DNA extraction

For the analysis of population genetic structure, one population per locality (total of ten) throughout the whole geographical distribution of *C. brasiliense*, with different histories of disturb and fragmentation were surveyed (Fig. 1, Table 1): three continuous areas - AGE - Águas Emendadas Ecological Station, Brasilia; CNV - State Park of Caldas Novas, Goiás; GSV – Grandes Sertões Veredas National Park, Minas Gerais; two continuous area but with a recent history of anthropic disturb - FAL - Água Limpa Forestry Park, Brasilia; PNB – Brasília National Park; and five fragmented and isolated areas with high degree of anthropic disturb - CGR – Campus of the Federal University of Mato Grosso do Sul, Campo Grande; ITI – Itirapina Ecological Reserve, São Paulo; MTR - Rondonópolis, Mato Grosso; TOC - Porto Nacional, Tocantins; URU - Serra da

Mesa Hydroelectric Powerplant, Uruaçú, Goiás. In each of these populations all individuals (at least 30) were marked and expanded leaves were collected and stored at –80°C. Genomic DNA extraction followed standard CTAB procedure (Doyle & Doyle 1987).

Genetic analysis

Ten SSR loci previously developed and optimised for *C. brasiliense* were used to genotype all individuals. SSR loci were developed from a genomic library enriched for the dinucleotide sequence motif poly AG/poly TC. Development assay, optimisation and characterisation of these loci were published elsewhere (Collevatti *et al* 1999).

For all genotyping experiments, microsatellite amplifications were performed in a 13 µl volume containing 0.9 µM of each primer, 1unit Taq DNA polymerase (Gibco, MD), 200 µM of each dNTP, 1X reaction buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), DMSO 50% and 10.0 ng of template DNA. Amplifications were performed using a PT-100 thermal controller (MJ Research) with the following conditions: 96°C for 2 min (1 cycle), 94°C for 1 min, 54 to 56°C for 1 min (according to each locus), 72°C for 1 min (30 cycles); and 72°C for 7 min (1 cycle).

For genotype determination the amplified products were separated on 4% denaturing polyacrylamide gels stained with silver nitrate (Bassam *et al.* 1991) and sized by comparison to a 10 bp DNA ladder standard (Gibco, MD) on a computer screen. Allele sizes were estimated using the software SEQAID II (Rhoads & Roufa 1990) taking into consideration the expected allelic series in base pairs from the primer designed and the original DNA clone from which the SSR locus was developed.

Statistical analysis

SSR loci were characterised for number of alleles per locus and expected and observed heterozygosities under Hardy-Weinberg for each locus and population (Nei 1978). Because we were dealing with highly multiallelic marker loci, departures from Hardy-Weinberg expected heterozygosity were tested with a Fisher exact test, to prevent against type I error due to low frequencies of some alleles (Weir 1996). Population were characterised for expected and observed heterozygosities and inbreeding coefficient (f) overall loci.

Population genetic structure was assessed using correlation of gene frequencies: θ – the correlation between allele frequencies of different individuals in the same population; F – the correlation of alleles within random individuals from different populations; f – the correlation of alleles within individual within populations (Cockerham 1969). These correlations may be equivalent to Wright's F-statistics as following: $\theta=F_{ST}$, the fixation index, $F=F_{IT}$ the total inbreeding coefficient and $f=F_{IS}$, the inbreeding coefficient (Wright 1951, 1965; Cockerham 1969). The estimation followed an analysis of variance, which was carried out using the software GDA (Lewis & Zaykin 1999). Significance tests of correlations were performed by bootstrapping over loci with 95% nominal confidence interval (Weir 1996). To test the hypothesis of "isolation by distance" a matrix of Slatkin's pairwise linearized F_{ST} (Slatkin 1991, 1995) was obtained by the AMOVA procedure using the software Arlequin (Schneider *et al.* 1997) and correlated to a geographical distance matrix by Mantel test (Mantel 1967).

Since alleles that are close in size may be confused with stutter bands or not distinguished, heterozygotes may be scored as homozygotes, leading to an overestimate of inbreeding. On the other hand, stutter bands may be scored as a different allele

leading to overestimate of number of alleles per locus and heterozygosity. In this manner, an additional more conservative analysis was performed, estimating correlation of gene frequencies, in which alleles with low frequencies were consolidated with the immediately close allele.

Results

All ten microsatellite loci used in this work displayed high polymorphism both within and among populations (Fig. 2). By genotyping a total of 314 individuals from 10 populations, the least and the most variable loci displayed respectively 20 (cb5 and cb6) and 27 alleles (cb23) (Table 2). Mean expected and observed heterozygosity, for all populations and loci, ranged from 0.129 to 0.924 and from 0.226 to 1.000, respectively. For many loci, observed heterozygosity was lower than expected under Hardy-Weinberg equilibrium. However, some loci displayed an excess of heterozygotes: cb12 and cb23 in AGE; cb1 and cb3 in CGR; cb6, cb12 and cb23 in CNV; cb11, cb20 and cb23 in FAL; cb3 in GSV and ITI; cb11 and cb12 in MTR; cb3 and cb6 in PNB; cb12 and cb20 in TOC, cb1, cb6, cb9, cb11, cb12, cb20 and cb23 in URU.

Fragmented populations did not present lower observed heterozygosities and higher inbreeding coefficient, than continuous populations (Table 3). Although CGR and ITI presented higher inbreeding coefficient, for MTR and URU there was no significant inbreeding. Additionally, PNB a continuous but disturbed area presented high inbreeding coefficient.

Overall, a significant amount of inbreeding was found ($f>0$), hence alleles within groups did not unite at random (Table 4). Likewise, values of θ and F were different

from zero i.e. alleles are not randomly distributed among or within populations, indicating population subdivision (Table 4). As expected under the isolation by distance model, pairwise F_{ST} values were significantly and positively correlated with geographical distance (Fig. 3, r=0.518, t=1.429, p=0.0396).

Correlations of gene frequencies did not change when a more conservative analysis was performed.

Discussion

The ten SSR loci used for genetic analyses in this work detected high levels of allelic variation confirming the high expected genetic information content of these markers for the study of *C. brasiliense* population structure and gene flow (Collevatti *et al.* 1999).

The broad range observed and expected heterozygosity values results from the broad variation in the number of alleles per locus, and allele frequency distribution within populations. Loci with smaller numbers of alleles or with a skewed frequency distribution, i.e. a few alleles with high frequency, tend to have low heterozygosity values (Nei 1978). Furthermore, observed departures from Hardy-Weinberg at some loci may most likely derive from statistical sampling (Weir 1996), resulting from the fact that the SSR marker loci used are highly multiallelic – much more than typical isozyme loci - and the number of individuals genotyped per population (~30) is relatively limited to efficiently sample all possible genotypes at a locus.

For some loci, an excess of heterozygotes was found and for others an excess of homozygotes. Excess of heterozygous individuals and negative values of f may indicate

that selective forces within population are acting at these loci (Lewontin & Cockerham 1959). Nevertheless, inbreeding coefficient (f) indicated that, in general, alleles within populations are not united at random, when all loci combined were analysed, or when populations where analysed separately overall loci. This result suggests that mating between close relatives is playing an important role in the determination of the genetic structure of these populations (Weir 1996). This result was not an artefact of unbiased allele size determination, since results for a more conservative allele size determination (consolidating alleles with low frequency) was virtually the same as the result obtained for all detected alleles. Additionally, there was not a tendency of higher inbreeding values for more variable loci, as found by Aldrich *et al.* (1998).

Our results indicate that alleles are not randomly distributed among and within groups, since θ and F were significantly different from zero (Cockerham 1969). As we expected, populations are differentiated by genetic drift or differential selective forces, since $\theta \neq 0$ (Weir 1996). In spite of that, θ presented low values and $F > \theta$ indicating little differentiation among populations due to non-random distribution of alleles, but a system in which mates are more related than expected by random distribution (Cockerham 1969). Nevertheless, because microsatellites are highly variable and subject to high mutation rates ($4N_e u > 1$), they usually display high levels of within-population heterozygosity. Measures of differentiation (θ or F_{ST}) tend therefore to be small. In other words, F_{ST} may be strongly influenced by the amount of variation at a locus (Hedrick 1999).

Mean number of migrants per generation (Nm), estimated under Wright's equilibrium island model of migration (Wright 1943), using F_{ST} overall loci, was greater than one (3.32). Clearly, the significant spatial genetic structure and positive correlation

between population differentiation and geographical distance observed in our study violate Wright's island model of migration (Whitlock & McCauley 1999), so that these estimates of Nm are most likely not legitimate. Consequently, one could expected higher values between closer populations and lower values between more distant populations.

Our genetic data suggest that *C. brasiliense* populations in the Cerrado of Central Brazil are differentiated by a process of "isolation by distance". This hypothesis is supported by a significant Mantel test between a matrix of Slatkin's pairwise linearized F_{ST} and a matrix of geographical distance. Although the species is widely distributed, "interbreeding is restricted to small distances by the occurrence of only short range means of dispersal" (Wright 1943). This isolation and restriction in gene flow may increase the effects of genetic drift and maybe selective forces, leading to non-zero values of θ.

Paradoxically however, there are evidences from pollination tests (Gribel & Hay 1993) and from genetic data analyses (Collevatti *et al.* 2000, in press) that *C. brasiliense* has a preferentially allogamous mating system with high outcrossing rate. We suggest that, as indicated by the estimated inbreeding coefficient and by mating system study (Collevatti *et al.* 2000, in press) gametes do not unit at random not because of self-pollination, but because of outcrossed mating between relatives. So, as predicted by Wright's isolation by distance model, *C. brasiliense* tends to present small panmitic units or demes on account of restriction in gene flow. The restriction in gene flow may occur mainly by restricted seed dispersal. Although seeds could be dispersed by greater-rhea (*Rhea americana*), the great majority of seeds are dispersed by gravity, and tend to remain under the mother tree canopy (JD Hay, unpublished). Furthermore, "Cerrado"

fragmentation could increased gene flow restriction, isolating fragments with a surrounding matrix of crops and encouraging human access to fragments for fruit gathering. Additionally, bats that pollinate *C. brasiliense* belong to territorial small sized species and have low flight range (Gribel & Hay 1993). Fragmentation may be isolating these mammals and keeping them inside fragments (Kearns *et al.* 1998) increasing the rate of mating between close relatives.

Inbreeding coefficients were higher in more disturbed populations, such as ITI and CGR, but not in MTR, TOC and URU, and not significantly different from zero in more preserved and continuous areas, such as GSV and AGE. Furthermore, although large and continuous, some specific populations also displayed a high inbreeding coefficient. In PNB, a large area of National Park inside Brazilian Federal District, events of fire are frequent, and there is an intense and illegal fruit gathering by the local population. Additionally, there is evidence that this area was used for wood harvesting during the recent settlement of Brazilian Federal District, which occurred only 50 to 60 years ago. Therefore, the results of population subdivision observed may have been highly influenced by the process of fragmentation and anthropic disturbs. FAL has the same characteristics of PNB, nevertheless, f was not significant different form zero. MTR and TOC are highly fragmented and disturbed areas only with isolated remnant individuals in a pasture. In these areas only old adult individuals were found (circumference at 30 cm above ground higher than 1.0m) and f was low or not significant. As fragmentation of Cerrado is a recent event (around 60 years) for a species with long life cycle such as *C. brasiliense*, most likely these individuals are significantly older than the first anthropic disturbances. In other populations, on the other hand, a large number of individuals sampled were juveniles, with less than 10 cm of circumference at 30 cm

above ground (e.g. 25% in CGR; 80% in ITI; 50% in PNB). In these areas, the effect of fragmentation could already be detected by a significant inbreeding and reduction of heterozygosity. Additionally, GSV presented the higher but not significant f. We suggest that, as discussed above, because of the high polymorphism of the SSR loci used in this work, results may be highly influenced by sampling effect, since sample size may be relatively limited to efficiently sample all possible genotypes at each locus.

Evidences that fragmentation may induce changes in genetic structure, gene flow and mating structure of tropical tree populations were found by other authors (e.g. Hall 1994; Nason & Hamrick 1997; Aldrich *et al.* 1998; Aldrich & Hamrick 1998). Nevertheless, we suggest that fragmentation effect on *C. brasiliense* populations are difficult to be detected because this species presents a deme structure, determined by mating between close relatives, as indicated by high value of f. Additionally, studies of mating system indicated that, although highly outcrossed, *C. brasiliense* presents high biparental inbreeding. Thus, as fragmentation is still recent, adequate hypothesis testing on the effect of habitat fragmentation will require the recurrent analysis of juveniles across generations in these fragmented versus non fragmented areas.

However, we suggest that fragmentation may lead to isolation among *C. brasiliense* populations, decreasing effective population size (Wang & Caballero 1999), increasing inbreeding, which potentially lead to local extinction. Therefore, the maintenance of non-isolated populations in many and large preserved areas, providing means for gene flow by pollen and seed to occur, may be necessary to maintain species viability in Cerrado biome. Additionally, these areas will play an important role in the maintenance of small populations in highly disturbed areas, as a source of pollen and seeds. Likewise,

these large areas should guarantee the viability of populations of pollinators and seed disperses.

Acknowledgements

This work was supported by WWF (World Wildlife Fund), CNPq/PADCT (Brazilian Ministry of Science and Technology) competitive grant 62.00059/97-4 to D.G., and a fellowship to R.G.C., whose assistance we gratefully acknowledge. We wish to thank Célia M. T. Cordeiro for assistance in statistical analysis, Rosana P.V. Brondani for continued support in the laboratory, Paul Lewis and Dmitri Zaykin for the authorization for the use of the GDA software as well as Aécio A. Santos, Antônio Dimas Costa Jr., Bruno Walter, Eugenia C. Amaral, Fabian Borghetti, Flavia Pinto, Frederico S. Lopes, Galdino R. Santos, João B. Pereira, Maura C. Almeida, Neilton R. Oliveira, Sebastião P.Cordovil, for their kind help in field work. We also acknowledge the support of IBAMA (Brazilian Ministry of Environment), IEF-SP (State Forest Institute of São Paulo), FEMAGO (State Environmental Intitute of Goiás) and SEMATEC (Environmental Institute of Brasília) for authorizing plant collections in Brasília and Grandes Sertões Veredas National Parks, in Itirapina Ecological Station, in State Park of Caldas Novas, and in Águas Emendadas Ecological Station.

References

Aldrich P, Hamrick JL (1998)Reproductive dominance of pasture trees in a fragmented tropical mosaic. *Science* **281**, 103-105.

Aldrich P, Hamrick JL, Chavarriaga P, Kochert G (1998) Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree *Symphonia globulifera*. *Molecular Ecology* **7**, 933-944.

Araújo FD (1994) A review of *Caryocar brasiliense* Camb. (Caryocaraceae) – an economically valuable species of the Central Brazilian cerrados. *Economic Botany* **49**, 40-48.

Bassam BJ, Caetano-Anolles G, Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels *Annals of Biochemical* **196**, 80-83.

Cocherham CC (1969) Variance of gene frequencies *Evolution* **23**, 72-84.

Collevatti RG, Brondani RVP, Grattapaglia D (1999) Development and characterization of microsatellite markers for genetic analysis of a Brazilian endangered tree species *Caryocar brasiliense Heredity* **83**, 748-756.

Collevatti RG, Grattapaglia D, Hay JD (2000) High resolution microsatellite based analysis of mating system allows the detection of significant biparental inbreeding in populations of *Caryocar brasiliense*, an endangered tropical tree species. *Heredity,* in press.

Doyle JJ, Doyle JL (1987) Isolation of plant DNA from fresh tissue *Focus* **12**, 13-15.

Gilpin ME, Soulé ME (1986) Minimum viable populations: process of species extinction. In: *Conservation Biology, the Science of Scarcity and Diversity* (ed. Soulé ME), pp. 19-34. Sinauer Associates, Sunderlands, MA.

Gribel R, Hay JD (1993) Pollination ecology of *Caryocar brasiliense* (Caryocaraceae) in Central Brazil cerrado vegetation. *Journal of Tropical Ecology* **9**, 199-211.

Hedrick PW (1999) Highly variable loci and their interpretation in evolution and conservation *Evolution* **53**, 313-318.

Kearns CA, Inouye DW, Waser NM (1998) Endangered mutualisms: The conservation of plant-pollinator interaction *Annual Review of Ecology and Systematics* **29**, 83-112.

Krebs JR, Davies NB (eds) (1997) *Behavioural Ecology*. 4th ed. Blackwell Science, Oxford.

Laurance WF, Bierregaard RO (eds) (1997) *Tropical Forest Remnants*. Chicago University Press, Chicago.

Lewis P, Zaykin D (1999) *Genetic data analysis: computer program for the analysis of allelic data. Version 1.0 (d12)*. Free program distributed by the authors over the internet from the GDA Home Page at http://chee.unm.edu/gda/

Lewontin RC, Cockerham CC (1959) The goodness-of-fit test for detecting natural selection in random mating populations *Evolution* **13**, 561-564.

Mantel NA (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209-220.

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics *Plant Journal* **3**, 175-182.

Nason JD, Hamrick JL (1997) Reproductive and genetic consequences of forest fragmentation: two case studies of Neotropical canopy trees *Journal of Heredity* **88**, 264-276.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individual *Genetics* **89**, 583-590.

Pickett STA, White PS (eds) (1985) *The Ecology of Natural Disturbance and Patch Dynamics*. Academic Press, San Diego, CA.

Ratter JA, Ribeiro JF, Bridgewater S (1997) The Brazilian cerrado vegetation and threats to its biodiversity *Annals of Botany* **80**, 223-230.

Rhoads DD, Roufa DJ (1990) *Seqaid II 3.80*. Molecular Genetics Laboratory, Kansas State University, Manhattan.

Schneider S, Kueffer JM, Roessli D, Excoffier L (1997) *Arlequin Ver. 1.1: A software for population genetic data analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Slatkin M (1991) Inbreeding coefficients and coalescence times *Genetic Research of Cambridge* **58**, 167-175.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies *Genetics* **139**, 457-462.

Wang J, Caballero A (1999) Developments in predicting the effective size of subdivided populations *Heredity* **82**, 212-226.

Weir BS (1996) *Genetic Data Analysis II*. Sinauer Associates, Sunderland, Massachusetts.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{st} \neq 1/(4Nm+1)$ *Heredity* **82**, 117-125.

Wright S (1931) Evolution in mendelian population *Genetics* **16**, 97-159.

Wright S (1943) Isolation by distance *Genetics* **28**, 114-138.

Wright S (1951) The genetic structure of populations *Annual Eugenics* **15**, 323-354.

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating *Evolution* **19**, 395-420.

Wright S (1969) *Evolution and the Genetics of Populations,* vol. II. The University of Chicago Press, London.

Author Information Box

This study is part of the Doctoral Thesis of Rosane G. Collevatti who is involved in the application of molecular markers to population genetics, phylogeny and conservation of tropical tree species. Dr. Dario Grattapaglia is a geneticist involved in the development and use of molecular markers for population genetics and genomic mapping. Dr. John D. Hay is an ecologist and works with demography and dynamics of tropical tree species. This work is part of a project of the EMBRAPA - Genetic Resources and Biotechnology Centre, involving the development of microsatellites markers for tropical tree species as a tool for population genetic structure, mating system and phylogeny, and to generate useful information for conservation strategies.

Figure Legends

Fig. 1. Localisation of the ten areas in which analysed populations of *C. brasiliense* were surveyed. Grey area represents the Cerrado biome. Thick lines are the main rivers of Cerrado, thin lines are State divisions. See text for population legends.

Fig. 2 Microsatellite polymorphisms at locus cb12 for 30 unrelated individuals of CNV population of *C. brasiliense*. Silver stained polyacrylamide denaturing gel; first lane is a 10 bp ladder size standard (Gibco, MD).

Fig. 3 Relationship between Slatkin's linearized pairwise F_{st} and geographical distance in km among populations of *C. brasiliense* (Mantel test of correlation, r=0.518, t=1.429, p=0.0396).

Table 1. Characterisation of the ten *C. brasiliense* localities and populations sampled. Sample areas consisted of rectangles. Ni – number of individuals sampled in each population.

Population locality	Area (ha)	Locality characteristics	Sample Size (m^2)	N_i
AGE	10,547	continuous	2,500	30
CGR	32	fragmented/isolated/disturbed	50,000	30
CNV	12,000	continuous	6,000	30
FAL	4,000	continuous/disturbed	2,500	41
GSV	84,000	continuous	1,200	31
ITI	2,300	fragmented/isolated/disturbed regeneration	8,800	30
MTR	5	fragmented/isolated/disturbed remnants in a pasture	50,000	30
PNB	28,000	continuous/disturbed	8,000	32
TOC	10	fragmented/isolated/disturbed remnants in a pasture	100,000	30
URU	5	fragmented/isolated/disturbed	2,500	30

Table 2. Genetic characterisation of ten SSR loci of *C. brasiliense*, pooling individuals from ten populations. N – number of individual plants; A – total number of alleles; He – expected heterozygosity; Ho - observed heterozygosity. He and Ho range across populations.

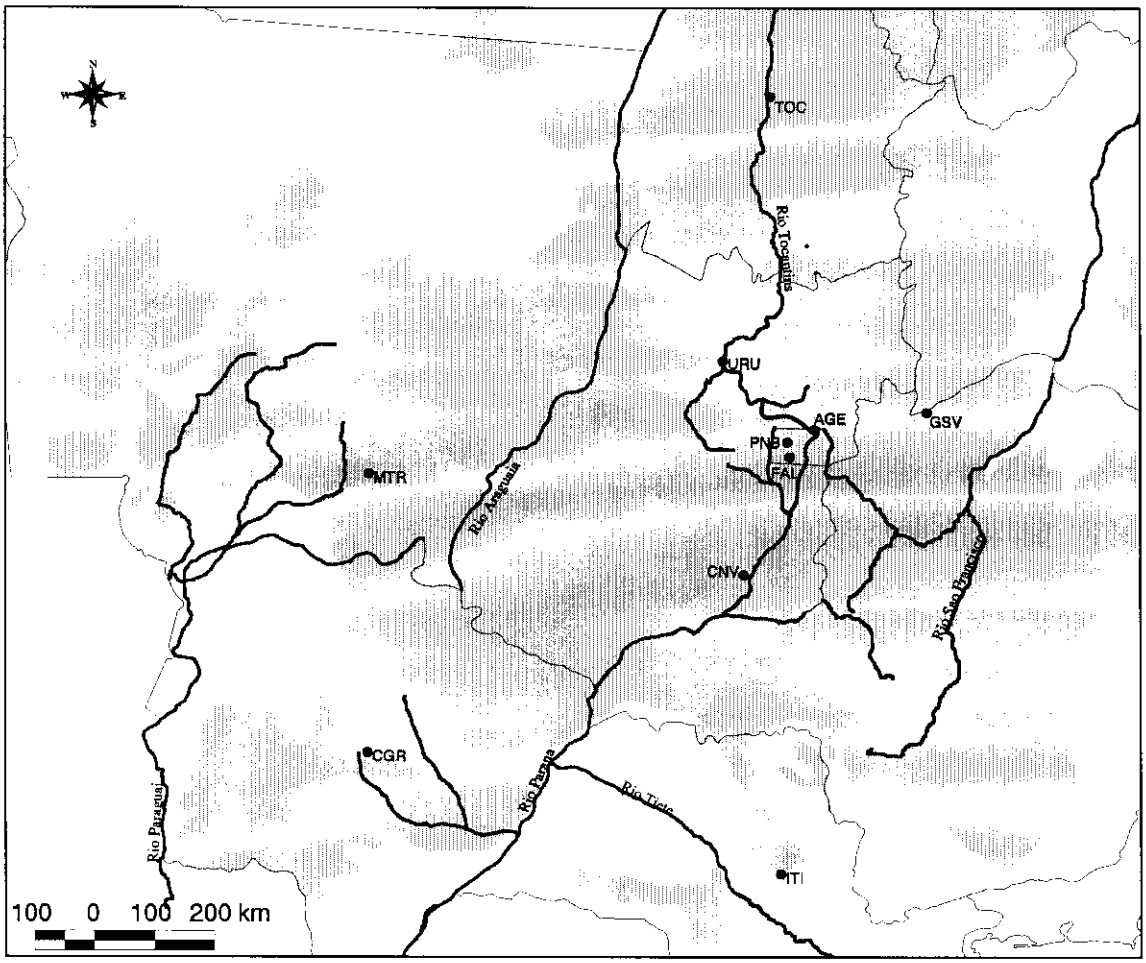
Locus	N	A	He	He Range	Ho	Ho Range
cb1	284	26	0.785	0.853-0.890	0.705	0.633-0.967
cb3	314	26	0.881	0.850-0.909	0.819	0.512-1.000
cb5	314	20	0.866	0.746-0.921	0.751	0.500-0.909
cb6	284	20	0.778	0.812-0.915	0.762	0.733 -1.000
cb9	314	24	0.789	0.557-0.902	0.758	0.400-0.969
cb11	314	23	0.879	0.822-0.913	0.824	0.645-0.951
cb12	314	26	0.898	0.842-0.924	0.767	0.226-0.967
cb13	314	23	0.731	0.129-0.899	0.516	0.067-0.709
cb20	314	22	0.882	0.756-0.917	0.783	0.533-0.967
cb23	314	27	0.907	0.861-0.924	0.823	0.452-1.000
Overall loci			0.856		0.765	

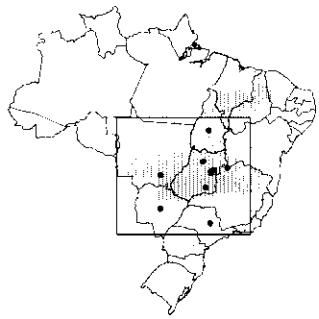
Table 3. Characterisation of ten populations of *C. brasiliense* based on ten SSR loci, pooling all loci. N – number of individual plants; A – mean number of alleles; He – expected heterozygosity; Ho - observed heterozygosity; f – inbreeding coefficient, overall loci; C.I. nominal confidence interval = 95% obtained by bootstrapping over loci (999 replicates).

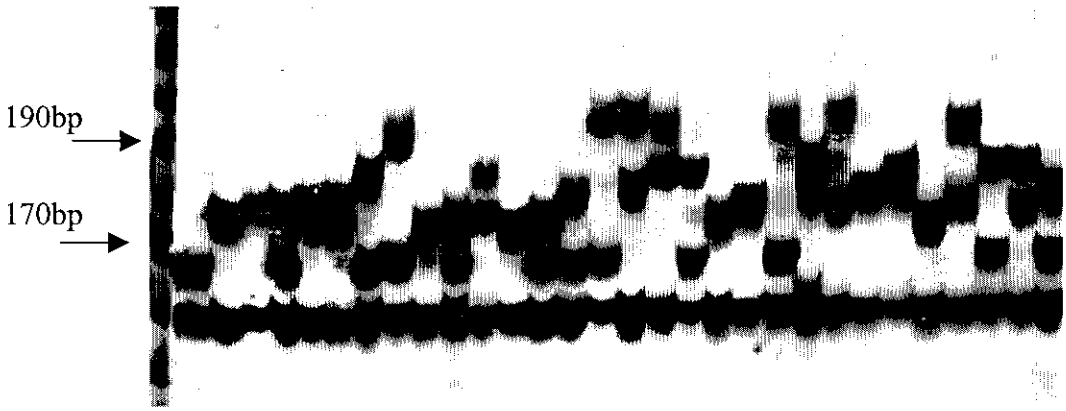
Population	N	A	He	Ho	f	C.I.
AGE	30	10.0	0.865	0.800	0.076^{ns}	-0.004 - 0.175
CGR	30	10.1	0.844	0.726	0.142	0.035 - 0.259
CNV	30	11.0	0.871	0.790	0.095	0.019 - 0.197
FAL	41	11.0	0.879	0.800	0.092^{ns}	-0.005 - 0.195
GSV	31	10.6	0.855	0.709	0.172^{ns}	-0.024 - 0.329
ITI	30	10.4	0.839	0.697	0.172	0.049 - 0.301
MTR	30	10.1	0.751	0.719	0.044^{ns}	-0.078 - 0.128
PNB	32	11.5	0.882	0.766	0.134	0.023 - 0.242
TOC	30	10.9	0.889	0.807	0.095	0.019 - 0.168
URU	30	10.2	0.881	0.840	0.047^{ns}	-0.026 - 0.127
Overall population	31.4	10.6	0.856	0.765		

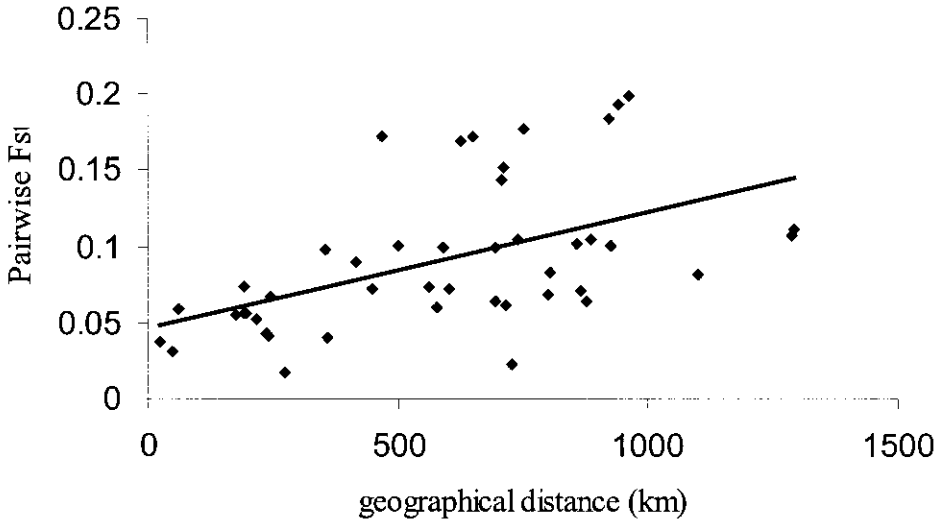
Table 4. Population genetic structure of *C. brasiliense*, based on an analysis of variance of allele frequencies for each SSR locus. f – inbreeding coefficient; F – total inbreeding coefficient; θ – fixation index. Upper and lower bounds obtained by bootstrapping over loci, number of replicates=999, nominal confidence interval=95%.

Locus	f	F	θ
cb1	0.10	0.16	0.06
cb3	0.08	0.13	0.06
cb5	0.13	0.15	0.02
cb6	0.02	0.06	0.03
cb9	0.04	0.19	0.17
cb11	0.06	0.09	0.03
cb12	0.15	0.19	0.04
cb13	0.29	0.44	0.21
cb20	0.11	0.15	0.05
cb23	0.10	0.17	0.07
Overall loci	0.11	0.17	0.07
Upper Bound	0.148	0.244	0.123
Lower Bound	0.067	0.123	0.040


TOC
Rio Tocantins
URU
AGE
GSV
PNG
FAL
MTR
Rio Araguaia
CNV
Rio São Francisco
CGR
Rio Paraguai
Rio Paranaíba
Rio Tietê
ITI
100 0 100 200 km




190bp
170bp


0.25
0.2
0.15
0.1
0.05
0
Pairwise Fst
0
500
1000
1500
geographical distance (km)

Capítulo 3

High Resolution Microsatellite Based Analysis of Mating System Allows the Detection of Significant Biparental Inbreeding in *Caryocar brasiliense*, an Endangered Tropical Tree Species

Artigo aceito pela Heredity em Janeiro de 2000.

Autores: Rosane Garcia Collevatti, Dario Grattapaglia & John DuVall Hay,

Resumo

Nesse trabalho investigamos o sistema de cruzamento da espécie arbórea tropical *Caryocar brasiliense*, utilizando dados genéticos de dez locos microsatélites. Foram amostrados oito a dez famílias de meios-irmãos de polinização aberta, por população, e dezesseis meios-irmãos por família. Os parâmetros do sistema de acasalamento foram estimados considerando o modelo de acasalamento misto, implementado no programa MLTR. A taxa de fecundação cruzada "singlelocus" (t_s) variou entre locos e populações, mas a taxa fecundação cruzada "multilocus" (t_m) foi igual a 1,0 para todas as populações. Entretanto, a endogamia biparental (t_m-t_s) foi diferente de zero para todas as populações, indicando que os eventos de polinização cruzada podem ocorrer entre indivíduos aparentados. Nossos resultados indicam que o alto polimorfismo dos marcadores microsatélites possuem uma resolução extraordinária para discriminar precisamente eventos de auto-polinização e eventos de polinização cruzada entre indivíduos aparentados. Os resultados sugerem que a fragmentação do Cerrado pode limitar o fluxo gênico isolando os dispersores de sementes e as espécies de morcegos polinizadores de pequeno tamanho corporal e territoriais dentro dos fragmentos e aumentando a taxa de cruzamento entre indivíduos aparentados. A conservação de populações não isoladas em grandes áreas preservadas deve ser necessário para permitir eventos de cruzamento entre indivíduos não aparentados e manter a viabilidade da espécie.

High resolution microsatellite based analysis of the mating system allows the detection of significant biparental inbreeding in *Caryocar brasiliense*, an endangered tropical tree species

ROSANE GARCIA COLLEVATTI[1,3], DARIO GRATTAPAGLIA[2] AND JOHN DUVALL. HAY[1]

[1]*Departamento de Ecologia, Universidade de Brasília, Brasília, D.F. Brasil, 70910-900. jdhay@guarany.unb.br*

[2]*Laboratório de Genética de Plantas, Recursos Genéticos e Biotecnologia – EMBRAPA, C.P.02372, Brasília, D.F. Brasil. 70770-900.*

E-mail: dario@cenargen.embrapa.br

[3]*Pós-Graduação em Biotecnologia e Recursos Genéticos, Universidade Católica de Brasília, Brasília, DF. E-mail: rosanegc@pos.ucb.br*

Send proofs to: Rosane Garcia Collevatti, *Pós-Graduação em Biotecnologia e Recursos Genéticos, Universidade Católica de Brasília, SGAN 916, Mod. B, Asa Norte, Brasília, DF. 70790-160. E-mail:* rosanegc@pos.ucb.br Phone and FAX: +55613405550.

Key words: *Caryocar brasiliense,* Caryocaraceae, fragmentation, mating system, microsatellites, tropical tree

Running title: Mating system in *Caryocar brasiliense*

Word count: Main text - 2.850 words

Summary - 217 words

Summary

In this work we investigated the mating system of four populations of the endangered tropical tree species *Caryocar brasiliense*, using genetic data from ten microsatellite loci. Eight to ten open-pollinated progeny arrays of 16 individuals, together with their mother tree, were sampled per population. Mating system parameters were estimated under the mixed mating model, implemented by the software MLTR. The single-locus outcrossing rate (t_s) varied among loci and populations, but multilocus outcrossing rates (t_m) were equal to one for all four populations. Nevertheless, biparental inbreeding (t_m-t_s) was different from zero for all populations, indicating that outcrossing events may occur between relatives. Our results also indicate that the high polymorphism of microsatellite markers provide an extraordinary resolution to discriminate precisely selfing events from outcrossing events between close relatives. Our results indicated that, although highly outcrossed, *C. brasiliense* presents high levels of biparental inbreeding, most likely due to the limited flight range of pollinators and restriction in seed dispersal. Furthermore these results suggest that Cerrado fragmentation could limit gene flow by isolating seed dispersers and territorial small sized bat pollinators inside fragments, increasing the rate of mating between close relatives. The conservation of non-isolated populations in large preserved areas may be necessary to foster

outcrossing events between unrelated individuals and thus maintain species viability.

Introduction

The Brazilian Cerrado biome covers nearly 22% of the Brazilian territory (2 million km^2) consisting of a very heterogeneous vegetation. This biome displays high biodiversity, with nearly 160,000 species, including plants, animals and fungi. The number of trees and shrubs in savannah-like vegetation (cerrado *sensu stricto*) exceeds 800 species, nearly 40% of them are endemic (Ratter *et al.*, 1997). Cerrado has been intensively fragmented in the last 60 years because of agricultural expansion resulting in continental islands of wild habitat surrounded by an "ocean" of crops, jeopardising species viability.

Despite the high biodiversity and endemism and the high rate of deforestation and fragmentation, studies in reproductive biology of cerrado species are scarce (but see Gribel & Hay, 1993; Oliveira & Silva, 1993). Likewise, virtually no information about population genetic structure and mating system is available (but see Collevatti *et al.*, 2000, submitted).

Caryocar brasiliense Camb. (Caryocaraceae) is a widely distributed but endangered Brazilian Cerrado tree species, locally distributed in well delimited patches. Flowers are hermaphroditic and pollination is done mainly by small sized glossophagine bats (*Glossophaga soricina* and *Anoura geoffroyi*) (Gribel & Hay, 1993). The seeds are surrounded by a woody endocarp coated with a yellow fleshy mesocarp rich in oil and

vitamin A, and are eaten by several wild animals, such as birds, greater-rhea (*Rhea americana*), macaws (*Ara* spp.), parrots (*Amazona* spp.), pampas deer (*Ozotocerus berzoarticus*) and paca (*Agouti paca*) (Gribel & Hay, 1993). Furthermore, *C. brasiliense* is a source of raw material for small and middle-sized industries, playing an important role in economy of the inhabitants of Central Brazil (Araújo, 1994). Despite its high ecological and economic importance, Cerrado fragmentation and the higher frequency of fire due to agricultural practices have affected recruitment and ultimately population size and dynamics of this species, augmented by the intense commerce of fruits (Araújo, 1994).

Mating system may determine the role of inbreeding in genetic differentiation among populations (Wright 1940). In plants, mating system is determined mainly by (1) reproductive system features such as self-incompatible mechanisms and degree of protogyny and protandry in hermaphroditic and plant gender in dioecious and monoecious species; (2) foraging behaviour of pollinators; (3) selective abortion by maternal regulation of seed quality or by sibling rivalry; (4) flowering phenology and individual density, which affect pollinator behaviour and, ultimately, outcrossing rate (e.g. Shaanker *et al.*, 1988; Marshall & Folsom, 1991).

Until the 70´s it was believed that many tropical tree species were self-compatible, and self-pollination would predominate over cross-pollination, since co-specific individuals were highly distant (e.g. Federov, 1966). High

levels of inbreeding and genetic drift could lead to genetic differentiation among populations and, eventually, to speciation. Allozyme data have demonstrated, however, that tropical trees are generally highly outcrossed (e.g. O´Malley & Bawa, 1987; Murawski & Hamrick, 1991; Boshier *et al.*, 1995; James *et al.*, 1998; Loveless *et al.*, 1998), even though some species may present significant levels of self-fertilisation, such the Bombabaceae species *Ceiba pentandra* (Murawski & Hamrick, 1992) and the Dipterocarpaceae *Shorea congestiflora* and *S. trapezifolia*. (Murawski *et al.*, 1994). In addition, reproductive biology studies have shown the high frequency of dioecy and self-incompatibility in hermaphroditic species (reviewed in Bawa, 1992). Additionally, microsatellite data for *Pithecellobium elegans* have demonstrated that the great majority of the observed mating events resulted from long distance gene flow (Chase *et al.*, 1996). The same result was observed for allozyme data from other species, such as *Cecropia obtusifolia* (Kaufman *et al.* 1998) and *Ficus* spp. (Nason *et al.*, 1998).

Essentially, all studies of mating system of Neotropical trees have been carried out for species in tropical forests of Central America, particularly Panamá, Costa Rica and Mexico, or even Asian tropics, and few studies to date were published for Brazilian tropical species (e.g. O'Malley *et al.*, 1988; Franceschinelli & Bawa, 2000).

We are interested in understanding the mating system of *Caryocar brasiliense*, and how it could be affected by fragmentation, isolation and anthropic disturbs, in order to generate useful information for conservation programs. In this work, we employed highly informative microsatellite (Simple Sequence Repeats or SSR) markers to estimate mating system parameters in four populations of *C. brasiliense* with contrasting histories of human disturbance.

Materials and methods

Populations, sampling and DNA extraction

Four populations of *C. brasiliense* with contrasting histories of human disturbance were surveyed (Fig. 1): CNV - State Park of Caldas Novas, Goiás, a 12,000 ha of continuous cerrado; FAL - Água Limpa Forestry Park, Brasilia – 4,000 ha of cerrado surrounded by crops; PNB – Brasília National Park, Federal District, a 28,000 ha of continuous cerrado; URU - Uruaçú, Goiás – a cerrado fragment of 5 ha affected by the construction of the Serra da Mesa Hydroelectric Powerplant. In each of these populations, eight to ten families of open pollinated half-sibs were sampled in a rectangular area of 5,000 m^2 in FAL, 6,000 m^2 in CNV, 8,000 m^2 in PNB and 10,000 m^2 in URU. Although the density of reproductive individuals was low, we avoided to sample individuals at large distances. We sampled all

reproductive individuals available within a radius of 100 m from every tree sampled. From each mother tree, expanded leaves and sixteen seeds were collected and stored at –80°C.

Genomic DNA extraction from leaves followed standard CTAB procedure (Doyle & Doyle, 1987). For the progeny arrays, DNA was extracted directly from the embryo, because of the very low germination potential of dormant seeds. For DNA extraction from embryos we used the Fast DNA™ Kit H, and a FP120 FastPrep Cell Disruptor™ (BIO101/Savant Instruments Inc., CA), according to manufacturers' instructions.

Microsatellite marker analysis

Ten SSR loci previously developed and optimised for *Caryocar brasiliense* were used to genotype the mother tree and an open-pollinated progeny array of 16 individuals. SSR loci were developed from a genomic library enriched for the dinucleotide sequence motif poly AG/poly CT. Development assay, optimisation and characterisation of these loci were published elsewhere (Collevatti *et al.*, 1999).

For all genotyping experiments, microsatellite amplifications were performed in a 13 µl volume containing 0.9 µM of each primer, 1unit Taq (Gibco, MD) DNA polymerase, 200 µM of each dNTP, 1X reaction buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), DMSO 50% and 10.0 ng of template DNA. Amplifications were performed using a PT-100

thermal controller (MJ Research) with the following conditions: 96°C for 2 min (1 cycle), 94°C for 1 min, 54 to 56°C for 1 min (according to each primer pair annealing temperature), 72°C for 1 min (30 cycles); and 72°C for 7 min (1 cycle).

For genotype determination and estimation of allele sizes, the amplified products were separated on 4% denaturing polyacrylamide gels stained with silver nitrate (Bassam *et al.*, 1991) and sized by comparison to a 10 bp DNA ladder standard (Gibco, MD) on a computer screen. Allele sizes were estimated using the software SEQAID II (Rhoads & Roufa, 1990) taking into consideration the expected allelic series in base pairs from the primer designed and the original DNA clone from which the SSR locus was developed.

Inheritance of all microsatellite marker loci was confirmed by verifying the presence of one of the maternal alleles in each progeny individual.

Statistical analysis

To estimate single and multilocus outcrossing rate (probability that each offspring of a maternal plant is an outcross), populations were analysed separately, under the mixed mating model of Ritland & Jain (1981) and Ritland (1989), implemented by the software MLTR (Ritland, 1996).

Single and multilocus outcrossing rates were estimated by maximum likelihood, fitting the observed proportions of genotypes descended from a

known maternal genotype to the proportions expected under the mixed mating model. The mixed mating model assumes that: (1) each mating represents a random event of outcross or self-fertilisation, with probabilities equal to t and (1-t), respectively; (2) no mutation and selection following fertilisation may occur; (3) there is no assortative mating (probability of outcross is independent of maternal or paternal genotypes) or variability in pollen pool frequencies (Ritland & Jain, 1981)

As MLTR accepts only up to eight alleles, when the number of alleles exceeded eight, which occurred for all loci but cb1 and cb13, alleles with low frequency (under 0.05) were pooled into a single class. This process was performed separately in each population, since allele frequency distributions were quite different among populations (Collevatti *et al.*, 1999).

As the maternal genotypes were known, the following parameters were estimated for each population: single locus (t_s) and multilocus (t_m) outcrossing rates; average single locus inbreeding coefficient of maternal parents (f); the correlation of outcrossing rate within progeny arrays or normalised variation of outcrossing rate among progeny arrays (r_t); the correlation of outcrossed paternity within progeny (r_p) or the probability that a randomly chosen pair of progeny from the same array are full sibs. Additionally, pollen and ovule allele frequencies were estimated.

Expectation–Maximization method was used for maximising the likelihood equation to estimate all parameters and correlations, and 100 bootstraps were performed and standard errors were obtained for each parameter.

To test for heterogeneity of allele frequency distribution in the pollen pool among maternal trees in each area, contingency tables for each locus and area were constructed and a Fisher exact test was performed using a Markov chain method, implemented by the software GENEPOP (Raymond & Rousset, 1995).

Results

In all progeny arrays, all sibs displayed at least one of the maternal alleles confirming Mendelian inheritance and suggesting no seed contamination (Fig. 2). Loci cb6 and cb12 however showed more than two alleles in the profile suggesting locus duplication possibly because of an ancient polyploid nature of the *Caryocar* genome. Nevertheless, locus interpretation of polymorphisms was not a problem as the second locus was monomorphic.

Single locus outcrossing rate (t_s) varied among loci and populations (Table 1). For CNV, cb9 and cb13 presented the lowest values of t_s, but parental inbreeding coefficient (f) was positive and different from zero only

for cb13. For progeny arrays in FAL, cb13 and cb23 presented the lowest values of t_s but f was higher and different from zero only for cb20. For PNB and URU cb23 and cb13 presented the lowest values of t_s and the highest value of f.

Although multilocus outcrossing rates (t_m) were equal to one for all populations (i.e. 100% outcrossing), mean t_s was lower than t_m for all population, and was highest for FAL (Table 2). Consequently, the difference t_m-t_s (biparental inbreeding) was lower for this population, but not different among the other populations (Table 2). Parental inbreeding coefficient (f) was significantly different from zero only for URU, where density of reproductive individuals was lower (Table 2). When families were analysed, again t_m was equal to one for all half-sib families in all populations. Heterogeneity of pollen allele frequencies among maternal individuals was detected for all loci ($p<0.001$, $SE<0.001$, for all loci and populations), so violation of the mixed mating model occurred, indicating non-random sampling of the pollen pool by each maternal tree.

Both, r_t (correlation of outcrossing rate within progeny arrays) and r_p (correlation of outcrossed paternity within progeny arrays) presented low values indicating that outcrossing rate is independent of progeny arrays and probability of full-sibship within progeny arrays is very low.

Discussion

This is one of the first reports to use highly informative microsatellite markers to estimate mating system parameters in a tropical tree. Although gene flow and paternity analyses were carried out with this class of markers (e.g. Chase *et al.*, 1996; Aldrich & Hamrick, 1998), no study to date has explored the power of this marker class for the detailed understanding of mating systems in tropical trees. Our results indicate that the high polymorphism of microsatellite markers provide an extraordinary resolution to precisely discriminate selfing events from outcrossing events even between close relatives. As these markers become more commonly used for this type of genetic investigation in tropical trees, complementing the popular isozymes, we should not be surprised to learn that for some species what had been estimated as selfing events could turn out to be more precisely interpreted as outcrossing events between relatives. Allozyme markers with limited numbers of alleles, spiky allele frequencies and expected heterozygosities in the 20 to 30% range may not provide the necessary resolution to discriminate such events in particular situations.

Low values of r_p indicate no correlation of paternal parentage, or low probability of full-sibship from the same half-sib family. Multiple paternity was expected since *C. brasiliense* is pollinated by bat species that might promote a high pollen carryover (Bawa, 1990). Additionally, abortion of

ovules or initiated seeds was frequently observed – although fruits may sire four seeds, fruits with one or two developed seeds were in fact the most commonly seen. This strongly suggests early inbreeding depression leading to selective abortion, caused by nutrient limitation or by maternal regulation of seed quality or by sibling competition, or pollen limitation leading to low seed set (Shaanker *et al.*, 1988). Another evidence of selective abortion due to inbreeding depression in *C. brasiliense* was the high rate of fruit and seed abortion observed under controlled self-pollination experiments (Gribel & Hay, 1993).

The high multilocus outcrossing rate observed for *C. brasiliense* is comparable to values found for other tropical tree species (e.g. O'Malley & Bawa, 1987; Murawski & Hamrick, 1991; Boshier *et al.*, 1995; 1996; James *et al.*, 1998; Loveless *et al.*, 1998). This result suggests that this species may have an efficient self-incompatible mechanism in place and that selective abortion may increase the outcrossing rate (t_m) by eliminating self-pollinated seeds.

Biparental inbreeding, represented by the observed difference t_m-t_s (Ritland & Jain, 1981), was an order of magnitude higher than values found so far for other outcrossing tropical tree species using allozyme data (e.g. James *et al.*, 1998). The high values of t_m-t_s (biparental inbreeding) suggest that, although outcrossing is absolute, leading to a t_m of 100%, some cross fertilisation events occur between close relatives. This is quite important in

substructured populations where random matings occur within demes, and between closely related individuals (Ritland, 1985, 1988). In fact, *C. brasiliense* presents highly subdivided populations and *f* (inbreeding coefficient) is as high as 0.11, suggesting that inbreeding is an important force structuring populations (Collevatti *et al.*, 2000, submitted). Nevertheless, heterogeneity of allele frequencies in pollen pool among maternal trees was observed for all loci. This may result from matings among near-neighbour individuals. In fact, bat pollinators tend to forage in groups remaining in the same tree or group of trees for long time (Gribel & Hay, 1993). Furthermore this heterogeneity may also result from statistical sampling derived from the very high multiallelism observed at the SSR marker loci used coupled with the relatively limited number of individuals genotyped per family (~16 seeds) leading to a limited efficiency in sampling all possible genotypes at a locus.

Results of this study as well as other data on the population genetic structure (Collevatti *et al.*, 2000, submitted) indicate that *C. brasiliense* tends to present small panmitic units or demes as a result of restriction in gene flow. Besides the fact that bats that pollinate *C. brasiliense* are small, territorial and with low flight range (Gribel & Hay, 1993), restriction in gene flow may also be caused by a limited seed dispersal. Although seeds could be dispersed by a large bird (greater-rhea), the great majority of seeds are dispersed by gravity keeping seeds under the mother tree canopy. In fact,

we and others (J.D. Hay, personal communication) have observed many seedling under mother trees of *C. brasiliense*. "Cerrado" fragmentation could also be restricting gene flow, by isolating small mammals into fragments (Kearns *et al.*, 1993) and thus increasing the rate of mating between closely related individuals. Furthermore, *C. brasiliense* individuals are spatially distributed in clumps and bat pollinators tend to forage inside these clumps, increasing the rate of mating between relatives. In fact, areas with lower density of reproductive individuals tended to present higher levels of biparental inbreeding.

Although a high resolution of mating events was achieved with microsatellite markers, our data did not provide clear-cut evidences of the effect of fragmentation on mating system parameters. Fragmented area such as URU displayed the same level of outcrossing rate and biparental inbreeding (t_m-t_s) as continuous areas such as CNV and PNB. We hypothesise that no pattern could be detected mainly due to the deme structure and restricted gene flow of *C. brasiliense* that lead to a naturally high level of biparental inbreeding. Additionally, Cerrado fragmentation is a relatively recent event (~60 years) in relation to the species life cycle. So, the deme structure and the rate of biparental inbreeding currently found may very well have existed prior to beginning of fragmentation.

Our present results indicate however that fragmentation may favour the isolation among *C. brasiliense* populations, decreasing population size and

increasing inbreeding, ultimately resulting in local extinction. Therefore, the maintenance of non-isolated populations in many and large preserved areas may be necessary for species viability in Cerrado biome, providing means for gene flow by pollen and seed. Additionally, these areas play an important role in the maintenance of small populations in highly disturbed areas, as a source of pollen and seeds, and contribute to the viability of pollinator and dispersers populations.

Acknowledgements

This study is part of the Doctoral Thesis of the first author, and has been supported by the Universidade de Brasília, EMBRAPA-Genetic Resources and Biotechnology, WWF (World Wildlife Fund), CNPq/PADCT (Brazilian Ministry of Science and Technology) competitive grant 62.00059/97-4 to D.G. and a fellowship to R.G.C. whose assistance we gratefully acknowledge. We wish to thank Célia M. T. Cordeiro for assistance in statistical analysis, Rosana V. Brondani for continued support in laboratory, Kermit Ritland for the authorization for the use of the MLTR software as well as Aécio A. Santos, Antônio Dimas Costa Jr., Bruno Walter, João B. Pereira, Neilton R. Oliveira, Sebastião P.Cordovil, for their kind help in field work. We also acknowledge the support of IBAMA (Brazilian Ministry of Environment) and FEMAGO (State Environmental Institute of

Goiás) for authorizing plant collections in Brasília National Park and in State Park of Caldas Novas and the suggestions of two anonymous reviewers that helped improving this manuscript.

References

ALDRICH, P. R. AND HAMRICK, J. L. 1998. Reproductive dominance of pasture trees in a fragmented tropical forest mosaic. *Science*, **281**,103-105.

ARAÚJO, F. D. 1994. A review of *Caryocar brasiliense* Camb. (Caryocaraceae) – an economically valuable species of the Central Brazilian cerrados. *Economic Botany*, **49**, 40-48.

BASSAM, B. J., CAETANO-ANOLLES, G. AND GRESSHOFF, P. M. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.*, **196**, 80-83.

BAWA, K. S. 1990. Plant-pollinator interactions in tropical rain forests. *Ann. Rev. Ecol. Syst.*, **21**, 399-422.

BAWA, K. S. 1992. Mating systems, genetic differentiation and speciation in tropical rain forest plants. *Biotropica*, **24**, 250-255.

BOSHIER, D. H., CHASE, M. R. AND BAWA, K. S. 1995. Population genetics of *Cordia alliodora* (Boraginaceae), a Neotropical tree. 2. Mating system. *Am. J. Bot.*, **82**, 476-483.

CHASE, M. R., MOLLER, C., KESSELI, R. AND BAWA, K. S. 1996. Distant gene flow in tropical trees. *Nature*, **383**, 398-399.

COLLEVATTI, R. G., BRONDANI, R. V. P. AND GRATTAPAGLIA, D. 1999. Development and characterization of microsatellite markers for genetic analysis of a Brazilian endangered tree species *Caryocar brasiliense*. *Heredity*, **83**, 748-756.

DOYLE, J. J. AND DOYLE, J.L. 1987. Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13-15.

FEDEROV, A. A. 1966. The structure of the tropical rain forest and speciation in the humid tropics. *J. Ecol.*, **54,** 1-11.

FRANCESCHINELLI, E. V. AND BAWA, K. S. 2000. The effect of ecological factors on the mating system of a South American shrub species (*Helicteres brevispira*). *Heredity*, **84**, 116-123.

GRIBEL, R. AND HAY, J.D. 1993. Pollination ecology of *Caryocar brasiliense* (Caryocaraceae) in Central Brazil cerrado vegetation. *J. Trop. Ecol.*, **9**, 199-211.

JAMES, T., VEGE, S., ALDRICH, P. AND HAMRICK, J. L. 1998. Mating system of three tropical dry forest tree species. *Biotropica*, **30**, 587-594.

KAUFMAN, S. R., SMOUSE, P. E AND ALVAREZ-BUYLLA, E. R. 1998. Pollen-mediated gene flow and differential male reproductive success in a tropical pioneer tree, *Cecropia obtusifolia* Bertol. (Moraceae): a paternity analysis. *Heredity*, **81**, 164-173.

KEARNS, C. A., INOUYE, D. W. AND WASER, N. M. 1998. Endangered mutualisms: The conservation of plant-pollinator interaction *Ann. Rev. Ecol. Syst.*, **29**, 83-112.

LOVELESS, M. D., HAMRICK, J. L. AND FOSTER, R. B. 1998. Population structure and mating system in *Tachigali versicolor*, a monocarpic Neotropical tree. *Heredity*, **81**, 134-143.

MARSHALL, D. L. AND FOLSOM, M. W. 1991. Mate choice in plants: an anatomical to population perspective. *Ann. Rev. Ecol. Syst.*, **22**, 37-63.

MURAWSKI, D. A. AND HAMRICK, J. L. 1991. The effect of the density of flowering individuals on the mating systems of nine tropical tree species. *Heredity*, **67**, 167-174.

MURAWSKI, D. A. AND HAMRICK, J. L. 1992. Mating system and phenology of *Ceiba pentandra* (Bombacaceae) in Central Panama. *J. Hered.*, **83**, 401-404.

MURAWSKI, D. A., DAYANANDAN, B. AND BAWA, K. S. 1994. Outcrossing rate of two endemic *Shorea* species from Sri lankan tropical rain forest. *Biotropica*, **26**, 23-29.

NASON, J. D., HERRE, E. A. AND HAMRICK, J. L. 1998. The breeding structure of a tropical keystone plant resource. *Nature*, **391**, 685-687.

OLIVEIRA, P. E. AND SILVA, J. C. S. 1993. Reproductive biology of two species of *Kielmeyera* (Guttiferae) in the cerrados of Central Brazil. *J. Trop. Ecol.*, **9**, 67-79.

O'MALLEY, D. M. AND BAWA, K. S. 1987. Mating system of a tropical rain forest tree species. *Am. J. Bot.*, **74**, 1143-1149.

O'MALLEY, D. M., BUCKLEY, D. P., PRANCE, G. T. AND BAWA, K. S. 1988. Genetics of Brazil nut (Bertholletia excelsa Humb. & Bonpl.: Lecythidaceae), 2. Mating system. *Theor. Appl. Genet.*, **76**, 929-932.

RATTER, J. A., RIBEIRO, J. F. AND BRIDGEWATER, S. 1997. The Brazilian cerrado vegetation and threats to its biodiversity. *Annals of Botany*, **80**, 223-230.

RAYMOND, M. AND ROUSSET, F. 1995. GENEPOP (version 1.2.): population genetics software for exact tests and ecumenism. J. Heredity, 86, 248-249.

RITLAND, K. 1985. The genetic mating structure of subdivided populations. I. Open mating model. *Theor. Pop. Biol.*, **27**, 51-74.

RITLAND, K. 1988. The genetic mating structure of subdivided populations. II. Correlated mating models. *Theor. Pop. Biol.*, **34**, 320-346.

RITLAND, K. 1989. Correlated matings in the partial selfer *Mimulus guttatus*. *Evolution*, **43**, 848-859.

RITLAND, K. 1996. *Multilocus mating system program – MLTR.* Available free of charge via ftp at 128.100.165.100.

RITLAND, K. AND JAIN, S 1981. A model for the estimation of outcrossing rate and gene frequencies using n independent loci. *Heredity*, **47**, 35-52.

RHOADS, D. D., ROUFA, D. J. 1990. *Seqaid II 3.80.* Molecular Genetics Laboratory, Kansas State University, Manhattan.

SHAANKER,R. U, GANESHAIAH, K. N. AND BAWA, K. S. 1988. Parent-offspring conflict, sibling rivalry, and brood size patterns in plants. *Ann. Rev. Ecol. Syst.*, **19**, 177-205.

WRIGHT, S. 1940. Breeding structure of populations in relation to speciation. *Am. Nat.*, **74**, 232-248.

Table 1. Singlelocus outcrossing rate (t_s) and maximum likelihood estimates of pollen and ovule frequency of the most common allele in *Caryocar brasiliense* for four populations. A – number of alleles; f – parental inbreeding coefficient; N_F – number of families analysed; N_P – number of progeny individuals.

Pop	Locus	A	Pollen	Ovule	t_s±SE	f±SE	N_F	N_P
CNV	cb1	8	0.168	0.381	0.669±0.100	0.099±0.122	9	134
	cb3	8	0.208	0.364	0.899±0.089	0.000±0.000	10	147
	cb5	8	0.318	0.286	0.578±0.047	0.010±0.039	10	147
	cb6	8	0.301	0.286	0.770±0.053	0.084±0.099	10	147
	cb9	8	0.120	0.286	0.547±0.073	0.000±0.000	10	147
	cb11	8	0.254	0.182	0.954±0.046	0.000±0.000	10	147
	cb12	8	0.309	0.150	0.610±0.063	0.000±0.000	10	147
	cb13	6	0.740	0.792	0.562±0.105	0.630±0.340	10	147
	cb20	8	0.208	0.304	0.821±0.113	0.000±0.000	10	147
	cb23	8	0.249	0.182	0.950±0.056	0.000±0.000	10	147
FAL	cb1	8	0.276	0.333	0.778±0.143	0.000±0.000	6	96
	cb3	8	0.392	0.278	0.831±0.102	0.000±0.000	8	126
	cb5	8	0.406	0.294	0.808±0.090	0.039±0.094	8	126
	cb6	8	0.233	0.294	0.834±0.087	0.000±0.000	8	126
	cb9	8	0.229	0.188	0.886±0.053	0.000±0.000	8	126

Table 1. continuing...

Pop	Locus	A	Pollen	Ovule	t_s±SE	f±SE	N_F	N_P
FAL	cb11	8	0.245	0.222	0.795±0.068	0.000±0.000	7	112
	cb12	8	0.286	0.235	0.876±0.073	0.001±0.004	8	126
	cb13	8	0.254	0.211	0.589±0.092	0.000±0.000	8	126
	cb20	8	0.312	0.278	0.903±0.039	0.175±0.167	8	126
	cb23	8	0.114	0.444	0.644±0.171	0.000±0.000	8	126
PNB	cb1	8	0.140	0.409	0.868±0.044	0.050±0.113	4	51
	cb3	8	0.238	0.211	0.826±0.077	0.000±0.000	9	108
	cb5	8	0.304	0.263	0.704±0.097	0.087±0.087	9	108
	cb6	8	0.180	0.333	0.636±0.054	0.000±0.000	9	108
	cb9	8	0.258	0.222	0.822±0.092	0.000±0.000	9	108
	cb11	8	0.081	0.300	0.679±0.081	0.112±0.143	9	108
	cb12	8	0.217	0.111	0.696±0.124	0.005±0.007	9	108
	cb13	8	0.241	0.316	0.527±0.126	0.116±0.127	9	108
	cb20	8	0.243	0.105	0.840±0.064	0.000±0.000	9	108
	cb23	8	0.250	0.222	0.486±0.089	0.190±0.141	9	108

Table 1. continuing...

Pop	Locus	A	Pollen	Ovule	t_s±SE	f±SE	N_F	N_P
URU	cb1	6	0.417	0.190	0.785±0.049	0.147±0.187	2	32
	cb3	8	0.214	0.238	0.841±0.072	0.119±0.149	10	154
	cb5	8	0.143	0.333	0.633±0.074	0.252±0.177	10	153
	cb6	8	0.190	0.273	0.711±0.090	0.000±0.000	10	154
	cb9	8	0.453	0.200	0.650±0.079	0.000±0.000	10	154
	cb11	8	0.171	0.435	0.733±0.099	0.045±0.094	10	153
	cb12	8	0.130	0.286	0.796±0.078	0.000±0.000	10	153
	cb13	8	0.276	0.091	0.243±0.079	0.332±0.200	10	153
	cb20	8	0.263	0.455	0.791±0.112	0.001±0.004	10	153
	cb23	8	0.236	0.136	0.776±0.082	0.134±0.111	10	152

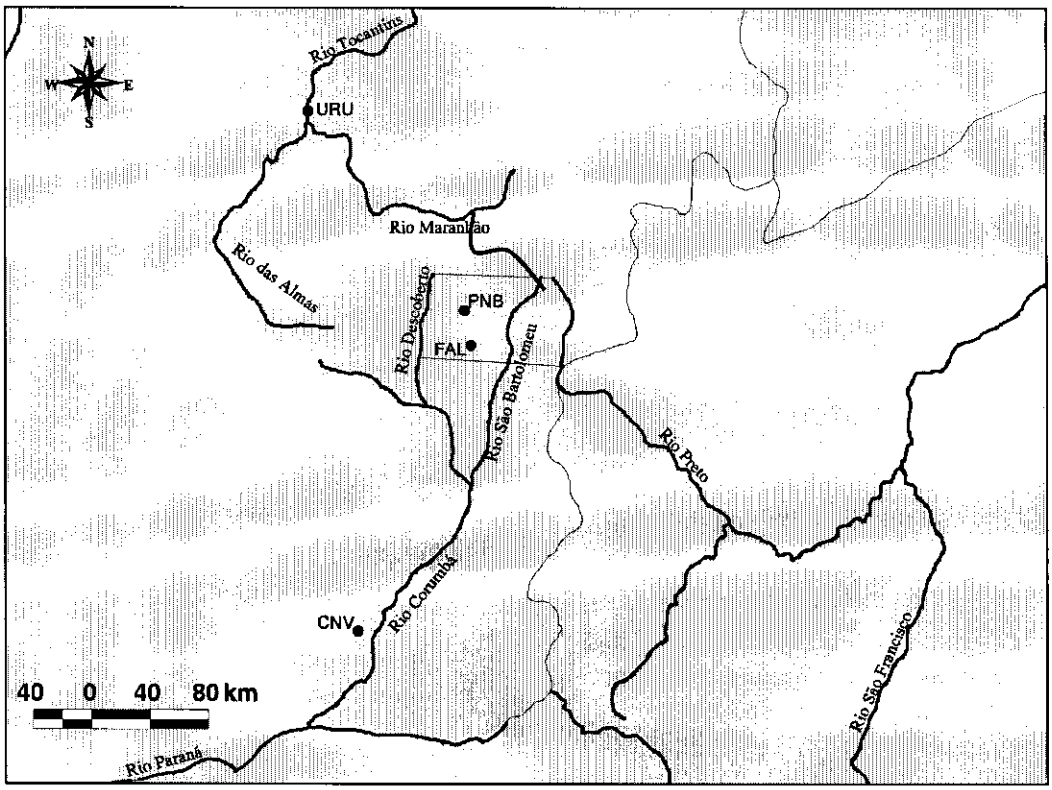
Table 2. Multilocus outcrossing rate (t_m) and mean single locus outcrossing rate (t_s) in four populations of *Caryocar brasiliense*. f – parental inbreeding coefficient; r_t – correlation of t estimates, r_p – correlation of p estimates N_F – number of families analysed; N_r – number of reproductive individuals per hectare; N_P – number of progenies.

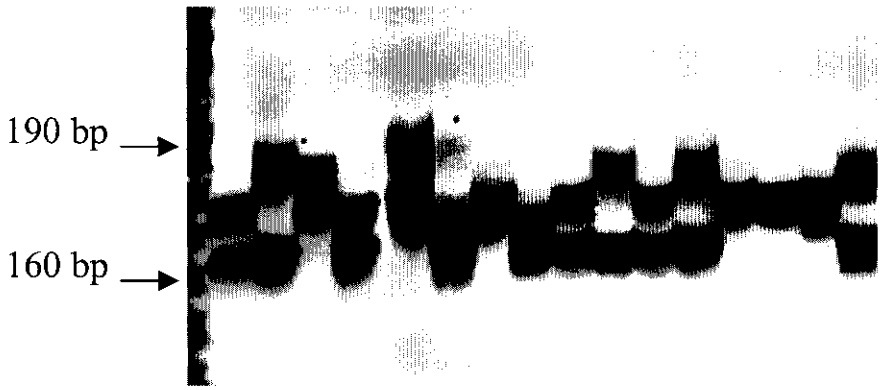
Population	CNV	FAL	PNB	URU
t_m±SE	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000
t_s±SE	0.811±0.023	0.869±0.025	0.771±0.029	0.769±0.028
t_m-t_s±SE	0.189±0.023	0.131±0.025	0.229±0.029	0.231±0.028
f±SE	0.000±0.000	0.000±0.000	0.006±0.006	0.020±0.017
r_t±SE	0.110±0.000	0.110±0.000	0.110±0.000	0.110±0.000
r_p±SE	0.205±0.042	0.086±0.022	0.141±0.051	0.193±0.193
N_F	10	8	9	10
N_r	16.7	16.0	11.3	10.0
N_P	147	126	108	154

Figure captions

Fig. 1. Localisation of the four areas in which analysed populations of *C. brasiliense* were surveyed. Grey area represents Cerrado biome. Thick lines are the main rivers of the region. Thin lines are State divisions. See text for population legends.

Fig. 2 Inheritance and segregation in a open-pollinated half-sib family of *C. brasiliense*, for locus cb20, visualised in silver-stained denaturing polyacrylamide gels. First lane is a 10 bp ladder size standard (Gibco, MD). Lane 2 maternal tree followed by 15 progeny individuals.


N
W
E
S
Rio Tocantins
URU
Rio Maranhão
Rio das Almas
Rio Descoberto
PNB
FAL
Rio São Bartolomeu
Rio Preto
Rio Corumbá
CNV
Rio São Francisco
40 0 40 80 km
Rio Paraná


190 bp
160 bp

Capítulo 4

Phylogeography of the Endangered Brazilian Tree Species *Caryocar brasiliense* Based on Variability at Chloroplast Microsatellite Loci: Genetic Structure and a Comparison with Nuclear Microsatellite Loci

Artigo submetido a Evolution em Fevereiro de 2000

Autores: Rosane Garcia Collevatti, Dario Grattapaglia & John DuVall Hay

Resumo

Nesse trabalho estudamos a filogeografia da espécie arbórea *Caryocar brasiliense*, utilizando dez locos microsatélites de cloroplasto, e comparamos a estrutura genética de populações para o genoma de cloroplasto e nuclear. A diferenciação entre populações para o genoma de cloroplasto foi maior que para o genoma nuclear. Para o genoma nuclear, a diferenciação é correlacionada com a distância entre populações, mas não para o genoma de cloroplasto. Além disso, as árvores enraizadas baseadas nos dados de cloroplasto tiveram uma topologia diferente do fenograma obtido para dados de microsatélites nuclear. O fenograma baseado no polimorfismo a nível de genoma nuclear tendeu a agrupar populações geograficamente mais próximas. A genealogia a nível de genoma do cloroplasto mostrou uma quebra filogeográfica não relacionada a barreiras geográficas. Esse resultado sugere uma restrição na dispersão de sementes, assim como uma origem comum das populações atuais, a partir de antigos refúgios do último período glacial (Wisconsin).

PHYLOGEOGRAPHY OF THE ENDANGERED BRAZILIAN TREE SPECIES *Caryocar brasiliense* BASED ON VARIABILITY AT CHLOROPLAST MICROSATELLITE LOCI: GENETIC STRUCTURE AND A COMPARISON WITH NUCLEAR MICROSATELLITE LOCI

ROSANE G. COLLEVATTI[1,3], DARIO GRATTAPAGLIA[2] AND JOHN V. HAY[1]

[1]*Departamento de Ecologia, Universidade de Brasília, Brasília, D.F. Brasil, 70910-900. E-mail: rosanegc@hotmail.com*

[2]*Laboratório de Genética de Plantas, Recursos Genéticos e Biotecnologia –EMBRAPA, C.P.02372, Brasília, D.F. Brasil. 70770-900*

[3]*Pós-Graduação em Biotecnologia e Recursos Genéticos, Universidade Católica de Brasília, Brasília, DF. E-mail: rosanegc@pos.ucb.br*

Send proofs to: Rosane Garcia Collevatti, *Pós-Graduação em Biotecnologia e Recursos Genéticos, Universidade Católica de Brasília, SGAN 916, Mod. B, Asa Norte, Brasília, DF. 70790-160. E-mail:* rosanegc@pos.ucb.br Phone and FAX: +55613405550.

Running title: Phylogeography of *C. brasiliense*

Author's name: Rosane G. Collevatti et al.

Abstract

In this work we report the phylogeography of the endangered tree species *Caryocar brasiliense* based on variability at ten chloroplast microsatellite loci, and compare population genetic structure at chloroplast and nuclear genome. Populations differentiation for chloroplast genome was greater than for nuclear genome. For nuclear genome, differentiation is correlated with distance, but not for chloroplast genome. Additionally, rooted tree based on chloroplast data had a different topology from nuclear UPGMA phenogram. Unrooted tree based on nuclear genome polymorphism tended to group geographically closer populations. Chloroplast gene genealogy showed a high phylogeography break not related with geographic barriers. This result strongly suggests restriction in seed dispersal, as long as a founder effect with common origin of current populations from ancient relic population refugia of the last Wisconsin glaciation.

Key words: *Caryocar brasiliense,* Caryocaraceae, phylogeography, microsatellites, cpDNA, tropical tree,

Introduction

Species that undergo long-term biogeographic barriers to gene flow may be composed of geographic populations that belong to different gene genealogies (Avise et al 1987). In this manner, genetic breaks or subdivision may be correlated to geographical boundaries (Avise 1992, 1994). Phylogeography is based on the spatial distribution of gene genealogies, and provides a way of detecting the correlation between geographical distribution of haplotypes and their genealogical relationships (Avise et al. 1987). This area of study relies on variability on more conserved cytoplasmatic genomes with low levels of mutation rate and no recombination, such as chloroplast for plants (cpDNA), and mitochondrial genome (mtDNA) for animals (Birky, 1988; Swofford and Olsen 1990; Avise 1994).

Spatial genetic structure for nuclear genome may be caused by patterns of current gene flow, differential selection among patches, genetic drift and mating system that determine inbreeding strength (Wright 1931, 1943). However, genetic structure of maternally inherited organelle genome is more influenced by historical relationship and ancient gene flow among population, as long as historical events such as glaciations and climatic changes over the geological time (Avise et al 1987; Avise 1994; Schaal et al 1998). Additionally, because of the haploid nature and mode of inheritance, effective population size for chloroplast is expected to be one-half that for nuclear genome, leading to a stronger effect of genetic drift in organelle gene frequencies than on nuclear genes (Birky et al. 1983; Ennos, 1994). Additionally, the analysis of nuclear and cytoplasmatic genome, with different modes of inheritance, may provide insights about

the relative importance of pollen and seed in gene flow in structuring plant populations (McCauley 1995; Schall et al. 1998).

Nevertheless, because of the conserved property of chloroplat genome, it is difficult to discriminate between closely related individuals in intraspecific phylogeography analysis. In this manner a new cpDNA marker system has recently been developed, based on the existence of microsatellites on chloroplast genome (Powell et al. 1995a,b; Vendramin et al. 1996; Weising and Gardner 1999), providing a polymorphic system for evaluating plant phylogeography and maternal gene flow.

Caryocar brasiliense Camb. (Caryocaraceae) is a widely distributed but endangered Brazilian Cerrado tree species. Flowers are hermaphroditic and pollination is done mainly by small sized glossophagine bats, *Glossophaga soricina* and *Anoura geoffroyi* (Gribel and Hay 1993). The seeds are surrounded by a woody endocarp coated with a yellow fleshy mesocarp rich in oil and vitamin A, and are eaten by several wild animals, such as birds, greater-rhea (*Rhea americana*), macaws (*Ara* spp.), parrots (*Amazona* spp.), pampas deer (*Ozotocerus berzoarticus*) and paca (*Agouti paca*). Additionally, *C. brasiliense* is a source of raw material (mesocarp, flowers, bark and leaves) for small and middle-sized industries, playing an important role in the economy of the inhabitants of Central Brazil (Araújo 1994).

Caryocaraceae is an exclusively Neotropical family, consisting of 23 species in two genera: *Caryocar*, with 15 species and *Anthodiscus*, with eight species. All *Anthodiscus* species occur in Amazonian Forest and from the 15 species of *Caryocar* only three occur outside Amazonian Forest: *C. edule*, in Atlantic Forest. *C. coriaceum*, in Caatinga and *C. brasiliense*, in Cerrado (Prance and Silva 1973).

In spite of an increased interest in plant phylogeography, essentially all studies to date have been carried out with temperate species and rarely discuss intraspecific phylogeography (e.g. Dumolin-Lapègue et al 1997; Soltis et al 1997; Comes and Abbott 1999; Latta and Mitton 1999). Surprisingly, no studies to date were published for Neotropical species, in spite of the high species diversity in this region and recognised conservation importance (Wilson, 1988).

In this study we report the phylogeography of *C. brasiliense* using chloroplast microsatellite loci and compare population genetic structure and gene flow for cpDNA and nuclear microsatellite data.

Materials and Methods

Populations, sampling and DNA extraction

Ten populations, throughout the whole geographical distribution of *C. brasiliense*, were surveyed (Fig. 1): AGE - Águas Emendadas Ecological Station, Brasilia; CGR – Campus of the Federal University of Mato Grosso do Sul, Campo Grande; CNV - State Park of Caldas Novas, Goiás; FAL - Água Limpa Forestry Park, Brasilia; GSV – Grandes Sertões Veredas National Park, Minas Gerais; ITI – Itirapina Ecological Reserve, São Paulo; MTR - Rondonópolis, Mato Grosso; PNB – Brasília National Park, Federal District; TOC - Porto Nacional, Tocantins; URU - Serra da Mesa Hydroelectric affected area, Uruaçú. In each of these populations 16 individuals were marked and expanded leaves were collected and stored at –80°C. Genomic DNA extraction followed standard CTAB procedure (Doyle and Doyle 1987).

Chloroplast microsatellite analysis

Ten chloroplast SSR loci developed by Weising and Gardner (1999) for dicotyledoneous angiosperms were used to genotype 16 individuals in each population.

Each primer pair was initially screened for product polymorphism, and the annealing temperature was later optimised for *C. brasiliense* to produce clear and robust DNA band amplification in all loci. Microsatellite amplifications were performed in a 13 µl volume containing 0.9 µM of each primer, 1unit Taq DNA polymerase (Gibco, MD), 200 µM of each dNTP, 1X reaction buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), DMSO 50% and 10.0 ng of template DNA. Amplifications were performed using a PT-100 thermal controller (MJ Research) with the following conditions: 96°C for 2 min (1 cycle), 94°C for 1 min, 56°C for 1 min, 72°C for 1 min (30 cycles); and 72°C for 7 min (1 cycle). Analysis of amplified fragments produced by these amplifications were carried out in 4% denaturing polyacrylamide gels stained with silver nitrate (Bassam et al. 1991) and sized by comparison to a 10 bp DNA ladder standard (Gibco, MD) on a computer screen.

Chloroplast microsatellites that amplified clear, specific and polymorphic products across several genotypes of *C. brasiliense* were labelled with fluorescent dyes. Four of them were labelled with 6-FAM (o) dye (ccmp1, ccpm2, ccpm5 and ccpm6), three with TET (and) dye (ccpm4, ccpm7 and ccpm8) and three with HEX (@) dye (ccpm3, ccpm9 and ccpm10). For all genotyping experiments of fluorescent microsatellites, PCR amplifications were performed in a 10 µl volume containing 10.0 µM of each primer, 1unit Taq DNA polymerase (Gibco, MD), 200 µM of each dNTP, 1X reaction buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), BSA mg/ml and 10.0 ng of

template DNA. Amplifications were performed using a PT-100 thermal controller (MJ Research) with the following conditions: 96°C for 2 min (1 cycle), 94°C for 1 min, 56°C for 1 min, 72°C for 1 min (30 cycles); and 72°C for 7 min (1 cycle). PCR amplifications were performed separately for each locus. Reactions were then diluted for 1:5 in three multiplexes with three loci each, labelled with different fluorescent dyes, except for locus ccmp2, which was diluted separately. One microliter of the 1:5 diluted reaction was added to 0.25 μL GeneScan 500 internal lane standard (ROX, Perkin-Elmer, CA), 0.45 μL of loading buffer (25mM EDTA and 50 mg/ml Blue-Dextran) and 2.3 μL deionized formamide. The reactions were then heated to 95°C for 3 min, chilled on ice and electrophoresed in 5% denaturing polyacrylamide gel in an ABI Prism 377 automated DNA sequencer (Perkin-Elmer, CA). Fluorescent PCR products were automatically sized using Genescan software (Perkin-Elmer, CA).

Because of inaccurate fragment size determination generated by Taq polymerase slippage, and by comparative sizing determination method, using size standards (Haberl and Tautz 1999), we used two individuals from CNV population as a control in all PCR amplifications and electrophoresis. Additionally, for all microsatellite loci, PCR reactions and electrophoresis were repeated for four individuals of each population, to verify allele sizing variation. Twenty-four additional individuals from one population (FAL) were genotyped, resulting in 40 individuals from this population, to verify the existence of other haplotypes. Additionally, we examined the segregation of maternal haplotype in four open-pollinated half-sib families, to verify conservation of maternal haplotype in progeny. Four seeds were collected from four mother trees and DNA was extracted directly from the embryo, because of the low germination of dormant seeds.

For DNA extraction from embryos we used Fast DNA™ Kit H, from BIO101 and FP120 FastPrep Cell Disruptor™ (BIO101/Savant Instruments Inc., CA), according to manufacture's instructions. PCR amplification and visualisation of haplotype segregation followed the same protocols used for leaves.

Additionally, five individuals of *Caryocar villosum* from Amazonian Forest, Manaus, Brazil, were genotyped to be used as outgroup in the phylogeography analysis (see below).

Genetic structure analysis

As cpDNA is a non-recombinant maternally inherited genome, chloroplast microsatellite genotypes were interpreted as haplotypes. Population genetic structure was assessed using correlation (θ) between allele frequencies of different individuals in the same population, which may be equivalent to Wright's F_{ST} (Cockerham 1969). The estimation followed an analysis of variance, which was carried out using the software GDA (Lewis and Zaykin 1999). Significance test of correlation was performed by bootstrapping over loci with 95% nominal confidence interval (Weir 1996).

To test the hypothesis of "isolation by distance" for chloroplast microsatellite data a matrix of Slatkin's pairwise linearized F_{ST} (Slatkin 1991, 1995) was obtained by the AMOVA procedure using the software Arlequin (Schneider et al. 1997) and correlated to a geographical distance matrix by a Mantel test (Mantel 1967). Chloroplast SSR data were compared to nuclear SSR data described earlier (Collevatti et al 2000a, submitted).

Phylogeography analysis

Chloroplast microsatellite data were scored as unordered characters with equal weights and analysed using Phylogenetic Analysis Using Parsimony (PAUP, Swofford 1993). Under the maximum parsimony optimality criterion of Fitch, optimal unrooted trees where searched using the exhaustive search procedure. Afterwards, *C. villosum* was added as outgroup and trees were rooted considering the outgroup paraphyletic with respect to the monophyletic ingroup. A bootstrap analysis was performed with 100 replicates. Populations were held constant and character were sampled with replacement to build new data sets of the same size as the original data. A heuristic search method was used to find most parsimonious trees and a 50%majority-rule consensus tree and confidence level associated with the groups (the frequency of occurrence of each group or bootstrap support values) were obtained.

Additionally, Cavalli-Sforza and Edwards (1967) genetic distances were estimated for nuclear microsatellite data and a UPGMA phenogram based on those distances was obtained to compare with phylogeography obtained from chloroplast data.

Results

Chloroplast microsatellites characterization

Fluorescent PCR amplifications resulted in a high resolution analysis (Fig. 2a,b,c). Fragment size automatically called by Genescan and Genotyper varied between PCR reactions in a range of ±1.0, but there was no variation in allele sizing between gels, for the same PCR reaction. Additionally, allele sizing was uniform within population. Considering the ten populations analysed in this work, and a total of 160 individuals, no

within population variation was found for all loci, but eight haplotypes could be recognised for *C. brasiliense* populations (Table 1). The outgroup *C. villosum* displayed an haplotype different from those found for *C. brasiliense* (Table 1). No different haplotype was found when an additional sample of 24 individuals were analysed for the FAL population. Loci ccpm2 and ccpm9 showed two alleles in the profile suggesting locus duplication (ccpm2 and ccpm2', ccpm9 and ccpm9'), but interpretation of polymorphisms was not a problem as the second locus polymorphism could be interpreted independently (Fig. 2d), resulting in 12 analysed loci. From the 12 analysed loci, one was monomorphic (ccpm2', Table 1) and 11 were polymorphic and parsimony-informative. Conservation of maternal haplotype was verified for all 12 SSR loci by analysing a mother tree and its open-pollinated half-sib family. All sibs displayed the maternal allele, as an evidence of maternal inheritance, also confirming the precision of allele sizing.

Population genetic structure

Chloroplast genome showed high differentiation among populations (Table 2) - θ was significantly different from zero, consequently genes are not randomly distributed among populations. Although significantly different from zero, differentiation was much lower for nuclear genome (Table 2). Additionally, a significant amount of inbreeding was found in the analysis of nuclear genome ($f>0$), showing that genes within populations did not unite at random (Table 2). Although highly differentiated, pairwise F_{ST} is not correlated to geographical distance for chloroplast genome (Fig. 3a, r=-0.099, t=-0.499, p=0.4059). However, as expected under the isolation by distance model, these

variables were significantly and positively correlated for nuclear genome (Figure, 3b, r=0.518, t=1.429, p=0.0396).

Phylogeography

Exhaustive search evaluated 34,459,425 trees and retained 63 equally parsimonious trees. Rooted trees had a total length of 23 steps and consistency index (CI) equal to 0.913 (rescaled consistecy, RC=0.830), retention index (RI) equal to 0.909, and homoplasy index (HI) equal to 0.087 (Fig. 4).

The rooted tree obtained from chloroplast data and the UPGMA phenogram of nuclear data showed different topologies. With nuclear data, populations tended to be grouped according to distance (Fig. 5) as expected under isolation by distance. For nuclear data, ITI is more similar to CGR and AGE is similar to CNV and PNB. This last group is connected to FAL (Fig. 5). For chloroplast data, however, GSV and ITI had a common maternal ancestor and diverge from the other populations in the first node (node 18, Fig 4). CGR has a common ancestor with MTR and AGE with FAL. PNB, TOC and the group of ((CGR,MTR)(AGE,FAL)) was unresolved as a polytomy, even in the consensus tree. Nevertheless, the group (AGE, FAL) is supported at 100% and (CGR,MTR) at 79% in 100 bootstrap replicates (Fig. 4).

Discussion

All chloroplast microsatellite loci developed by Weising and Gardner (1999) were transferable to *C. brasiliense*, and detected intraspecific polymorphism providing a

powerful tool for phylogeography analysis. This is one of the first reports of phylogeography analysis in a tropical tree, comparing data from chloroplast and nuclear microsatellites.

High levels of genetic differentiation were detected for the chloroplast genome (θ=0.99), in sharp contrast to the nuclear genome (θ=0.07). Additionally, overall loci inbreeding coefficient for the nuclear genome (f) indicated that, in general, genes within populations are not united at random.

The genetic structure based on nuclear genome indicates that alleles are not randomly distributed among and within populations, since θ and F were significantly different from zero (Cockerham 1969). As pointed by Collevatti et al. (2000a, submitted), θ presented low values and $F>\theta$, indicating little differentiation among populations due to non-random distribution of alleles. Therefore, population differentiation for nuclear genome may be the outcome of mating between relatives.

As expected, population differentiation for maternally inherited genome (cpDNA) was greater than for the biparentally inherited nuclear genome (Ennos 1994). In this manner, under the migration-drift equilibrium of island model (Wright, 1943) F_{STm} (θ_m), for maternally inherited cpDNA, will exceed F_{STb} (θ_b), for maternally and paternally inherited nuclear genome (Ennos 1994). Mean number of migrants per generation (Nm), at migration-drift equilibrium for nuclear genome, using F_{STb} (θ overall loci), $F_{STb}=1/(4Nm_b+1)$, was equal to 3.32. For maternally inherited chloroplast genome, assuming migration-drift equilibrium of island model to obtain gene flow, $F_{STm} = 1/(2Nm_m+1)$ (Birky et al. 1983; Ennos 1994), Nm=0.005. If we calculate the ratio of pollen and seed flow for inbreeding population since f was statistically different from

zero for *C. brasiliense* (f=F_{IS}=0.11), as proposed by Ennos (1994) - (pollen flow)/(seed flow)={[(1/F_{STb})-1]x(1+F_{IS})-2x[(1/F_{STm})-1]}/ [(1/F_{STm})-1] =144.01. These results show that, considering migration drift-equilibrium, pollen movement is significantly greater than seed movement. Therefore the low population differentiation for the nuclear genome is most likely due to pollen movement and not to seed dispersal.

Additionally, *C. brasiliense* populations in the Cerrado of Central Brazil are differentiated by a process of "isolation by distance", that was confirmed with a Mantel test, for nuclear genome (Collevatti et al. 2000a, submitted). Nevertheless, although highly differentiated, the genetic divergence for chloroplast genome is not due to isolation by distance, as pairwise F_{ST} was not correlated to distance. This result suggests that differentiation for chloroplast genome is independent of distance and restriction in seed movement may be determined by other factors than distance among populations.

Our results strongly support that the connection among populations of *C. brasiliense* is mainly due to pollen flow, although *C. brasiliense* is pollinated by territorial small sized bat species with low flight range (Gribel and Hay 1993). Although seeds could be dispersed by greater-rhea (*Rhea americana*), the great majority of seeds are dispersed by gravity, so that seeds tend to remain under the mother tree canopy (J. D. Hay, unpublished). In fact, Collevatti et al. (2000b, in press), pointed out that in spite of the high outcrossing rate (t_m=1.0), *C. brasiliense* tends to present small panmitic units or demes due to restriction in gene flow, indicated by a high biparental inbreeding and f.

The nuclear genome variability is, however, highly influenced by current gene flow by pollen, population fragmentation and isolation. On the other hand, pattern of cpDNA genealogy was unrelated with distance among populations and phylogenetic breaks did

not agree with geographic boundaries, which are represented by hydrography (Fig.1). The common origin of GSV and ITI populations, situated in such geographycally different and distant localities, as long as the polytomy may be caused by the lack of polymorphism to differentiate GSV and ITI and to resolve the polytomy. Thus it would be necessary more cpDNA SSR to resolve these questions, or other cpDNA class of molecular marker, such as cpDNA sequences.

Nevertheless, phylogeography break may be correlated not only with current patterns in seed flow, but with ancient gene flow and historical changes in species distribution. We hypothesise that the restriction in seed movement and the high chloroplast population subdivision may be explained by the megafaunal fruit hypothesis (Janzen and Martin 1982), which argues that many living fruits are botanical anachronisms adapted for dispersal by an extinct megafauna. *Caryocar* appeared in the fossil record in the Tertiary, in the middle Eocene (nearly 50-44 m.y. BP) in Venezuela Amazonian Forest (Muller 1981). As the Amazonian Forest is the diversity centre of Caryocaraceae, it is supposed that *Caryocar brasiliense* is derived from an Amazonian ancestral. The mammalian megafauna evolved in the late Tertiary (nearly 20 m.y BP). It is possible that, as hypothesised for other species, *C. brasiliense* large fruits, a drupe with one to four seeds, evolved in an environment where megafauna predominated. Under the fleshy and conspicuous yellow mesocarp, a hardy and woody endocarp coated on the outer surface with a lot of spines, protects the seed. The only current potential disperser is the great-rhea (*Rhea americana*, Rheidae) a large terrestrial and flightless bird, which could defecate viable scarified seeds (R. Gribel, unpublished). The last megafauna was extinct in the Quaternary (ca. 10,000 BP) probably due to environmental

deterioration caused by the last Wisconsin glaciation, or by overhunting (Martin and Wright 1967; Lewin 1983). In this manner, current patterns of chloroplast population subdivision may be caused by the lost of ancient dispersers, leading to current restriction in seed flow by gravity, keeping seeds under the mother tree canopy, and isolating maternal genealogies.

On the other hand, glaciations changed the South-American landscape, especially the Brazilian savanna-like vegetation, such as the Cerrado (e.g. Prance 1982; Behling 1998; Ledru et al. 1998; Salgado-Labouriau et al. 1998). The last glaciation (Wisconsin) coincides with a drier climate in Southern and Central Cerrado domain (12000-8000 yr B. P.). Savanna landscapes were dominated by grasslands and frequent fires were recorded. In the fossil pollen record, arboreal and "vereda" (Cerrado swamp) elements are rare, and restricted to semi-deciduous forest taxa (Salgado-Laboriau et al. 1998). It is possible that *C. brasiliense*, as some other arboreal savanna taxa, became restricted by to sites with moist climatic conditions, that behaved as refugias. After 7000 yr B.P., the climate became moister and arboreal pollen record dominated the savanna vegetation. After 4000 yr B.P., cerrado and semi-deciduous forests attained their modern distribution and taxa composition, and seasonality continued to increase until reaching the current pattern (Salgado Laboriau et al 1998). We hypothesise that the current pattern in chloroplast haplotypes distribution is mainly determined by the common origin of maternal genealogies or by founder effect. Populations restricted to refugias in the last glaciation may have spread and dispersed to favourable areas in the last 7000 yr, attaining the present geographical distribution. Because megafauna was already extinct by 7000 yr B.P., *C. brasiliense* may have then lost seed dispersers. Seed movement was

then restricted to surrounding areas, leading to the high phylogeographic break observed in this study, and a founder effect.

Acknowledgements

This study is part of the Doctoral Thesis of the first author, and has been supported by WWF (World Wildlife Fund), CNPq/PADCT (Brazilian Ministry of Science and Technology) competitive grant 62.00059/97-4 to D.G., and a fellowship to R.G.C., whose assistance we gratefully acknowledge. We wish to thank Rosana P.V. Brondani for continued support in laboratory, Paul Lewis and Dmitri Zaykin for the authorization for the use of the GDA software as well as Aécio A. Santos, Antônio Dimas Costa Jr. Bruno Walter, Eugenia C. Amaral, Fabian Borghetti, Flavia Pinto, Frederico S. Lopes, Galdino R. Santos, João B. Pereira, Maura C. Almeida, Neilton R. Oliveira, Sebastião P. Cordovil, for their kind help in field work. We also acknowledge the support of IBAMA (Brazilian Ministry of Environment), IEF-SP (State Forest Institute of São Paulo), FEMAGO (State Environmental Intitute of Goiás) and SEMATEC (Environmental Institute of Brasília) for authorizing plant collections in Brasília and Grandes Sertões Veredas National Parks, in Itirapina Ecological Station, in State Park of Caldas Novas, and in Águas Emendadas Ecological Station.

References

ARAÚJO, F.D. 1994. A review of *Caryocar brasiliense* Camb. (Caryocaraceae) – an economically valuable species of the Central Brazilian cerrados. Economic Botany 49:40-48.

AVISE, J. C. 1992. Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. Oikos 63:62-76.

AVISE, J. C. 1994. Molecular markers, natural history and evolution. Chapman and Hall, New York.

AVISE, J. C., J. ARNOLD, R. M. BALL, E. BERMINGHAM, T. LAMB., J. E. NEIGEL, C. A. REEB, AND N. C. SAUNDERS. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Ann. Rev. Ecol. Syst. 18:489-522.

BASSAM, B.J., G. CAETANO-ANOLLES, AND P. M. Gresshoff. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels Annals of Biochemical 196:80-83.

BEHLING, H. 1998. Late quaternary vegetational and climatic changes in Brazil. Review of Paleobotany and Palynology 99:143-156.

BIRKY, C. W. 1988. Evolution and variation in plant chloroplast and mitochondrial genomes. Pp. 23-55 *in* L. D. Gottlieb and S. K. Jain, eds. Plant evolutionary biology. Chapman and Hall, London.

BIRKY, C. W., T. MARUYAMA, AND P. FUERST. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplast, and some results. Genetics 103:513-527.

CAVALLI-SFORZA, L. L. AND A. W. F. EDWARDS. 1967. Phylogenetic analysis: Models and estimation procedures. Evolution 21:550-570.

COCHERHAM, C. C. 1969. Variance of gene frequencies Evolution 23:72-84.

COLLEVATTI, R.G., D. GRATTAPAGLIA, AND J. D. HAY. 2000a. Population genetic structure of the endangered tropical tree species *Caryocar brasiliense*, based on variability at microsatellite loci. Mol. Ecol. (submitted).

COLLEVATTI, R.G., D. GRATTAPAGLIA, AND J. D. HAY. 2000b. High resolution microsatellite based analysis of mating system allows the detection of significant biparental inbreeding in *Caryocar brasiliense*, an endangered tropical tree species. Heredity (in press).

COMES, H. P., AND R. J. ABBOTT. 1999. Population genetic structure and gene flow across arid versus mesic environments: a comparative study of two parapatric Senecio species from the near east. Evolution 53:36-54.

DOYLE, J. J., AND J. L. DOYLE. 1987. Isolation of plant DNA from fresh tissue Focus 12:13-15.

DUMOLIN-LAPEGUE, S., B. DEMESURE, S. FINESCHI, V. LE CORRE, AND R. J. PETIT. 1997. Phylogeographic structure of white oaks throughout the European Continent. Genetics 146:1475-1487.

ENNOS, R.A. 1994. Estimating the relative rates of pollen and seed migration among plant populations. Heredity 72:250-259.

FAEGRI, K., AND L. van der PIJL. 1971. The principles of pollination ecology. Pergamon Press, Oxford.

GRIBEL, R., AND J. D. HAY. 1993. Pollination ecology of *Caryocar brasiliense* (Caryocaraceae) in Central Brazil cerrado vegetation. J. Trop. Ecol. 9:199-211.

HABERL, M., AND D. TAUTZ. 1999. Comparative allele sizing can produce inaccurate allele size differences for mcirosatellites. Mol. Ecol. 8:1347-1350.

HARRIS, S. A., AND R. INGRAM. 1991. Chloroplast DNA and biosystematics: The effect of intraspecific diversity and plasmid transmission. Taxon 40:393-412.

JANZEN, D. H., AND P. S. MARTIN. 1982. Neotropical anachronisms: the fruits the gomphotheres left behind. Science 215:19-27.

LATTA, R.G., AND J. B. MITTON. 1999. Historical separation and present gene flow through a zone of secondary contact in ponderosa pine. Evolution 53:769-776.

LEDRU, M. P., M. L. SALGADO-LABOURIAU, AND M. L. LORSCHEITTER. 1998. Vegetation dynamics in southern and central Brazil during the last 10,000 yr B.P. Review of Paleobotany and Palynology 99:131-142.

LEWIN, R. 1983. What killed the giant mammals? Science 221:1036-1037.

LEWIS, P., AND D. ZAYKIN. 1999. Genetic data analysis: computer program for the analysis of allelic data. Version 1.0 (d12). Free program distributed by the authors over the internet from the GDA Home Page at http://chee.unm.edu/gda/

MANTEL, N. A. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27:209-220.

MARTIN, P. S., AND H. E. WRIGHT Jr. 1967. Pleistocene Extinctions: the search for a cause. Yale University Press, New Haven.

MCCAULEY, D.E. 1995. The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends Ecol. Evol. 10:198-202.

MULLER, J. 1981. Fossil pollen records of extant angiosperms. Bot. Rev. 47:1-142.

POWELL, W., M. MORGANTE, R. McDEVITT, G. G. VENDRAMIN, AND J.A. RAFALSKI. 1995a. Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. Proc. Natl. Acad. Sci. U.S.A. 92:7759-7763.

POWELL, W., M. MORGANTE, C. ANDRE, J. W. McNICOL, G. C. MACHRAY, J. J. DOYLE, S. V. TINGEY, AND J. A. RAFALSKI. 1995b. Hypervariable variable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. Curr. Biol. 5:1023-1029.

PRANCE, G. T. 1982. A review of the phytogeographic evidences for Pleistocene climatic changes in the Neotropics. Ann. Missouri Bot. Gard. 69:594-624.

PRANCE, G. T., AND M. F. SILVA. 1973. Flora Neotropica: Caryocaraceae. Monograph n° 12. Hafner, New York.

RATTER, J.A., J. F. RIBEIRO, AND S. BRIDGEWATER. 1997. The Brazilian cerrado vegetation and threats to its biodiversity. Annals of Botany 80:223-230.

SALGADO-LABOURIAU, M. L., M. BARBERI, K. R. FERRAZ-VICENTINI, AND M. G. PARIZZI. 1998. A dry climatic event during the late Quaternary of tropical Brazil. Review of Paleobotany and Palynology 99:115-129.

SCHAAL, B. A., D. A. HAYWORTH, K. M. OLSEN, J. T. RAUSCHER, AND W. A. SMITH. 1998. Phylogeographic studies in plants: problems and prospects. Mol. Ecol. 7:465-474.

SCHNEIDER, S., J. M. KUEFFER, D. ROESSLI, AND L. EXCOFFIER. 1997. Arlequin Ver. 1.1: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

SLATKIN, M. 1991. Inbreeding coefficients and coalescence times. Genetic Research of Cambridge 58:167-175.

SLATKIN, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462.

SOLTIS, D. E., M. A. GITZENDANNER, D. D. STRENGE, AND P. S. SOLTIS. 1997. Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. Plant Syst. Evol. 206:353-373.

SWOFFORD, D. L. 1993. PAUP: phylogenetic analysis using parsimony Ver. 3.1.1. Illinois Natural History Survey, Champaign, IL.

SWOFFORD, D. L., AND G. J. OLSEN. 1990. Phylogeny reconstruction. Pp. 411-501 *in* D. M. Hills, and C. Moritz, eds. Molecular systematics. Sinauer, Sunderland, MA.

VENDRAMIN, G. G., L. LELLI, P. ROSSI, AND M. MORGANTE. 1996. A set of primers for the amplification of 20 chloroplast microsatellite in Pinaceae. Mol. Ecol. 5:595-598.

WEIR, B.S. 1996. Genetic Data Analysis II. Sinauer, Sunderland, MA.

WEISING, K., AND R. C. GARDNER. 1999. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42:9-19.

WILSON, E. O. (ed.). 1988. Biodiversity. National Academic Press, New York.

WRIGHT, S. 1931. Evolution in mendelian population. Genetics 16:97-159.

WRIGHT, S. 1943. Isolation by distance. Genetics 28:114-138.

Table 1. Chloroplast DNA haplotypes (Hapl.) observed for ten populations of *C. brasiliense*, and five individuals of *C. villosum*, for ten SSR loci.

Hapl.	Population	Allele size (bp)											
		ccpm1	**ccpm2**	**ccpm2'**	**ccpm3**	**ccpm4**	**ccpm5**	**ccpm6**	**ccpm7**	**ccpm8**	**ccpm9**	**ccpm9'**	**ccpm10**
A	AGE/FAL	119	99	185	101	118	119	87	117	119	87	117	118
B	CNV	120	99	185	103	119	116	87	117	116	87	117	115
C	CGR	121	98	185	106	120	116	87	117	116	87	117	115
D	MTR	121	98	185	106	120	117	87	117	116	87	117	115
E	PNB	121	99	185	103	119	117	87	117	116	87	117	115
F	TOC	121	99	185	103	119	117	87	117	117	87	117	115
G	URU	121	99	185	103	119	116	87	117	116	87	117	115
H	GSV/ITI	122	99	185	105	120	116	80	114	116	80	114	114
I	*C. villosum*	120	99	185	103	120	116	87	116	116	87	116	114

Table 2. Population genetic structure of *C. brasiliense* based on an analysis of variance of allele frequencies for chloroplast and nuclear (in Collevatti *et al.*, 1999, in press) SSR loci. f – inbreeding coefficient; F – total inbreeding coefficient; θ – fixation index, σ^2_g - expected mean square for alleles within individuals; σ^2_P – expected mean square among populations. Bootstrapping over loci - number of replicates=999; nominal confidence interval=95%.

Nuclear microsatellites				**Chloroplast microsatellites**			
Locus	**f**	**F**	θ	**Locus**	θ	σ^2_g	σ^2_p
cb1	0.10	0.16	0.06	**ccpm1**	1.00	0.00	0.73
cb3	0.08	0.13	0.06	**ccpm2**	1.00	0.00	0.36
cb5	0.13	0.15	0.02	**ccpm2'**	-	0.00	0.00
cb6	0.02	0.06	0.03	**ccpm3**	1.00	0.00	0.80
cb9	0.04	0.19	0.17	**ccpm4**	0.98	0.01	0.70
cb11	0.06	0.09	0.03	**ccpm5**	1.00	0.00	0.69
cb12	0.15	0.19	0.04	**ccpm6**	1.00	0.00	0.36
cb13	0.29	0.44	0.21	**ccpm7**	1.00	0.00	0.36
cb20	0.11	0.15	0.05	**ccpm8**	1.00	0.00	0.51
cb23	0.10	0.17	0.07	**ccpm9**	1.00	0.00	0.36
				ccpm9'	1.00	0.00	0.36
				ccpm10	1.00	0.00	0.62

Table 2. cont.

	Nuclear microsatellites			Chloroplast microsatellites		
	f	**F**	θ	θ	σ^2_g	σ^2_p
Overall loci	0.11	0.17	0.07	0.99	0.01	5.84
Upper Bound	0.148	0.244	0.123	1.00	-	-
Lower Bound	0.067	0.123	0.040	0.99	-	-

Figure legends

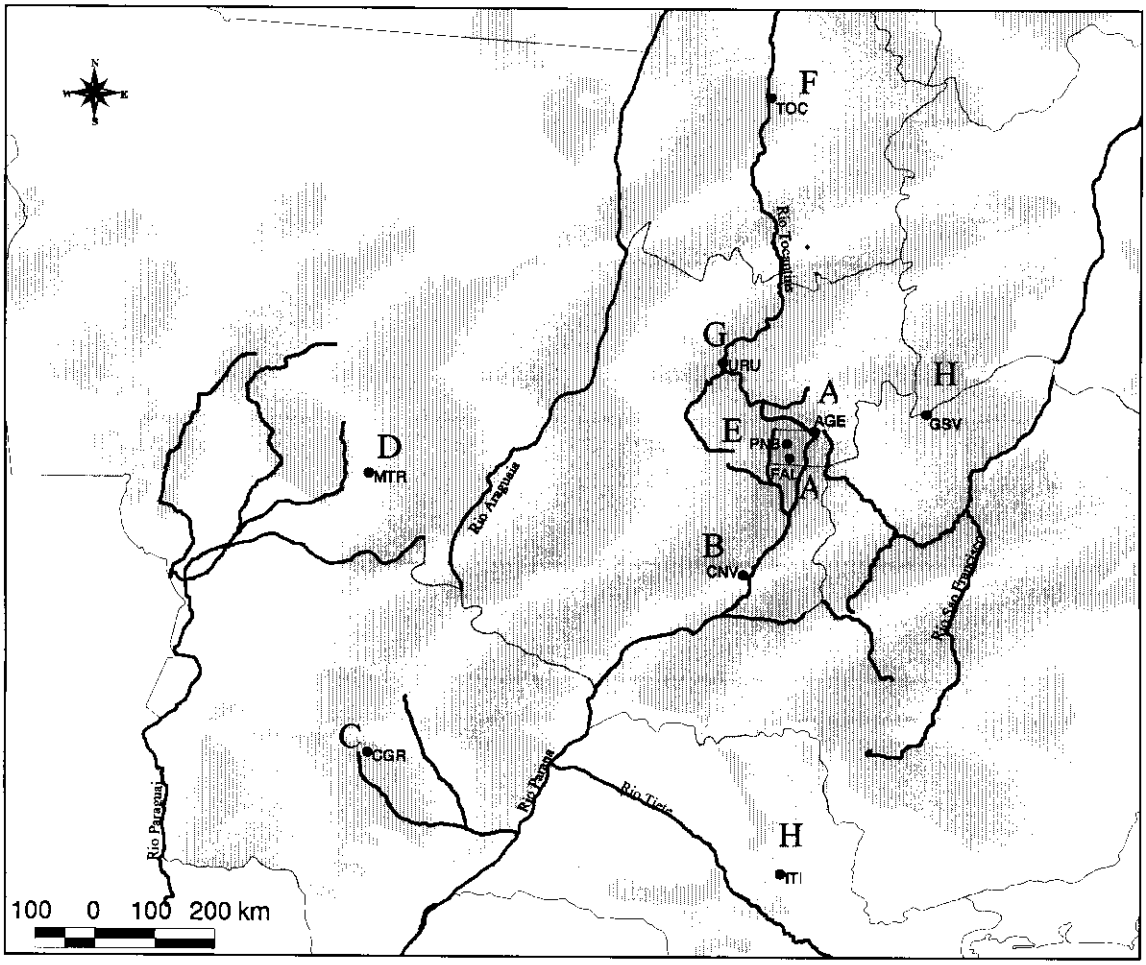
Fig. 1. Localisation of the 10 areas in which analysed populations of *C. brasiliense* were surveyed. Thick lines are the main rivers of Cerrado region. Thin lines are State division. See the text for population legends. Labels near each population are haplotypes describe on Table 1.

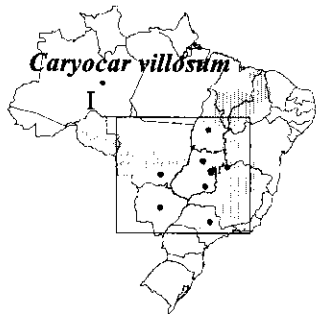
Fig. 2. Electropherogram of one individual from CNV population, for four chloroplast microsatellite loci. (a) loci ccpm1 labelled with 6-FAM (o) dye; (b) loci ccpm4 labelled with TET (and) dye; (c) loci ccpm3 labelled with HEX (@) dye; (d) duplicated loci ccpm2 labelled with 6-FAM (o) dye showing ccpm2 region (allele 99) and ccpm2' (allele 185).

Fig 3. Relationship between Slatkin's linearized pairwise F_{st} and geographical distance among populations of *C. brasiliense*: (a) chloroplast microsatellite data, Mantel test r=-0.099, t=0.499, p=0.4059; (b) nuclear microsatellite data, Mantel test for correlation for, r=0.518, t=1.429, p=0.0396 (in Collevatti, 1999, in press)

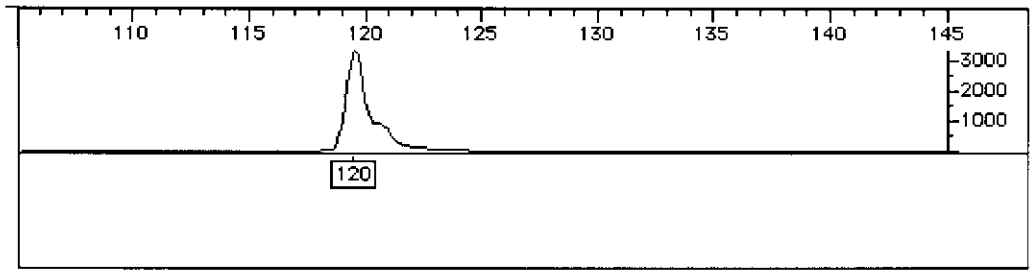
Fig. 4. One of 63 equally rooted parsimonious trees (length=23 steps) found by exhaustive search procedure of PAUP 3.1, for chloroplast microsatellite haplotypes of *C. brasiliense*, considering *C. villosum* (VIL) as a paraphyletic outgroup. Branch lengths (bellow respective branch) are absolute number of character changes over 12 characters analysed. Bootstraping support value (over respective branch) were obtained with 100 replicates.

Fig. 5. Relationship among ten populations of *C. brasiliense*, based on ten nuclear microsatellite data. The unrooted phenogram is based on a UPGMA analysis of interpopulation genetic distances (Cavalli-Sforza and Edwards 1967).

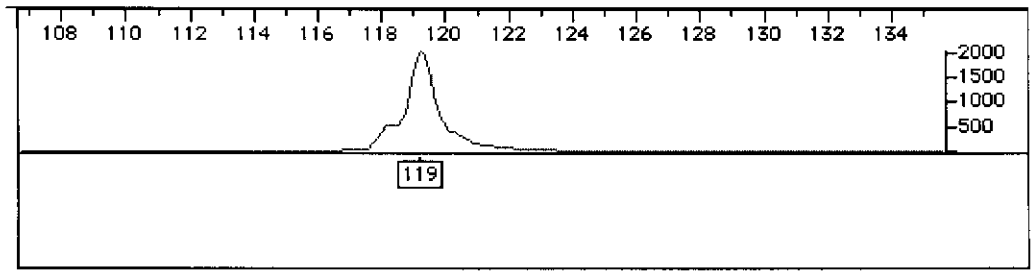

F
TOC
Rio Tocantins
G
URU
A
AGE
H
GSV
E
PNB
FAL
A'
D
MTR
Rio Araguaia
B
CNV
Rio São Francisco
C
CGR
Rio Paraguai
Rio Paraná
Rio Tietê
H
ITI
100 0 100 200 km


Caryocar villosum
I

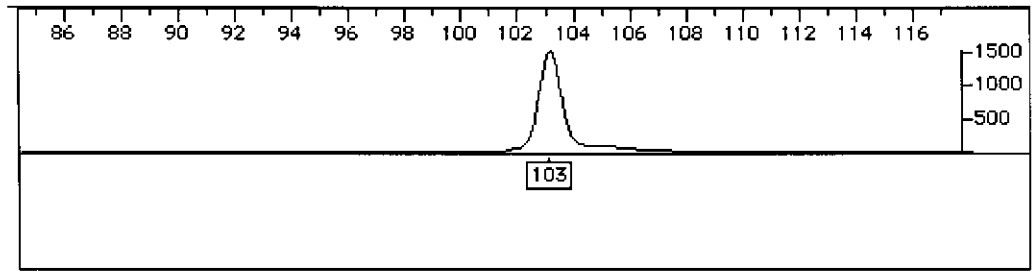
a


110 115 120 125 130 135 140 145
3000
2000
1000
120

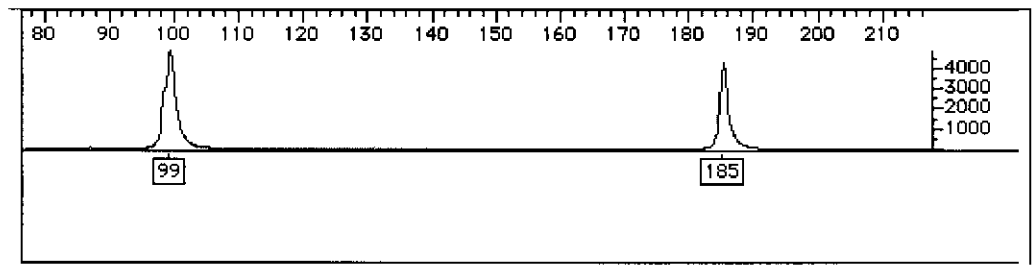
b

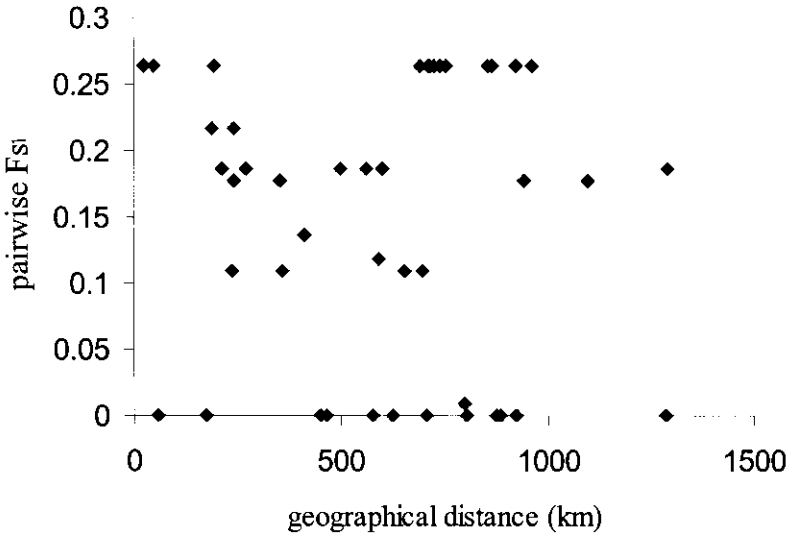

108 110 112 114 116 118 120 122 124 126 128 130 132 134
2000
1500
1000
500
119

c

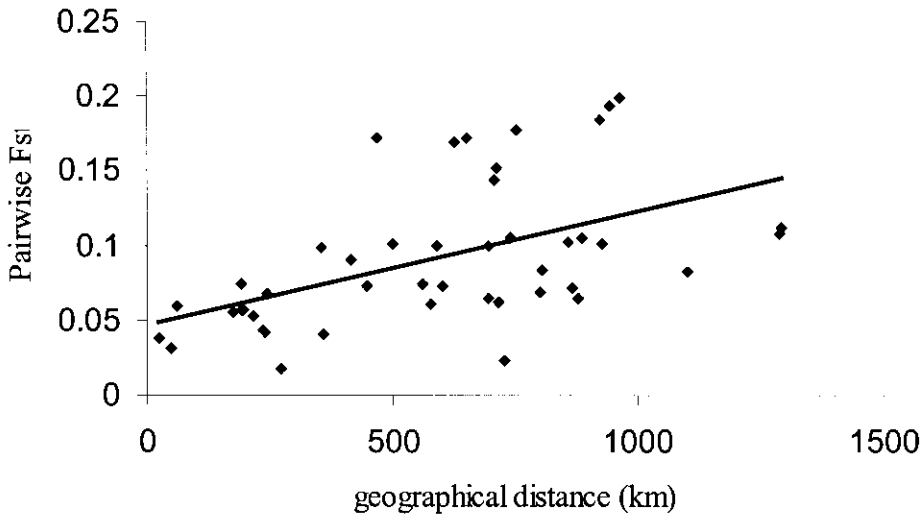

86 88 90 92 94 96 98 100 102 104 106 108 110 112 114 116
1500
1000
500
103

d

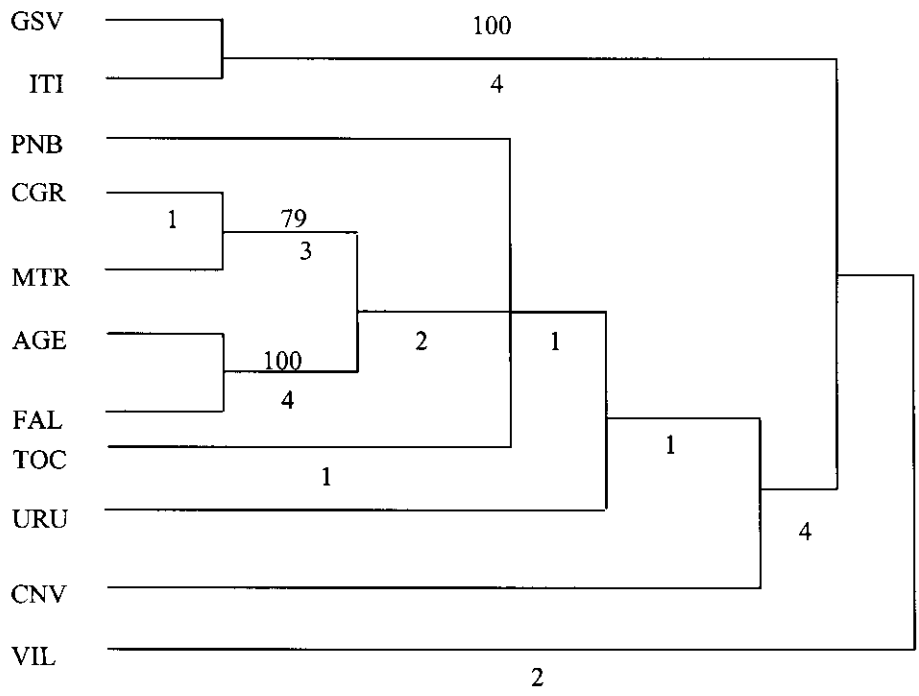

80 90 100 110 120 130 140 150 160 170 180 190 200 210
4000
3000
2000
1000
99
185

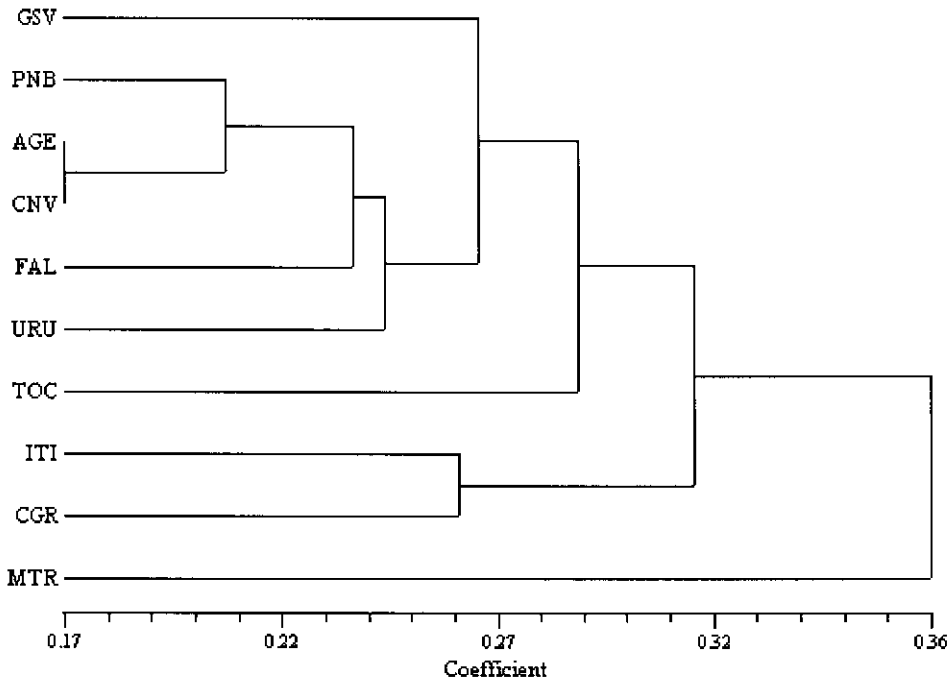

0.3
0.25
0.2
0.15
0.1
0.05
0
pairwise Fst
0
500
1000
1500
geographical distance (km)

(a)


0.25
0.2
0.15
0.1
0.05
0
Pairwise Fst
0
500
1000
1500
geographical distance (km)

(b)


GSV
ITI
PNB
CGR
MTR
AGE
FAL
TOC
URU
CNV
VIL
100
4
1
79
3
2
1
100
4
1
1
4
2


GSV
PNB
AGE
CNV
FAL
URU
TOC
ITI
CGR
MTR
0.17
0.22
0.27
0.32
0.36
Coefficient

Conclusão Geral

- ✓ A eficiência do desenvolvimento dos marcadores SSR, usando enriquecimento de biblioteca genômica, foi de 14,4%, do sequenciamento à obtenção de locos polimórficos com amplificação clara e robusta;
- ✓ Os dez locos SSR desenvolvidos para *C. brasiliense* foram totalmente transferidos para cinco espécies do mesmo gênero: *C. coriaceum*, *C. edule*, *C. glabrum*, *C. pallidum* e *C. villosum*, indicando um alto nível de homologia genômica entre essas espécies, que permitirá a aplicação destes mesmos marcadores para estudos genéticos nas espécies em questão;
- ✓ Os dez locos apresentaram altos níveis de polimorfismo. O número médio de alelos por loco (16 para os dez locos) e a amplitude de heterozigosidade esperada (0,84-0,94) foi maior que valores encontrados para outras espécies tropicais;
- ✓ O alto valor de probabilidade de identidade genética combinada para os dez locos desenvolvidos mostra que os genótipos SSR multilocos são únicos e capazes de discriminar prontamente os indivíduos de *C. brasiliense*. Esta capacidade de discriminação é uma ferramenta poderosa na identificação de clones em populações naturais;
- ✓ Por sua vez, o alto poder de exclusão de paternidade combinada indica que estes marcadores permitem estudos detalhados de parentesco em populações naturais, mesmo quando ambos os pais são desconhecidos;
- ✓ Os microsatélites desenvolvidos e caracterizados nesse trabalho abrem uma nova perspectiva de geração de dados fundamentais para desenvolver estratégias de conservação e manejo, tanto do *C. brasiliense*, quanto de outras espécies do mesmo gênero;

- ✓ Os resultados de estrutura genética indicam que as populações de *C. brasiliense* estão subdivididas segundo o modelo de "isolamento por distância", evidenciado pela correlação positiva entre distância geográfica entre populações e F_{ST} entre pares de populações;
- ✓ Os alelos não estão distribuídos aleatoriamente dentro e entre as populações, uma vez que θ e F são significativamente diferentes de zero. Entretanto, o baixo valor de θ, F>θ e o alto valor de f indicam que a diferenciação é determinada principalmente pelo sistema de acasalamento onde a freqüência de acasalamento entre indivíduos aparentados é maior que o esperado pelo acaso;
- ✓ Sugere-se que, embora *C. brasiliense* seja amplamente distribuído, deve apresentar pequenas unidades panmíticas ou demes, devido à restrição no fluxo gênico. Esta restrição deve ocorrer principalmente por restrições no fluxo de semente. Embora as sementes possam ser dispersas por ema (*Rhea americana*), a maioria das sementes é dispersa por gravidade, permanecendo debaixo da copa da árvore mãe;
- ✓ A fragmentação do Cerrado pode ter aumentado a restrição ao fluxo gênico, isolando os fragmentos devido à matriz de plantações e conglomerados urbanos, dificultando o fluxo de pólen promovido por morcegos que podem permanecer dentro dos fragmentos, aumentando a taxa de acasalamento entre indivíduos aparentados;
- ✓ O fluxo gênico estimado pelo modelo de migração de ilhas foi igual a Nm=3,32, um valor considerado alto o suficiente para contrabalançar os efeitos de deriva e seleção. A existência de uma estrutura espacial segundo o modelo de isolamento por distância viola o modelo de migração de ilhas. A estimativa de fluxo gênico pode, portanto, apresentar um erro superestimando o número de migrantes por populações para determinados pares de populações ou subestimando para outros pares;

- ✓ O estudo do sistema de cruzamento indicou que os marcadores microsatélites proporcionam uma alta resolução para discriminar eventos de auto-polinização e eventos de polinização cruzada entre indivíduos aparentados;
- ✓ A baixa probabilidade de encontrar irmãos germanos em uma mesma família pode ser conseqüência do sistema de polinização de *C. brasiliense*, que é polinizado por morcegos que podem promover fluxo de pólen a grandes distâncias, ou pela existência de aborto seletivo;
- ✓ Embora a taxa de fecundação cruzada tenha sido alta, a taxa de endogamia biparental foi alta, principalmente quando comparada a outras espécies tropicais consideradas alógamas. Esse resultado indica que eventos de polinização cruzada devem ocorrer com alta freqüência entre indivíduos aparentados;
- ✓ O genoma de cloroplasto apresentou maior diferenciação genética que o genoma nuclear, conforme o esperado. Consequentemente, o fluxo gênico via semente, expresso pelo genoma de cloroplasto herdado maternalmente (Nm=0,005), foi menor que o encontrado considerando o genoma nuclear, com herança biparental (3,32). Além disso, a razão entre o fluxo via pólen e semente foi alta (144,01), mostrando que o movimento de pólen é maior que o de semente. Assim, a baixa diferenciação entre populações para o genoma nuclear deve ser atribuído, principalmente, ao fluxo de pólen. Este resultado corrobora os resultados encontrados nos estudos de estrutura genética e sistema de cruzamento;
- ✓ *C. brasiliense* apresentou uma forte estrutura filogeográfica. Porém, a filogeografia não pode ser explicada por divisões geográficas, como por exemplo as Bacias Hidrográficas;
- ✓ Sugere-se que o padrão filogeográfico encontrado seja conseqüência da extinção dos dispersores à longa distância. O gênero *Caryocar* evoluiu no Eoceno (50-44 m.a. A.P.). É possível que os frutos de *C. brasiliense* tenham evoluído em resposta à predação pela megafauna, extinta no último período glacial (ca. 10.000 a. A.P.). O único potencial

dispersor atualmente é a ema (*Rhea americana*, Rheidae) uma grande ave terrestre. Assim, o padrão filogeográfico atual pode ter sido causado pela perda dos dispersores, levando a uma restrição no fluxo de sementes, que atualmente ocorre principalmente por gravidade, mantendo as sementes sob a copa da árvore mãe, isolando as linhagens maternas;

Por outro lado, a última glaciação coincide com um clima mais seco que o atual na região sudeste e central do cerrado, com predomínio de vegetação de campo. É possível que *C. brasiliense,* assim como outras espécies arbóreas de savana, tenham ficado restritas a refúgios, representados por locais com clima mais úmido. Assim, as populações atuais devem ser oriundas das poucas e isoladas populações destes refúgios, que recolonizaram o Cerrado após o restabelecimento de um clima mais úmido (ca. 7000 a A.P.). Como a megafauna já estava extinta, a dispersão das linhagens maternas pode ter sido restrita levando a uma alta subdivisão filogeográfica atual;

Nossos resultados indicam que a fragmentação do cerrado pode levar ao isolamento entre populações de *C. brasiliense* aumentando a endogamia, que pode, potencialmente, levar a uma extinção local. Assim, a manutenção de populações não isoladas, em grandes áreas de preservação, pode ser necessária para manter a viabilidade da espécie no Cerrado, proporcionando meios para que haja fluxo gênico por pólen e semente. Além disso, estas áreas de preservação podem desempenhar um importante papel na manutenção de pequenas populações fragmentadas ou em áreas que sofrem altos níveis de distúrbios, funcionando como doadoras de pólen e semente, além de garantir a viabilidade das populações de polinizadores e potenciais dispersores.

Referência Bibliográfica

Alvarez-Buylla, E.R. & Garay, A.A. 1994. Population genetic structure of *Cecropia obtusifolia*, a tropical pioneer tree species. **Evolution 48**:437-453.

Araújo, F.D. 1994. A review of *Caryocar brasiliense* Camb. (Caryocaraceae) – an economically valuable species of the Central Brazilian cerrados. **Economic Botany 49:**40-48.

Ashton, P.S. 1969. Speciation among tropical forest trees: some deductions in the light of recent evidence. **Biol. J. Linn. Soc. 1**:155-196.

Avise, J.C. 1992. Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. **Oikos 63**:62-76.

Avise, J.C. 1994. **Molecular markers, natural history and evolution**. New York, Chapman & Hall.

Avise, J.C.; Arnold, J.; Ball, R.M.; Bermingham, E.; Lamb, T.; Neigel, J.E.; Reeb, C.A. & Saunders, N.C. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. **Ann. Rev. Ecol. Syst. 18**:489-522.

Barrett, S.C.H. & Kohn, J.R. 1991. Genetic and evolutionary consequences of small population size in plants: implications for conservation. In: Falk,D.A. & Holsinger,K.E., **Genetics and conservation of rare plants**. Oxford, Oxford University Press. pp. 3-30.

Bassam, B.J., Caetano-Anolles, G. & Gresshoff, P.M. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. **Annals of Biochemical 196**:80-83.

Bawa, K.S. 1992. Mating systems, genetic differentiation and speciation in tropical rain forest plants. **Biotropica 24**:250-255.

Birky, C. W. 1988. Evolution and variation in plant chloroplast and mitochondrial genomes. Pp. 23-55 *in* Gottlieb, L.D. & Jain, S.K. eds. **Plant evolutionary biology**. Chapman and Hall, London.

Boshier, D.H.; Chase, M.R. & Bawa, K.S. 1995. Population genetics of *Cordia alliodora* (Boraginaceae), a Neotropical tree. 2. Mating system. **Am. J. Bot. 82:**476-483.

Cavalli-Sforza, L.L. & Edwards, A.W.F. 1967. Phylogenetic analysis: Models and estimation procedures. **Evolution 21**:550-570.

Charlesworth, D. & Charlesworth, B. 1987. Inbreeding depression and its evolutionary consequences. **Ann. Rev. Ecol. Syst. 18**:237-268.

Chase, M.R.; Bosher, D.H. & Bawa, K.S. 1995. Population genetics of *Cordia alliodora* (Boraginaceae), a neotropical tree. 1. Genetic variation in natural populations. **Am. J. Bot. 82**:468-475.

Cocherham, C.C. 1969. Variance of gene frequencies. **Evolution 23**:72-84.

Collevatti, R.G.; Brondani, R.V.P. & Grattapaglia D. 1999. Development and characterization of microsatellite markers for genetic analysis of a Brazilian endangered tree species *Caryocar brasiliense*. **Heredity 83**:748-756.

Dowling, T.E.; Moritz, C. & Palmer, J.D. 1990. Nucleic acids II: restriciton site analysis. In: Hillis, D.M. & Moritz, C. **Molecular systematics**. Sunderland, Sinauer Associates, Inc., Publ. P.250-317.

Doyle, J.J. & Doyle, J.L. 1987. Isolation of plant DNA from fresh tissue. **Focus 12:**13-15.

Ellstrand, N.C. & Elam, D.R. 1993. Population genetic consequences of small population size: implications for plant conservation. **Ann. Rev. Ecol. Syst. 24**:217-242.

Federov, A.A. 1966. The structure of the tropical rain forest and speciation in the humid tropics. **J. Ecol. 54:**1-11.

Felfili, J.M., Filgueiras, T.S., *et al.* 1994. Projeto biogeografia do Bioma Cerrado: Vegetação e solos. **Cad. Geoc. 12** (out/dez):75-166.

Franklin, I.R. 1980. Evolutionary changes in small populations. In: Soulé, M.E. & Wilcox, B.A. (eds), **Conservation biology, an evolutionary-ecological perspective**. Sunderlands, Sinauer Associates, Inc. pp. 135-149.

Furley, P.A. & Ratter, J.A. 1988. Soil resource and plant communities of the central brazilian cerrado and their development. **Journal of Biogeography 15**:97-108.

Gilpin, M.E. & Soulé, M.E. 1986. Minimum viable populations: process of species extinction. In: Soulé, M.E. (ed.), **Conservation biology, the science of scarcity and diversity.** Sunderlands, Sinauer Associates, Inc. pp. 19-34.

Gribel, R. 1986. **Ecologia da polinização e da dispersão de Caryocar brasiliense Camb. (Caryocaraceae) na região do Distrito Federal**. Tese M. Sc., Brasília, Universidade de Brasília.

Gribel, R. & Hay, J.D. 1993. Pollination ecology of Caryocar brasiliense (Caryocaraceae) in Central Brazil cerrado vegetation. **J. Trop. Ecol. 9**:199-211.

Hamrick, J.L. & Loveless, M.D. 1986. Isozyme variation in tropical trees: procedures and preliminary results. **Biotropica 18**:201-207.

Hamrick, J.L. & Loveless, M.D. 1989. The genetic structure of tropical tree populations: associates with reproductive biology. In: BOCK, J.H. & Linhart, Y.B. **The evolutionary ecology of plants.** Boolder, Westview Press. p. 129-146.

Hanski, I. 1989. Metapopulation dynamics: does it help to have more of the same? **Trends in Ecology and Evolution** 4:113-114.

Hastings, A. & Harrison, S. 1994. Metapopulation dynamics and genetics. **Ann. Rev. Ecol. Syst.** 25:167-188.

Haberl, M. & Tautz, D. 1999. Comparative allele sizing can produce inaccurate allele size differences for mcirosatellites. **Mol. Ecol. 8**:1347-1350.

James, T.; Vege, S.; Aldrich, P. & HAMRICK, J.L. 1998. Mating system of three tropical dry forest tree species. **Biotropica 30**:587-594.

Kimura, M. 1953. "Stepping stone" model of population. **Ann. Rept. Nat. Inst. Genetics, Japan 3:**62-63.

Kimura, M. & Maruyama, T. 1971. Pattern of neutral polymorphism in a geographically structured population. **Genet. Res. 18**:125-131.

Lewis, P. & Zaykin, D. 1999. **Genetic data analysis: computer program for the analysis of allelic data. Version 1.0 (d12).** Free program distributed by the authors over the internet from the GDA Home Page at http://chee.unm.edu/gda/

Loiselle, B.A.; Sork, V.L. & Graham, C. 1995a. Comparison of genetic variation in bird-dispersed shrubs of a tropical wet forest. **Biotropica 27**:487-494.

Loiselle, B.A.; Sork, V.L.; Nason, J. & Graham, C. 1995b. Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). **Am. J. Bot. 82**:1420-1425.

Loveless, M.D.; Hamrick, J.L. & Foster, R.B. 1998. Population structure and mating system in *Tachigali versicolor*, a monocarpic Neotropical tree. **Heredity 81:**134-143.

Macedo, J. 1995. Os cerrados brasileiros: alternativa para a produção de alimentos no limiar do século XXI. **Revista de Política Agrícola, IV**(2):11-18.

Mantel, N.A. 1967. The detection of disease clustering and a generalized regression approach. **Cancer Research 27:**209-220.

Marshall, D.L. & Folsom, M.W. 1991. Mate choice in plants: an anatomical to population perspective. **Ann. Rev. Ecol. Syst. 22**:37-63.

McCauley, D.E. 1995. The use of chloroplast DNA polymorphism in studies of gene flow in plants. **Trends Ecol. Evol. 10**:198-202.

McCauley, D.E.; Stevens, J.E.; Peroni, P.A. & Raveill, J.A., 1996. The spatial distribution of chloroplast DNA and allozyme polymorphisms within a population of *Silene alba* (Caryophyllaceae). **Am. J. Bot. 83:**727-731.

Murawski, D.A. & Hamrick, J.L. 1991. The effect of the density of flowering individuals on the mating systems of nine tropical tree species. **Heredity 67:**167-174.

Murawski, D.A. & Hamrick, J.L. 1992. Mating system and phenology of *Ceiba pentandra* (Bombacaceae) in Central Panama. **J. Hered. 83:**401-404.

Murawski, D.A.; Dayanandan, B. & Bawa, K.S. 1994. Outcrossing rate of two endemic *Shorea* species from Sri lankan tropical rain forest. **Biotropica 26:**23-29.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individual. **Genetics 89**:583-590.

O'Malley, D.M. & Bawa, K.S. 1987. Mating system of a tropical rain forest tree species. **Am. J. Bot. 74:**1143-1149.

Pennington, R.T. 1995. Cladistic analysis of chloroplast DNA restriciton site characters in *Andira* (Leguminosae: Dalbergieae). **Am. J. Bot. 82**:526-534.

Prance, G.T. & Silva, M.F. 1973. **Flora Neotropica: Caryocaraceae**. Monograph n° 12. Hafner, New York.

Ratter JA, Ribeiro JF, Bridgewater S (1997) The Brazilian cerrado vegetation and threats to its biodiversity **Annals of Botany 80**, 223-230.

Raymond, M. & ROUSSET, F. 1995. GENEPOP (version 1.2.): population genetics software for exact tests and ecumenism. **J. Heredity 86:**248-249.

Ritland, K. 1989. Correlated matings in the partial selfer *Mimulus guttatus*. **Evolution 43:**848-859.

Ritland, K. 1996. **Multilocus mating system program – MLTR**. Available free of charge via ftp at 128.100.165.100.

Ritland, K. & Jain, S. 1981. A model for the estimation of outcrossing rate and gene frequencies using n independent loci. **Heredity 47:**35-52.

Rhoads, D.D. & Roufa, D.J. 1990. **Seqaid II 3.80**. Molecular Genetics Laboratory, Kansas State University, Manhattan.

Schaal, B.A.; Hayworth, D.A.; Olsen, K.M.; Rauscher, J.T. & Smith, W.A. 1998. Phylogeographic studies in plants: problems and prospects. **Mol. Ecol. 7**:465-474.

Schneider, S.; Kueffer, J.M.; Roessli, D. & Excoffier, L. 1997. **Arlequin Ver. 1.1: A software for population genetic data analysis.** Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Shaanker, R.U; Ganeshaiah, K.N. & Bawa, K.S. 1988. Parent-offspring conflict, sibling rivalry, and brood size patterns in plants. **Ann. Rev. Ecol. Syst. 19:**177-205.

Silva, F.C. 1990. **Compartilhamento de nutrientes em diferentes componentes da biomassa aérea em espécies arbóreas de um cerrado**. Tese M.Sc. Brasília, Universidade de Brasília.

Slatkin, M. 1985. Gene flow in natural populations. **Ann. Rev. Ecol. Syst. 16**:393-430.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. **Science 236**:787-792.

Slatkin, M. 1991. Inbreeding coefficients and coalescence times. **Genetic Research of Cambridge 58:**167-175.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. **Genetics 139**:457-462.

Slatkin, M. & Barton, N.H. 1989. A comparison of three indirect methods for estimating average levels of gene flow. **Evolution 43:**1349-1368.

Smith, J.F. & Doyle, J.J. 1995. A cladistic analysis of chloroplast DNA restriciton site variation and morphology for the genera of the Junglandaceae. **Am. J. Bot. 82**:1163-1172.

Swofford, D.L. 1993. **PAUP: phylogenetic analysis using parsimony Ver. 3.1.1.** Illinois Natural History Survey, Champaign, IL.

Swofford, D.L. & OLSEN, G.J. 1990. Phylogeny reconstruction. Pp. 411-501 *in* Hills, D.M. & Moritz, C. eds. **Molecular systematics**. Sinauer, Sunderland, MA.

Terborgh, J. & Winter, B. 1980. Some causes of extinction. In: Soulé, M.E. & Wilcox, B.A. (eds), **Conservation biology, an evolutionary-ecological perspective**. Sunderlands, Sinauer Associates, Inc. pp. 119-133.

Weir, B.S. 1996. **Genetic Data Analysis II**. Sinauer Associates, Sunderland, MA.

Weising, K. & Gardner, R.C. 1999. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. **Genome 42**:9-19.

Whitlock, M.C. & McCauley, D.E. 1999. Indirect measures of gene flow and migration: $F_{st} \neq 1/(4Nm+1)$. **Heredity 82**:117-125.

Wilson, E.O. 1994. **Diversidade da vida**. São Paulo, Companhia das Letras.

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

Wright, S. 1940. Breeding structure of populations in relation to speciation. **Am. Nat. 74**:232-248.

Wright, S. 1943. Isolation by distance. **Genetics 28:**114-138.

Wright, S. 1951. The genetic structure of populations. **Annual Eugenics 15:**323-354.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. **Evolution 19**:395-420.

Wright, S. 1968. **Evolution and the genetics of populations. Vol. 1 Genetic and biometric foundations**. London, The University of Chicago press.

Wright, S. 1978. **Evolution and the genetics of populations.** Vol. IV. London, The University of Chicago press.