Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Análises Filogenéticas no Gênero *Anacardium*

Alexandre Franco Garcia

Tese apresentada para obtenção do título de Doutor em
Agronomia. Área de Concentração: Genética e
Melhoramento de Plantas

Piracicaba

2009

Alexandre Franco Garcia
Engenheiro Agrônomo

Análises Filogenéticas no Gênero *Anacardium*

Orientador:
Prof. Dr. GIANCARLO CONDE XAVIER OLIVEIRA

Tese apresentada para obtenção do título de Doutor em
Agronomia. Área de Concentração: Genética e Melhoramento
de Plantas

Piracicaba
2009

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Garcia, Alexandre Franco

Análises filogenéticas no gênero *Anacardium* / Alexandre Franco Garcia. - - Piracicaba, 2009.

76 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2009.
Bibliografia.

1. Caju 2. Filogenia 3. Sequenciamento genético 4. Variação genética vegetal I. Título

CDD 634.573
G216a

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

Ofereço todo meu esforço, toda minha alegria e angústia na realização dessa etapa importante da minha vida a minha família. Zaíra sabe muito bem o papel que tem. Minha mãe e meu irmão sabem muito bem o quanto são essenciais. Minha irmã sabe muito bem o quanto faz falta.

Amélie, nós sabemos exatamente o que ninguém sequer imagina. Você consegue imaginar o que significa estarmos aqui agora? E agora? Amo muito você.

Agradecimentos

Agradeço a Universidade de São Paulo e à Escola Superior de Agricultura “Luiz de Queiroz” pela oportunidade oferecida.

Particularmente, ao Departamento de Genética da Esalq e ao CNPq por todo suporte na minha formação.

Agradeço, especialmente, ao prof. Dr. Giancarlo C. X. Oliveira, pela orientação, liberdade, convívio excepcional (embora raro de minha parte). Giancarlo fez por mim muito mais do que pode supor e, se precisar, tem um amigo.

Agradeço ao prof. Dr. José Baldin Pinheiro e a profa. Dra Anete Pereira de Souza pela disponibilidade sempre pronta em me auxiliar no término do trabalho.

Muitas pessoas contribuíram de muitas maneiras diferentes e a elas sou extremamente grato. José Galvão e família pela acolhida em Mato Grosso, ao Marcelo Mollinari e ao Fernando Piotto por tornar o caminho muito mais divertido. Ao pessoal da Embrapa Fortaleza pela imensa disposição em me ajudar.

SUMÁRIO

RESUMO.....	9
ABSTRACT.....	11
1 INTRODUÇÃO.....	13
2 Desenvolvimento	15
2.1 Revisão Bibliográfica.....	15
2.1.1 Família, Gênero e Espécie.....	15
2.1.2 Origem, Distribuição e Diversidade.....	16
2.1.3 Análises Filogenéticas.....	18
2.1.3.1 Filogenia.....	18
2.1.3.2 Sequências de DNA.....	19
2.1.3.3 Distâncias Evolutivas.....	20
2.1.3.4 Alinhamento das Sequências de Nucleotídeos.....	23
2.1.3.5 Árvores Filogenéticas.....	23
2.1.3.6 Filogenias Moleculares e Diversidade em Anacardiaceae.....	27
2.2 Material e Métodos.....	28
2.2.1 Material Vegetal.....	28
2.2.2 Isolamento e Amplificação de DNA.....	28
2.2.3 Sequenciamento de DNA	30
2.2.4 Análise das Sequências.....	30
2.2.5 Blast das Sequências obtidas.....	30
2.3 Resultados e Discussão.....	32
2.3.1 Coletas.....	32
2.3.2 Sequenciamento.....	39
2.3.3 Variação Intraespecífica.....	59
2.3.4 Variação Interespecífica.....	64
3 CONSIDERAÇÕES FINAIS.....	69
REFERÊNCIAS.....	71

RESUMO

Análises filogenéticas no gênero *Anacardium* L.

O gênero *Anacardium* L. é composto por um número relativamente pequeno de espécies arbóreas e arbustivas, dentre elas *Anacardium occidentale* L., única espécie cultivada e distribuída em todo o mundo tropical. O Brasil é o provável centro de origem, com dois centros de diversidade identificados. Neste trabalho, duas regiões gênicas comumente empregadas (ITS1 e trnL) foram utilizadas para avaliar a diversidade intra e interespecífica do gênero. Há evidências de variação intraespecífica principalmente para a região ITS1, que dificulta as reconstruções filogenéticas, uma vez que relativamente pouca variação interespecífica foi detectada para as espécies avaliadas. Para a região trnL, há uma evidência de que essas regiões podem identificar diferenças entre os centros de diversidade do gênero. Os agrupamentos observados nas árvores sugerem o mesmo tipo de agrupamento já descrito na literatura por taxonomia numérica. No entanto, mais dados devem ser avaliados para testar essa evidência.

Palavras chave: Caju, *Anacardium*, filogenia.

ABSTRACT

Phylogenetic analysis of the genus *Anacardium* L.

The genus *Anacardium* L. is composed by a relatively small number of species, being *A. occidentale* the only species that is cultivated and distributed in all tropical regions. Brasil is probably the center of origin, with two different diversity centers. Here, two genic regions commonly used (ITS1 e trnL) were employed to evaluate the diversity present between and within species of the genus. The results showed evidences of the existence of intraspecific variation, mainly for the ITS1 region, which can difficult the phylogenetic reconstructions, since small variation was observed among the analysed species. The trnL region showed an evidence that this region could be used to discriminate the two diversity centers. The groups obtained suggest the same groups observed in the literature by numerical taxonomy. However more data is necessary to test those evidences.

Key words: Cashew, *Anacardium*, phylogeny

1 INTRODUÇÃO

O gênero *Anacardium* L. é composto por um número relativamente pequeno de espécies arbóreas e arbustivas (MITCHELL; MORI, 1987), dentre elas *Anacardium occidentale* L., única espécie cultivada e distribuída em todo o mundo tropical (BARROS; CRISÓSTOMO 1995).

Dentre os países com grande concentração da planta estão Índia, Tanzânia, Quênia e Moçambique. No Brasil, o cajueiro é encontrado em praticamente todos os estados brasileiros, mas adapta-se melhor às condições ecológicas do litoral do Nordeste (BARROS; CRISÓSTOMO, 1995, BARROS et al., 2002)

Provas circunstanciais indicam que a origem da espécie pode ter ocorrido na região que inclui o Norte da América do Sul e parte da América Central. Mais especificamente, indicam que o centro de origem mais provável é o Brasil (BARROS; CRISÓSTOMO, 1995, BARROS et al., 2002), sendo a Amazônia o centro de diversidade primário e o Planalto o centro de diversidade secundário (MITCHELL; MORI, 1987).

Embora a espécie cultivada tenha ampla distribuição e utilização comercial (PESSOAL; LEITE; PIMENTEL, 1995), há relativamente pouca informação sobre as outras espécies do gênero e suas inter-relações. Pela taxonomia clássica (JOHNSON, 1973, 1974) foram identificadas 21 espécies, posteriormente reduzidas a 10 pela taxonomia numérica (MITCHELL; MORI, 1987), que, além disso, identificou três agrupamentos principais (clados), com base em informações morfológicas.

Agrupamentos desse tipo são importantes, pois permitem identificar relações filogenéticas entre espécies, como, por exemplo, a relação de espécies domesticadas com seus ancestrais selvagens. Ainda, podem direcionar o estabelecimento de conjuntos gênicos, que refletem a capacidade de intercruzamento das espécies. Tais informações são relevantes para o melhoramento genético, para o manejo de recursos genéticos e conservação, pois permitem definir a utilização do germoplasma e a sistematização de coletas.

Nesse sentido, os estudos filogenéticos baseados em seqüências de DNA têm fornecido resultados bastante promissores (CLEGG; ZURAWSKI, 1992; SOLTIS; SOLTIS; MILLIGAN, 1992; SOLTIS; SOLTIS, 1998; AVISE, 2000; SAVOLAINEN; CHASE, 2003; AVISE, 2004),

como, por exemplo, o entendimento de padrões de ramos filogenéticos difíceis de elucidar pelas metodologias anteriores (SAVOLAINEN; CHASE, 2003). Dentre as vantagens da metodologia, está a capacidade de avaliar grande quantidade de informação com significado evolutivo e padrão relativamente regular no tempo (NEI; KUMAR, 2000).

De forma geral, são escolhidas regiões das seqüências de DNA com variação suficiente para investigar os organismos em suas respectivas posições taxonômicas (famílias, gêneros, etc.) (FIJII, 2007). Contudo, para muitas espécies assume-se que a variação intraespecífica dessas seqüências é ausente ou desprezível, embora poucos estudos tenham sido feitos para avaliar tais variações (ROSSELLO et al., 2007).

Dessa forma, o objetivo deste trabalho foi estabelecer relações filogenéticas de espécies de caju, levando em consideração a variabilidade do gênero no Brasil, isto é, considerando as variações das seqüências de DNA de indivíduos de espécies distintas e também de indivíduos dentro da mesma espécie.

2 DESENVOLVIMENTO

2.1 Revisão Bibliográfica

2.1.1 Família, Gênero e Espécie

A família Anacardiaceae é composta de 82 gêneros e mais de 700 espécies. É economicamente importante e tem distribuição ampla, concentrando-se principalmente nos trópicos e sendo ausente apenas em latitudes muito altas e algumas regiões áridas. É subdividida em cinco tribos: Anacardieae, à qual pertence o gênero *Anacardium*; Spondiadeae, à qual pertence o gênero do cajá (*Spondias*); Semecarpeae, à qual pertence a espécie indiana *Semecarpus anacardium*; Rhoeae, tribo à qual pertence o gênero *Rhus*; e Dobineae (PELL, 2004).

A família apresenta diversos registros fósseis, devido à ampla distribuição mundial e ao porte arbóreo. Pólen e madeira foram considerados pertencentes ao Paleoceno (65 a 55 milhões de anos atrás) (HSU, 1983; MULLER, 1984), enquanto a origem da ordem Sapindales, à qual a família pertence, foi datada em cerca de 84 a 65 milhões de anos atrás. (KNOBLOCH; MAI, 1986; MAGALLÓN; SANDERSON, 2001; WIKSTRÖM et al., 2001). Por outro lado, há poucos registros fósseis do gênero. Fósseis datados encontrados na América do Sul posicionam o gênero nos períodos Eoceno e Oligoceno (BARROS; CRISÓSTOMO, 1995).

Os gêneros da família são cultivados em muitas regiões do mundo e têm uso diverso. De acordo com as características das plantas, são aproveitados frutos, sementes, madeira e compostos medicinais. Além do gênero *Anacardium*, há outros gêneros conhecidos como o *Mangifera*, ao qual pertence a manga, o *Pistacia* (pistache) e o *Spondias* (cirigüela e marula) (BARROS; CRISÓSTOMO, 1995; PELL, 2004).

O número de espécies do gênero *Anacardium* é relativamente pequeno. No entanto, a identificação das espécies não é fácil, além de que algumas delas inter cruzam. Pela taxonomia clássica (JOHNSON, 1973, 1974) foram identificadas 21 espécies, posteriormente reduzidas a 10 pela taxonomia numérica (MITCHELL; MORI, 1987). Dentre essas espécies, está *Anacardium occidentale* L., única espécie cultivada e distribuída em todo o mundo tropical (BARROS; CRISÓSTOMO, 1995).

2.1.2 Origem, Distribuição e Diversidade

Diversas provas circunstanciais indicam que a origem da espécie pode ter ocorrido na região que inclui o Norte da América do Sul e parte da América Central ou, mais especificamente, no Brasil: (i) referências antigas da planta; (ii) distribuição geográfica das espécies; (iii) comportamento ecológico; (iv) padrões de variação; (v) registros fósseis e (vi) atividades humanas (BARROS; CRISÓSTOMO, 1995).

A referência mais antiga do gênero foi feita pelo naturalista francês André de Thevet, após sua visita pela costa do Nordeste e Norte do Brasil (THEVET, 1558). É uma ilustração que retrata a colheita de frutos de um cajueiro (Figura 1)

Figura 1 – Primeiro registro conhecido do gênero *Anacardium*, representando a colheita de frutos de um cajueiro.

Em relação à distribuição geográfica, das 21 espécies de *Anacardium* descritas pela taxonomia clássica, apenas três não ocorrem no território nacional e, dentre essas três, apenas uma não ocorre na América do Sul (JOHNSON, 1973, 1974). Mitchell e Mori (1987) descrevem uma distribuição natural do gênero desde o sul de Honduras até o Estado do Paraná-BR e leste do Paraguai.

Quanto ao comportamento ecológico, de maneira geral, todas as espécies do gênero ocorrem em matas e cerrados no Brasil e em matas fora do território nacional. A espécie *A. occidentale* L é a única dispersa em todo mundo tropical, dos Estados Unidos à África do Sul.

Contudo, as referências sobre essa dispersão de *A. occidentale* L. são sempre posteriores a 1558, data da publicação de Thevet e da presença de portugueses e espanhóis no Brasil, sendo que não há referências sobre *Anacardium* nas crônicas espanholas da época da colonização da América Central (BARROS; CRISÓSTOMO, 1995). Ainda, a origem africana ou asiática do gênero foi descartada, quando foram avaliadas as relações entre cultivo e origem geográfica (CANDOLLE, 1959). Assim, a consideração de todos esses aspectos em conjunto sugere que o cajueiro foi introduzido nas regiões fora do Brasil (BARROS; CRISÓSTOMO, 1995).

Outro indicativo de introdução do cajueiro fora do território nacional é quanto ao padrão de variação de *A. occidentale*. A espécie é a que apresenta maior variabilidade em relação aos caracteres de interesse do homem, sendo que, no Brasil, há grande variabilidade para esses caracteres e nos demais países com grande concentração da planta, há maior uniformidade de tais caracteres (BARROS; CRISÓSTOMO, 1995).

Conforme já comentado, há poucos registros fósseis do gênero. Contudo, considerando as informações disponíveis, há registros de fósseis na Bahia, Equador e Peru, isto é, todos na América do Sul (JOHNSON, 1974).

O Brasil é região do mundo onde há maior tradição de usos e costumes relacionados ao cajueiro, que pode indicar maior convivência do homem com a espécie (MARCGRAVE, 1942). Além disso, há indícios de que os termos indígenas que denominam o cajueiro e o caju (acaiuba e acaiu, respectivamente) foram transformados nos termos em Português que denominam a planta e o fruto (LIMA et al., 1952; GUANDAVO, 1965). Posteriormente, os termos em Português foram derivados para as outras línguas (BARROS; CRISÓSTOMO, 1995).

O gênero *Anacardium* tem dois centros de diversidade: a Amazônia Central, centro de diversidade primário contendo quatro espécies, e o Planalto do Brasil, centro de diversidade secundário contendo três espécies. A espécie com distribuição mais ampla é *A. occidentale*, que apresenta populações disjuntas no Planalto do Brasil, nas restingas do Nordeste, nas savanas da Amazônia e “llanos” da Colômbia e Venezuela, embora a distribuição natural da espécie seja difícil de estabelecer, considerando a grande distribuição de áreas cultivadas tanto no Velho quanto no Novo Mundo (MITCHELL; MORI, 1987).

2.1.3 Análises Filogenéticas

2.1.3.1 Filogenia

A filogenética estuda as relações evolutivas entre diferentes grupos de organismos. Cada unidade de comparação é denominada táxon e cada grupo de comparação é denominado taxa. Assim, a filogenética estuda a história evolutiva e a relação de táxons biológicos (SAVOLAINEN; CHASE, 2003).

Desde Charles Darwin, pesquisadores almejam reconstruir a história evolutiva de todos os organismos na Terra e expressar essa reconstrução na forma de uma árvore filogenética (HAECKEL, 1866). A estratégia ideal para isso seria a utilização de registros fósseis, mas uma vez que registros fósseis são raros e incompletos, pesquisadores têm empregado métodos morfológicos e fisiológicos comparativos, o que permitiu a inferência sobre aspectos importantes da história evolutiva dos organismos (NEI; KUMAR, 2000).

Contudo, mudanças evolutivas de caracteres morfológicos e fisiológicos são tão complexas, que não fornecem quadros evolutivos claros e as árvores filogenéticas obtidas normalmente são controversas (NEI; KUMAR, 2000). Ainda, caracteres morfológicos podem apresentar adaptações que não refletem relações genealógicas com fidelidade, tais como convergências e paralelismos (SOLTIS; SOLTIS, 1998).

Os avanços da biologia molecular tornaram possível estudar as variações das moléculas de DNA, que contêm as informações essenciais de (quase) todos os organismos. Dessa forma, é possível estudar as relações evolutivas dos organismos comparando suas moléculas de DNA (NEI; KUMAR, 2000).

Essa abordagem apresenta algumas vantagens em relação à abordagem anterior: (i) permite comparar quaisquer grupos de organismos (por exemplo, bactérias, plantas e animais), uma vez que (quase) todos os organismos armazenam suas informações da mesma maneira; (ii) permite a comparação de organismos distantes do ponto de vista evolutivo, pela possibilidade de modelar as mudanças na molécula de DNA, que segue um padrão relativamente regular e (iii) permite acessar uma quantidade de informação filogenética maior que a de caracteres

morfológicos, uma vez que os genomas dos organismos são constituídos de longas seqüências de nucleotídeos (NEI; KUMAR, 2000).

Uma vez que todo caráter morfológico ou fisiológico de um organismo é controlado pela informação genética contida em seu DNA, qualquer mudança nesse caráter é devida a algum tipo de mudança na molécula de DNA (NEI; KUMAR, 2000)

2.1.3.2 Seqüências de DNA

Há diversas ferramentas moleculares que podem ser empregadas para inferências filogenéticas, dentre elas análises de sítio de restrição, seqüenciamento comparativo, análise de rearranjos de DNA e perdas de genes e íntrons, além de diversas técnicas baseadas em PCR (SOLTIS; SOLTIS, 1998). Nos últimos anos, contudo, a utilização de seqüências de DNA aumentou bastante, basicamente devido à rapidez e à facilidade de obtenção dos dados. Conseqüentemente, o número de genes e regiões do DNA disponíveis para utilização em análises filogenéticas também aumentou. Enquanto no início da década de 90 os estudos filogenéticos eram realizados predominantemente com seqüências *rbcL* e alguns RNAs ou DNAs ribossomais, atualmente muitas outras seqüências tem sido utilizadas, tais como ITS, *nadhF*, *matK*, genes de RNA 18S, dentre muitas outras (SOLTIS et. al., 1998).

De forma geral, estudos filogenéticos em plantas envolvem a utilização do genoma do cloroplasto e algumas seqüências gênicas nucleares. As vantagens da utilização de genomas de cloroplasto em estudos filogenéticos incluem o fato de esse genoma ser pequeno, tornando relativamente fácil o estudo do genoma completo. Além disso, a maioria dos genes de cloroplasto é de cópia única, diferentemente da maioria dos genes nucleares. Ainda, a evolução desse genoma é considerada conservativa, embora diferentes porções do genoma evoluam em diferentes taxas. Dessa forma, há muitas possibilidades para resolver as relações entre grupos, pelo uso de dados obtidos a partir de genomas de cloroplasto (STRAND; LEEBENS-MACK; MILLIGAN, 1997; SOLTIS; SOLTIS, 1998).

A taxa de evolução varia bastante entre genes e seqüências, portanto, a escolha do gene ou região mais adequada dependem da espécie em questão e também do nível de estudo (ordens, famílias ou gêneros, por exemplo). Assim, pode-se estudar as relações evolutivas de virtualmente qualquer nível de classificação de organismos (por exemplo: reino, filo, família, gênero, espécie e

populações intraespecíficas) através do uso de diferentes genes ou segmentos de DNA (WILSON et al., 1977; DAYHOFF et al., 1978). Por exemplo, McDade e Moody (1999) utilizaram a região trnL-F e conseguiram estabelecer relações filogenéticas entre gêneros, porém não obtiveram bons resultados em níveis taxonômicos mais baixos, enquanto Rossello et al. (2007) observaram que há variação intraespecífica para a região ITS.

Muitos outros estudos utilizam diferentes regiões, para objetivos específicos. De forma geral, a região rbcL tem sido usada para grupos de níveis mais elevados, (famílias, etc), bem como IGS, 26S-25R, pois são altamente conservados, enquanto matk tem se mostrado apropriado para resolver relações intergenéricas ou interespecíficas em plantas, devido à alta taxa de evolução que apresenta, bem como ITS5 e ITS2, utilizados para comparar espécies e gêneros proximamente correlacionados. Ainda, as regiões trnLF têm se mostrado úteis em estudos intrafamiliares (HAM et al., 1994, GIELLY; TABERLET, 1994; RENNER, 1998; KUSUMI et al., 2000), mesmo com essas regiões apresentando taxas de evolução mais que três vezes maiores que as taxas de rbcL (SOLTIS et al., 1998, PELL, 2004).

2.1.3.3 Distâncias Evolutivas

A causa primária de evolução é a mudança de genes pela mutação. A mutação num gene ou, de maneira mais geral, numa sequência de DNA pode ser causada por alterações nos nucleotídeos por substituições, inserções, deleções, recombinações e conversões gênicas. Essas alterações podem ser distribuídas na população por deriva genética e/ou seleção natural. Eventualmente, também podem ser fixadas nas espécies (NEI, 1987; HARTL; CLARK, 1997).

Se um gene mutante produz um novo caráter morfológico ou fisiológico, esse caráter será herdado pelos descendentes das espécies, a menos que sofra novas mutações. Assim, se for possível estabelecer uma árvore filogenética válida para um grupo de espécies, é possível identificar linhagens de espécies nas quais um determinado caráter apareceu por mutação (NEI; KUMAR, 2000).

Há diferentes formas de considerar as alterações observadas nas sequências de DNA. A partir das suposições feitas sobre essas alterações, foram desenvolvidos diferentes métodos estatísticos para medir as distâncias evolutivas entre as sequências (SOLTIS; SOLTIS, 1998).

Quando duas seqüências de DNA são derivadas de uma seqüência ancestral comum, as seqüências descendentes gradualmente divergem por substituição de nucleotídeos. Uma forma simples de medir essa divergência é através da distância **p**, que mede a proporção de nucleotídeos diferentes entre as seqüências (número de bases diferentes pelo tamanho total da seqüência). Assim, quanto maior o valor de **p**, mais divergentes são as seqüências, o que sugere que as diferenças entre duas seqüências tendem a aumentar à medida que o tempo de divergência (**t**) entre elas aumenta (NEI; KUMAR, 2000).

Contudo, a relação entre **p** e **t** não é linear, uma vez que mais de uma substituição pode ocorrer na mesma posição. Assim, pode-se corrigir essa relação utilizando a distribuição de Poisson. Nesse caso, tem-se a distância corrigida de Poisson (**PC**). Ainda, nos cálculos anteriores assume-se que a taxa de substituição nucleotídica (**r**) é constante para todas as posições, situação que raramente ocorre na prática. Análises estatísticas de **r** em diferentes posições sugerem que sua variação aproxima-se da distribuição gama, informação que pode ser considerada nos cálculos de distância. (NEI; KUMAR, 2000).

Uma vez que há diferentes tipos de substituição, outras medidas úteis são a frequência de pares de nucleotídeos diferentes entre duas seqüências e a relação transição/transversão (**R**). A distância **p** fornece uma estimativa do número de substituições por posição quando as seqüências são intimamente relacionadas. Contudo, quando o valor de **p** é grande, esse valor é subestimado, pois não considera as substituições paralela e reversa. Assim, há diversos modelos de substituição disponíveis.

Um dos modelos mais simples para estimar o número de substituições nucleotídicas (**d**) é o método de Jukes e Cantor (JUKES; CANTOR, 1969), que assume que a taxa de substituição de nucleotídeos (**r**) ocorre com frequências iguais em qualquer posição e que em cada posição um nucleotídeo é trocado por um dos outros três com a mesma probabilidade. No entanto, em dados reais, a taxa de substituição nucleotídica por transição é normalmente maior que a taxa de substituição nucleotídica por transversão. Assim, o método de Dois Parâmetros de Kimura (KIMURA, 1980) leva em conta essa observação. Além disso, permite estimar o número de substituições nucleotídicas por transição (**s**) e o número de substituições nucleotídicas por transversão (**v**).

O método de Tajima e Nei (FELSENSTEIN, 1981; TAJIMA; NEI, 1982) utiliza frequências nucleotídicas (**g**) para estimar (**d**) e, por isso, é aparentemente menos sensível a fatores que prejudicam a estimativa do número de substituições. Na prática, esse método fornece valores maiores de **d**, uma vez que as taxas de substituição nucleotídicas não são iguais.

O método de Tamura (TAMURA, 1992) é uma extensão do método de Dois Parâmetros de Kimura. No método de Tamura, **d** é estimado levando em consideração o conteúdo de GC das espécies (**θ**) que normalmente difere de 0,5.

Um dos modelos frequentemente utilizados em análises filogenéticas é o método de máxima verossimilhança de Hasegawa (HASEGAWA; KISHINO; SAITOU, 1991), que leva em consideração a relação transição/transversão, além das frequências nucleotídicas e do conteúdo de GC. O método de Tamura e Nei (TAMURA; NEI, 1993) utiliza o mesmo tipo de modelo para calcular **d**.

Na prática, é possível observar que quando o número de substituições nucleotídicas (**d**) é baixo (**d** < 0,25) os métodos fornecem resultados semelhantes. Mesmo o método de Jukes-Cantor fornece resultados semelhantes aos demais quando **p** < 0,1. Dessa forma, não necessariamente medidas sofisticadas de distância fornecem os melhores padrões de árvore (topologia). Contudo, na estimativa dos comprimentos dos ramos, as medidas de distância que melhor se ajustam aos dados fornecem os dados mais confiáveis (NEI; KUMAR, 2000).

Ainda sobre as distâncias, todos os métodos descritos anteriormente não consideram a distribuição gama de **r** nos cálculos de distância. As distâncias gama são mais realistas que as não-gama e fornecem melhores resultados de distância. Contudo, apresentam também maior variância. Assim, não necessariamente fornecem os melhores resultados para inferência filogenética, a não ser que o número de nucleotídeos (**n**) empregados na análise seja grande (NEI; KUMAR, 2000).

Há também os métodos numéricos de estimação de distâncias evolutivas (todos os métodos anteriores são analíticos), que permitem, por exemplo, estimar parâmetros dos próprios dados, mesmo para métodos com vários parâmetros para estimar, como, por exemplo, o método Tamura-Nei utilizando **r** com distribuição gama. Para isso, utilizam máxima verossimilhança.

Ainda, permitem selecionar os modelos de substituição mais apropriados, através do teste de razão de verossimilhança (NEI; KUMAR, 2000).

Assim como os métodos gama, os métodos numéricos necessitam de valores grandes de **n**. Quando **n** é pequeno, não necessariamente fornecem boas estimativas para obtenção da topologia (NEI; KUMAR, 2000).

2.1.3.4 Alinhamento das Seqüências de Nucleotídeos

Normalmente, o número de nucleotídeos das seqüências utilizadas em análises filogenéticas é diferente entre as seqüências, pela presença de inserções ou deleções, que introduzem falhas nas seqüências. Quando a divergência entre as seqüências é pequena e o número de seqüências também é pequeno, inferir a posição das falhas é relativamente fácil. Em outras situações há programas computacionais que realizam a tarefa, de acordo com diferentes métodos, dentre eles aqueles baseados em índice de similaridade (NEDLEMAN; WUNSCH, 1970) e na distância de alinhamento, que insere penalidades em parâmetros de alinhamento (SELLERS, 1974).

A presença de falhas de alinhamento nas seqüências introduz complicações nas estimativas das distâncias evolutivas, além de que falhas na seqüência podem ser resultado de falhas experimentais. Normalmente, essas posições são ignoradas durante as estimativas. Contudo, há duas formas de tratá-las. Na opção denominada deleção completa, todas as falhas são retiradas das análises de dados, o que normalmente é válido, uma vez que diferentes regiões de DNA evoluem diferentemente. Na outra opção, denominada deleção par a par, são ignoradas apenas as falhas de cada par de seqüência. Essa opção é válida se o número de falhas envolvido é pequeno e se as falhas estão distribuídas de forma aproximadamente aleatória (NEI; KUMAR, 2000).

2.1.3.5 Árvores Filogenéticas

A reconstrução de árvores filogenéticas utilizando métodos estatísticos iniciou-se independentemente na taxonomia numérica, em avaliações de caracteres morfológicos (SOKAL; SNEATH, 1963) e na genética de populações, em avaliações de dados de frequência gênica

(CAVALLI-SFORZA; EDWARDS, 1964). Alguns desses métodos ainda são usados em análises filogenéticas, porém, muitos outros métodos têm sido desenvolvidos.

As relações filogenéticas dos genes ou organismos são apresentadas em forma de árvore, que podem ou não apresentar raiz, além de serem bifurcadas ou multifurcadas. O padrão de ramos das árvores é denominado topologia. Há um grande número de possíveis topologias com e sem raiz para um grande número de diferentes taxa. À medida que o número de taxa (m) aumenta, rapidamente aumenta o número de topologias possíveis. Por exemplo, para $m=10$ há 34.459.425 possíveis árvores bifurcadas. Embora muitas dessas topologias possam ser descartadas, por apresentarem relações evolutivas improváveis, é uma tarefa difícil selecionar a topologia verdadeira quando o valor de m é grande (AMORIM, 2002).

Teoricamente, uma sequência de DNA se divide em duas seqüências durante a especiação ou numa duplicação gênica. Contudo, quando uma sequência relativamente curta é considerada, alguns ramos interiores podem não apresentar substituições nucleotídicas, surgindo um nó multifurcado. Árvores desse tipo são denominadas multifurcadas. A maioria dos métodos constrói árvores bifurcadas, mas a árvore obtida pode ser reduzida a uma árvore multifurcada, pela eliminação de ramos com comprimento zero. Ainda, é possível que mesmo que a árvore verdadeira seja bifurcada, a árvore reconstruída seja multifurcada, devido a erros estatísticos. Na prática, é difícil distinguir os dois casos (AMORIM, 2002).

Árvores filogenéticas que representam a história evolutiva de um grupo de espécies ou populações são denominadas árvores de espécie e de população, respectivamente. Numa árvore de espécie o tempo de divergência entre duas espécies refere-se ao tempo em que tais espécies tornaram-se reprodutivamente isoladas. Árvores filogenéticas construídas a partir de genes dessas espécies (árvores de genes) não necessariamente são parecidas com a árvore da espécie, nem tampouco o padrão de ramos dessas árvores são necessariamente iguais (NEI, 1986; NEI, 1987).

A chance de que as árvores de genes e de espécies sejam diferentes é relativamente grande para espécies proximamente relacionadas. A obtenção de árvores confiáveis para espécies próximas pode ser obtida pela utilização de grande número de genes com evolução independente (SAITOU; NEI, 1986; PAMILO; NEI, 1988). Ainda, mesmo que o padrão de divisão gênica seja parecido com o da divisão das espécies, o padrão de uma árvore gênica reconstruída pode ser

diferente, se um pequeno número de nucleotídeos for examinado. Isso ocorre porque o número de substituições nucleotídicas ocorre de forma estocástica. Assim, para evitar esse tipo de erro é importante examinar um grande número de nucleotídeos.

Supõe-se, na teoria da inferência filogenética, que as seqüências de DNA estudadas são bastante longas (teoricamente infinitas) e que um grande número de nucleotídeos que representa uma amostra aleatória da grande seqüência foi amostrado. Enquanto essas suposições simplificam as análises estatísticas das seqüências de DNA, na prática o interesse é reconstruir a história evolutiva a partir de seqüências curtas.

Ainda, se o número esperado de substituições aumentar linearmente com o tempo, a árvore filogenética obtida com as substituições atuais, pode ser muito diferente da que se poderia esperar intuitivamente. Nesse caso, mesmo a topologia atual poderia ser diferente daquela obtida de seqüências longas. A árvore que pode ser construída utilizando seqüências infinitas ou o número de substituições esperado de cada ramo é denominada árvore esperada, enquanto a árvore obtida com o número de substituições atuais é denominada árvore realizada (NEI, 1987). Normalmente ambas são diferentes da árvore reconstruída a partir dos dados observados de seqüência. Os métodos de construção de árvore, de maneira geral, tentam reconstruir árvores realizadas. À medida que o número de nucleotídeos examinado (**n**) aumenta a árvore realizada aproxima-se da árvore esperada.

Há diversos métodos estatísticos que podem ser usados para reconstruir árvores filogenéticas a partir de dados moleculares. Os métodos mais comumente utilizados podem ser classificados em três grupos principais: (i) métodos de distância; (ii) métodos de parcimônia e (iii) métodos de verossimilhança.

É comum considerar a reconstrução de uma árvore filogenética como a inferência estatística da árvore filogenética verdadeira, que é desconhecida. Há dois processos envolvidos nessa inferência: a “estimativa” da topologia e a estimativa do comprimento dos ramos para uma dada topologia. Quando a topologia é conhecida, estimar o comprimento dos ramos é relativamente simples e há muitos métodos disponíveis, tais como mínimos quadrados e máxima verossimilhança. O problema é a reconstrução da topologia, pois normalmente há um grande número topologias possíveis, o que torna difícil encontrar a topologia verdadeira.

Na inferência filogenética, normalmente são empregados certos princípios de otimização, tais como máxima verossimilhança ou o princípio da evolução mínima, que produzem bons resultados se n (tamanho da seqüência) for grande. Quando n é pequeno e o número de seqüências é grande, esses princípios tendem a fornecer topologias incorretas.

Há diferentes abordagens nas considerações sobre a topologia. Alguns autores consideram a topologia como um parâmetro na estimação estatística e o método de construção da árvore como uma estatística (ou estimador) para estimar o parâmetro. Ainda, utilizam o conceito de inconsistência para argumentar sobre a inferioridade dos métodos de parcimônia em relação aos de máxima verossimilhança sob certas condições (FELSENSTEIN, 1978, 1988).

De forma análoga, é possível considerar se um determinado método de construção de árvore fornece a topologia correta quando o mesmo processo evolutivo é repetido infinitas vezes com um número finito de n . Se o método fornecer a topologia correta é considerado um “estimador não viesado”. Se essa definição for aplicada, todos os métodos de construção de árvores não são “estimadores não viesados”, uma vez que fornecem topologias incorretas (NEI et al., 1998).

Assim, até o momento, a metodologia de reconstrução filogenética é controversa, basicamente por três motivos. Primeiro, pesquisadores treinados como sistematas utilizando caracteres morfológicos tendem a suspeitar de métodos baseados em modelos matemáticos, por considerarem que as mudanças evolutivas de caracteres morfológicas são complexas demais para seguir uma única regra, preferindo métodos de parcimônia, que requerem o mínimo de suposições; pesquisadores treinados como geneticistas ou biólogos moleculares preferem utilizar estratégias analíticas, mas não confiam em modelos altamente sofisticados; pesquisadores treinados como matemáticos e estatísticos, tentam entender a construção de árvores filogenéticas mais como um problema matemático que como um problema prático. Uma vez que as estratégias adotadas pelos três grupos são diferentes, ocorrem controvérsias. Segundo, há pesquisadores interessados em evolução de curto prazo, dentro de espécies ou espécies próximas, enquanto outros estão interessados em evolução de longo prazo, dentro de reinos, por exemplo. Como as metodologias diferem, também ocorrem controvérsias. Terceiro, em análise filogenética a árvore verdadeira é quase sempre desconhecida e, assim, é difícil testar a acurácia das árvores obtidas

pelos diferentes métodos. Ainda, os diversos métodos para testar a acurácia das árvores são baseados em suposições simplificadas.

Assim, considerando os diversos aspectos, cada método estatístico tem pontos fortes e fracos e a escolha do método deve ser feita de acordo com o objetivo do estudo, visando minimizar os erros.

2.1.3.6 Filogenias Moleculares e Diversidade em Anacardiaceae

A primeira investigação molecular da família Anacardiaceae empregou seqüências do gene de cloroplasto *rbcL*, além de dados morfológicos e de anatomia da madeira para interpretar a filogenia da família (TERRAZAS, 1994). Alguns de seus resultados foram também observados em outros trabalhos que envolveram dados bioquímicos, anatômicos e morfológicos (GUNDERSEN, 1950; CRONQUIST, 1981; THORNE, 1992), além de dados obtidos por sequenciamento de DNA (GADEK et al., 1996; APG, 1998, 2003; SAVOLAINEN et al., 2000a, 2000b). No entanto, interpretando filogenias em níveis de ordem e acima, outros estudos chegaram a resultados diferentes (CHASE et al., 1993; GADEK et al., 1996; APG, 1998, 2003; BREMER et al., 1999; SAVOLAINEN et al., 2000a, 2000b). Num estudo mais recente, Pell (2004) interpretou filogenias em níveis de família, utilizando dados de sequenciamento de DNA, a partir dos genes de plastídeo *matK*, *rps16*, o íntron *trnL*, o éxon 3' e o espaçador intergênico entre *trnL* e *trnF*.

Análises filogenéticas moleculares também foram realizadas por Yi et al (2004) para o gênero *Rhus*, pertencente à família Anacardiaceae e distribuído na Ásia, Europa, Havaí e nas Américas do Norte e Central. No estudo, foram empregados o espaçador transcrito interno de DNA ribossomal nuclear (ITS), o gene de cloroplasto *ndhF* e as regiões *trnL-F* de cloroplasto.

Em relação à diversidade da espécie, Syed et al. (2005) empregaram marcadores do tipo SSAP e obtiveram altos níveis de polimorfismo para *Anacardium occidentale*. Archak et al. (2003) avaliaram 19 acessos de *A. occidentale*, empregando marcadores do tipo RAPD, ISSR e AFLP e encontraram alta variabilidade, principalmente para os marcadores AFLP. Dhanaraj et al (2002) e Mneney et al. (2001) também empregaram marcadores RAPD no estudo de acessos do Banco de Germoplasma Indiano e dentro e entre populações de *A. occidentale* e também detectaram variabilidade intraespecífica.

2.2 Material e Métodos

2.2.1 Material Vegetal

Foram utilizadas amostras de folhas adultas pertencentes ao gênero *Anacardium* e uma amostra pertencente à espécie *Mangifera indica*, utilizada como *outgroup*. As amostras foram obtidas de três formas: (i) em viagens de coleta, (ii) em Banco de Germoplasma e (iii) via colaboradores. A viagem de coleta foi feita no município de General Carneiro – MT, enquanto as amostras de banco de germoplasma são provenientes da coleção da Embrapa Agroindústria Tropical, localizada em Fortaleza – CE e as amostras enviadas por colaboradores foram obtidas em diferentes locais.

As amostras coletadas foram armazenadas em sacos plásticos vedados contendo o dessecante sílica gel, sendo posteriormente congeladas em nitrogênio líquido, liofilizadas e moídas em liquidificador. Após moagem, as amostras foram armazenadas a -20°C.

2.2.2 Isolamento e Amplificação de DNA

O DNA foi isolado a partir de 60mg de material seco utilizando a metodologia descrita por Hoisington *et al.* (1994), com poucas modificações. As amostras foram quantificadas em gel de agarose 1%, coradas com brometo de etídeo, por comparação ao padrão λ 500 (Invitrogen) em diferentes concentrações.

Foram utilizadas duas regiões comumente utilizadas para inferências filogenéticas. A região ribossomal nuclear composta pelos genes 18S, 5,8 S e 25 S, dois espaçadores internos (ITS1 e ITS2) e o espaçador intergênico IGS, formam a unidade que é processada para produzir RNAs maduros que formam os ribossomos citoplasmáticos dos organismos vivos. A região de

cloroplasto consiste de três partes: espaçador intergênico entre os éxons trnT-L 5', 35 pb de trnL 5'exon e trnL intron.

A região de cloroplasto trnL íntron e a sequência de RNA ribossomal ITS 1 (Internal Transcribed Spacer) foram amplificadas por PCR (Polymerase Chain Reaction) para diferentes grupos de amostras (Tabela com quem tem ITS e quem tem TRN). As reações de PCR foram feitas em um volume final de 100µl, contendo 1X tampão de PCR (Invitrogen), 200µM dNTPs (Fermentas), 1,5mM MgCl₂ (Invitrogen), 1 µM de cada primer, 0,625U de Taq polimerase *platinum* (Invitrogen) e 40ng de DNA genômico. Para a amplificação da região ITS 1 foram utilizados os primers 1 (5'TCCGTAGGTGAACCTGCGG3') e 2 (5'GCTACGTTCTTCATCGATGC3') (Figura 1a), enquanto para amplificação da região trnL íntron foram utilizados os primers c (5'CGAAATCGGTAGACGCTACG3') e d (5'GGGGATAGAGGGACTTGAAC3') (Figura 1b). O programa de amplificação para a região trnL íntron consistiu de uma desnaturação inicial a 97°C por 2 min, seguida de 30 ciclos de desnaturação a 94°C por 1min, anelamento a 48°C por 2min; extensão a 72°C por 2min e uma extensão final a 72°C por 16 minutos conforme descrito por Pell, 2004 para Anacardiaceae. Já para a amplificação da região ITS 1, o programa consistiu de uma desnaturação inicial a 97°C por 2 min, seguida de 35 ciclos de desnaturação a 94°C por 1min, anelamento a 58°C por 1min; extensão a 72°C por 2min e uma extensão final a 72°C.

Figura 2 – Mapa de *primers* para amplificação das regiões ITS (a) e trn (b).

2.2.3 Sequenciamento de DNA

As amostras de PCR foram avaliadas em gel de agarose 2% e purificadas utilizando o QIAGEN PCR Purification Kit, empregando o protocolo sugerido pelo fabricante. A quantificação dos produtos de amplificação foi realizada utilizando o marcador Low Mass DNA Ladder (Invitrogen). As amostras foram diluídas para 20 ng/μl e enviadas ao laboratório do Centro de Estudos do Genoma Humano para sequenciamento, sendo que cada amostra foi sequenciada em ambos sentidos, visando a obtenção de uma sequência mais robusta. As sequências foram obtidas usando a metodologia do DYEnamic ET Dye Terminator Kit (com Thermo Sequenase™ II DNA Polimerase) em um sequenciador MegaBace 1000 (GE Healthcare).

2.2.4 Análise das Sequências

Para cada região, as sequências obtidas com os primers direto e reverso foram alinhadas e editadas a fim de se obter uma sequência consenso, utilizando o programa SeqMan (DnaStar Inc.). As regiões ITS 1 e trnL intron foram analisadas em separado utilizando o programa MEGA 4. O alinhamento das sequências foi realizado utilizando o programa Clustal W incluído no

MEGA, utilizando parâmetros múltiplos gap opening de 15 e gap extension de 0,30 e parâmetros par a par gap opening de 10 e gap extension de 0,10 de acordo com Shneider, 2007. As estimativas das distâncias e respectivos erros padrões foram calculadas utilizando o método Tamura Nei, sendo que os desvios foram calculados através de 5000 bootstrap e método analítico. A árvore filogenética foi construída com o método Neighbor-Joining, com deleção par a par.

2.2.5 Blast das Sequências obtidas

Foi realizada a comparação das sequências de ambas regiões com o banco de dados através do Blastn. O Blast é uma ferramenta que permite comparar sequências de nucleotídeos ou proteínas com os bancos de dados e calcular a significância dos alinhamentos encontrados. Há diversos tipos de alinhamentos disponíveis, entre eles o nucleotide blast (Blastn), que compara uma sequência de DNA com as bases de dados de sequências de nucleotídeos. A significância de um alinhamento encontrado pelo Blast é dada pelo E-value, que é a probabilidade de outro alinhamento ter uma similaridade maior do que o alinhamento encontrado devido simplesmente ao acaso. Sendo assim, quanto mais o E-value se aproxima de zero, maior a significância do alinhamento.

2.3 Resultados e Discussão

2.3.1 Coletas

Uma vez que o objetivo do trabalho era avaliar tanto a variação entre espécies como dentro das espécies, o critério para as coletas foi amostrar o maior número possível de espécies e; quando disponível, amostrar mais de um indivíduo da mesma espécie. Além disso, tentou-se coletar amostras dos dois centros de diversidade do gênero: Planalto Brasileiro e Amazônia (MITCHELL; MORI, 1987).

Basicamente, as amostras da região de diversidade primária (Amazônia) foram coletadas por colaboradores, enquanto as amostras da região de diversidade secundária (Planalto) foram realizadas no município de General Carneiro – MT. Além disso, foram coletadas amostras do banco de germoplasma da Embrapa, que continha acessos coletados de diversas regiões, principalmente dos dois centros de diversidade. Assim, algumas amostras foram obtidas no próprio local de ocorrência, enquanto outras de um único local, representativo de muitos locais de ocorrência (Tabela 1).

Tabela 1 – Diferentes coletas, regiões amostradas, número de amostras e espécies/tipos de *Anacardium*.

Coletas	Regiões amostradas	Número de Amostras	Espécies e Tipos
General Carneiro – MT	MT	38	2
Banco de Pacajus – CE	Diversas	100	5
Colaboradores	Diversas	75	7

Na coleta no município de General Carneiro – MT (janeiro de 2007) foram obtidas 38 amostras (Tabela 1), sendo que a maioria delas foi identificada como *A. humile* e algumas como *A. occidentale* (Tabela 2), de acordo com os critérios de Mitchell e Mori (1987).

Tabela 2 - Regiões de coleta do município de General Carneiro e respectivas espécies e número de amostras coletadas

Região	Espécie	Número de amostras
Assentamento A	<i>A. occidentale</i>	1
Assentamento A	<i>A. humile</i>	4
Assentamento B	<i>A. occidentale</i>	3
Assentamento B	<i>A. humile</i>	1
Fazenda Caracilho	<i>A. humile</i>	21
Fazenda Jarvis	<i>A. occidentale</i>	2
Fazenda Jarvis	<i>A. humile</i>	6

Uma vez que todas as amostras coletadas em General Carneiro foram georreferenciadas por GPS, as regiões de coleta e a localização das amostras podem ser vistas na Figura 3

Figura 3- Mapa das regiões de coleta do município de General Carneiro. No detalhe, os locais de coleta

No banco de germoplasma de Pacajus-CE foram coletadas 100 amostras (novembro de 2007), representando três espécies identificadas, um gênero identificado (*Anacardium* sp), e amostras não identificadas (“Desconhecida”), totalizando cinco tipos diferentes de amostras (Tabela 1). As espécies coletadas e os locais de coleta identificados no banco de germoplasma podem ser vistos na Tabela 3.

Tabela 3. Local de origem, estado, país, espécie e número de amostras referentes à coleta no Banco de Germoplasma.

Local	Estado	País	Espécie	Número de amostras
AltoParaíso	GO	Brasil	<i>Anacardium humile</i>	3
Cavalcante	GO	Brasil	<i>Anacardium humile</i>	4
Jaciara	GO	Brasil	<i>Anacardium humile</i>	3
Posse	GO	Brasil	<i>Anacardium humile</i>	2
SãoJoãodaAliança	GO	Brasil	<i>Anacardium humile</i>	4
BarradoGarça	GO	Brasil	<i>Anacardium occidentale</i>	1
Colider	MT	Brasil	<i>Anacardium occidentale</i>	1
Desconhecido	Desconhecido	Brasil	<i>Anacardium occidentale</i>	1
Itaituba	PA	Brasil	<i>Anacardium occidentale</i>	3
SantaTerezinha	MT	Brasil	<i>Anacardium occidentale</i>	1
SimonRodrigues-ElTigre	Anzoátegui	Venezuela	<i>Anacardium occidentale</i>	2
Valinhos	SP	Brasil	<i>Anacardium occidentale</i>	2
Ullal	Desconhecido	Índia	<i>Anacardium occidentale</i>	2
ÁguaBoa	MT	Brasil	<i>Anacardium othonianum</i>	1
AltoParaíso	GO	Brasil	<i>Anacardium othonianum</i>	1
Arraias	GO	Brasil	<i>Anacardium othonianum</i>	4
CamposBelos	GO	Brasil	<i>Anacardium othonianum</i>	3
Cavalcante	GO	Brasil	<i>Anacardium othonianum</i>	3
Galheiros	GO	Brasil	<i>Anacardium othonianum</i>	2
MonteAlegre	GO	Brasil	<i>Anacardium othonianum</i>	3
SãoDomingos	GO	Brasil	<i>Anacardium othonianum</i>	3
Tucuruí	PA	Brasil	<i>Anacardium othonianum</i>	3
Cristalina	GO	Brasil	<i>Anacardium sp</i>	3
CristinoCastro	PI	Brasil	<i>Anacardium sp</i>	3
EliseuMartins	PI	Brasil	<i>Anacardium sp</i>	3
EstTaianokm20	RR	Brasil	<i>Anacardium sp</i>	3
Desconhecido	Desconhecido	Brasil	<i>Anacardium sp</i>	5
Desconhecido	Desconhecido	Brasil	Desconhecida	23
Desconhecido	GO	Brasil	Desconhecida	4
Desconhecido	Desconhecido	Desconhecido	Desconhecida	1
EstradaNormandiakm25	RR	Brasil	Desconhecida	1
Igapora	BA	Brasil	Desconhecida	2

A distribuição das amostras de acordo com os locais de coleta identificados no banco de germoplasma pode ser vista na Figura 4. Na Figura 5 está a distribuição das amostras na América do Sul.

Figura 4 - Distribuição das amostras coletadas no banco de germoplasma.

Figura 5 - Distribuição das amostras coletadas no banco de germoplasma na América do Sul.

As coletas enviadas por colaboradores totalizaram 75 amostras (coletadas em diferentes datas entre janeiro de 2007 a novembro de 2007), representando cinco espécies identificadas, um

gênero identificado e amostras não identificadas, totalizando sete tipos diferentes de amostras (Tabela 1). As espécies coletadas e os locais de coleta podem ser vistos na Tabela 4.

Tabela 4 – Local de coleta, estado, espécie e número de amostras coletadas por colaboradores.

Local	Estado	Espécie	Número de Amostras
Chapada dos Guimarães	MT	<i>Anacardium corymbosum</i>	5
Bragança	PA	<i>Anacardium giganteum</i>	2
Desconhecido	Desconhecido	<i>Anacardium giganteum</i>	1
Tucuruí	PA	<i>Anacardium giganteum</i>	6
Caldas	MG	<i>Anacardium nanum</i>	5
Desconhecido	RR	<i>Anacardium occidentale</i>	1
Goiânia	GO	<i>Anacardium othonianum</i>	6
Bragança	PA	<i>Anacardium</i> sp	2
Bragança	PA	Desconhecida	10
Desconhecido	Desconhecido	Desconhecida	24
Desconhecido	PA	Desconhecida	1
Desconhecido	RR	Desconhecida	5
Desconhecido	Desconhecido	Desconhecida	6

A distribuição das amostras de acordo com o local de coleta das amostras pode ser vista na Figura 6.

Figura 6 – Distribuição das amostras coletadas por colaboradores.

Assim, considerando todas as amostras coletadas, foi possível amostrar diversos locais dos centros de diversidade primário e secundário (Figura 7) (MITCHELL; MORI, 1987).

Figura 7 – Distribuição de todas as amostras coletadas considerando os centros de diversidade primário e secundário.

O centro de diversidade primário (Amazônia) abrange a Bolívia, o Brasil, a Colômbia, o Equador, a Guiana, a Guiana Francesa, o Peru, o Suriname e a Venezuela. No Brasil, abrange áreas do Planalto das Guianas (área em cinza), da Planície Amazônica e do Planalto Brasileiro (Estados de Amazonas, Pará, Maranhão, Rondônia, Mato Grosso e Tocantins). O centro de diversidade secundário (Planalto Brasileiro) está delimitado pela linha cinza (Figura 7).

Os totais de espécies e de amostras coletadas por espécie, bem como de Estados e de locais por espécie podem ser vistos na Tabela 5.

Tabela 5 – Total de espécies, número de estados, número de locais e número de amostras coletadas.

Espécies	Número de Estados	Número de Locais	Número de amostras
<i>A. corymbosum</i>	1	1	5
<i>A. giganteum</i>	1	2	11
<i>A. humile</i>	2	6	48
<i>A. nanum</i>	1	1	5
<i>A. occidentale</i>	6	8	20
<i>A. othonianum</i>	3	10	29
<i>Anacardium</i> sp	4	5	19
Desconhecida	4	3	76

2.3.2 Seqüenciamento

Embora tenha sido possível coletar diversas espécies e indivíduos de cada espécie (Tabela 5), representativos dos centros de diversidade de *Anacardium* (Figura 7), poucas amostras efetivamente forneceram DNA em quantidade e qualidade suficientemente para o seqüenciamento.

De todas as amostras coletadas, foram efetivamente enviadas para seqüenciamento aquelas em que foi possível obter DNA de qualidade suficiente para amplificação via PCR (Tabela 4). Dessa forma, nem todas as amostras foram representadas pelas duas regiões escolhidas para a análise filogenética (Tabela 5).

Em março de 2007, três amostras de folhas adultas preservadas em sílica gel provenientes da coleta de General Carneiro foram avaliadas quanto à qualidade e quantidade de DNA. Nessas avaliações foram considerados dois protocolos diferentes para isolamento do DNA (HOISINGTON et al., 1994; BONATO et al., 2002), um deles para plantas que contêm grande quantidade de compostos secundários (BONATO et al., 2002). A partir do DNA obtido, foram testadas diferentes condições de PCR (BILLOTTE, 1999; CREGAN, 1999; PELL, 2004).

Para o teste de seqüenciamento foram escolhidas as regiões do íntron trnL e o espaçador intergênico trnL-trnF do cloroplasto (Figura 8A). Os resultados da amplificação, purificação da amplificação e alinhamento também podem ser vistos na mesma figura.

Os dois protocolos de isolamento forneceram DNA de boa quantidade e em boa quantidade, sendo que o protocolo de Hoisington et al. (1994) forneceu uma quantidade de DNA

ligeiramente superior ao de Bonato et al. (2002) (Figura 8B). Quanto às amplificações, todas funcionaram nas diferentes condições de PCR. Assim, foram enviadas para seqüenciamento as amostras dos protocolos de Hoisington et al (1994) e de Pell (2004), este último porque as condições de PCR haviam sido previamente empregadas em Anacardiaceae (PELL, 2004) (Figura 8C).

Figura 8 – Resultados do teste preliminar. A: região trn; B: amplificação do DNA; C: purificação do PCR; D: alinhamento das três seqüências da região trnL. E: alinhamento das três seqüências da região trnL-trnF

O seqüenciamento apresentou picos bem definidos, mostrando que a metodologia empregada no isolamento, amplificação e purificação do DNA foi adequada. O alinhamento das seqüências mostrou que na região intergênica trnL (Figura 8D) houve uma inserção para uma das amostras (ou deleção para duas amostras), enquanto na região do íntron trnL-trnF houve uma substituição do tipo transverso em um das amostras (Figura 8E).

Das três seqüências empregadas nos testes, aquela que apresentou as duas diferenças foi a de uma amostra identificada como *A. occidentale*, distante cerca de 5,3 km das outras duas amostras indetificadas como *A. humile* (Figura 9), o que pareceu indicar (de maneira preliminar), variabilidade nessas regiões amplificadas para essas duas espécies.

Figura 9 – locais de coleta e distância das amostras. No local FazCaracilho, há duas amostras sobrepostas.

Contudo, os isolamentos e as amplificações de DNA posteriores a esse teste não forneceram resultados tão bons. Assim, diversas modificações foram empregadas, desde testes de isolamento a partir de amostras sob diferentes formas de armazenamento (armazenadas em sílica de General Carneiro e Pacajus, liofilizadas de Pacajus e folhas frescas recém coletadas de *Anacardium* sp em Campinas-SP) até modificações nos protocolos de isolamento e amplificação, bem como a utilização de novos protocolos e de kit de extração.

A mudança mais significativa foi a modificação do protocolo de Hoisington et al. (1994) pela introdução de proteinase K numa das etapas iniciais, o que permitiu obter DNA em boa quantidade. No entanto, a qualidade das amplificações continuou a oscilar. As condições que forneceram os melhores resultados foram as descritas nesse trabalho.

A partir de todas as amostras avaliadas, foram obtidas 57 seqüências, 21 para a região nuclear ITS1 (Figura 2A) e 36 para a região trnL de cloroplasto (Figura 2B), as quais foram submetidas à busca por seqüências homólogas no banco de dados NCBI utilizando a ferramenta Blastn (Tabelas 6 e 7, respectivamente).

Tabela 6 – Resultado obtido através do Blastn para as sequências de ITS1. São mostrados o nome da sequência, a espécie à qual a sequência pertence, o número dos três acessos do GenBank com maior homologia à sequência e suas respectivas espécies e E-values.

Sequência	Espécie	Homologias	Espécie no GenBank	E-value*
Asp-ITS-RR-ESItaiano-C2014	<i>Anacardium sp</i>	AB071690.1	<i>Anacardium occidentale</i>	4,00E-86
		EU436905.1	<i>Oxalis ambigua</i>	5,00E-10
		EU436904.1	<i>Oxalis luteola</i>	5,00E-10
Asp-ITS-GO-Cristalina-C2195	<i>Anacardium sp</i>	AB071690.1	<i>Anacardium occidentale</i>	1,00E-96
		AY641512.1	<i>Schinus molle</i>	6,00E-49
		AY531202.1	<i>Schinus areira</i>	6,00E-49
Ahum-ITS-MT-General-Ass01636	<i>A. humile</i>	AB071690.1	<i>Anacardium occidentale</i>	2,00E-108
		AY641512.1	<i>Schinus molle</i>	3,00E-47
		AY531202.1	<i>Schinus areira</i>	3,00E-47
Desc-ITS-1429	Desconhecida	DQ787387.1	<i>Byrsonima vacciniifolia</i>	9,00E-92
		DQ787385.1	<i>Byrsonima gardneriana</i>	4,00E-90
		DQ787393.1	<i>Byrsonima crassifolia</i>	2,00E-68
Aocc-ITS-BAG170	<i>A. occidentale</i>	AB071690.1	<i>Anacardium occidentale</i>	1,00E-117
		AY641488.1	<i>Rhus michauxii</i>	4,00E-52
		EF193078.1	<i>Pistacia atlantica</i>	2,00E-49
Ananum-ITS-CaldasG9-3	<i>A. nanum</i>	AB071690.1	<i>Anacardium occidentale</i>	5,00E-110
		AY641512.1	<i>Schinus molle</i>	7,00E-49
		AY531202.1	<i>Schinus areira</i>	7,00E-49
Aoth-ITS-GO-Goiânia-UFG003	<i>A. othonianum</i>	AB071690.1	<i>Anacardium occidentale</i>	2,00E-119
		AY641488.1	<i>Rhus michauxii</i>	2,00E-54
		EF193078.1	<i>Pistacia atlantica</i>	3,00E-53
Aoth-ITS-GO-SãoDomingos-C2117	<i>A. othonianum</i>	AB071690.1	<i>Anacardium occidentale</i>	1,00E-91
		EU436905.1	<i>Oxalis ambigua</i>	5,00E-10
		EU436904.1	<i>Oxalis luteola</i>	5,00E-10
Ahum-ITS-GO-Jaciara-C2229	<i>A. humile</i>	AB071690.1	<i>Anacardium occidentale</i>	2,00E-108
		AY641512.1	<i>Schinus molle</i>	7,00E-49
		AY531202.1	<i>Schinus areira</i>	7,00E-49
Ahum-ITS-MT-General-fazCar669	<i>A. humile</i>	AB071690.1	<i>Anacardium occidentale</i>	2,00E-124
		AY641488.1	<i>Rhus michauxii</i>	5,00E-61
		EF193078.1	<i>Pistacia atlantica</i>	7,00E-60
Desc-ITS-1625	<i>A. giganteum</i>	AB071690.1	<i>Anacardium occidentale</i>	2,00E-99
		AY531202.1	<i>Schinus areira</i>	4,00E-46
		AY864894.1	<i>Schinus terebinthifolius</i>	4,00E-46
Aocc-ITS-C6001	<i>A. occidentale</i>	AB071690.1	<i>Anacardium occidentale</i>	4,00E-111
		EF193088.1	<i>Pistacia mexicana</i>	4,00E-41
		EF193087.1	<i>Pistacia mexicana</i>	4,00E-41
Ananum-ITS-CaldasG9-4	<i>A. nanum</i>	AB071690.1	<i>Anacardium occidentale</i>	2,00E-119
		AY641488.1	<i>Rhus michauxii</i>	4,00E-52

Aoth-ITS-GO-Goiânia-UFG002	<i>A. othonianum</i>	EF193078.1	<i>Pistacia atlantica</i>	5,00E-51
		AB071690.1	<i>Anacardium occidentale</i>	5,00E-110
		AY641512.1	<i>Schinus molle</i>	9,00E-48
		AY531202.1	<i>Schinus areira</i>	9,00E-48
Aocc-ITS-MT-Colider-C2192	<i>A. occidentale</i>	AB071690.1	<i>Anacardium occidentale</i>	3,00E-102
		AY531202.1	<i>Schinus areira</i>	7,00E-44
		AY864894.1	<i>Schinus terebinthifolius</i>	7,00E-44
Aoth-ITS-PA-Tucuruí-C2279	<i>A. othonianum</i>	AB071690.1	<i>Anacardium occidentale</i>	7,00E-79
		AY641512.1	<i>Schinus molle</i>	2,00E-49
		AY531202.1	<i>Schinus areira</i>	2,00E-49
		AB071690.1	<i>Anacardium occidentale</i>	4,00E-117
Ahum-ITS-MT-General-fazCar677	<i>A. humile</i>	AY641488.1	<i>Rhus michauxii</i>	4,00E-52
		EF193078.1	<i>Pistacia atlantica</i>	5,00E-51
		AB071690.1	<i>Anacardium occidentale</i>	3,00E-76
Aocc-ITS-BAG84	<i>A. occidentale</i>	AY531202.1	<i>Schinus areira</i>	5,00E-44
		AY864894.1	<i>Schinus terebinthifolius</i>	5,00E-44
		AB071690.1	<i>Anacardium occidentale</i>	1,00E-120
Aocc-ITS-C6015	<i>A. occidentale</i>	EF193088.1	<i>Pistacia mexicana</i>	6,00E-45
		EF193087.1	<i>Pistacia mexicana</i>	6,00E-45
		AB071690.1	<i>Anacardium occidentale</i>	1,00E-116
Aoth-ITS-GO-Goiânia-UFG004	<i>A. othonianum</i>	AY864894.1	<i>Schinus terebinthifolius</i>	2,00E-45
		AY641512.1	<i>Schinus molle</i>	6,00E-45
		AJ890466.1	<i>Mangifera indica</i>	3,00E-159
Mangifera-SP-Campinas-ITS	<i>Mangifera indica</i>	AJ890469.1	<i>Mangifera zeylanica</i>	6,00E-156
		AJ890467.1	<i>Mangifera indica</i>	3,00E-154

A procura por seqüências homólogas da região ITS1 no NCBI encontrou seqüências homólogas de *A. occidentale* (AB071690.1) para todas as amostras (Tabela 6), exceto para Desc-ITS-1429, que foi retirada das análises, e para Mangifera-ITS-SP-Campinas, cuja busca encontrou seqüências homólogas de *Mangifera* (AJ890466.1), como esperado.

De todas as amostras seqüenciadas, apenas duas delas não alinharam com *A. occidentale*: a seqüência de manga (*Mangifera indica*) escolhida como outgroup (Tabelas 6 e 7) e a amostra 1429 identificada por colaboradores como desconhecida (Tabela 6). Uma vez que a amostra 1429 não representava nenhuma espécie de caju, foi retirada das análises.

Tabela 7 – Resultado obtido através do Blast de nucleotídeos para as seqüências da região tRNA-Leu (trn-L). São mostrados o nome da seqüência, a espécie à qual a seqüência pertence, o número dos três acessos do GenBank com maior homologia à seqüência e suas respectivas espécies e E-values.

Seqüência	Espécie	Homologias	Espécie no GenBank	E-value*
Ahum-TRN-MT-General-FazCar-1	<i>A. humile</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640456.1	<i>Rhus schiedeana</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
Ahum-TRN-MT-General-FazCar-3	<i>A. humile</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640456.1	<i>Rhus schiedeana</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
Asp-TRN-PI-CrisCastro-C2009	<i>Anacardium sp.</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
Asp-TRN-PI-CrisCastro-C2014	<i>Anacardium sp.</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
Aocc-TRN-MT-General-26	<i>A. occidentale</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
Ahum-TRN-GO-Cavalcante-C2079	<i>A. humile</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		EF089151.1	<i>Faguetia falcata</i>	1,00E-178
		EF682862.1	<i>Searsia lancea</i>	3,00E-174
Aoth-TRN-GO-SãoDomingos-C2117	<i>A. othonianum</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		EF089151.1	<i>Faguetia falcate</i>	9,00E-164
		EF682862.1	<i>Searsia lancea</i>	2,00E-160
Aoth-TRN-GO-Galheiros-C2147	<i>A. othonianum</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		EF089151.1	<i>Faguetia falcata</i>	1,00E-162
		EF682862.1	<i>Searsia lancea</i>	9,00E-159
Ahum-TRN-GO-AltoParaíso-C2163	<i>A. humile</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0

		AY640451.1	<i>Rhus integrifolia</i>	0
		DQ131556.1	<i>Anacardium occidentale</i>	0
Aocc-TRN-MT-Colider-C2192	<i>A. occidentale</i>	AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
		DQ131556.1	<i>Anacardium occidentale</i>	0
Asp-TRN-GO-Cristalina-C2195	<i>Anacardium sp.</i>	AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
		DQ131556.1	<i>Anacardium occidentale</i>	0
Asp-TRN-PI-EliseuMartinsC2213	<i>Anacardium sp.</i>	EF089151.1	<i>Faguetia falcate</i>	4,00E-177
		AY640452.1	<i>Rhus kearneyi</i>	2,00E-175
		DQ131556.1	<i>Anacardium occidentale</i>	0
Ahumile-TRN-GO-Jaciara-C2229	<i>A. humile</i>	AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
		DQ131556.1	<i>Anacardium occidentale</i>	0
Aoth-TRN-PA-Tucuruí-C2279	<i>A. othonianum</i>	AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
		DQ131556.1	<i>Anacardium occidentale</i>	0
Aoth-TRN-MT-ÁguaBoa-C2281	<i>A. othonianum</i>	AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
		DQ131556.1	<i>Anacardium occidentale</i>	0
Ahum-TRN-MT-General-Ass01636	<i>A. humile</i>	AY640456.1	<i>Rhus schiedeana</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		DQ131556.1	<i>Anacardium occidentale</i>	0
Ahum-TRN-MT-General-fazCar664	<i>A. humile</i>	AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
		DQ131556.1	<i>Anacardium occidentale</i>	0
Ahum-TRN-MT-General-fazCar669	<i>A. humile</i>	AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
		DQ131556.1	<i>Anacardium occidentale</i>	0
Ahum-TRN-MT-General-fazCar677	<i>A. humile</i>	AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
		DQ131556.1	<i>Anacardium occidentale</i>	0
Ahum-MT-General-fazJarvis680	<i>A. humile</i>	AY640456.1	<i>Rhus schiedeana</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		DQ131556.1	<i>Anacardium occidentale</i>	0
Desc-TRN-1495	desconhecida	AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
		DQ131556.1	<i>Anacardium occidentale</i>	0
Desc-TRN-1635	<i>A. giganteum</i>	AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
		DQ131556.1	<i>Anacardium occidentale</i>	0
Aocc-TRN-BAG84	<i>A. occidentale</i>	AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0

Aocc-TRN-BAG166	<i>A. occidentale</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
Aocc-TRN-BAG167	<i>A. occidentale</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
Aocc-TRN-BAG169	<i>A. occidentale</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
Aocc-TRN-BAG170	<i>A. occidentale</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
Desc-TRN-RR-EstNormandia-C4-040	desconhecida	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
Aocc-TRN-C6-015	<i>A. occidentale</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	9,00E-174
		AY640451.1	<i>Rhus integrifolia</i>	9,00E-174
Aocc-TRN-C6-135	<i>A. occidentale</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
Ananum-TRN-CaldasG9-3	<i>A. nanum</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640456.1	<i>Rhus schiedeana</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
Ananum-TRN-CaldasG9-4	<i>A. nanum</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640456.1	<i>Rhus schiedeana</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
India-TRN-Ullal	<i>A. occidentale</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
Aoth-TRN-GO-Goiânia-UFG002	<i>A. othonianum</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
Aoth-TRN-GO-Goiânia-UFG003	<i>A. othonianum</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
Aoth-TRN-GO-Goiânia-UFG004	<i>A. othonianum</i>	DQ131556.1	<i>Anacardium occidentale</i>	0
		AY640452.1	<i>Rhus kearneyi</i>	0
		AY640451.1	<i>Rhus integrifolia</i>	0
Manga-TRN-SP-Campinas	<i>Mangifera indica</i>	EF089151.1	<i>Faguetia falcate</i>	2,00E-165
		EF089152.1	<i>Heeria argentea</i>	1,00E-162
		AY640470.1	<i>Searsia undulata</i>	1,00E-162

Todas as seqüências TRN (Tabela 7) encontraram homologia com seqüências de *A. occidentale* (DQ131556.1), exceto Mangifera-TRN-SP-Campinas, que encontrou seqüências homólogas de outras espécies que não *Mangifera*. Nesse caso, provavelmente não foram encontradas as seqüências homólogas esperadas porque não haviam seqüências de *Mangifera* depositadas no banco para essa região. Como a mesma planta (Mangifera-SP-Campinas) foi utilizada para amplificar tanto a região ITS1 quanto a região trnL, a seqüência não foi retirada das análises.

Dessa forma, foram avaliadas, 56 seqüências, 20 para a região nuclear ITS1 e 36 para a região de cloroplasto trnL. A distribuição das seqüências pelos centros de diversidade pode ser vista na Figura 10. Os totais de espécies e de amostras coletadas por espécie, bem como de Estados e de locais por espécie para as duas regiões seqüenciadas podem ser vistos na Tabela 8.

Figura 10 – Distribuição das seqüências avaliada nos centros de diversidade

Tabela 8 - Número de sequências avaliadas para ambas regiões por estado e locais.

Espécies	ITS1			TrnL		
	Número de Estados	Número de locais	Número de amostras	Número de Estados	Número de locais	Número de amostras
<i>A. giganteum</i>	1	1	1	1	5	1
<i>A. humile</i>	2	2	4	2	3	9
<i>A. nanum</i>	1	1	2	1	1	2
<i>A. occidentale</i>	1	1	3	2	3	6
<i>A. othonianum</i>	2	3	5	3	5	7
<i>Anacardium</i> sp	2	2	2	3	4	5
Desconhecida	...	...	3	1	1	6

As 20 seqüências da região ITS1 e sua distribuição pelos centros de diversidade podem ser vistas na Tabela 9 e na Figura 11, respectivamente.

Tabela 9 – Código das sequências com identificação de cada espécie, estado, local e região.

Identificação da seqüência	Espécie	Estado	Local	Região	Amostra
Mangifera-ITS-SP-Campinas	<i>Mangifera</i> sp	SP	Campinas	Unicamp	Manga
Aocc-ITS-C6001	<i>A. occidentale</i>	desc.	desconhecido	desconhecida	C6-001
Aocc-ITS-C6015	<i>A. occidentale</i>	desc.	desconhecido	desconhecida	C6-015
Aocc-ITS-MT-Colider-C2192	<i>A. occidentale</i>	MT	Colider	desconhecida	C2-192
Ananum-ITS-MG-Caldas-G93	<i>A. nanum</i>	MG	Caldas	desconhecida	G9-3
Ananum-ITS-MG-Caldas-G94	<i>A. nanum</i>	MG	Caldas	desconhecida	G9-4
Ahum-ITS-MT-General-fazCar677	<i>A. humile</i>	MT	General Carneiro	Fazenda Caracilho	677
Ahum-ITS-MT-General-fazCar669	<i>A. humile</i>	MT	General Carneiro	Fazenda Caracilho	669
Ahum-ITS-MT-General-Ass01636	<i>A. humile</i>	MT	General Carneiro	Assentamento 01	636
Ahum-ITS-GO-Jaciara-C2229	<i>A. humile</i>	GO	Jaciara	desconhecida	C2-229
Aoth-ITS-PA-Tucuruí-C2279	<i>A. othonianum</i>	PA	Tucuruí	desconhecida	C2-279
Aoth-ITS-GO-Goiania-UFG002	<i>A. othonianum</i>	GO	Goiânia	desconhecida	UFG-002
Aoth-ITS-GO-Goiania-UFG003	<i>A. othonianum</i>	GO	Goiânia	desconhecida	UFG-003
Aoth-ITS-GO-Goiania-UFG004	<i>A. othonianum</i>	GO	Goiânia	desconhecida	UFG-004
Aoth-ITS-GO-SaoDomingos-C2117	<i>A. othonianum</i>	GO	São Domingos	desconhecida	C2-117
Asp-ITS-RR-EstTaiano-C2014	<i>Anacardium</i> sp	RR	Estrada Taiano	desconhecida	C2-014
Asp-ITS-GO-Cristalina-C2195	<i>Anacardium</i> sp	GO	Cristalina	desconhecida	C2-195
Desc-ITS-1625	<i>A. giganteum</i>	desc.	desconhecido	desconhecida	1665
Desc-ITS-BAG170	Desconhecida	desc.	desconhecido	desconhecida	BAG-170
Desc-ITS-BAG084	Desconhecida	desc.	desconhecido	desconhecida	BAG-084

Figura 11 – distribuição geográfica das seqüências ITS1 avaliadas

As 36 seqüências da região *trnL* e sua distribuição pelos centros de diversidade podem ser vistas na Tabela 10 e na Figura 12, respectivamente.

Tabela 10 – Código das amostras para cada espécie, estado, local e região

Identificação da seqüência	Espécie	Estado	Local	Região	Amostra
Mangífera-TRN-SP-Campinas	<i>Mangífera</i> sp	SP	Campinas	Unicamp	Manga
Aocc-TRN-C6015	<i>A. occidentale</i>	desc.	Desconhecido	Desconhecida	C6-015
Aocc-TRN-C6135	<i>A. occidentale</i>	desc.	Desconhecido	Desconhecida	C6-135
Aocc-TRN-MT-General-26	<i>A. occidentale</i>	MT	General Carneiro	Assentamento B	26
Aocc-TRN-MT-Colider-C2192	<i>A. occidentale</i>	MT	Colider	Desconhecida	C2-192
Aocc-TRN-BAG166	<i>A. occidentale</i>	desc.	desconhecido	Desconhecida	BAG-166
India-TRN-Ullal	Desconhecida	Ullal	Ullal	Desconhecida	Índia
Ahum-TRN-MT-General-Ass01636	<i>A. humile</i>	MT	General Carneiro	Assentamento A	636
Ahum-TRN-MT-General-FazCar-1	<i>A. humile</i>	MT	General Carneiro	Fazenda Caracilho	1
Ahum-TRN-MT-General-FazCar-3	<i>A. humile</i>	MT	General Carneiro	Fazenda Caracilho	3
Ahum-TRN-MT-General-fazCar664	<i>A. humile</i>	MT	General Carneiro	Fazenda Caracilho	664

Ahum-TRN-MT-General-fazCar669	<i>A. humile</i>	MT	General Carneiro	Fazenda Caracilho	669
Ahum-TRN-MT-General-fazCar677	<i>A. humile</i>	MT	General Carneiro	Fazenda Caracilho	677
Ahum-TRN-MT-General-fazJarvis680	<i>A. humile</i>	MT	General Carneiro	Fazenda Jarvis	680
Ahum-TRN-GO-Cavalcante-C2079	<i>A. humile</i>	GO	Cavalcante	desconhecida	C2-079
Ahum-TRN-GO-AltoParaíso-C2163	<i>A. humile</i>	GO	Alto Paraíso	desconhecida	C2-163
Aoth-TRN-GO-Galheiros-C2147	<i>A. othonianum</i>	GO	Galheiros	desconhecida	C2-147
Aoth-TRN-GO-SaoDomingos-C2117	<i>A. othonianum</i>	GO	São Domingos	desconhecida	C2-117
Aoth-TRN-PA-Tucuruí-C2279	<i>A. othonianum</i>	PA	Tucuruí	desconhecida	C2-279
Aoth-TRN-MT-AguaBoa-C2281	<i>A. othonianum</i>	MT	Água Boa	desconhecida	C2-281
Aoth-TRN-GO-Goiania-UFG002	<i>A. othonianum</i>	GO	Goiânia	desconhecida	UFG-002
Aoth-TRN-GO-Goiania-UFG003	<i>A. othonianum</i>	GO	Goiânia	desconhecida	UFG-003
Aoth-TRN-PA-Tucuruí-C2279	<i>A. othonianum</i>	PA	Tucuruí	desconhecida	C2-279
Ananum-TRN-MG-Caldas-G93	<i>A. nanum</i>	MG	Caldas	desconhecida	G9-3
Ananum-TRN-MG-Caldas-G94	<i>A. nanum</i>	MG	Caldas	desconhecida	G9-4
Asp-TRN-PI-CrisCastro-C2009	<i>Anacardium</i> sp	PI	Cristino Castro	desconhecida	C2-009
Asp-TRN-PI-CrisCastro-C2014	<i>Anacardium</i> sp	PI	Cristino Castro	desconhecida	C2-014
Asp-TRN-GO-Cristalina-C2195	<i>Anacardium</i> sp	GO	Cristalina	desconhecida	C2-195
Asp-TRN-PI-EliseuMartins-C2213	<i>Anacardium</i> sp	PI	Eliseu Martins	desconhecida	C2-213
Asp-TRN-1495	<i>Anacardium</i> sp	desc.	Desconhecido	desconhecida	1495
Desc-TRN-1635	<i>A. giganteum</i>	desc.	Desconhecido	desconhecida	1635
Desc-TRN-RR-EstNormandia-C4-040	Desconhecida	RR	Estrada Normandia	desconhecida	C4-040
Desc-TRN-BAG084	Desconhecida	desc.	Desconhecido	desconhecida	BAG-084
Desc-TRN-BAG167	Desconhecida	desc.	Desconhecido	desconhecida	BAG-167
Desc-TRN-BAG169	Desconhecida	desc.	Desconhecido	desconhecida	BAG-169
Desc-TRN-BAG170	Desconhecida	desc.	Desconhecido	desconhecida	BAG-170

Figura 12 – Distribuição geográfica nos centros de diversidade das seqüências trnL avaliadas

As 20 seqüências ITS1 foram alinhadas de maneira relativamente simples (Tabela 11), com apenas três seqüências apresentando regiões notáveis de inserções/deleções (indels), ao contrário do encontrado para muitos autores (GIELLY; TABERLET, 1994). A seqüência Asp-ITS-RR-EstTaiano-C2014 apresentou uma região pequena de indel no início do alinhamento, as seqüências Aoth-ITS-PA-Tucurui-C2279 e Desc-ITS-BAG084 apresentaram uma região indel grande cada, no meio do alinhamento e no final do alinhamento, respectivamente (Tabela 11, em negrito). O alinhamento das 20 seqüências ITS pode ser visto na Tabela 11.

Tabela 11 – Alinhamento das seqüências ITS1

Seqüências	Bases alinhadas de 1 a 80							
Mangifera-SP-Campinas-ITS	AGATAATGAA	AGGAGGCGTT	GCGTCCCGCA	CGCACGCACC	GTGTCCGGGG	GCGACGGGAG	CGAGCCCTCT	CGTTAACAAG
Aocc-ITS-C6001	T..C.....	AAC	..T.....	...-?.....	.A.CA.-...			.AAGT.AG.T
Aocc-ITS-C6015	T..C.....	AAC	..T.....	...-?.....	.A.CA.-...			.AAGT.AG.T
Aocc-ITS-MT-Colider-C2192	...C.....	AAC	..T.....	...-?.....	.A.CA.-...			.AAGT.G..T
Ananum-ITS-Caldas-G93	...C.....	AAC	..T.....	...-?.....	.A.CA.-...			.AAGT.G..T
Ananum-ITS-Caldas-G94	...C.....	AAC	..T.....	...-?.....	.A.CA.-...			.AAGT.G..T
Ahum-ITS-MT-General-fazCar677	...C.....	AAC	..T.....	...-?.....	.A.CA.-...			.AAGT.G..T
Ahum-ITS-MT-General-fazCar669	...C.....	AAC	..T.....	...-?.....	.A.CA.-...			.AAGT.G..T
Ahum-ITS-MT-General-AssA636	...C.....	AAC	..T.....	...-?.....	.A.CA.-...	.-.....		.AAGT.G..T
Ahum-ITS-GO-Jaciara-C2229	...C.....	AAC	..T.....	...-?.....	.A.CA.-...			.AAGT.G..T
Aoth-ITS-PA-Tucurui-C2279	...C.....	AAC	..T.....	...-?.....	.A.CA.-...			.AAGT.G..T
Aoth-ITS-GO-Goiania-UFG002	...C.....	AAC	..T.....	...-?.....	.A.CA.-...			.AAGT.G..T
Aoth-ITS-GO-Goiania-UFG003	...C.....	AAC	..T.....	...-?.....	.A.CA.-...			.AAGT.G..T
Aoth-ITS-GO-Goiania-UFG004	...C.....	AAC	..T.....	...-?.....	.A.CA.-...			.AAGT.G..T
Aoth-ITS-GO-SaoDomingos-C2117	...C.....	AAC	..T.....	...-?.....	.A.CA.-...			.AAGT.G..T
Asp-ITS-RR-EstTaiano-C2014	??????-..	AAC	..T.....	...-?.....	.AACA.-...	A.....	-	.AAGT.G..T
Asp-ITS-GO-Cristalina-C2195	...C.....	AAC	..T.....	...-?.....	.A.CA.-...			.AAGT.G..T
Desc-ITS-1625	...C.....	...G...AAC	..T.....	...-?.....	.A.CA.-...			..CGT.G..T
Desc-ITS-BAG170	T..C.....	AAC	..T.....	...-?.....	.AACA.-...	A.....		.AAGT.G..T
Desc-ITS-BAG084	-..C.....	AAC	..T.....	...-?.....	.A.CA.-...	.-.....		.AAGT.G..T

Tabela 11 - Alinhamento das seqüências ITS1

Seqüências	Bases alinhadas de 81 a 160							
Mangifera-SP-Campinas-ITS	TCCTTGGCGC	AATTCGCGCC	GGGGTTGGTT	AAGGCACCGC	GAGGGGCGAA	CGGGCACGCA	CACGCAGGGG	TGCACGCGCA
Aocc-ITS-C6001		G.C.....	C...	G.C..G....	.GC.....G.	GA..C.G.???	???????????	???-A-...?
Aocc-ITS-C6015		G.C.....	C...	G.C..G....	.GC.....G.	GA..C.G.???	???????????	???-A-...?
Aocc-ITS-MT-Colider-C2192		G.C.....	C...	G.C..G....	.GC.....G.	GA..A.G.???	???????????	???-A-...?
Ananum-ITS-Caldas-G93		G.C.....	C...	G.C..G....	.GC.....G.	GA..A.G.???	???????????	???-.....?
Ananum-ITS-Caldas-G94		G.C.....	C...	G.C..C....	.GC.....	GA..A.G.???	???????????	???-.....?
Ahum-ITS-MT-General-fazCar677		G.C.....	C...	G.C..C....	.GC.....	GA..A.G.???	???????????	???-.....?
Ahum-ITS-MT-General-fazCar669		G.C.....	C...	G.C..G....	.GC.....	GA..A.G.???	???????????	???-.....?
Ahum-ITS-MT-General-AssA636		G.C.....	C...	G.C..G....	.GC.....G.	GA..A.G.???	???????????	???-.....?
Ahum-ITS-GO-Jaciara-C2229		G.C.....	C...	G.C..G....	.GC.....G.	GA..A.G.???	???????????	???-.....?
Aoth-ITS-PA-Tucuruí-C2279		G.C.....	C...	G.C..G....	.GC.....??	???-?????	???????????	???????????
Aoth-ITS-GO-Goiania-UFG002		G.C.....	C...	G.C..G....	.GC.....	GA..A.G.???	???????????	???-.....?
Aoth-ITS-GO-Goiania-UFG003		G.C.....	C...	G.C..C....	.GC.....	GA..A.G.???	???????????	???-.....?
Aoth-ITS-GO-Goiania-UFG004		G.C.....	C...	G.C..C....	.GC.....	GA..A.G.???	???????????	???-.....?
Aoth-ITS-GO-SaoDomingos-C2117		G.C.....	C...	G.C..G....	.GC.....G.	GA..A.G.???	???????????	???-.....?
Asp-ITS-RR-EstTaiano-C2014		G.C.....	C...	G.C..G....	.GC.....T.	GA..A.G.???	???????????	???-.....?
Asp-ITS-GO-Cristalina-C2195		G.C.....	C...	G.C..G....	.GC.....G.	GA..A.G.???	???????????	???-.....?
Desc-ITS-1625		G.C.....	C...	G.C..G....	.GC.....G.	GA..C.G.???	???????????	???-A-...?
Desc-ITS-BAG170		G.C.....	C...	G.C..G....	.GC.....T.	GA..A.G.???	???????????	???-.....?
Desc-ITS-BAG084		G.C.....	C...	G.C..G....	.GC.....G.	GA..A.G????	???????????	???????????

Tabela 11 - Alinhamento das seqüências ITS1

Seqüências		Bases alinhadas de 181 a 218				
Mangifera-SP-Campinas-ITS	ACGAAAGCCC	AACGCGACGC	GAGCGAGGGC	GCGCAAGCAC	AAGGCCACG	CGTCCCCG
Aocc-ITS-C6001	??????????	?-...CG...	.GTG.G....	A...GG..G.	T.....G..	-.CT...C
Aocc-ITS-C6015	??????????	?-..A.CG...	.GTG.G....	A...GG..G.	T.....G..	-.CT...C
Aocc-ITS-MT-Colider-C2192	??????????	?-...AG...	.GTG.G....	A...GG..G.	G.....G..	-.CT...C
Ananum-ITS-Caldas-G93	??????????	?-..A..GA..	.GTG.G....	A...GG..G.	G.....G..	-.CT...C
Ananum-ITS-Caldas-G94	??????????	?-..A..GA..	.GTG.G....	A...GG..G.	G.....G..	-.CT...C
Ahum-ITS-MT-General-fazCar677	??????????	?-..A..GA..	.GTG.G....	A...GG..G.	G.....G..	-.CT...C
Ahum-ITS-MT-General-fazCar669	??????????	?-..A..GA..	.GTG.G....	A...GG..G.	G.....G..	-.CT...C
Ahum-ITS-MT-General-AssA636	??????????	?-..A..GA..	.GTG.G....	A...GG..G.	G.....G..	-.CT...C
Ahum-ITS-GO-Jaciara-C2229	??????????	?-..A..GA..	.GTG.G....	A...GG..G.	G.....G..	-.CT...C
Aoth-ITS-PA-Tucurui-C2279	??????????	??????????	??????????	??????????	??????????	?..CT...C
Aoth-ITS-GO-Goiania-UFG002	??????????	?-..A..GA..	.GTG.G....	A...GG..G.	G.....G..	-.CT...C
Aoth-ITS-GO-Goiania-UFG003	??????????	?-..A..GA..	.GTG.G....	A...GG..G.	G.....G..	-.CT...C
Aoth-ITS-GO-Goiania-UFG004	??????????	?-..A..GA..	.GTG.G....	A...GG..G.	G.....G..	-.CT...C
Aoth-ITS-GO-SaoDomingos-C2117	??????????	?-..A..GA..	.GTG.G....	A...GG..G.	G.....G..	-.CT...C
Asp-ITS-RR-EstTaiano-C2014	??????????	?-..A..GA..	.GTG.G....	AT..GG..G.	G.....G..	-.CT...C
Asp-ITS-GO-Cristalina-C2195	??????????	?-..A..GA..	.GTG.G....	A...GG..G.	G.....G..	-.CT...C
Desc-ITS-1625	??????????	?-...AGA..	AGTG.G....	A...GG..G.	G.....G..	-.CT...C
Desc-ITS-BAG170	??????????	?-..A..GA..	.GTG.G....	AT..GG..G.	G.....G..	-.CT...C
Desc-ITS-BAG084	??????????	??????????	??????????	??????????	??????????	??????-?

As 36 seqüências *trnL* também foram alinhadas de maneira simples (Tabela 12), sendo que ainda menos *indels* foram observados. Nenhum *indel* maior que três pares de base (pb) foi observado.

Tabela 12 – Alinhamento das seqüências trnL

Seqüências	Bases alinhadas de 1 a 80							
#Mangifera-TRN-SP-Campinas	GGGACTCTAT	CTTTATTCTC	GTCCGATTGA	TTAATTAGTT	CTTCAAAAGA	AAAACAAAAA	AGGATTCGGG	TCGTCATTTT
#Aocc-TRN-C6015			??	?-.....			-.....	..A.....
#Aocc-TRN-C6135			??	?-.....			-.....	..A.....
#Aocc-TRN-MT-General-26			??	?-.....			-.....	..A.....
#Aocc-TRN-MT-Colider-C2192			??	?-.....			-.....	..A.....
#Aocc-TRN-BAG166			??	?-.....			-.....	..A.....
#India-TRN-Ullal			??	?-.....			-.....	..A.....
#Ahum-TRN-MT-General-Ass01636			??	?-.....			-.....	..A.....
#Ahum-TRN-MT-General-FazCar-1			??	?-.....			-.....	..A.....
#Ahum-TRN-MT-General-FazCar-3			??	?-.....			-.....	..A.....
#Ahum-TRN-MT-General-fazCar664			??	?-.....			-.....	..A.....
#Ahum-TRN-MT-General-fazCar669			??	?-.....			-.....	..A.....
#Ahum-TRN-MT-General-fazCar677			??	?-.....			-.....	..A.....
#Ahum-MT-General-fazJarvis680			??	?-.....			-.....	..A.....
#Ahum-TRN-GO-Cavalcante-C2079			??	?-.....			-.....	..A.....
#Ahum-TRN-GO-AltoParaíso-C2163			??	?-.....			-.....	..A.....
#Aoth-TRN-GO-Galheiros-C2147			??	?-.....			-.....	..A.....
#Aoth-TRN-GO-SaoDomingos-C2117			??	?-.....			-.....	..A.....
#Aoth-TRN-PA-Tucurui-C2279			??	?-.....			-.....	..A.....
#Aoth-TRN-MT-AguaBoa-C2281			??	?-.....			-.....	..A.....
#Aoth-TRN-GO-Goiania-UFG002			??	?-.....			-.....	..A.....
#Aoth-TRN-GO-Goiania-UFG003			??	?-.....			-.....	..A.....
#Aoth-TRN-GO-Goiania-UFG004			??	?-.....			-.....	..A.....
#Ananum-TRN-Caldas-G93			??	?-.....			-.....	..A.....
#Ananum-TRN-Caldas-G94			??	?-.....			-.....	..A.....
#Asp-TRN-PI-CrisCastro-C2009			??	?-.....			-.....	..A.....
#Asp-TRN-RR-EstTaiano-C2014			??	?-.....			-.....	..A.....
#Asp-TRN-GO-Cristalina-C2195			??	?-.....			-.....	..A.....
#Asp-TRN-PI-EliseuMartinsC2213			??	?-.....			-.....	..A.....
#Ahumile-TRN-GO-Jaciara-C2229			??	?-.....			-.....	..A.....
#Asp-TRN-1495			??	?-.....			-.....	..A.....
#Desc-TRN-1635			??	?-.....			-.....	..A.....
#Desc-TRN-RR-EstNormandia-C4040			??	?-.....			-.....	..A.....
#Desc-TRN-BAG084			??	?-.....			-.....	..A.....
#Desc-TRN-BAG167			??	?-.....			-.....	..A.....
#Desc-TRN-BAG169			??	?-.....			-.....	..A.....
#Desc-TRN-BAG170			??	?-.....			-.....	..A.....

Tabela 12– Alinhamento das seqüências trnL

Seqüências	Bases alinhadas de 81 a 160							
#Mangifera-TRN-SP-Campinas	TGAGATCCTT	TTTCATTACG	TATACATAGT	GTATTTTTC	ATTACTTATA	CATAGTGTA-	TATAGG-CTT	ATCCTTCATC
#Aocce-TRN-C6015						A	-	
#Aocce-TRN-C6135						A	-	
#Aocce-TRN-MT-General-26						A	-	
#Aocce-TRN-MT-Colider-C2192						A	-	
#Aocce-TRN-BAG166						A	-	
#India-TRN-Ullal						A	-	
#Ahum-TRN-MT-General-Ass01636						A	-	
#Ahum-TRN-MT-General-FazCar-1						A	-	
#Ahum-TRN-MT-General-FazCar-3						A	-	
#Ahum-TRN-MT-General-fazCar664						A	-	
#Ahum-TRN-MT-General-fazCar669						A	-	
#Ahum-TRN-MT-General-fazCar677						A	-	
#Ahum-MT-General-fazJarvis680						A	-	
#Ahum-TRN-GO-Cavalcante-C2079					...C.....	A	-	
#Ahum-TRN-GO-AltoParaiso-C2163					...C.....	A	-	
#Aoth-TRN-GO-Galheiros-C2147						A	-	
#Aoth-TRN-GO-SaoDomingos-C2117						A	-	
#Aoth-TRN-PA-Tucurui-C2279						A	-	
#Aoth-TRN-MT-AguaBoa-C2281						A	G.....	
#Aoth-TRN-GO-Goiania-UFG002						A	-	
#Aoth-TRN-GO-Goiania-UFG003						A	-	
#Aoth-TRN-GO-Goiania-UFG004						A	-	
#Ananum-TRN-Caldas-G93						A	-	
#Ananum-TRN-Caldas-G94						A	-	
#Asp-TRN-PI-CrisCastro-C2009						A	-	
#Asp-TRN-RR-EstTaiano-C2014						A	-	
#Asp-TRN-GO-Cristalina-C2195						A	-	
#Asp-TRN-PI-EliseuMartinsC2213						A	-	
#Ahumile-TRN-GO-Jaciara-C2229					...C.....	A	-	
#Asp-TRN-1495						A	-	
#Desc-TRN-1635						A	-	
#Desc-TRN-RR-EstNormandia-C4040						A	-	
#Desc-TRN-BAG084						A	-	
#Desc-TRN-BAG167						A	-	
#Desc-TRN-BAG169						A	-	
#Desc-TRN-BAG170						A	-	

Tabela 12– Alinhamento das seqüências trnL

Seqüências	Bases alinhadas de 161 a 240							
#Mangifera-TRN-SP-Campinas	CGTTATGGCC	TTTCGATATT	CGATAGAAGA	ATTCTTTTCT	TTACCCAAAA	AAGGCAATCA	ACTCCATTG	TTAGAACAGC
#Aocc-TRN-C6015	.T...AA..			...C...	..C...C...			
#Aocc-TRN-C6135	.T...AA..			...C...	..C...C...			
#Aocc-TRN-MT-General-26	.T...AA..			...C...	..C...C...		.T.....	
#Aocc-TRN-MT-Colider-C2192	.T...AA..			...C...	..C...C...		.T.....	
#Aocc-TRN-BAG166	.T...AA..			...C...	..C...C...			
#India-TRN-Ullal	.T...AA..			...C...	..C...C...			
#Ahum-TRN-MT-General-Ass01636	.T...AA..		..A.....	...C...	..C...C...			
#Ahum-TRN-MT-General-FazCar-1	.T...AA..		..A.....	...C...	..C...C...			
#Ahum-TRN-MT-General-FazCar-3	.T...AA..		..A.....	...C...	..C...C...			
#Ahum-TRN-MT-General-fazCar664	.T...AA..				..C...C...			
#Ahum-TRN-MT-General-fazCar669	.T...AA..			...C...	..C...C...		.T.....	
#Ahum-TRN-MT-General-fazCar677	.T...AA..				..C...C...			
#Ahum-MT-General-fazJarvis680	.T...AA..		..A.....	...C...	..C...C...			
#Ahum-TRN-GO-Cavalcante-C2079	.T...AA..			...C...	..C...C...		.T.....	
#Ahum-TRN-GO-AltoParaíso-C2163	.T...AA..			...C...	..C...C...		.T.....	
#Aoth-TRN-GO-Galheiros-C2147	.T...AA..			...C...	..C...C...		.T.....	
#Aoth-TRN-GO-SaoDomingos-C2117	.T...AA..			...C...	..C...C...			
#Aoth-TRN-PA-Tucurui-C2279	.T...AA..			...C...	..C...C...			
#Aoth-TRN-MT-AguaBoa-C2281	.T...AA..			...C...	..C...C...			
#Aoth-TRN-GO-Goiania-UFG002	.T...AA..			...C...	..C...C...		.T.....	
#Aoth-TRN-GO-Goiania-UFG003	.T...AA..			...C...	..C...C...		.T.....	
#Aoth-TRN-GO-Goiania-UFG004	.T...AA..			...C...	..C...C...		.T.....	
#Ananum-TRN-Caldas-G93	.T...AA..		..A.....	...C...	..C...C...			
#Ananum-TRN-Caldas-G94	.T...AA..		..A.....	...C...	..C...C...			
#Asp-TRN-PI-CrisCastro-C2009	.T...AA..			...C...	..C...C...			
#Asp-TRN-RR-EstTaiano-C2014	.T...AA..			...C...	..C...C...			
#Asp-TRN-GO-Cristalina-C2195	.T...AA..			...C...	..C...C...			
#Asp-TRN-PI-EliseuMartinsC2213	.T...AA..			...C...	..C...C...			
#Ahumile-TRN-GO-Jaciara-C2229	.T...AA..			...C...	..C...C...		.T.....	
#Asp-TRN-1495	.T...AA..				..C...C...	C		
#Desc-TRN-1635	.T...AA..				..C...C...	C		
#Desc-TRN-RR-EstNormandia-C4040	.T...AA..			...C...	..C...C...			
#Desc-TRN-BAG084	.T...AA..			...C...	..C...C...			
#Desc-TRN-BAG167	.T...AA..			...C...	..C...C...			
#Desc-TRN-BAG169	.T...AA..			...C...	..C...C...			
#Desc-TRN-BAG170	.T...AA..			...C...	..C...C...			

Tabela 12– Alinhamento das seqüências trnL

Seqüências	Bases alinhadas de 241 a 311							
#Mangifera-TRN-SP-Campinas	TTCCATTGAG	TCTCTGCACC	TATCCTTTT	TT---CTCCC	GTCTGACCC	CTTGTTTGTT	TTTGTCTCG	T
#Aocce-TRN-C6015				.T--.....		.C.....		.
#Aocce-TRN-C6135				.T--.....		.C.....		.
#Aocce-TRN-MT-General-26				.TT-.....		.C.....		.
#Aocce-TRN-MT-Colider-C2192				.TT-.....		.C.....		.
#Aocce-TRN-BAG166				.T--.....		.C.....		.
#India-TRN-Ullal				.T--.....		.C.....		.
#Ahum-TRN-MT-General-Ass01636				.T--.....		.C.....		.
#Ahum-TRN-MT-General-FazCar-1				.T--.....		.C.....		.
#Ahum-TRN-MT-General-FazCar-3				.T--.....		.C.....		.
#Ahum-TRN-MT-General-fazCar664	...G.....			.T--.....		.C.....		.
#Ahum-TRN-MT-General-fazCar669				.TT-.....		.C.....		.
#Ahum-TRN-MT-General-fazCar677	...G.....			.T--.....		.C.....		.
#Ahum-MT-General-fazJarvis680				.T--.....		.C.....		.
#Ahum-TRN-GO-Cavalcante-C2079				.TT-.....		.C.....		.
#Ahum-TRN-GO-AltoParaíso-C2163				.TT-.....		.C.....		.
#Aoth-TRN-GO-Galheiros-C2147				.TTT.....		.C.....		.
#Aoth-TRN-GO-SaoDomingos-C2117				.T--.....		.C.....		.
#Aoth-TRN-PA-Tucurui-C2279				.T--.....		.C.....		.
#Aoth-TRN-MT-AguaBoa-C2281				.T--.....		.C.....		.
#Aoth-TRN-GO-Goiania-UFG002				.TT-.....		.C.....		.
#Aoth-TRN-GO-Goiania-UFG003				.TT-.....		.C.....		.
#Aoth-TRN-GO-Goiania-UFG004				.TT-.....		.C.....		.
#Ananum-TRN-Caldas-G93				.T--.....		.C.....		.
#Ananum-TRN-Caldas-G94				.T--.....		.C.....		.
#Asp-TRN-PI-CrisCastro-C2009				.T--.....		.C.....		.
#Asp-TRN-RR-EstTaiano-C2014				.T--.....		.C.....		.
#Asp-TRN-GO-Cristalina-C2195				.T--.....		.C.....		.
#Asp-TRN-PI-EliseuMartinsC2213				.T--.....		.C.....		.
#Ahumile-TRN-GO-Jaciara-C2229				.TT-.....		.C.....		.
#Asp-TRN-1495				.T--.....		.C.C.....		.
#Desc-TRN-1635				.T--.....		.C.C.....		.
#Desc-TRN-RR-EstNormandia-C4040				.T--.....		.C.....		.
#Desc-TRN-BAG084				.T--.....		.C.....		.
#Desc-TRN-BAG167				.T--.....		.C.....		.
#Desc-TRN-BAG169				.T--.....		.C.....		.
#Desc-TRN-BAG170				.T--.....		.C.....		.

A baixa ocorrência de falhas nas seqüências dos dois alinhamentos permitiu utilização do método de deleção par a par na obtenção das estimativas das distâncias evolutivas, em vez de

avaliar os dados com e sem *indels*, atribuindo notas empíricas às inserções/deleções (MACDADE; MOODY, 1999). No método par a par, são ignoradas apenas as falhas das duas seqüências envolvidas no cálculo das estimativas das distâncias evolutivas, o que tende a aumentar os valores das distâncias (NEI; KUMAR, 2000).

Embora o tamanho das seqüências da região *trnL* tenham sido maiores que o tamanho das seqüências ITS1, a variabilidade foi menor (Tabela 13).

Tabela 13 – Valores médios das porcentagens de bases e de tamanho, além de coincidência e razão coincidência/tamanho para as duas seqüências.

Região	T%	C%	A%	G%	Tamanho	Coincidências	Razão
ITS1	9,7	33,8	17,3	39,2	179,7	164	0,91
TrnL	40,1	22,3	24,7	12,9	303,4	301	0,99

Dessa forma, foram encontradas mais diferenças intraespecíficas entre as seqüências ITS1 que entre as seqüências *trnL*, conforme mostrado a seguir.

2.3.3 Variação Intraespecífica

Embora diversos autores encontrem variação intraespecífica (ROSSELLO et al, 2007), poucos investigam tal variação quando avaliam a variação interespecífica (MACDADE; MOODY,). Assim, selecionam indivíduos que representem a variação da espécie e estabelecem, então, as relações filogenéticas entre espécies.

Nesta avaliação, todos os indivíduos de uma mesma espécie foram avaliados com a seqüência *outgroup* na tentativa de identificar padrões de variação. O número de indivíduos e de locais de coleta de cada espécie variou de acordo com o DNA de boa qualidade obtido, conforme já comentado. Também foram incluídas as amostras identificadas pelo gênero ou identificadas como “Desconhecida”, para avaliar se essas amostras poderiam pertencer a alguma das espécies avaliadas, de acordo com os resultados.

A escolha dos métodos para obtenção das árvores depende de vários aspectos. Por exemplo, do nível taxonômico que se pretende avaliar, da região genômica escolhida, do tamanho

das seqüências consideradas e do método para o cálculo das estimativas de distância (NEI; KUMAR, 2000).

Todas as árvores foram reconstruídas utilizando *neighbor-joining* utilizando distâncias evolutivas obtidas pelo método de Tamura-Nei e deleção par a par, com 5000 *bootstraps*. A seguir, todas as topologias das árvores originais e as topologias consenso obtidas a partir de 5000 *bootstraps* para as diferentes combinações de amostras.

Para *A. occidentale*, foram obtidas três amostras para a região ITS1, uma delas de Colider-MT e as outras duas sem local de coleta indentificado (C6-001 e C6-015). Na Figura 13 podem ser vistas as topologias obtidas para essa espécie.

Figura 13 – Topologias obtidas para as seqüências Aocc.

Em todas as topologias, sempre há a formação de um grupo bem suportado (de 92 a 99%) das seqüências Aocc-ITS-C6001 e Aocc-ITS-C6015. Nas topologias que incluem apenas as seqüências Aocc (*A. occidentale*) Asp (*Anacardium* sp), estas últimas formam um grupo isolado com bom suporte (82%), indicando que essas seqüências podem não ser de *A. occidentale*. Quando foram incluídas também as seqüências Desc, as seqüências Asp foram separadas e uma delas (Asp-ITS-RR-EstTaiano-C2014) foi agrupada com uma Desc (Desc-ITS-BAG170) com

grande suporte (99%). Ainda assim, as seqüências Asp permaneceram num grupo maior, também com bom suporte (88%). Quando as seqüências Desc foram incluídas sem as amostras Asp, não se agruparam com as amostras Aocc. A seqüência Aocc-ITS-MT-Colider-C2192 não formou grupos consistentes com nenhuma das outras seqüências, a seqüência Desc-ITS-BAG084 mostrou-se muito diferente de todas as demais e a seqüência Desc-ITS-1625 não formou agrupamentos consistentes com nenhuma das outras seqüências, indicando que essa seqüência pode ser de outra espécie.

Considerando todos os resultados das avaliações das seqüências Aocc, Asp e Desc, é possível obter algumas informações: (i) há variação entre indivíduos de *A. occidentale*, sendo que duas seqüências agruparam com bom suporte, enquanto que outra não formou agrupamentos bem suportados. Considerando que *A. occidentale* é a espécie com maior distribuição, é possível haver maior variabilidade entre indivíduos. Além disso, é possível também que as duas seqüências que se agruparam com bom suporte sejam de um local diferente de MT, (ii) de forma geral, as seqüências Asp e Desc parecem não ser de *A. occidentale*, uma vez que não formam grupos consistentes com essa espécie. Exceção a esse comportamento geral é a seqüência Desc-ITS-1625, que embora não forme agrupamentos consistentes, tem um comportamento semelhante ao da seqüência Aocc de MT. Entre as seqüências Desc, uma delas é bastante diferente de todas as outras seqüências, indício de que possa ser uma espécie diferente das demais. A outra seqüência Desc (Desc-ITS-BAG170) e as seqüências Asp, formam um grupo consistente que pode ser outra espécie com variação intraespecífica.

Para *A. othonianum*, foram obtidas cinco amostras para a região ITS1, de dois Estados diferentes e três locais distintos. As topologias obtidas para essa espécie podem ser vistas na Figura 14.

Figura 14

Entre as seqüências de Aoth, a do Estado do PA apresentou comportamento bastante diferente das demais. Dentre as seqüências de GO, dois grupos foram formados com bom suporte (95%). A introdução das seqüências Asp e Desc não melhora a resolução dos grupos (pelo contrário), apenas enfatiza que Aoth-ITS-PA-Tucuruí-C2279 é bastante diferente das demais Aoth (e das outras seqüências também) e que parece haver variação dentro do grupo das seqüências Aoth UFG, com a subdivisão desses grupos (principalmente quando observadas somente as seqüências Aoth). As amostras Asp RR e Desc BAG170, continuaram a se agrupar com bom suporte (99%).

Considerando todos os resultados é possível observar que: (i) Aoth-ITS-PA-Tucuruí-C2279 é diferente das demais Aoth; (ii) parece haver variabilidade dentro de *A. othonianum*, além daquela que diferencia a seqüência de Tucuruí das demais, que divide os indivíduos em dois grupos. Observando que pela taxonomia numérica (MITCHELL;MORI, 1987) *A. othonianum* e *A. occidentale* foram agrupadas numa mesma espécie, pode haver ampla distribuição de *A.*

othonianum, o que pode aumentar a chance de ocorrência de maior variabilidade dentro da mesma espécie; (iii) as seqüências Asp RR e Desc BAG170 parecem pertencer a uma mesma espécie.

Para *A. nanum*, foram obtidas duas amostras para a região ITS1, de dois Estados diferentes e três locais distintos. As topologias obtidas para essa espécie podem ser vistas na Figura 15

Figura 15 – Topologias obtidas para as seqüências Ananum

A seqüência Ananum-ITS-Caldas-G94 parece ser diferente da demais, incluindo a outra seqüência Ananum do mesmo local. Há bom suporte (91%) para diferenciar Ananum-ITS-Caldas-G93 (e, conseqüentemente, de Ananum-ITS-Caldas-G94) das seqüências Asp. Novamente, Asp RR e BAG170 agrupam-se com bom suporte.

Para *A. humile*, foram obtidas quatro amostras para a região ITS1, de dois Estados diferentes e dois locais distintos. As topologias obtidas para essa espécie podem ser vistas na Figura 16.

Figura 16 – Topologias obtidas para as seqüências Ahum

As seqüências de Ahum não formaram agrupamentos com bom suporte, nem entre si nem com as seqüências Asp e Desc. A partir desses resultados, é possível observar que: (i) As seqüências Asp e Desc parecem não pertencer à mesma espécie das seqüências Ahum, pois não formam agrupamentos com bom suporte.

Conforme já comentado, a variabilidade observada entre as seqüências *trnL* foi menor que a observada entre as seqüências para a região ITS1, permitindo identificar ainda menos padrões de variabilidade intraespecífica. Esses padrões serão discutidos no item a seguir.

2.3.4 Variação Interespecífica

Considerando todas as avaliações em conjunto, não parece haver padrões muito claros de variação das seqüências da região ITS1, que permitam estabelecer relações filogenéticas confiáveis entre as espécies (AMORIM, 2002), isto é, não é possível decidir claramente quais

indivíduos de cada espécie poderiam ser selecionados para a reconstrução filogenética (MACDADE; MOODY, 1999).

A árvore considerando todas as seqüências pode ser vista na Figura 17.

Figura 17 – Árvore reconstruída (esquerda) e bootstrap (esquerda) obtida com todas as seqüências ITS1.

Nas árvores, é possível observar que apenas dois nós são bem suportados, aquele com as seqüências Aocc (96%) e aquele com as seqüências Asp RR e BAG 170 (99%). Todas as demais relações são fracamente suportadas, com mistura de espécies em diferentes grupos.

Conforme já comentado, a reconstrução filogenética para as seqüências *trnL* forneceu árvores com ainda menos variabilidade intraespecífica (e interespecífica) (Figura 18)

Figura 18 - Árvore reconstruída (esquerda) e bootstrap (esquerda) obtida com todas as sequências trnL íntron.

Para as sequências trnL, há apenas um grupo bem suportado (Asp-TRN-1495 e Desc-TRN-1635) (87%), isolado de todas as outras sequências. Embora para todas as outras sequências não seja possível estabelecer relações filogenéticas, considerando apenas essa região, o alto suporte desse agrupamento permite diferenciar amostras entre os dois centros de diversidade, uma vez que a única amostra pertencente ao centro de diversidade primário identificada como uma espécie dessa região (*A. giganteum* - Desc-1635) agrupou com outra amostra dessa mesma região e diferenciou-se de todas as outras espécies do outro centro de diversidade.

Esse único resultado foi bastante surpreendente, uma vez que foi obtido para uma região que não é a mais indicada para estudos desse tipo (VIJVERBERG; BACHMANN, 1999; VARGAS; MORTON; JURY, 1999), a não ser que haja variação interespecífica suficiente, que foi a proposta desse trabalho.

Ainda mais interessante é que esse dado reflete, juntamente com os resultados obtidos com as sequências ITS1, o mesmo padrão de agrupamento encontrado por Mitchell e Mori (1987), por taxonomia numérica. Nesse trabalho, foram agrupadas as espécies *A. occidentale*, *A. humile* e *A. nanum*. Ainda, por essa classificação, *A. occidentale* e *A. othonianum* são consideradas uma mesma espécie (*A. occidentale*). Assim, todas as relações inferidas pelas sequências ITS, estariam sendo estabelecidas dentro de um grupo de espécies muito próximo. A junção dessas informações ao resultado obtido com as sequências trnL, corrobora os resultados de Mitchell e Mori (1987), com muito menos informações disponíveis para análise.

Assim, esses resultados parecem indicar um uso promissor dessas regiões, para análises desse tipo, desde que, evidentemente, mais dados sejam avaliados.

3 CONSIDERAÇÕES FINAIS

- 1) Utilizando poucos dados disponíveis, há evidências de diferenças intraespecíficas, que necessitam mais aprofundamento
- 2) Foi identificada maior variação para a região ITS1 em relação a região trnL.
- 3) A informação mais relevante, contudo, foi obtida de uma única sequência de trnL, que sugere que os centros de diversidade poderiam ser avaliados por essas regiões.
- 4) As informações das duas regiões permitiram obter agrupamentos semelhantes aos descritos na literatura para *Anacardium* sp

REFERÊNCIAS

- AMORIM, DS. **Fundamentos de sistemática filogenética**. Ribeirão Preto: Holos, 2002. 156p.
- ANGIOSPERM PHYLOGENY GROUP (APG). An ordinal classification for the families of flowering plants. **Annals of the Missouri Botanical Gardens.**, Saint Louis, v. 85, p. 531–553, 1998.
- ANGIOSPERM PHYLOGENY GROUP (APG). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. **Botanical Journal of the Linnean Society**, Londres, v. 141, p. 399–436, 2003.
- ARCHAK, S; GAIKWAD, A.D; GAUTAM, D.; RAO, E.V.V.B.; SWAMY, K.R.M.; KARIHALOO, J.L. Comparative assessment of DNA fingerprinting techniques (RAPD, ISSR and AFLP) for genetic analysis of cashew (*Anacardium occidentale* L.) accessions of India. **Genome**, Toronto, v. 46, p. 362-369, 2003.
- AVISE, J.C. **Phylogeography**. London: Harvard University Press, 2000. 447p
- AVISE, J.C. **Molecular markers, natural history, and evolution**. 2ed. Sunderland: Sinauer, 2004. 541p
- BARROS, L.M.; CRISÓSTOMO, J.R. Melhoramento genético do cajueiro. In: ARAÚJO, J.P.P.; SILVA, V.V. **Cajucultura: Modernas Técnicas de Produção**. Fortaleza: Embrapa-Cnpq, 1995. p.73-93.
- BARROS, L.M.; PAIVA, J.R.; CAVALCANTI, J.J.V.; ARAÚJO, J.P.P. Cajueiro. In: BRUCKNER, C.H. **Melhoramento de Fruteiras Tropicais**, Viçosa: UFV, 2002. p.159-176.
- BILLOTTE, N.; LAGODA, P.J.L.; RISTERUCCI, A.M.; BAURENS, F.C.(1999). Microsatellite-enriched libraries: applied methodology for the development of SSR in tropical crops. *Fruits*, v.54. p 277-288.
- BONATO, A.L.V.; VERZIGNASSI, J.R.; RESENDE, R. M.S.; FERNANDES, C. D.; LEGUIZAMÓN, G. O. C. *Extração de DNA genômico de Stylosanthes spp.* Comunicado técnico 78. Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), Embrapa Gado de Corte. Campo Grande, 2002.
- BREMER, K., BREMER, B.; THULIN, M. **Introduction to Phylogeny and Systematics of Flowering Plants**. Uppsala: Uppsala University. 1999. 91p.
- CANDOLLE, A P. **Prodromus Systematis Naturalis Regni Vegetabilis**. v.2. Paris: Treuttel et Würtz, 1825. 644p.
- CAVALLI-SFORZA, L.L; EDWARDS, A.W.F. Analysis of human evolution. In: PROC. INTL. CONG. GENET., 11, 1964. Pergamon, 1964.
- CHASE, M. W. et al. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcl*. **Annals of the Missouri Botanical Garden**, St Louis, v. 80, p. 528-580, 1993.

CLEGG, M.T.; ZURAWSKI, G. Chloroplast DNA and the study of plant phylogeny: present status and future prospects. In: SOLTIS, P.S; SOLTIS, D.E.; DOYLE, J.J. (eds) **Molecular systematics of plants**. New York: Chapman and Hall, 1992, p. 1–13.

CREGAN, P.B.; JARVIK, T.; BUSH, A.L.; SHOEMAKER, R.C.; KAHLER, A.L.; KAYA, N.; VAN TOAI, T.T.; LOHNES, D.G.; CHUNG, J.; SPECHT, J.E. An Integrated Genetic Linkage Map of the Soybean Genome. *Crop Science*. Stanford, v. 39, p. 1464–1490, 1999.

CRONQUIST, A. **An integrated system of classification of flowering plants**. New York: Columbia University Press, 1981. 1262p.

DAYHOFF, M.O.; SCHWARTZ, R.M.; ORCUTT, B.C. A model of evolutionary change in proteins. In: DAIHOFF, M.O. **Atlas of protein sequence and structure**. Silver Spring: National Biomedical Research Foundation, 1978. p. 345–352.

DHANARAJ, A.L.; RAO, E.V.V.B.; SWAMY, K.R.M.; BHAT, M.G.; PRASAD, D.T.; SONDUR, S.N. Using RAPDs to assess the diversity in Indian cashew (*Anacardium occidentale* L.) germplasm. **Journal of horticultural science and biotechnology**, Ashford, v. 77, p. 41–47, 2002.

FELSENSTEIN, J. Cases in which parsimony or compatibility methods will be positively misleading. **Systematic Zoology**, Amsterdã, v. 27, p. 401–410, 1978.

FELSENSTEIN, J. Evolutionary trees from DNA sequences: A maximum likelihood approach. **Journal of Molecular Evolution**, New York, v. 17, p. 368–376, 1981.

FIJII, N. Chloroplast DNA phylogeography of *Pedicularis* ser. *Glorisae* (Orobanchaceae) in Japan. **Journal of Plant Research**, Japão, v. 120, p. 491–500, 2007.

GADEK, P.A.; FERNANDO, E.S.; QUINN, C.J.; HOOT, S.B.; TERRAZAS, T.; SHEAHAN, M.C.; CHASE, M.W. Sapindales: Molecular delimitation and infraordinal groups. **American Journal of Botany**, St. Louis, v. 83, p. 802–811, 1996.

GIELLY, L.; TABERLET, P. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. **Molecular Biology and Evolution**, Oxford, v. 11, p. 769–777, 1994.

GUANDAVO, P.M. **Tratado da província do Brasil**. Brasília: MEC/INL, 1965. p. 167–168.

GUNDERSEN, A. **Families of dicotyledons**. **Chronica Botanica Company**, New York: Stechert-Hafner, 1950 237 p.

HAECKEL, E. **Generelle morphologie der organismen**. Berlim: Georg Riemer, 1866. 479p.

HAM, R.C.H.J.; HART, H.T.; MES, T.H.M.; SANDBRINK, J.M. Molecular evolution of noncoding regions of the chloroplast genome in the Crassulaceae and related species. **Current Genetics**, Berlim, v. 25, p. 558–566, 1994.

HARTL, D.L.; CLARK, A.G. **Principles of population genetics**. Sunderland: Sinauer Associates, 1997. 682p.

- HASEGAWA, M.; KISHINO, H.; SAITOU, N. On the maximum likelihood method in molecular phylogenetics. **Journal of Molecular Evolution**, New York, v. 32, p. 443-445, 1991.
- HOISINGTON, D; KHAIRALLAH, M.; LEON, G.. **Laboratory protocols CIMMYT Applied Molecular Genetics Laboratory:2**. México: CIMMYT, 1994, 74p.
- HSU, J. Late Cretaceous and Cenozoic Vegetation in China, emphasizing their connections with North America. **Annals of the Missouri Botanical Garden**, St Louis, v.70, p. 490-508, 1983.
- JOHNSON, D.V. The botany, origin and spread of cashew (*Anacardium occidentale* L.). **Journal of plantation crops, Kasaragod**, v. 1, p. 1-7, 1973.
- JOHNSON, D.V. **O caju no Nordeste do Brasil: um estudo geográfico**. Fortaleza: BNB/ETENE, 1974. 169p.
- JUKES, T.H.; CANTOR, C.R. Evolution of protein molecules. In: MUNRO, H.N. **Mammalian protein metabolism**, Burlington: Academic Press, 1969. p. 21-132.
- KNOBLOCH, E., MAI, D.H. Monographie der Früchte und Samen in der Kreide von Mitteleuropa. Rozpr. Ustred. **Ustavu Geologickeho**, Praga, v. 47, p. 1-219, 1986.
- KUSUMI, J.; TSUMURA, Y.; YOSHIMARU, H.; TACHIDA, H. 2000. Phylogenetic relationships in Taxodiaceae and Cupressaceae sensu stricto based on matK gene, chlL gene, trnL-trnF IGS region, and trnL intron sequences. **American Journal of Botany**, St Louis, v. 87, p. 1480-1488, 2000.
- LIMA, O.E.; MAGALHÃES NETO, B.; FARIAS, L.; ALBUQUERQUE, J.L.; SIMÕES FILHO, S. **Introdução ao estudo dos cajus de Pernambuco**. Recife: Universidade de Recife/Escola de Química de Pernambuco, 1952. 28p.
- MAGALLÓN, S.; SANDERSON, M. J. Absolute diversification rates in angiosperm clades. **Evolution**, São Francisco, v. 55, p. 1762-1780, 2001.
- MARCGRAVE, J. **História natural do Brasil**. São Paulo: Museu Paulista, 1942. 297p.
- McDADE, L.A.; MOODY, M.L. Phylogenetic relationships among Acanthaceae: evidence from noncoding trnL-trnF chloroplast DNA sequences. **American Journal of Botany**, St. Louis, v. 86, p. 70-80. 1999.
- MITCHELL, J.D.; MORI, S.A. **The cashew and its relatives (Anacardium: Anacardiaceae)**. New York: Memoirs of the New York Botanical Garden, 1987. vol. 42, 76p.
- MNENEY, E.E.; MANTELL, S.H; BENNETT, M. Use of random amplified polymorphic DNA (RAPD) markers to reveal genetic diversity within and between populations of cashew (*Anacardium occidentale* L.). **J. Horticultural Science and Biotechnology**, Ashford, v. 76, p. 375-383, 2001.
- MULLER, J. Significance of fossil pollen for angiosperm history. **Annals of the Missouri Botanical Garden**, St. Louis, v. 71, p. 419-443, 1984.

NEEDLEMAN, S.G.; WUNSCH, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. **Journal of Molecular Biology**, New York, v. 48, p. 443-453, 1970.

NEI, M. Stochastic errors in DNA evolution and molecular phylogeny. In: GERHOWITZ, H. **Evolutionary perspectives and the new genetics**, New York: Alan R. Liss, 1986. p. 133-147.

NEI, M. **Molecular evolutionary genetics**. New York: Columbia University Press, 1987, 512p.

NEI, M.; KUMAR, S. **Molecular evolution and phylogenetics**. New York: Oxford, 2000. 333p.

NEY, M.; KUMAR, S.; TAKAHASHI, K. The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. **Proceedings of the National Academy of Science**, Stanford, v. 95, p. 390-397, 1998.

PAMILO, P.; NEI, M. Relationships between gene trees and species trees. **Molecular Biology and Evolution**, Oxford, v. 5, p.568-583, 1988.

PELL, S.K. Molecular systematics of the cashew family (Anacardiaceae). 2004, 193p. Tese (Doutorado em Ciências Biológicas) - Louisiana State University, Louisiana, 2004.

PESSOAL, P.F.A.P.; LEITE, L.A.S.; PIMENTEL, C.R.M. Situação atual e perspectivas da agroindústria do caju. In: **Cajucultura: Modernas Técnicas de Produção**. Fortaleza:Embrapa-Cnpq, 1995. p.23-33.

RENNER, S.S. Phylogenetic affinities of Monimiaceae based on cpDNA gene and spacer sequences. **Perspectives in Plant Ecology, Evolution and Systematics**, New York, v. 1, p. 61-77, 1998.

ROSSELLO, J.P.; LÁZARO, A.; COSÍN, R.; MOLINS, A. A Phylogeographic Split in *Buxus balearica* (Buxaceae) as Evidenced by Nuclear Ribosomal Markers: When ITS Paralogues Are Welcome. **Journal of Molecular Evolution**, New York, v. 64, p. 143-157, 2007.

SAITOU, N.; NEI, M. The number of nucleotides required to determine the branching order of three species, with special reference to the human-chimpanzee-gorilla divergence. **Journal of Molecular Evolution**, New York, v. 24, p. 189-204, 1986.

SAVOLAINEN, V.; CHASE, M. W. A decade of progress in plant molecular phylogenetics. **Trends in Genetics**, New York, v19, p. 717-724, 2003.

SAVOLAINEN, V.; CHASE, M.W.; HOOT, S.B.; MORTON, C.M.; SOLTIS, D.E.; BAYER, C.; FAY, M.F.; BRUIJN, A.Y.; SULLIVAN, S.; QIU, Y.L. Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcL* sequences. **Systematic Biology**, Oxford, v. 49, p. 306-362, 2000(a).

SAVOLAINEN, V.; FAY, M.F.; ALBACH, D.C.; BACKLUND, A.; van der BANK, M.; CAMERON, K.M.; JOHNSON, S.A.; LLEDÓ, M.D.; PINTAUD, J.C.; POWELL, M.; SHEAHAN, M.C.D.; SOLTIS, E.; SOLTIS, P.S.; WESTON, P.; WHITTEN, W.M.; WURDACK, K.J.; CHASE, M.W. Phylogeny of the eudicots: a nearly complete familial analysis based on *rbcL* gene sequences. **Kew Bulletin**, Kew, v. 55, p. 257-309, 2000(b).

SELLERS, P.H. On the theory and computation of evolutionary distances. **SIAM Journal on Applied Mathematics**, Filadélfia, v. 26, p. 787-793, 1974.

SOKAL, R.R.; SNEATH, P.H.A. **Principles of numerical taxonomy**. San Francisco: Freeman, 1963. 359p.

SCHNEIDER, H. **Métodos de análise filogenética**. Ribeirão Preto: Holos, 2007. 200p.

SOLTIS, D.E.; SOLTIS, P.S. Choosing an approach and an appropriate gene for phylogenetic analysis. In: SOLTIS, D.E.; SOLTIS, P.S., DOYLE, J.J. (eds) **Molecular systematics of plants II**, Massachusetts: Kluwer, 1998. p 1–42.

SOLTIS, D.E.; SOLTIS, P.S.; MILLIGAN, B.G. Intraspecific chloroplast DNA variation: systematics and phylogenetic implication. In: SOLTIS, D.E.; SOLTIS, P.S., DOYLE, J.J. (eds) **Molecular systematics of plants**, New York: Chapman and Hall, 1992. p 117–150.

SOLTIS, D.E.; SOLTIS, P.S.; MORT, M.E.; CHASE, M.W.; SAVOLAINEN, V.; HOOT, S.B.; MORTON, C.M. Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. **Systematic Biology**, Oxford, v. 47, p. 32–42, 1998.

STRAND, A.E.; LEEBENS-MACK, J.; MILLIGAN, B.G. Nuclear DNA-based markers for plant evolutionary biology. **Molecular Ecology**, São Francisco, v. 6, p. 113-118, 1997.

SYED, N.H.; SURESHSUNDAR, S.; WILKINSON, M.J.; BHAI, B.S.; CAVALCANTI, J.J.V.; FLAVELL, A.J. Ty1-*copia* retrotransposon-based SSAP marker development in cashew (*Anacardium occidentale* L.). **Theoretical and Applied Genetics**, Berlim, v. 110, p. 1195-1202, 2005.

TAJIMA, F.; NEI, M. Biases of the estimates of DNA divergence obtained by restriction enzyme technique. **Journal of Molecular Evolution**, New York, v. 17, p. 115-120, 1982.

TAMURA, K. Estimation of number nucleotide substitutions when there are strong transition-transversion and G+C-content biases. **Molecular Biology and Evolution**, Oxford, v.9, p. 678-687, 1992.

TAMURA, K.; NEI, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. **Molecular Biology and Evolution**, Oxford, v. 10, p. 512-526, 1993.

TERRAZAS, T. **Wood anatomy of the Anacardiaceae: Ecological and phylogenetic interpretation**. 1994, 321p. Tese (Doutorado). University of North Carolina, Chapel Hill, 1994.

THEVET, A. **Singularidades da França antártica**. São Paulo: Companhia Editora Nacional, 1958, 504p.

THORNE, R. F. Classification and geography of the flowering plants. **Botanical Review**, New York, v. 58, p. 225-348, 1992.

VARGAS, P; MORTON, CM; JURY, SL. Biogeographic patterns in mediterranean and macaronesian species of saxifragaceae inferred from phylogenetic analyses of ITS sequences. **American Journal of Botany**, St. Louis, v. 86, p. 724-734, 1999.

VIJVERBERG, K; BACHMANN, K. Molecular evolution of a tandemly repeated trnF(GAA) gene in the chloroplast genomes of *Microseris* (Asteraceae) and the use of structural mutations in phylogenetic analyses. **Molecular Biology and Evolution**, Oxford, v. 16, p. 1329-1340, 1999.

WIKSTRÖM, N., SAVOLAINEN, V.; CHASE, M.W. Evolution of the angiosperms: calibrating the family tree. **Proceeding of the Royal Society of London**, Londres, v. 268: p. 2211-2220, 2001.

WILSON, A.C.; CARLSON, S.S.; WHITE, T.J. Biochemical evolution. **Annual Review of Biochemistry**, Palo Alto, v. 46, p. 573-639, 1977.

YI, T.; MILLEN, A. J.; WEN, J. Phylogenetic and biogeographic diversification of *Rhus* (Anacardiaceae) in the Northern Hemisphere. **Molecular Phylogenetics and Evolution**, Nova Iorque, v. 33, p. 861-879, 2004.