Vitor Junji Suzaki

Filogenia e biogeografia do complexo *Croton pallidulus* (Euphorbiaceae), inferidas por sequências de DNA e marcadores AFLP

Phylogeny and biogeography of complex *Croton pallidulus* (Euphorbiaceae), inferred by DNA sequences and AFLP markers

São Paulo

2011

Vitor Junji Suzaki

Filogenia e biogeografia do complexo *Croton pallidulus* (Euphorbiaceae), inferidas por sequências de DNA e marcadores AFLP

Phylogeny and biogeography of complex *Croton pallidulus* (Euphorbiaceae), inferred by DNA sequences and AFLP markers

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

Orientadora: Profa. Dra. Maria Luiza Faria Salatino

São Paulo

2011

Suzaki, Vitor
Filogenia e biogeografia do complexo *Croton pallidulus* (Euphorbiaceae), inferidas por sequências de DNA e marcadores AFLP.

Número de páginas: 89
Dissertação (Mestrado) – Instituto de Biociências da Universidade de São Paulo. Departamento de Botânica, 2010.

1.Complexo *Croton pallidulus* 2.Filogenia molecular 3.AFLP. Universidade de São Paulo. Instituto de Biociências. Departamento de Botânica.

Comissão Julgadora

_____ _____

Prof(a.) Dr(a). Prof(a.) Dr(a).

_____

Profª Drª Maria Luiza Faria Salatino

(Orientadora)

Dedico este trabalho
a minha família que
sempre me apoiou
em todas as etapas.

Agradecimentos

Gostaria de agradecer primeiramente a meus pais que sempre contribuíram para o desenvolvimento deste projeto de pesquisa, em todos os momentos bons e principalmente nos ruins, muito obrigado. Um agradecimento especial também a minha namorada Fabiana, por ter me aturado todo este tempo, foram vários momentos difíceis, de muito trabalho, ansiedades e chateações porque às vezes um experimento não dava certo, mas que estava sempre do meu lado para me apoiar e me dar força.

Não tenho palavras pra agradecer todos os anos em que estive no laboratório, desde a iniciação científica, em que principalmente a minha orientadora Profa. Dra. Maria Luiza Faria Salatino me ajudou demais, me tratando como a um filho, e não vou conseguir agradecer o suficiente por todos estes momentos, só tenho a dizer muitíssimo obrigado. Assim como ao Prof. Dr. Antonio Salatino, que sempre foi muito prestativo para me ajudar nas análises, relendo relatórios juntamente comigo e a Profa. para resolver uma dúvida muito grande de como escrever objetiva e claramente a respeito da discussão de um ponto importante, muito obrigado também.

A todos os amigos do laboratório (e espero não esquecer todos): Adne, Alice, Ana Clara, Augusto, Bruno, Bruna, Caio, Carmen, Carol, Cris (que me ajudou muito, e se hoje consegui completar este projeto de pesquisa foi graças as coisas que você me ensinou), Dani, Diego, Edegar, Emerson, Fernanda, Janaína, Jóice, Josi, Kátia, Lilian, Liss, Mari coluna, Mari, Milene (que me ajudou muito nas fases finais do projeto), Priscila, Sarah, Vanessa. Muito obrigado pelo companheirismo, cada qual do seu jeito, com as brincadeiras, me desculpem qualquer besteira que tenha falado ou feito.

A técnica e amiga Mourisa que em tudo o que podia me ajudou sempre, às vezes reclamando (isso é uma brincadeira), e em muitos momentos de tanto eu perguntar eu até acreditava que incomodava muito, mas sempre continuava a perguntar antes de fazer algo errado para não tomar uma bronca maior depois. Pode-se perguntar de tudo a ela a respeito de um procedimento porque

ela domina o funcionamento do laboratório, e isto é muito importante ter alguém próximo que sempre está disponível para ajudar.

A técnica e amiga Paula por ser sempre tão alto astral, aparentemente, e por fazer um café forte pra acordar todos, além das conversas informais que em muito contribuíram, se não somente para conversar, entreter, também para desabafar sobre as experiências malsucedidas no laboratório.

Gostaria de agradecer também às técnicas Sílvia e Tatiana do Laboratório de Biologia Molecular de Plantas por terem ajudado em muito para a etapa de sequenciamento, além da utilização de alguns equipamentos.

As Profas. Déborah e Cláudia e a Dra. Lucimar pelas constantes ajudas não somente com os equipamentos, mas também me ensinando e zelando pelo bem estar do laboratório.

A Profa. Inês Cordeiro, minha eterna gratidão pelo auxílio na coleta das espécies utilizadas no projeto, e também ajudando posteriormente na identificação dos materiais que trouxe de uma coleta. Além disso, o contato com uma aluna do Rio Grande do Sul para o envio de indivíduos de uma espécie acrescentada no trabalho, novamente obrigado.

A todo o pessoal do Museu Botânico Municipal de Curitiba, em especial ao Osmar pela simpatia e auxílio, ao Joel e ao Eraldo que me ajudaram em muito na coleta, indicando os locais e identificando as espécies, e posteriormente também para o envio dos indivíduos de uma espécie, meu muito obrigado.

Aos nobres colegas de futsal da Bio que me proporcionaram momentos fantásticos, sejam em campeonatos, jogos, treinos ou concentrações, e foram importantíssimos para o meu bem estar tanto físico, mas também mental.

E por fim, mas não menos importante, à FAPESP pela concessão da bolsa de mestrado e financiamento do projeto *Croton*.

ÍNDICE

1. Introdução
1.1. *Croton* L. no Brasil..... 08
1.2. Sistemática Molecular..... 10
1.3. AFLP..... 11
1.4. Posicionamento de *Croton* em Euphorbiaceae s.l..... 15
1.5. *Croton pallidulus* Baill..... 17
1.6. Biogeografia..... 21

2. Objetivos..... 23

3. Materiais e Métodos
3.1. Materiais..... 23
3.2. Métodos..... 24
3.2.1. Extração do DNA para as análises moleculares..... 24
3.2.2. Sequenciamento de DNA..... 25
3.2.3. Análise por AFLP..... 27
3.2.3.1. Quantificação do DNA..... 27
3.2.3.2. Reação de restrição (digestão) do DNA..... 28
3.2.3.3. Reação de ligação..... 28
3.2.3.4. Reação de amplificação pré-seletiva..... 28
3.2.3.5. Reação de amplificação seletiva..... 28
3.2.3.6. Análise dos resultados..... 29
3.3. Biogeografia do complexo..... 30

4. Resultados e Discussão
4.1. Sequenciamento de DNA..... 31
4.2. Marcadores AFLP..... 45
4.3. Biogeografia do complexo..... 55

5. Resumo..... 61

6. Abstract..... 62

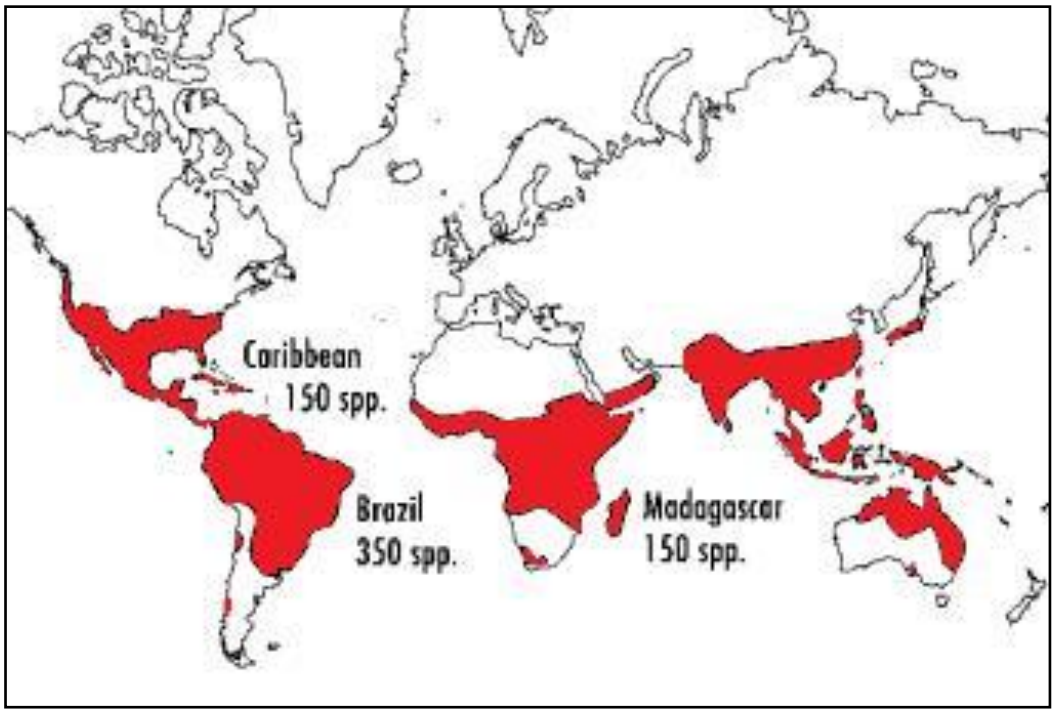
7. Referências Bibliográficas..... 63

8. Anexos..... 71

1. INTRODUÇÃO

1.1. *Croton* L. no Brasil

A flora brasileira é rica em representantes de Euphorbiaceae. Em muitas partes do mundo, representantes dessa família são muito conhecidos por serem tóxicos e/ou medicinais. O segundo maior gênero de Euphorbiaceae, *Croton* L., possui aproximadamente 1300 espécies, distribuídas em zonas tropicais e subtropicais, tanto no Novo quanto no Velho Mundo (Berry *et al*. 2005).



(Berry *et al*., 2005)

Figura 1. Mapa de distribuição do gênero *Croton* no mundo e principais centros de diversidade.

Os primeiros estudos taxonômicos completos das espécies brasileiras de *Croton* foram realizados por Martius na sua Flora Brasiliensis (Mueller, 1873). O gênero é composto por árvores, arbustos, ervas e raramente lianas; geralmente são plantas monóicas, algumas vezes dióicas, com tricomas simples (glandulares ou não), ou tricomas ramificados (estrelados, lepidotos, fasciculados, multirradiados, dentríticos e escamiformes). Em geral as inflorescências apresentam flores pistiladas na base e estaminadas no ápice, com estames dobrados no botão floral; nas flores pistiladas, as pétalas são reduzidas ou ausentes (Bentham, 1880; Barroso, 1991; Lucena, 2000; Carneiro *et al*., 2002; Secco, 2005; Guimarães, 2006). Em praticamente todos os

ecossistemas brasileiros encontram-se representantes de *Croton*. Apenas no estado do Rio de Janeiro, reconhecem-se 39 espécies (Pereira *et al*., 2002).

As espécies de *Croton* têm grande interesse pelas suas propriedades químicas e farmacológicas, comprovadas clinicamente como antiinflamatórias, cicatrizantes, anticancerígenas, antimicóticas, antimicrobianas e antivirais, sendo utilizadas para a cura de inúmeras doenças (Pieters & De Bruyne, 1995). No Brasil, espécies como *C. cajucara*, *C. celtidifolius*, *C. nepetaefolius*, *C. palanostigma*, *C. schiedeanus*, *C. urucurana* e *C. zehntneri* são usadas em medicina popular para tratamentos de uma ampla diversidade de males, tais como hipercolesterolemia e obesidade (*C. cajucara*); inflamações, úlcera e leucemia (*C. celtidifolius*); inapetência e cólicas (*C. nepetaefolius*); feridas e inflamação intestinal (*C. palanostigma*); reumatismo e câncer (*C. urucurana*); anorexia, distúrbios gastrointestinais, tendo suas folhas e cascas sido usadas também como adoçante (*C. zehntneri*) (Salatino *et al*., 2007).

Os estudos existentes, centrados em espécies de *Croton* são ainda escassos e dizem respeito às áreas de fitoquímica e farmacologia; mais raramente surgem aqueles dedicados à anatomia, levantamentos florísticos e fitossociológicos. Tratando-se de um gênero de taxonomia complexa, agravado pelo elevado número de espécies, são frequentes em *Croton* os problemas de delimitação específica, nomenclaturais e de polimorfismos. Surgem, portanto, dificuldades no reconhecimento dos táxons, culminando na descrição de espécies novas, quando na realidade trata-se de modificações morfológicas de táxons já conhecidos (Lucena, 2000).

A maioria das espécies de *Croton*, notadamente as herbáceas e arbustivas, têm crescimento rápido. De acordo com a mais bem sucedida teoria da ecologia bioquímica (alocação de recursos, Coley *et al*., 1985), plantas de crescimento rápido investem preferencialmente em defesas qualitativas, ou seja, substâncias com atividade fisiológica, como alcaloides e terpenoides. Não é surpreendente, portanto, que *Croton* seja uma fonte abundante de metabólitos secundários, com ampla diversidade de atividades farmacológicas (Salatino *et al.*, 2007).

O estudo desse representativo gênero propõe imenso desafio, com uma inigualável oportunidade de estudar fenômenos importantes relacionados à evolução das espécies, como as mudanças nos taxa de diversificações evolutivas, radiações adaptativas, principais inovações e rearranjos cromossômicos (Berry *et al.*, 2005).

1.2. Sistemática molecular

Em passado relativamente recente, foram criados sistemas de classificação de angiospermas, procurando incorporar caracteres de todas as fontes até então disponíveis, como morfologia, anatomia, ultraestrutura, fisiologia, bioquímica, citologia, palinologia e química (Judd *et al.*, 1999). Exemplos de tais sistemas são os de Cronquist (1981), Dahlgren *et al.* (1985) e Takhtajan (1991). Na atualidade, os resultados de todo esse esforço têm apenas interesse histórico (com exceção de alguma contribuição com relação à nomenclatura), devido ao enorme impacto imposto pelo uso combinado de evidências moleculares e métodos de análises filogenéticas. Os primeiros, viabilizados pela introdução de tecnologia que tornou possível o seqüenciamento de segmentos de DNA, e os últimos pelos avanços na tecnologia de computação, que culminaram com a criação de um novo sistema de classificação, assumido como metodologicamente moderno e adequado e, em conseqüência, filogeneticamente robusto.

O uso de caracteres moleculares viabilizou reconstruções filogenéticas antes impensáveis, como a comparação entre organismos de reinos distintos. O melhor exemplo é a ambiciosa construção da “árvore da vida”, baseada em seqüências de RNA ribossômico (Doolittle, 1999). Caracteres fenotípicos são escassos para comparações em níveis acima de família. Por essa razão, antes do advento da filogenia molecular, era muito difícil o estabelecimento de afinidades entre ordens, por exemplo, o que representava uma fonte de infinitas controvérsias entre os propositores de sistemas gerais de classificação de angiospermas. Os caracteres moleculares possibilitam o estabelecimento de relacionamentos consistentes entre grupos, em qualquer nível de hierarquia. O acúmulo de dados moleculares culminou na criação do sistema APG (Angiosperm Phylogeny Group). A versão mais recente é o APG III (APG 2009), baseado na combinação de seqüências de regiões de DNA nuclear e do cloroplasto. Os grupos terminais no sistema de classificação APG (APG 1998, 2003) têm nível de ordem, o que propicia a construção de um cladograma de todas as angiospermas com um número de terminais facilmente apreensível numa rápida análise e mnemonicamente conveniente.

Nas angiospermas realizam-se seqüenciamentos de regiões do DNA de três genomas: mitocondrial, nuclear e plastidial (Judd *et al.*, 1999). Pelo fato de apresentar cópias múltiplas, o *ITS* (*ITS*1 + 5.8 S + *ITS*2) é a região nuclear mais utilizada em filogenia. Para dar alguns poucos exemplos em que o *ITS* foi utilizado, pode-se citar as Euphorbiaceae (Steinmann & Porter, 2002), as Orchidaceae (Gravendeel *et al.*, 2001),

as Rubiaceae (Malcomber, 2002) e as Plantaginaceae (Albach & Chase, 2004). Outros marcadores nucleares, embora não apresentem cópias em número elevado como o *ITS*, também têm sido utilizados, como *rpb*2 (Oxelman & Bremer, 2000; Popp & Oxelman, 2001), *ncp*GS (Emshwiller & Doyle, 1999) e *PEPC* (Lohmann, 2006).

O genoma mais amplamente utilizado em filogenia de angiospermas é o do cloroplasto. Dentre as regiões transcritas pode-se citar o *rbc*L (Chase *et al.*, 1993; Olmstead & Reeves, 1995; Asmussen & Chase, 2001), o *atp*B (Hahn, 2002), o *ndh*F (Olmstead & Reeves, 1995) e o *mat*K (Gravendeel *et al.*, 2001). Regiões não transcritas, como íntrons e espaçadores, também são comumente usadas em filogenia, como *trn*L-F (que corresponde ao íntron *trn*L e o espaçador *trn*L-F) (Wallander & Albert, 2000; Salatino *et al.*, 2001; Ronsted *et al.*, 2002; Albach *et al.*, 2005), *atp*B-*rbc*L (Gonçalves, 2002), *trn*D-T e *trn*Q-*rps*16 (ambas regiões utilizadas por Hahn, 2002), *rps*16 (Albach & Chase, 2004) e *trnH-psbA* (Mello-Silva et. al., 2011).

1.3. AFLP

Uma série de técnicas e de marcadores moleculares tem sido desenvolvidos para se avaliar a diversidade genética, sendo esse estudo muito importante na sistemática, por responder a questões como, por exemplo, a relação genética entre indivíduos, a estrutura de populações, as relações filogenéticas e o mapeamento da quantidade de locos fenotipicamente característicos (QTL) (Mueller & Wolfenbarger, 1999).

Essas técnicas para a obtenção de impressões digitais gênicas de organismos, sem a necessidade de análises de seqüenciamento, síntese de iniciadores ou uso de sondas de DNA, fazem uso da amplificação de fragmentos de restrição através de PCR. Assim surgiu a técnica conhecida como RAPD ("random amplified polymorphic DNA") (Williams *et al.* 1990). Essa técnica apresenta como vantagens ter um desenvolvimento fácil, não requerer equipamentos caros e, portanto, ser relativamente barata, além de não requerer conhecimento prévio de sequências. Porém, apresentam várias desvantagens, tais como a falta de capacidade de reprodução, e a temperatura que pode influenciar muitos dos resultados.

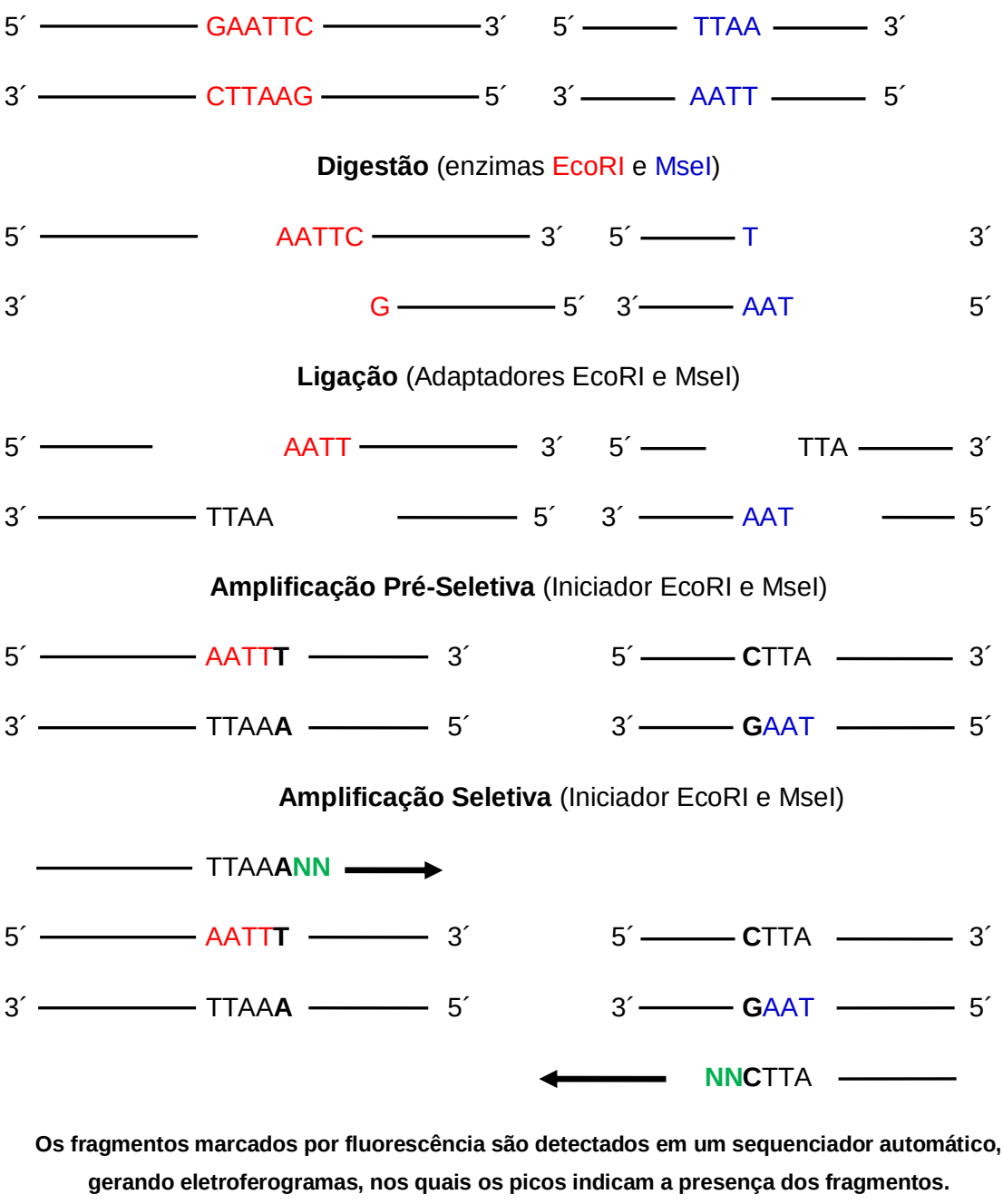
Uma técnica que aparece como ideal é a análise de polimorfismos de comprimento de fragmento amplificado (AFLP, sigla em inglês para "amplified fragment length polymorphism"), um método relativamente rápido, fácil e confiável para gerar centenas de marcadores genéticos informativos (Vos *et al.*, 1995; Vos &

Kuiper, 1997). O ponto principal da AFLP é a sua capacidade para realização de projeções simultâneas de diferentes regiões do DNA distribuídas randomicamente ao longo do genoma. Para se alcançar um alto grau de confiabilidade, o DNA genômico é preparado de tal forma a combinar a força de dois métodos, isto é, a replicação da análise dos fragmentos de restrição com o poder da PCR (Vos *et al.*, 1995). Em essência, a técnica de AFLP permite a detecção de polimorfismo de fragmentos de restrição genômicos através da amplificação por PCR. Essa técnica tem caráter universal em termos de aplicabilidade, sendo que os autores mostram resultados obtidos com análises de fragmentos genômicos de restrição de uma ampla variedade de organismos, como vírus, bactérias, fungos, espécie humana e diversos táxons de angiospermas de importância agrícola.

A técnica de AFLP compreende cinco etapas: 1) digestão do DNA com enzimas de restrição (em geral *Eco*RI e *Mse*I); 2) ligação das duas extremidades dos fragmentos a adaptadores que são oligonucleotídios com uma determinada seqüência de bases, um deles específico para o sítio de restrição da *Eco*RI e outro para *Mse*I; 3) amplificação pré-seletiva de fragmentos contendo os adaptadores, utilizando iniciadores com seqüências complementares aos adaptadores da *Eco*RI e *Mse*I, sendo que a base A (adenina) é adicionada ao iniciador *Eco*RI, assim como a base C (citosina) é adicionada ao iniciador *Mse*I, de modo a reduzir o número de fragmentos a serem amplificados nesta etapa (somente serão amplificados os fragmentos gerados cuja conformação tenha fragmentos contendo T – timina – associado à extremidade *Eco*RI e G – guanina – associado à extremidade *Mse*I); 4) amplificação seletiva, conseguida com iniciadores dotados de seqüências complementares aos adaptadores, acrescidas de três bases adicionais (duas além dos iniciadores pré-seletivos), sendo que necessariamente os fragmentos dos iniciadores *Eco*RI terão base adicional A da amplificação pré-seletiva e os fragmentos dos iniciadores *Mse*I terão base adicional C da amplificação pré-seltiva (Figura 2). Além disso, é possível para cada iniciador oito combinações possíveis, totalizando 64 combinações para a etapa de amplificação seletiva; 5) análise dos fragmentos amplificados por eletroforese.

Em sua proposta inicial, Vos *et al.* (1995) utilizaram iniciadores marcados radioativamente na amplificação seletiva. Atualmente, existe uma tecnologia disponível que prevê a utilização de iniciadores com marcador de fluorescência, semelhante ao que se usa em didesoxinucleotídios nas análises de seqüenciamento de DNA em seqüenciadores automáticos. É importante acrescentar que em AFLP as amplificações são feitas em condição de alta estringência, ou seja, temperaturas altas para acoplamento (“anelamento”) do iniciador com a região alvo. Essa condição é possível

porque os iniciadores são “perfeitos”, por terem a exata sequência complementar à sequência dos adaptadores. A alta estringência é importante para assegurar que se consiga a seletividade necessária.



Os fragmentos marcados por fluorescência são detectados em um sequenciador automático, gerando eletroferogramas, nos quais os picos indicam a presença dos fragmentos.

Figura 2. Esquema representativo das quatro etapas da técnica de AFLP.

O objetivo principal em AFLP é gerar o maior número possível de fragmentos na etapa de amplificação seletiva, com ênfase em fragmentos polimórficos. A técnica de AFLP utiliza vários recursos para que fragmentos com igual comprimento, oriundos de organismos distintos, sejam *homólogos*, isto é, provenientes da mesma região do

DNA. Os fragmentos têm um sítio *Eco*RI numa extremidade e um sítio *Mse*I na outra; há três bases adicionais às bases remanescentes dos sítios *Eco*RI e *Mse*I; as amplificações pré-seletiva e seletiva são realizadas em condições de altíssima estringência.

Inicialmente a técnica de AFLP foi desenvolvida para detecção de polimorfismos e determinação de perfis genéticos em cultivares, e usada principalmente por pesquisadores ligados à genética de plantas agrícolas e a programas de melhoramento genético, de tal maneira que se comparava os perfis eletroforéticos em nível infraespecífico. Mais recentemente têm aparecido abordagens utilizando AFLP para comparações em nível interespecífico, como a avaliação de afinidades entre espécies de *Salix* e seus híbridos (Beismann *et al.* 1997), entre espécies de *Manihot* (Roa *et al.* 1997) e até mesmo uma abordagem biossistemática e filogenética, como proposto por Kardolus *et al.* (1998), Hodkinson *et al.* (2000), Zhang *et al.* (2001), Schmidt-Lebuhn *et al.* (2005), Schmidt-Lebuhn *et al.* (2006) e Van Ee *et al.* (2006), entre outros.

Mas nem todos consideram a técnica de AFLP uma ferramenta importante para estudos de filogenia. As principais críticas a respeito da sua utilização referem-se às dificuldades para se estimar homologias e independência entre as bandas obtidas, assim como a inabilidade para identificar distinções nos sinais filogenéticos de diferentes partes do genoma (Van Ee *et al*., 2006). Por outro lado, essa técnica é importante na medida em que apresenta uma grande vantagem sobre o seqüenciamento, a de mostrar a evolução dos genes e não dos organismos. Também é interessante porque com essa técnica consegue-se analisar melhor as relações entre as espécies de um mesmo gênero, ou mesmo de populações de uma mesma espécie, por exemplo, em estudos de filogeografia.

Entretanto, os reagentes necessários para realizar-se a análise, como enzimas de restrição, adaptadores e iniciadores específicos são de custo relativamente alto, assim como o kit necessário para extração (de modo a tornar o DNA livre de impurezas e sem degradação). Outro problema que torna ainda mais cara esta técnica é a utilização de softwares específicos para AFLP.

A utilização de marcadores obtidos por AFLP é um poderoso método para o estudo de plantas relativamente próximas e em particular para variações infraespecíficas (Mueller & Wolfenbarger, 1999). O fato de os marcadores obtidos por AFLP serem geralmente distribuídos ao longo do genoma permite à técnica algumas vantagens sobre a análise de seqüenciamento para táxons relativamente próximos. O

seqüenciamento de uma única região de DNA pode proporcionar resultados enganosos por causa da hibridação (Riesberg & Soltis, 1991) e da presença de regiões parálogas de DNA (Baldwin *et al*., 1995). A técnica de AFLP limita-se a análise de espécies relativamente próximas e a táxons infraespecíficos (Karp *et al*., 1996). Acima desse nível, as impressões digitais gênicas tornam-se muito variáveis, e os marcadores obtidos serão menos propensos a serem homólogos (Mueller & Wolfenbarger, 1999).

O uso de AFLP juntamente com o seqüenciamento de regiões do DNA, principalmente *ITS*, corresponde a uma importante medida da validade daquela técnica em filogenia, pois avalia a congruência entre filogenias baseadas nos dois marcadores, refinando-as. Hodkinson *et al.* (2000), por exemplo, obtiveram excelentes resultados utilizando a técnica de AFLP em espécies de *Phyllostachys* (Poaceae). Koopman (2005) propôs um trabalho com *Lactuca* (Lactuceae, Asteraceae), no qual observou um alto grau de congruência entre os dados obtidos por seqüenciamento de *ITS* e a técnica de AFLP, sendo que esta apresentou um melhor poder de resolução. Schmidt-Lebuhn *et al.* (2005) compararam essa técnica com dados de seqüenciamento na evolução de um grupo de Acanthaceae, obtendo bons resultados na delimitação das relações entre as espécies. AFLP juntamente com seqüenciamento de regiões do DNA foram também utilizadas com sucesso em *Ilex* (Aquifoliaceae) (Gottlieb *et al*., 2005). Schmidt-Lebuhn *et al.* (2006) abordaram a evolução de *Polylepis* (Rosaceae, Sanguisorbeae) através do uso de AFLP e dados morfológicos do grupo. Outro estudo bem sucedido com esse enfoque foi realizado por Van Ee *et al.* (2006), com a filogenia e biogeografia de *Croton alabamensis* (Euphorbiaceae).

Existem alguns cuidados a serem tomados para a utilização dessa técnica, tais como a observação de que os marcadores AFLP são de evolução muito rápida, e, portanto, não se prestam a comparações entre organismos filogeneticamente distantes.

1.4. Posicionamento de *Croton* em Euphorbiaceae s.l.

O APG de 1998 claramente indicou que Euphorbiaceae s.l. não é um grupo monofilético. A família deve, portanto, ser separada em, pelo menos, cinco linhagens, na Ordem Malpighiales: Pandaceae; Putranjivaceae, Phyllanthaceae e Picrodendraceae, três grupos biovulados; e Euphorbiaceae, um grupo uniovulado, com as subfamílias Acalyphoideae, Crotonoideae e Euphorbioideae.

Croton está inserido na subfamília Crotonoideae de Euphorbiaceae, que possui cerca de 2400 espécies, distribuídas em 67 gêneros e 12 tribos (Webster, 1994). *Croton* possui a maior parte das espécies da subfamília, seguida por *Jatropha* (aproximadamente 185 espécies) e *Manihot* (aproximadamente 100 espécies). O Brasil aparece como o país que congrega o maior número de espécies (Allem, 1979), com aproximadamente 356, agrupadas em 26 das 40 seções propostas para o gênero por Webster (1993). Dessas, 172 espécies correspondentes a 21 seções, são encontradas nos quatro estados da região Sudeste, sendo, provavelmente essa região o centro de diversidade do gênero (Lima, 2006).

Uma sinapomorfia dessa subfamília é a presença de uma rara configuração do pólen: uma arquitetura particular da exina, com subunidades triangulares supratectais em uma malha de muri, e uma curta e irregular columela (Punt, 1961; Nowicke, 1994).

Diversos novos táxons de *Croton* têm sido propostos para o Brasil desde a revisão de Mueller (1873), geralmente em trabalhos esparsos, o que tem tornado ainda mais confusa a taxonomia do gênero, dificultando especialmente sua identificação. Alguns tratamentos para táxons na América do Sul, como o de Croizat (1944), ou floras regionais como os trabalhos de Lucena (2000), Carneiro *et al.* (2002), Secco (1992, 2005) e Guimarães (2006) ainda são insuficientes para que se possa avaliar precisamente os limites específicos de *Croton*. Além disso, as espécies brasileiras do gênero possuem distribuição restrita e grande diversidade morfológica intraespecífica (Lucena, 2000), o que dificulta suas delimitações, além de existirem problemas de nomenclatura e tipificação (Bentham, 1880; Lucena, 2000; Guimarães, 2006).

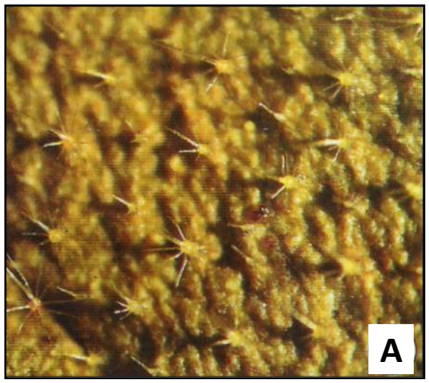
Um dos grandes estudiosos deste gênero, Paul Berry, publicou vários trabalhos, juntamente com outros pesquisadores, como Benjamin Van Ee e Ricarda Riina. O objetivo destes pesquisadores é classificar todas as espécies do gênero. O trabalho mais recente publicado a respeito trata do estudo de espécies pertencentes ao Novo Mundo, em que 712 espécies são classificadas em seções e subseções, além de um estudo de filogenia molecular com um marcador nuclear e um mitocondrial. Nesse recente artigo há também descrições morfológicas e chave de identificação de espécies de 31 seções e 10 subseções reconhecidas para o Novo Mundo (Van Ee *et al.*, 2011).

1.5. *Croton pallidulus* Baill.

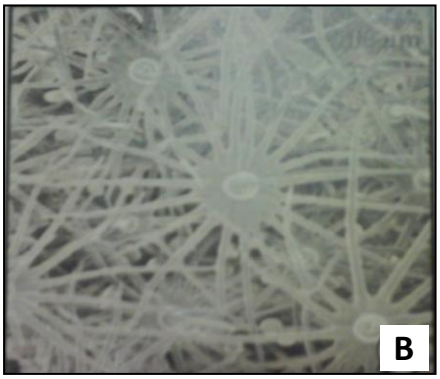
Croton pallidulus Baill. está inserida na seção *Lamprocroton*, inicialmente proposta por Mueller (1873) na *Flora brasiliensis*. Posteriormente, Webster (1993) publicou uma proposta de classificação infragenérica para *Croton* que, apesar de provisória, vem sendo utilizada nos trabalhos referentes ao gênero. De acordo com ela, *Lamprocroton* é composta por espécies predominantemente sul-americanas que incluem não somente aquelas com estiletes bífidos da série *Lamprocroton* de Mueller (1873), mas também as com estiletes bífidos, folhas sem glândulas e indumento (tricoma) lepidoto colocadas por Mueller (1873) na série *Argyrocroton*.

A distribuição dessa espécie no Brasil é de certo modo restrita, ocorrendo principalmente na região Sul do país. É bastante freqüente no Paraná onde ocorre em diversos municípios. Em São Paulo, foi coletada apenas no município de Campos do Jordão, e apenas uma vez no município de Itararé e na capital. Ocorre também em Minas Gerais, principalmente naqueles municípios próximos aos limites com o estado de São Paulo. Há somente uma coleção proveniente do Rio de Janeiro, contudo a ausência de dados na etiqueta da exsicata impossibilita saber em que município a planta teria sido coletada. Pode ser encontrada nos mais diversos tipos de ambientes como afloramentos rochosos, campo arenoso, cerrado pedregoso, campo úmido, campo sujo, capão de mata, capoeira, campo com palmeiras do gênero *Butia* Becc. (“butiazal”), e em borda de mata de galeria. Trata-se também de uma planta invasora, freqüente em regiões de campo degradado (Lima, 2006).

Essa espécie apresenta variedades: *Croton pallidulus* Baill. var. *pallidulus* (também conhecido popularmente como “velame-do-campo”) e *Croton pallidulus* var. *glabrus* L.R. Lima (Lima & Pirani, 2008). A diferenciação entre elas ocorre especialmente pela ausência, ou quase ausência, de indumento na face adaxial das folhas de *Croton pallidulus* var. *glabrus* L.R. Lima (Figura 3).



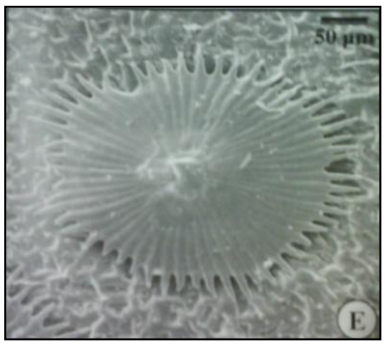
(Lima, 2006)



(Lima, 2006)

Figura 3. Diversidade de tricomas da espécie *Croton pallidulus*: var. *pallidulus* (**A**) apresentado na face adaxial, e a var. *glabrus* (**B**) apresentado somente na face abaxial.

Smith *et al.* (1988) já haviam proposto duas variedades para *C. pallidulus* baseadas principalmente na forma das folhas e no indumento da face adaxial desse órgão. Segundo esse trabalho, *Croton pallidulus* var. *pallidulus* caracterizava-se por possuir folhas elípticas até oblongo-ovais, muitas vezes mais alargadas próximo à base e com a face adaxial coberta por tricomas estrelados; *Croton pallidulus* var. *myrianthus* (Müll. Arg.) L.B. Smith & S.F. Smith, possuia folhas mais estreitas, atenuadas e freqüentemente mais largas no centro, com face adaxial, em geral, glabra. Entretanto, segundo Lima & Pirani (2008), esta última tem sido tratada como *Croton myrianthus* Müll. Arg., uma vez que apresenta caracteres suficientemente consistentes para ser mantida como um outro táxon, tais como tipo de tricomas das duas faces foliares (lepidoto como observado na Figura 4), tipo de inflorescência (em racemos unissexuados ao invés dos bissexuados presentes em *C. pallidulus*), ambiente em que ocorre, entre outros.



(Lima, 2006)

Figura 4. Detalhe do tricoma lepidoto presente nas faces adaxial e abaxial de *C. myrianthus*.

Croton pallidulus é extremamente semelhante a *C. ceanothifolius* Müll. Arg. quanto ao hábito, forma e tamanho das folhas, tipo de nervação, e inflorescência (Figura 5). Porém, enquanto na primeira, a face adaxial das folhas é completamente glabra ou com indumento formado por tricomas estrelados, na segunda, essa região

da planta é sempre coberta por tricomas simples (Figura 6). Tal característica pode ser utilizada, de maneira eficaz na distinção entre as duas espécies. De acordo com Lima (2006), apesar da maioria das coleções examinadas de *Croton pallidulus* var. *pallidulus* apresentar apenas tricomas estrelados constituindo o indumento da face adaxial das folhas, em alguns espécimes existem, entre os tricomas estrelados, tricomas simples. Em outros casos, o raio porrecto dos tricomas estrelados é muito longo em relação aos laterais, o que pode levar a uma falsa conclusão de que se trata de um tricoma simples (Lima, 2006).

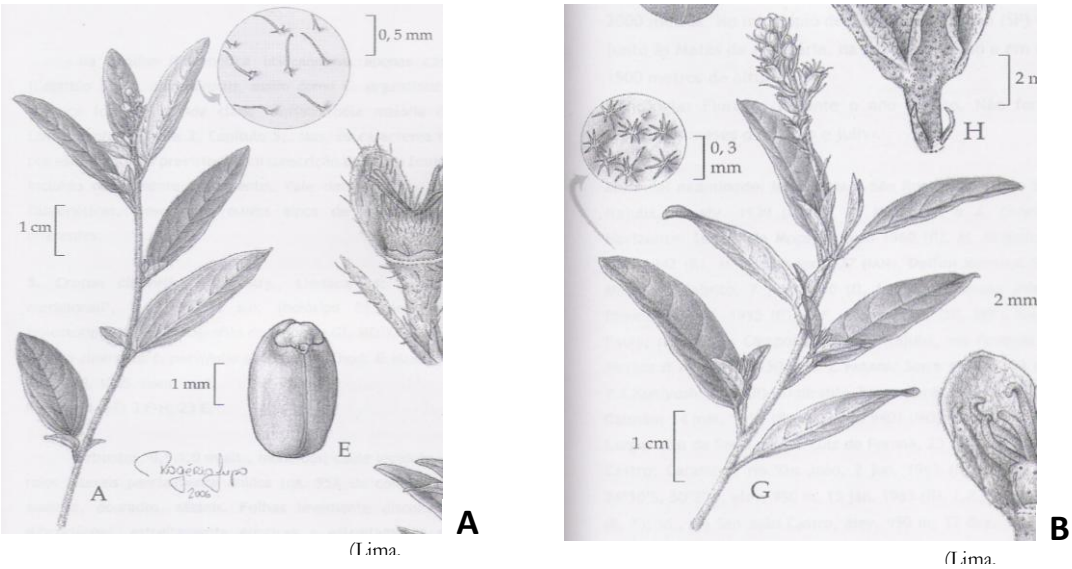


Figura 5. Hábito e detalhe do indumento da face adaxial das folhas de A = *Croton ceanothifolius*; B = *Croton pallidulus* var. *pallidulus*.

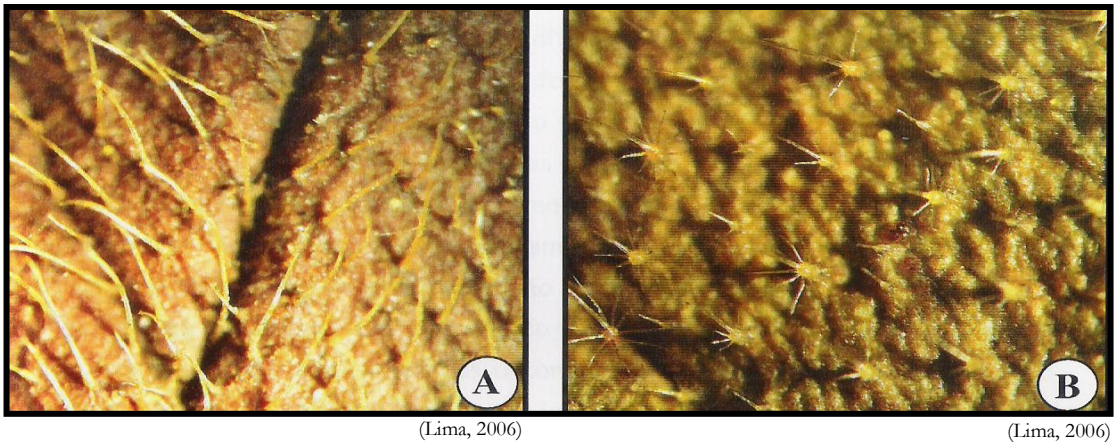


Figura 6. Detalhe dos tricomas na face adaxial vistos sob estereomicroscópio: simples em *C. ceanothifolius* (**A**) e estrelado em *C. pallidulus* var. *pallidulus* (**B**).

Croton pallidulus é um táxon que sempre surge como um representante da seção *Lamprocroton* em todas as análises filogenéticas; porém, as relações de parentesco entre as duas variedades da espécie diferem dependendo da análise. No

trabalho de Lima (2006), numa análise feita com o programa PAUP, somente com dados moleculares, *Croton pallidulus* var. *pallidulus* emerge como grupo irmão de *C. dichrous* Müll. Arg., em um clado distinto daquele onde está *Croton pallidulus* var. *glabrus*, o que poderia indicar a natureza distinta dessas duas variedades. Já na análise de evidência total utilizando-se o programa POY, esta mesma variedade tipo emerge na árvore, tendo como grupo irmão *Croton pallidulus* var. *glabrus*, em um clado no qual estão também *C. ceanothifolius* e *C. dusenii* Croizat. Nas demais análises as duas variedades emergem no mesmo clado, mas em politomias, não sendo possível estabelecer a relação de parentesco entre elas.

C. *ceanothifolius*, além dos motivos apresentados acima, é importante representativamente para a seção *Lamprocroton* por ser reconhecida como a espécie tipo da seção na *flora brasiliensis* (Mueller, 1873); é endêmica do estado de Minas Gerais (Lima & Pirani, 2008). *C. erythroxyloides* apresenta-se como uma das espécies mais polimórficas da seção, tendo sido subdividida em variedades desde o momento da sua descrição, primeiro por Baillon (1864) e em seguida por Mueller (1866); ocorre em regiões com altitudes superiores aos 1000 metros. *C. splendidus* também está entre as espécies da seção *Lamprocroton* mais fáceis de serem reconhecidas e pode ser identificada, mesmo em estado vegetativo, por meio das folhas fortemente discolores, densamente distribuídas nos ramos, com nervura primária bastante marcada e face adaxial das folhas coberta por indumento tomentoso, formado por tricomas estrelados, com os raios laterais totalmente livres (Lima & Pirani, 2008); desenvolve-se, preferencialmente, como espécie de altitude, entre 1000 e 1600 m, podendo ocorrer também ao nível do mar, nas restingas, mas segundo Smith *et al.* (1988), nesse tipo de ambiente e também nos campos litorâneos, ocorre esporadicamente como elemento raro e estranho.

Croton myrianthus é uma espécie bastante semelhante a *C. erythroxyloides* quanto ao hábito e tipo de indumento sendo, muitas vezes, difícil distinguir esses dois táxons vegetativamente, já que em ambos, as faces foliares apresentam indumento formado por tricomas tipicamente lepidotos (Lima & Pirani, 2008). Ocorre somente na região Sul do Brasil, nos estados do Rio Grande do Sul, Santa Catarina e Paraná.

Croton dichrous se distingui das demais espécies da seção devido, principalmente, às folhas fortemente discolores com a face adaxial, em geral, completamente glabra. Ocorre somente na região Sudeste do Brasil, nos estados de São Paulo, Rio de Janeiro e Minas Gerais, sendo sempre associada a localidades com

altitudes elevadas, superior aos 1.000 m, podendo ser encontrada em regiões com elevações superiores aos 2.500 m de altura (Lima & Pirani, 2008).

1.6. Biogeografia

Os estudos biogeográficos em sua fase inicial foram realizados buscando entender e estudar melhor a distribuição geográfica dos organismos no presente e no passado. Trata-se de uma ciência que procura explicar os padrões de distribuição de espécies e grupos taxonômicos superiores.

Um dos primeiros estudos, e que ainda hoje é bastante utilizado para se tentar entender estes padrões, divide o globo terrestre em regiões biogeográficas classificadas de acordo com as afinidades de flora e fauna territoriais. Essas regiões receberam os nomes de Paleártica, Neártica, Neotropical, Etiópica, Oriental, Australiana e Antártica. Foram observados, posteriormente, no trabalho de Takhtajan (1986), reinos florísticos, que poderiam ser divididos similarmente à classificação mencionada anteriormente, em Holártica (compreendendo as regiões biogeográficas Paleártica e Neártica), Neotropical, Paleotropical (região Etiópica e Oriental), do Cabo (região da África do Sul), Australiana e Antártica.

Essas diferentes regiões (ou reinos) são assim classificadas, de modo a contemplar grandes áreas, em que há congruência de dados referentes a espécies semelhantes que ocorrem, seja na flora e/ou na fauna. Há para cada uma delas, outras subdivisões de acordo com o tamanho da área a ser estudada.

Os eventos que podem desencadear uma especiação, ou seja, uma diferenciação de uma espécie ancestral em duas ou mais espécies posteriores, podem ser do tipo extinção, dispersão e vicariância. A primeira é autoexplicativa, pois uma espécie pode, após sofrer diferenciação, ser extinta, de acordo com o sucesso ecológico obtido pela espécie que a sucedeu. Já a segunda e a terceira podem ser explicadas pela existência de uma barreira, cuja espécie ancestral consegue ultrapassar e, posteriormente, sofrer diferenciação nas populações do novo habitat conquistado, gerando especiação. No caso da dispersão, essa barreira separa a espécie ancestral, ocasionando ocupação de áreas distantes. Na vicariância, a barreira “interrompe” a área habitada pelo ancestral pelo surgimento de um rio, por exemplo, ocasionando a separação das populações da espécie ancestral; com o passar do tempo e através de mutações cumulativas ao acaso, isso pode ocasionar

uma relativa distância genética, de modo a haver um isolamento reprodutivo em relação à espécie ancestral, ocorrendo então, a especiação.

A especiação pode ser dividida em três diferentes categorias:

1) Alopátrica: a população ancestral divide-se em duas populações geograficamente isoladas, devido, por exemplo, à fragmentação do hábitat, pelo aparecimento de uma cadeia montanhosa. Dessa maneira, as populações vão se diferenciar genotípica e/ou fenotipicamente. Após certo tempo, quando as barreiras do contato entre elas não mais existirem, os indivíduos que faziam parte de uma mesma espécie ancestral voltam a se encontrar, porém agora estarão reprodutivamente isolados e já não serão mais capazes de trocar genes entre si (Sadava *et al*., 2009).

2) Simpátrica: nesse caso, as populações divergem quando ainda ocupam a mesma área, sendo a poliploidia, o mecanismo mais comum que permite a especiação simpátrica. Esse fenômeno ocorre geralmente quando há ocorrência de autopoliploidia, na qual o organismo ancestral diplóide gera gametas também diplóides, e consequentemente, com a união desses gametas, indivíduos tetraplóides são gerados. Dessa maneira, as duas populações de indivíduos, isto é, aquela dos ancestrais diplóides, e aquela de indivíduos tetraploides serão separadas. Se houver um cruzamento entre as duas populações, será gerada uma prole estéril, pois terá um número de cromossomos incorreto, e, portanto, não viável.

3) Parapátrica: neste caso, não há separação geográfica completa entre as duas populações isoladas, implicando na ocorrência de algum fluxo gênico. Indivíduos das duas populações podem entrar em contato entre si, ou mesmo atravessar a barreira de tempos em tempos. Os híbridos, que porventura sejam formados, têm uma viabilidade reduzida, levando eventualmente ao reforço das barreiras à reprodução.

De acordo com a classificação da vegetação brasileira segundo o IBGE, adaptado de Veloso *et al.* (1991), existem 12 formações vegetais no Brasil, sendo que as espécies do Complexo que serão analisadas neste estudo estão predominantemente ora em áreas de vegetação de floresta ombrófila densa (Mata Atlântica) ora floresta ombrófila mista (Mata de Araucária).

2. OBJETIVOS

Uma vez que os estudos referentes à morfologia (Lima, 2006) não apresentaram evidências que sustentam a diferenciação das duas variedades, assim como o posicionamento dessa espécie dentro da seção Lamprocroton, este projeto tem como objetivo reconstruir as relações filogenéticas de *Croton pallidulus* Baill. e de táxons próximos, *C. ceanothifolius*, *C. dichrous*, *C. myrianthus* e *C. splendidus,* inseridos no "Complexo *Croton pallidulus*", além de *C. erythroxyloides*, utilizando seqüenciamento de DNA (região nuclear *ITS*) e dados obtidos através da técnica de AFLP, para avaliar a estrutura das populações e a diversidade genética das espécies. Com essas abordagens se pretende ter um quadro mais claro da delimitação das variedades em *Croton pallidulus*, bem como entre as diferentes espécies, avaliando, dessa maneira, as relações observadas nos estudos morfológicos. Finalmente, pretende-se avaliar a congruência entre a topologia da árvore obtida pelo seqüenciamento de *ITS* e aquela obtida através do dendrograma de AFLP.

3. MATERIAIS E MÉTODOS

3.1. Materiais

Diversas regiões de vários estados do Brasil foram visitadas para coletas de material suficiente para cobrir a distribuição do Complexo *Croton pallidulus.* Os nomes das espécies, bem como locais e números de coleta, encontram-se listados na Tabela 1. Foi utilizado como grupo externo, *Croton urucurana* Baill., uma espécie da seção *Cyclostigma*, morfologicamente diferente dos representantes de *Lamprocroton* na qual se encontram as espécies do complexo *Croton palidulus*. Exsicatas do material coletado estão depositadas nos Herbários do Departamento de Botânica - IBUSP (SPF), do Instituto de Botânica (SP) e do Museu Botânico Municipal de Curitiba (MBM).

Tabela 1. Espécies de *Croton* utilizadas para análises moleculares, com os respectivos locais e datas de coleta.

Espécie	Testemunho	Local e data de coleta
Croton ceanothifolius Müll. Arg.	Cordeiro 3153	Ouro Preto – MG – 11/2009
C. dichrous Müll. Arg. 1	Cordeiro 3074	Campos do Jordão – SP – 03/2009
C. dichrous Müll. Arg. 2	Cordeiro 3123	Pindamonhangaba – SP – 11/2009
C. erythroxyloides Baill.	Cordeiro 3216	Caldas – MG – 11/2009
C. aff. erythroxyloides	Cordeiro 3154	Ouro Preto – MG – 11/2009
C. myrianthus Müll. Arg.	Lorscheitter s.n. (SP 425771)	São Francisco Paula – RS – 10/2010
C. pallidulus Baill. *var. pallidulus 1*	Cordeiro 3072	Campos do Jordão – SP – 03/2009
C. pallidulus Baill. *var. pallidulus 2*	Cordeiro 3077	São José dos Pinhais – PR – 04/2009
C. pallidulus Baill. *var. pallidulus 3*	Cordeiro 3078	Tijucas do Sul – PR – 04/2009
*C. pallidulus var. glabrus*1 L.R.Lima	Suzaki 01	Curitiba – PR – 01/2011
C. pallidulus var. glabrus2 L.R.Lima*	Barbosa 3045	Garuva – SC – 02/2011
C. splendidus Mart. 1	Cordeiro 3079	Tijucas do Sul – PR – 04/2009
C. splendidus Mart. 2	Cordeiro 3149	Ouro Preto – MG – 11/2009
C. urucurana Spreng.	Cordeiro 3141	Mogi-Guaçu – SP – 11/2009

A espécie correspondente à coleção Cordeiro 3154 foi identificada por meio da chave de identificação de Lima & Pirani (2008) como *C. erythroxyloides* pela Dra Inês Cordeiro, que auxiliou este trabalho com a coleta e identificação dos materiais. Entretanto, devido a diferenças morfológicas entre este material e outros materiais anteriormente depositados no Herbário do IBt, esse nome foi modificado para *C. aff. erythroxyloides*, nome este que será utilizado neste trabalho.

A espécie correspondente à coleção Barbosa 3045 e grafada com asterisco (*) foi coletada em Santa Catarina e nos enviada como *C. dusenii* Croizat. Posteriormente, a Dra Inês examinando o material verificou tratar-se de *C. pallidulus* var. *glabrus*.

3.2. Métodos

3.2.1. Extração do DNA para as análises moleculares

O DNA total foi primeiramente extraído das folhas armazenadas em gel de sílica com indicador de umidade, usando-se tampão CTAB, segundo protocolo de Ferreira & Grattapaglia (1996), adaptado de Doyle & Doyle (1987). A qualidade do

DNA obtido e sua quantidade relativa foram avaliadas por eletroforese em gel de agarose 1% e a visualização foi feita com auxílio de “GelRed nucleic acid gel stain” 10.000X em DMSO, em transiluminador de luz UV. Notou-se que o DNA extraído com esse protocolo era de péssima qualidade, não sendo útil para análise por AFLP. Portanto, o DNA total foi novamente extraído utilizando-se o DNeasy Plant Mini Kit da QIAGEN. A qualidade do DNA obtido e sua quantidade relativa foram avaliadas novamente, da mesma maneira que anteriormente, verificando-se que a nova extração produziu um DNA de muito melhor qualidade.

3.2.2. Seqüenciamento de DNA

A amplificação da região *ITS* (nuclear) foi feita de acordo com os procedimentos de Berry *et al*. (2005). Posteriormente, foram realizadas as amplificações de duas outras regiões do cloroplasto: *trn*L-F e *rps*16.

Para a amplificação da região *ITS*, foram empregados os iniciadores LEU1 (GTC CAC TGA ACC TTA TCA TTT AG) e ITS4 (TCC TCC GCT TAT TGA TAT GC) (Urbatsch *et al.,* 2000; White *et al.,* 1990) em soluções de reação de 50 μL contendo 10 μL de tampão *Taq*, 1 μL de albumina de soro bovino (BSA), 2,5 μL de dimetil-sulfóxido (DMSO), 2,5 μL de dNTPs na concentração de 0,20 μM cada, 0,5 μL de iniciadores, 0,2 μL de *Taq* DNA polimerase e 30,8 μL de água Miliq, além de 2 μL DNA genômico total. A amplificação foi feita com a seguinte programação: 94°C, 5 min; 40 ciclos, 94°C-30 seg, 50°C-1 min, 72°C-1,30 min; 72°C-7 min e 15°C, por tempo indeterminado.

Para amplificação da região *trn*L-F, foram empregados os inciadores tabC (CGA AAT CGG TAG ACG CTA CG) e tabF (ATT TGA ACT GGT GAC ACG AG) (Taberlet *et al.*, 1991), e para *rps*16 foram utilizados os inciadores rpsR2 (TCG GGA TCG AAC ATC AAT TGC AAC) e rpsF (GTG GTA GAA AGC AAC GTG CGA CTT) (Oxelman *et. al.*, 1997), em soluções de 50 μL, contendo 10 μL de tampão *Taq*, 3 μL de PVP (polivinilpirrolidona), 3 μL de dNTPs na concentração de 0,20 μM cada, 0,5 μL de iniciadores, 0,25 μL de *Taq* DNA polimerase e 30,75 μL de água Milli-Q, além de 2 μL DNA genômico total. A amplificação foi feita com a seguinte programação: 95°C-4 min; 40 ciclos, 95°C-1min, 52°C-1min e 30seg, 72°C-2min; 72°C-7 min e 15°C por tempo indeterminado.

A qualidade do material amplificado obtido foi avaliada por eletroforese em gel de agarose 1% com observação em transiluminador, em presença de “GelRed”. A

purificação do material satisfatoriamente amplificado foi feita com o Kit “Concert Rapid PCR Purification System – Gibco BRL“, de acordo com protocolo recomendado pelo fabricante. Após a purificação, iniciou-se o processo de seqüenciamento, cujas amplificações foram feitas com os mesmos iniciadores, adicionando-se 2 μL de BigDye Terminator (ABI Systems) versão 3.0, 0,3 μL de iniciador a 10 nM, com a seguinte programação: 96°C-1 min; 40 ciclos, 96°C-10 seg, 52°C-20 seg, 60°C-4 min; e 10°C por tempo indeterminado. Para um volume final de 10 μL foi calculado uma quantidade de amostra de DNA correspondente à sua qualidade (geralmente, 5 μL das amostras foi suficiente), completando-se o volume com água Milli-Q.

Após a reação de amplificação para sequenciamento, o DNA foi precipitado com 25 μL de etanol 100%, 1μL de EDTA 125 nM pH 8,0 e 1μL de acetato de sódio 3M em uma placa de sequenciamento, por no mínimo 15 min, à temperatura ambiente (no escuro), sendo, posteriormente, centrifugada a 4000 rpm por 30 min. O sobrenadante foi descartado, sem perder o “pellet” e adicionado 35 μL de etanol 70%, centrifugando-se a 2300rpm por 15 min. O sobrenadante foi novamente descartado, sendo o processo repetido com etanol 70%; posteriormente, toda a água e etanol do resíduo foram eliminados em “speed vaccum”, por 5 min. O seqüenciamento foi conduzido nos sequenciadores automáticos ABI Prism 3100 e ABI Prism 3730 DNA (Applied Biosystems), do Laboratório de Biologia Molecular do Departamento de Botânica do Instituto de Biociências da USP.

As seqüências direta e reversa foram complementadas e posteriormente alinhadas em computador PC, através dos programas BioEdit e ClustalX. Para as inferências filogenéticas, os caracteres foram ‘não ordenados e sem pesagem’.

Foram construídos cladogramas pelo critério da máxima parcimônia com o programa PAUP 4.0b (Swofford, 2002), usando-se busca heurística e avaliação da consistência dos clados pelo método “bootstrap” (Felsenstein, 1985), utilizando-se 1000 réplicas para maior confiabilidade nos resultados (ver Anexo 1). Para dois dos marcadores utilizados (*ITS* e *trn*L-F) havia dados disponíveis no GenBank com espécies de *Croton*, coletados em território brasileiro, dos trabalhos de Berry *et. al* (2005) e de Riina *et. al* (2009). Posteriormente, foi realizada outra análise a fim de ampliar a confiabilidade da inferência filogenética, segundo o método bayesiano, baseado em probabilidades posteriores, com um milhão de gerações, por meio do programa MrBayes (Hall, 2001) (ver Anexo 2).

Os dados obtidos pelos diversos marcadores foram trabalhados isoladamente e depois reunidos de acordo com a viabilidade, a fim de inferir uma análise mais robusta. A avaliação do grau de congruência entre as topologias obtidas foi feita usando-se o teste ILD (Mason-Gamer & Kellogg 1996). Por este teste, as hipóteses iniciais são rejeitadas quando $p \leq 0,01$ (ver Anexo 3).

3.2.3. Análise por AFLP

3.2.3.1. Quantificação do DNA

As mesmas amostras usadas para análises de sequenciamento foram usadas para análises por AFLP. Foi feita uma estimativa de quantificação visual (adaptação de Sambrook *et al*. 1989), em transiluminador, utilizando-se para isso tampão TE (1µL), corante "GelRed" (2µL) e amostra de DNA (1µL). A quantificação visual foi feita, comparando-se as intensidades das fluorescências das amostras com as de uma curva padrão preparada a partir do DNA de fago λ que apresenta uma concentração já definida.

3.2.3.2. Reação de restrição (digestão) do DNA

Todos os procedimentos envolvendo reações de restrição, ligação e amplificação seguiram as instruções e adaptação do protocolo da Applied Biosystems "AFLP Plant Mapping", com pequenas modificações.

Para se determinar a eficácia do processo foi realizada a digestão de amostras do DNA com as enzimas de restrição *Eco*RI e *Mse*I e com a mistura de ambas. Após essa etapa, fez-se a análise em um gel de agarose 1,5%. A visualização foi feita com auxílio de GelRed nucleic acid gel stain 10.000X em DMSO, em transiluminador de luz UV.

A digestão dos DNAs foi realizada em tubos de 0,2 mL com os seguintes reagentes: 2 µL de tampão T4 DNA ligase 5X, 0,25 µL da enzima *Mse*I (5 U/µL) e 0,50 µL da enzima *Eco*RI(10 U/µL). Para um volume final de 10 µL foi calculado uma quantidade de amostra de DNA correspondente a 250 ng, sendo completado o volume final com água Milli-Q. A mistura foi colocada em termociclador por 2 h à 37ºC. As reações de digestão foram então armazenadas em freezer -20ºC até a próxima etapa.

3.2.3.3. Reação de ligação

Parte da amostra digerida (5,5 μL) foi submetida à reação de ligação com 2 μL de tampão T4 DNA ligase 5X, 1 μL de NaCl 0,5 M, 0,5 μL de BSA (1 mg/mL), 1 μL do adaptador de *Mse*I, 1 μL do adaptador de *Eco*RI e 1 μL da enzima T4 DNA ligase. A mistura foi submetida à seguinte programação do termociclador: 2 h à 37ºC. A reação posteriormente foi armazenada em freezer -20º C até a próxima etapa.

3.2.3.4. Reação de amplificação pré-seletiva

Após a ligação dos adaptadores, as amostras foram submetidas à reação de amplificação pré-seletiva. Em tubos de 0,2 mL foram misturados 2 μL do produto de ligação, 0,5 μL dos iniciadores específicos para as extremidades coesivas resultantes da digestão por *Eco*RI e *Mse*I (Applied Biosystems) e 7,5 μL da solução de "core mix" (Applied Biosystems). A reação então foi levada ao termociclador com a seguinte programação: 2 min. à 72ºC; 20 ciclos de 94ºC-20s, 56ºC-30s.; 72ºC-2 min.; 60ºC por 30 min. e 4ºC∞. A qualidade da reação foi verificada em gel de agarose 1,5% em tampão TBE 1X. As amostras com resultado positivo foram armazenadas em freezer para a etapa posterior.

3.2.3.5. Reação de amplificação seletiva

A reação de amplificação seletiva foi realizada com o produto da reação pré-seletiva diluída; dependendo da intensidade observada na eletroforese, as diluições variaram de 1:20 a 1:30. Em tubos de 0,2 mL foram misturados: 7,5 μL de "core mix", 0,5 μL de iniciador seletivo *Mse*I a 5 μM, 0,5 μL de iniciador seletivo *Eco*RI a 1 μM e 1,5 μL do produto de reação pré-seletiva diluído. A mistura foi levada ao termociclador seguindo a programação: 94ºC por 2 min.; 10 ciclos de 94ºC-20s, 66ºC-30s, reduzindo 1ºC/ciclo, 72ºC-2 min.; e 20 ciclos de 94ºC-20s, 56ºC-30s, 72ºC-2 min.; 60ºC-30 min; 10ºC∞.

Foram testadas várias combinações dos iniciadores seletivos *Mse*I e *Eco*RI, mas, muitas delas não se mostraram satisfatórias. Sendo assim, apenas duas combinações foram utilizadas: AAG – CAG e AGC – CTA. Foi realizada uma análise com mais uma combinação: ACA - CTG, a fim de aumentar o número de fragmentos polimórficos e, dessa maneira, tornar os resultados mais consistentes, porém nem todos os indivíduos mostraram fragmentos que pudessem ser utilizados.

Os fragmentos amplificados seletivamente foram analisados num sequenciador automático ABI Prism mod. 310, que trabalha com capilar de sílica fundida. A 2 μL da solução obtida na amplificação seletiva foram adicionados 0,2 μL da mistura de padrões internos para determinação dos comprimentos dos fragmentos (GeneScan-500 "Rox", contendo fragmentos com 35, 50, 75, 100, 139, 150, 160, 200, 250, 300, 340, 350, 400, 450, 490 e 500 pares de bases, marcados com indicador de fluorescência) e 11 μL de formamida. As misturas foram preparadas em tubos especiais para o seqüenciador, desnaturadas durante 5 min a 95°C, e então levadas ao seqüenciador automático.

3.2.3.6. Análise dos resultados

A análise dos tamanhos dos fragmentos foi feita com auxílio do programa Gene Mapper Software 4.1 (Applied Biosystems, 2009) que permite a determinação dos comprimentos dos fragmentos amplificados por comparação com os fragmentos componentes da mistura "GeneScan-500 Rox". Além disso, o programa possibilita o confronto entre os eletroferogramas de todas as amostras analisadas gerando um resultado preliminar de número de fragmentos obtidos e, conseqüentemente, revelando as melhores combinações a serem utilizadas para todos os terminais analisados.

Foram rejeitados os picos com intensidades abaixo de 100 e comprimentos inferiores a 75 pb e superiores a 500 pb. Nesse processo, foram eliminados da análise os fragmentos comuns a todas as amostras, gerando-se uma tabela de dados binários (presença e ausência) para os fragmentos polimórficos entre as amostras analisadas.

Foi realizada uma análise de evolução mínima, através do método de distância mínima (Neighbor Joining) de Saitou & Nei (1987), cujo princípio é identificar os pares de vizinhos mais próximos de forma a minimizar a longitude total da árvore. A medida de distância genética foi a de Nei & Li (1979). O programa utilizado nessas análises foi o PAUP versão 4.0 (Swofford, 2002), sendo que a consistência da monofilia dos ramos foi avaliada através da replicação "bootstrap", através de mil réplicas (Felsenstein, 1985).

3.3. Biogeografia do Complexo

Foram utilizados os mapas de Lima & Pirani (2008) das espécies analisadas neste trabalho. Esse cuidado foi tomado para que a localização das espécies fosse corretamente expressa. Por não pertencer à seção *Lamprocroton*, não se obteve o mapa de distribuição de *Croton urucurana*, usada como grupo externo.

Foi utilizado o programa DIVA, o qual estima a distribuição de ancestrais hipotéticos em nós internos de uma árvore completamente bifurcada baseando-se na distribuição dos nós terminais (Ronquist, 1997). Dessa maneira é possível utilizar o programa para inferir a biogeografia de um táxon mesmo quando um cladograma geral de áreas não está disponível. Resumidamente, o programa minimiza extinções e dispersões e assume vicariância como a hipótese nula (Ronquist, 1997). As incertezas que aparecem nos resultados de DIVA são provenientes de duas fontes: incertezas filogenéticas (por exemplo, politomias), e incertezas na otimização do DIVA (por exemplo como interpretar áreas ancestrais diferentes que foram reconstruídas como igualmente parcimoniosas). Portanto, torna-se necessário a utilização de árvores com boa resolução, ou seja, sem politomias, para que haja uma diminuição nas incertezas dos resultados a serem obtidos através do programa.

Para o estudo biogeográfico comparou-se os dados obtidos em filogenia molecular com os dados de localização das espécies. A reconstrução da história biogeográfica do Complexo *Croton pallidulus* foi feita por meio do programa RASP – *Reconstruct Ancestral State in Phylogenies* (Yu *et al.* 2011) sob o nome genérico de *Statistical Dispersal and Vicariance Analysis* (S-DIVA) (Yu *et al.* 2010) ou “Análise de Dispersão e Vicariância Estatística”, um programa em que está implementada cada alteração proposta ao DIVA original. Além disso, sua interface gráfica é de fácil utilização, está disponível para qualquer plataforma (Windows, Mac ou Linux) e é gratuito.

4. RESULTADOS E DISCUSSÃO

4.1. Sequenciamento de DNA

Foi realizada a extração do DNA de todas as amostras coletadas. A Fig. 7 mostra exemplos de análises por eletroforese em gel de agarose dos resultados referentes à extração, utilizando o método CTAB das amostras de DNA.

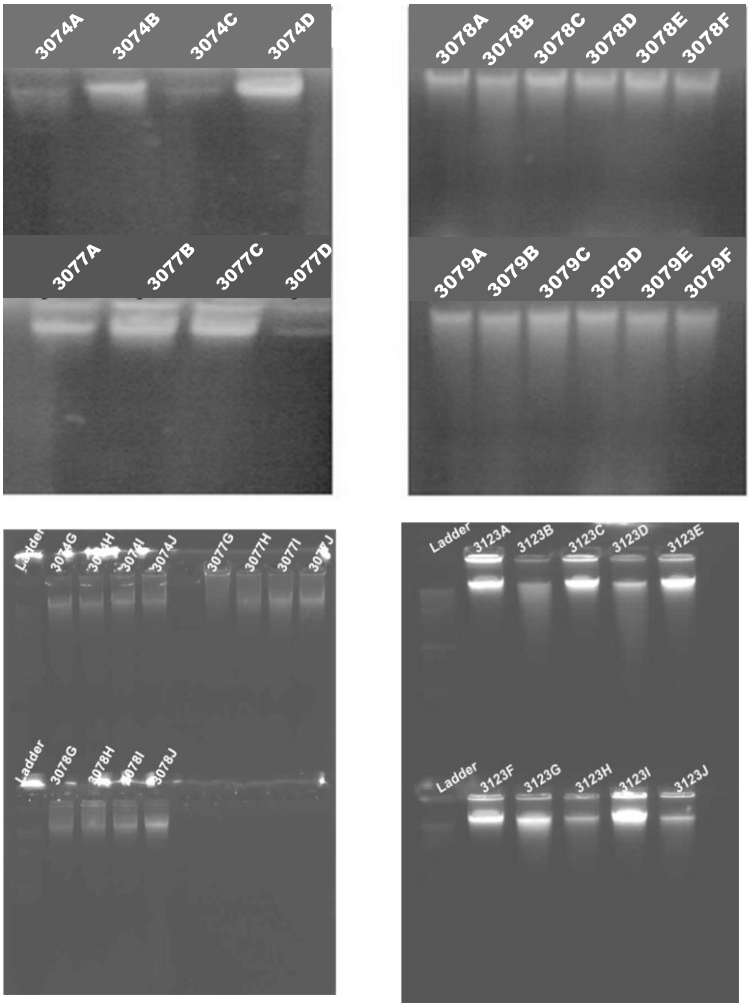


Figura 7. Análises por gel de agarose 1% de amostras de DNA extraídos de espécies de *Croton* através do método de CTAB para fins de sequenciamento e análise por AFLP. Os números referem-se às espécies coletadas apresentadas na Tabela 1. As letras correspondem aos indivíduos coletados em cada população.

Nos exemplos da Fig. 8, pode-se observar que a extração feita com o DNeasy Plant Mini Kit da QIAGEN se mostrou mais efetiva, pois as bandas aparecem mais íntegras, sem rastro, indicando menor degradação, seja por RNA ou por polissacarídeos, proteínas ou quaisquer outros contaminantes, que fariam com que a qualidade do DNA não fosse adequada para a análise por AFLP.

As amostras utilizadas para as amplificações foram selecionadas de acordo com a qualidade de seus DNAs, pois foram coletados vários indivíduos para cada espécie.

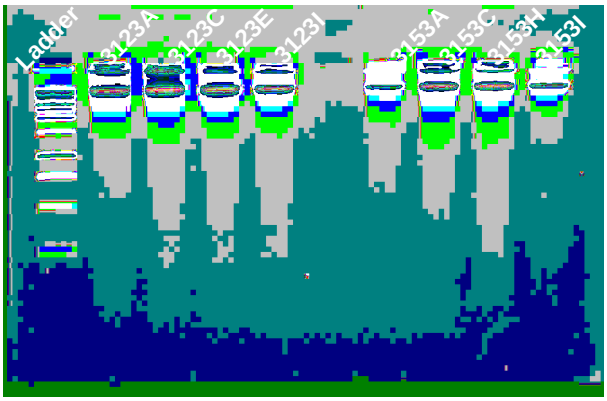


Figura 8. Análises por gel de agarose 1% de amostras de DNA extraídos de espécies de *Croton* através do DNeasy Plant Mini Kit para fins de análise por AFLP. Os números referem-se às espécies coletadas apresentadas na Tabela 1. As letras correspondem aos indivíduos coletados em cada população.

A Fig. 9 mostra o gel dos produtos de amplificação de algumas amostras.

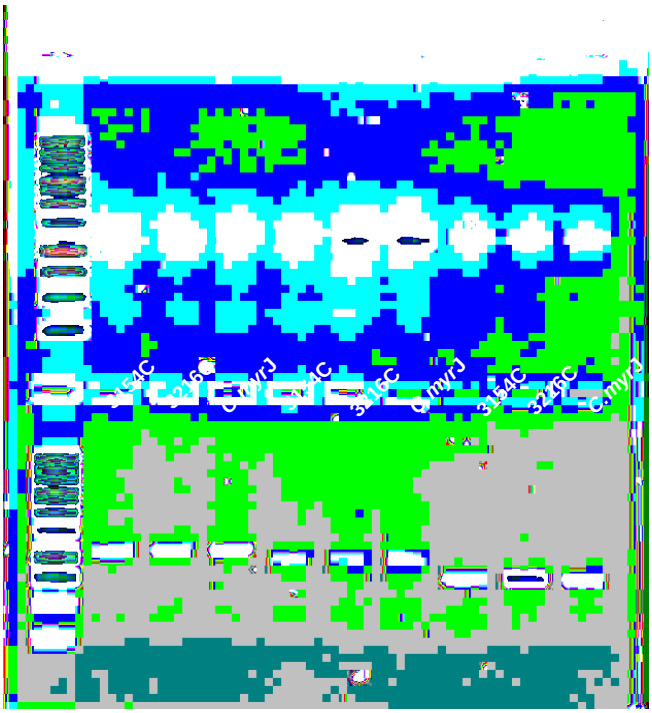


Figura 9. Produtos de amplificação em gel de agarose 1% de amostras de DNA de algumas espécies de *Croton*. Primeira fileira: amplificação da região *trn*L-F. Segunda fileira: amplificação da região *trn*L-F das 3 primeiras amostras; *rps*16, da quarta até a sexta amostras; *ITS*, da sétima até a nona amostras. Os números referem-se às espécies coletadas apresentadas na Tabela 1. As letras correspondem aos indivíduos coletados em cada população. *C.myr – Croton myrianthus.*

Pode-se observar na Fig. 9 que, de acordo com o número de pares de bases da região analisada, a banda amplificada está situada em alturas diferentes; quanto mais afastada do poço de deposição da amostra, menor o número de pares de bases (p. ex. *ITS*). As amplificações da região *trn*L-F foram as ficaram mais próximas aos poços.

Foi feito o sequenciamento da região nuclear *ITS* de todas as amostras das espécies do complexo (indicado na figura com o retângulo pontilhado). Sequências de *ITS* de espécies próximas às do complexo, além de algumas espécies mais distantes, foram obtidas do GenBank. Essas espécies foram escolhidas devido a sua ocorrência

no território brasileiro, não necessariamente na mesma área de abrangência do complexo *C. pallidulus*, que corresponde ao Sul-Sudeste do Brasil. Além das espécies brasileiras, foi incluída na análise *C. alabamensis*, que de acordo com Webster (1993) está definida na seção *Lamprocroton* devido ao seu tricoma lepidoto e pétalas nas flores femininas; porém, de acordo com Van Ee *et al*. (2011), ela é a espécie tipo da Seção *Alabamenses*.

A Fig. 10 corresponde à árvore de consenso estrito da análise por máxima parcimônia, tendo sido obtidas catorze árvores mais parcimoniosas com 728 passos, CI (índice de consistência) 0,6319 e RI (índice de retenção) 0,7893, apresentando 221 caracteres informativos. A árvore foi enraizada usando sequência de *Astraea praetervisa*. *Astraea* foi elevada a gênero por Berry *et al.* (2005), por segregação de *Croton* seção *Astraea* (Klotzsch) Baill. As espécies estudadas neste trabalho são indicadas por meio do retângulo pontilhado.

Figura 10. Árvore de consenso estrito de quatorze árvores mais parcimoniosas, baseadas em sequências de *ITS* e máxima parcimônia. Após os binômios, são indicados os números de depósitos de sequências obtidas do GenBank. Os números após as espécies estudadas neste trabalho correspondem às populações analisadas.

A resolução que se consegue entre espécies de *Croton* usando apenas sequências de *ITS* é baixa, principalmente entre as espécies do Complexo, estudadas neste trabalho. Os resultados apresentados na Fig. 10 indicam que *Brasiliocroton* é grupo irmão de *Croton*, com elevado grau de apoio (valor de "bootstrap" (BS) = 100). As espécies de *Croton* agrupam-se num clado com forte suporte (BS 100). No interior desse clado, o grupo representado pelo Complexo *C. pallidulus* (delimitado pela linha pontilhada) tem suporte bem elevado (BS 100). Os resultados sugerem que o complexo compreende dois grupos. *C. dichrous* e *C. aff. erythroxyloides* pertencem a um dos grupos. Como mencionado anteriormente, esta última espécie foi identificada como *C.* aff. *erythroxyloides*, por apresentar diferenças significativas de *C. erythroxyloides*, podendo ser uma possível espécie nova. Pode-se verificar no cladograma da Fig. 10 que ela está separada de *C. erythroxyloides*, que emerge com as outras espécies que formam o outro grupo do Complexo. Inicialmente, *C. erythroxyloides* não tinha sido cogitada para fazer parte do trabalho, mas analisando o artigo de Lima (2006), notou-se que ela sempre se posicionava próximo às espécies do Complexo *C. pallidulus*.

Ambos os grupos têm elevado suporte (BS=94 e BS=95) (Fig. 10). *C. cordiifolius* (indicada pela seta na Fig. 10) aparece como grupo-irmão do complexo, com um suporte bem elevado (BS 92). Pertence à Seção *Cordiifolii*, monotípica, restrita à vegetação do semiárido da caatinga da Bahia (Van Ee *et al*., 2011). De acordo com Riina *et al.* (2009) essa espécie não se relaciona com aquelas de *Lamprocroton*, subseção *Lamprocroton*, à qual pertencem todas as sete espécies do complexo analisadas no presente trabalho, cujo grupo é o mais diverso no Brasil, ocupando espaço em vegetação aberta e terrenos arenosos ou rochosos (Van Ee *et al.*, 2011). Lima (2006) também já havia notado essa aparente afinidade entre *C. cordiifolius* e as espécies do Complexo, não encontrando, entretanto, nenhuma evidente similaridade morfológica entre elas. O que pode ter havido, portanto, foi uma inadequada amostragem de *C. cordiifolius*, com uma seqüência pobre devido à baixa qualidade do DNA extraído.

Foi feita também a análise bayesiana para este marcador, com as mesmas espécies retiradas do GenBank, e os dados podem ser observados na Fig. 11, que representa o filograma da árvore de consenso de maioria para a análise bayesiana. A árvore foi enraizada usando sequência de *Astraea praetervisa*.

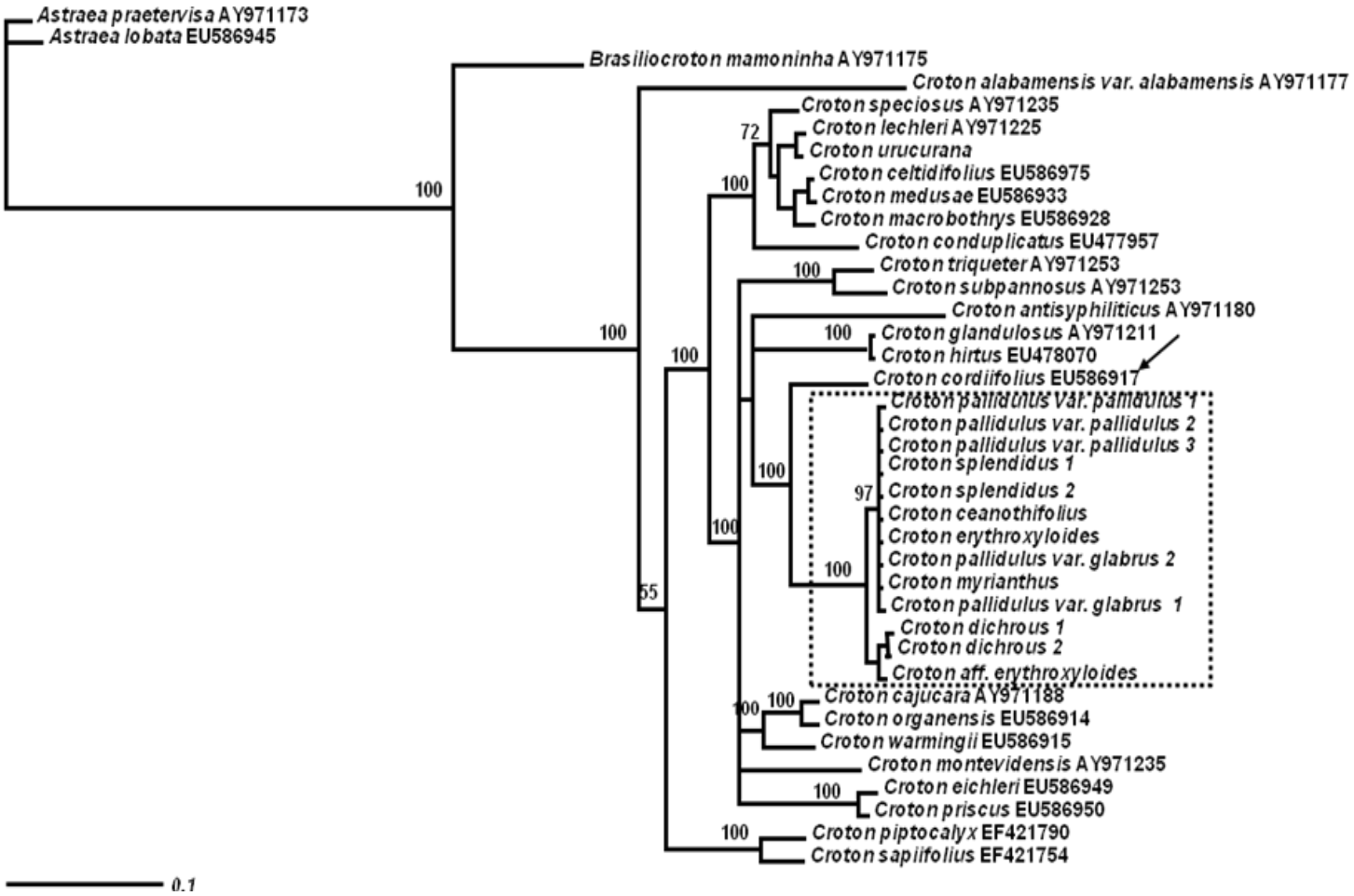


Figura 11. Árvore de consenso de maioria, baseada em sequências de *ITS* e análise bayesiana. Após os binômios, são indicados os números de depósitos de sequências obtidas do GenBank. Os números após as espécies estudadas neste trabalho correspondem às populações analisadas.

Novamente, observa-se que a resolução que se consegue entre espécies de *Croton* usando apenas sequências de ITS é baixa. Os resultados apresentados na Fig. 11 são basicamente bastante parecidos com os obtidos na análise de máxima parcimônia, sendo que *Brasiliocroton* aparece novamente como grupo irmão de *Croton*, com elevado grau de apoio (valor de probabilidades posteriores (PP) = 100). As espécies de *Croton* agrupam-se num clado com forte suporte (PP 100), e no interior desse clado, o grupo representado pelo complexo *C. pallidulus* (delimitado pela linha pontilhada) tem suporte bem elevado (PP 100). Os resultados novamente sugerem que o complexo compreende dois grupos. *C. dichrous* e *C. aff. erythroxyloides* pertencem a um dos grupos (não foi plotado por motivo de espaço os valores PP=100 para o agrupamento *C. dichrous* e *C. aff. erythroxyloides*, e PP=99 para as duas populações de *C. dichrous*). Observa-se mais uma vez que a espécie *C.*

aff. erythroxyloides está separada de *C. erythroxyloides*, que se posicionou junto às outras espécies que formam o outro grupo do complexo.

Em relação às outras espécies analisadas, podem-se observar os mesmos padrões observados na análise de máxima parcimônia, porém os valores de probabilidades posteriores obtidos na análise bayesiana são maiores do que os observados de bootstrap. Além disto, pode-se confirmar a presença de *C. cordiifolius* como espécie-irmã do complexo.

Foram feitas também amplificações e sequenciamentos da região *trn*L-F do cloroplasto de todas as amostras estudadas neste trabalho. Outras sequências foram obtidas do GenBank, a fim de que se pudesse situar o grupo estudado, como foi feito para a região *ITS*. A árvore foi enraizada usando sequência de *Astraea praetervisa* (Fig. 12). Foram obtidos 15 árvores mais parcimoniosas com 578 passos, CI 0,8668 e RI 0,9069, apresentando 327 caracteres informativos.

Pode-se observar que o clado formado pelo complexo (delimitado pela linha pontilhada) está muito mal resolvido, resultando numa politomia, embora altamente consolidado, com BS=100. Nota-se, entretanto, que existem, também aqui, dois grupos, mas as espécies que os delimitam são diferentes da anterior. *C. erythroxyloides* e *C. aff. erythroxyloides* se encontram na politomia, o mesmo acontecendo com os espécimes de *C. dichrous.* Um grupo com BS=88 é formado pelos espécimes de *C. pallidulus* var. *pallidulu*s. Os resultados apresentados na Fig. 12 novamente indicam que *Brasiliocroton* é grupo irmão de *Croton,* e que *C. cordiifolius* continua sendo grupo irmão do complexo, embora suportado com um valor menor (BS=70).

As espécies cujas sequências foram retiradas do GenBank pertencem a Seções diferentes daquela do complexo que está sendo estudada. Nota-se nas Figs 10 e 12 que as espécies de cada seção, normalmente estão agrupadas juntas. Por exemplo, na Fig. 12, *C. lechleri*, *C. urucurana*, *C. medusae*, *C. celtidifolius* e *C. macrobothrys*, pertencentes à Seção *Cyclostigma* (Van Ee *et al.*, 2011) estão unidas em um clado com suporte BS=73; *C. cajucara* e *C. organensis*, da Seção *Cleodora* apresentam um suporte de BS=92. *C. glandulosus e C. hirtus*, pertencentes à Seção *Geiseleria*, bem como *C. subpannosus* e *C. triqueter*, da Seção *Julocroton* apresentam um suporte de BS=100, na árvore de consenso estrito utilizando as sequências de *ITS* (Fg. 10). *C. sapiifolius* (Seção *Croton*, subgênero *Quadrilobi*; Riina *et al.*, 2010) e *C. piptocalyx* (Seção *Sampatik*; Van Ee *et al.*; 2011) são espécies que em todas as

filogenias feitas para o gênero aparecem como basais, o mesmo tendo ocorrido nas análises aqui apresentadas (Figs 10 e 12).

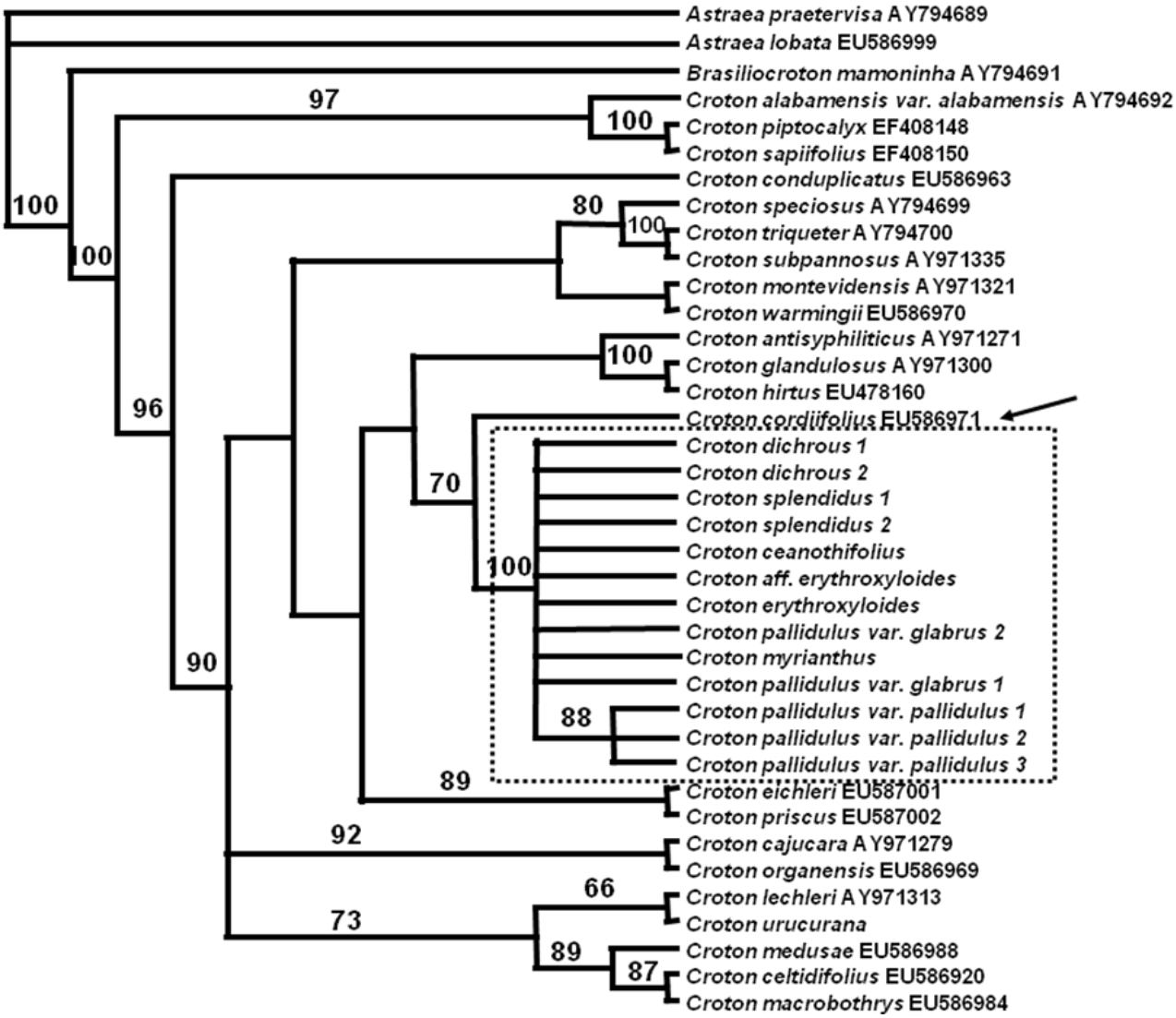


Figura 12. Árvore de consenso estrito, baseadas em sequências de *trn*L-F e máxima parcimônia. Após os binômios, são indicados os números de depósitos de sequências obtidas do GenBank. Os números após as espécies estudadas neste trabalho correspondem às populações analisadas.

A análise bayesiana foi realizada para este marcador também, com as mesmas espécies retiradas do GenBank, e os dados são apresentados na Fig. 13, que representa o filograma da árvore de consenso de maioria para a análise bayesiana. A árvore foi enraizada usando sequência de *Astraea praetervisa*.

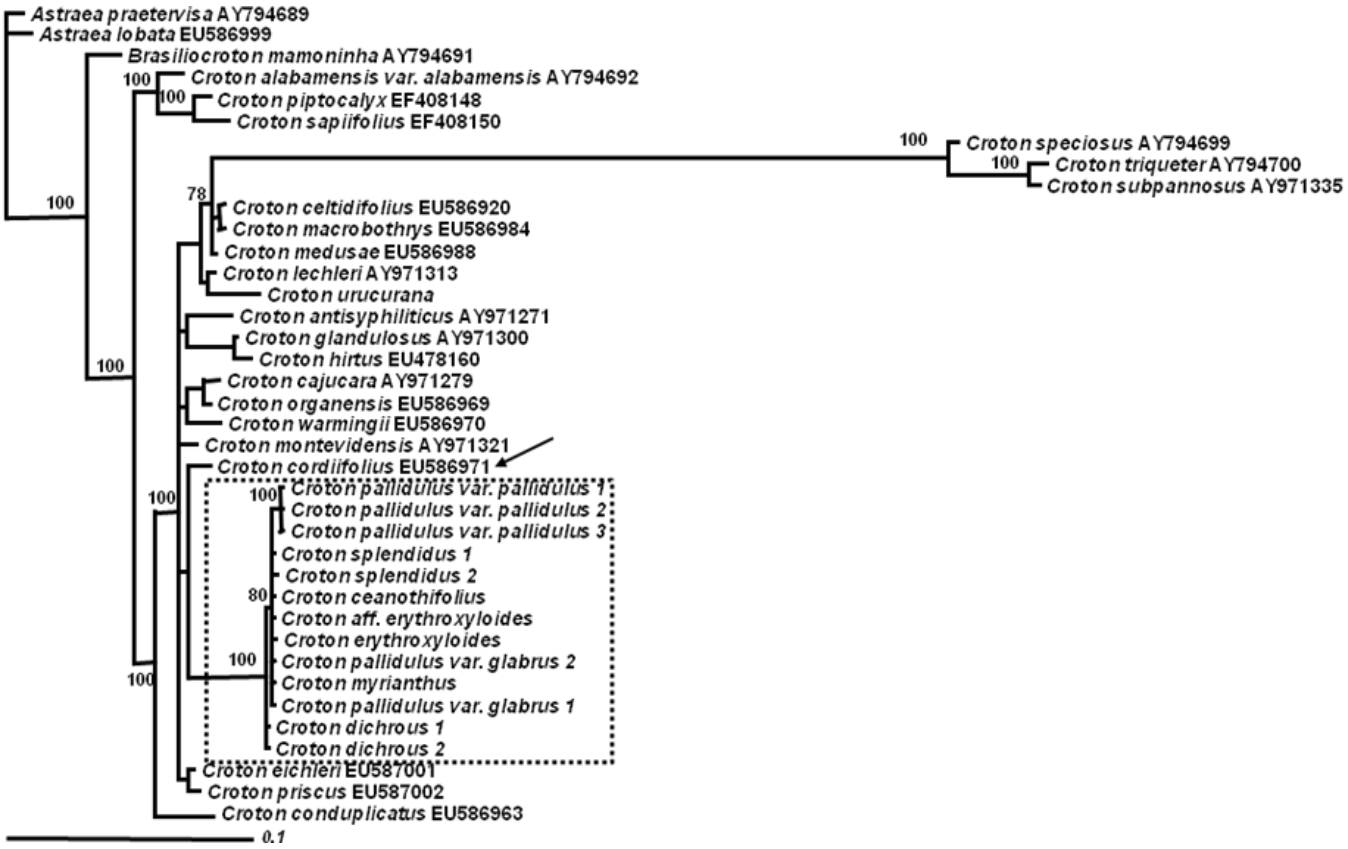


Figura 13. Árvore de consenso de maioria, baseadas em sequências de *trn*L-F e análise bayesiana. Após os binômios, são indicados os números de depósitos de sequências obtidas do GenBank. Os números após as espécies estudadas neste trabalho correspondem às populações analisadas.

Os resultados da Fig. 13, assim como os apresentados para a região *ITS*, são coincidentes para ambas as análises. Há, porém uma melhor resolução do Complexo (PP=100), sendo que os espécimes de *C. dichrous* emergem fora da politomia do Complexo, como um grupo aparentemente mais basal. Entretanto, uma vez que o valor de PP é baixo (80), este dado não tem sustentação. O grupo formado pelos espécimes de *C. pallidulus* var. *pallidulu*s também emergem da politomia com uma alta sustentação (PP=100). Os resultados apresentados na Fig. 13 novamente indicam que *Brasiliocroton* é grupo irmão de *Croton,* e que *C. cordiifolius* continua sendo grupo irmão do complexo, com elevado índice de suporte (PP=100, não plotado na árvore devido ao pequeno espaço).

Nota-se, portanto, que esses dois marcadores são bastante informativos quando se quer analisar espécies de *Croton.* Isso é mostrado pelos altos números de caracteres informativos. Entretanto, não são muito bons para separar espécies muito afins, como acontece com as analisadas neste trabalho. Em seus trabalhos, Van Ee *et*

al. (2011) utilizam esses dois marcadores, além de outros dois genes, *rps*3 (região mitocondrial) e a região nuclear *EMB2765*.

Foi feito o teste ILD para a verificação do grau de congruência entre as análises por ITS e *trn*L-F, utilizando as seqüências obtidas junto ao GenBank. O valor resultante foi menor do que 0,05 (P*value*= 0,01) indicando baixo grau de congruência.

Foi feita mais uma tentativa de obter um marcador que fosse interessante para separar as espécies aqui estudadas, e dessa maneira utilizou-se a região plastidial rps16. Uma vez que não havia no GenBank sequências para este marcador, a árvore mostrada na Fig. 14 apresenta apenas os dados das amostras do Complexo.

Pode-se notar que *rps*16 não é um bom marcador para a separação dessas espécies, não sendo informativo. O número de caracteres informativos (somente um), além do baixo número de passos (28), e CI e RI igual a 1, podem explicar o baixo desempenho desse marcador. *C.dichrous* forma uma politomia com a espécie utilizada como grupo externo (*C. urucurana*), juntamente com *C. ceanothifolius*. Entretanto, não há nenhuma semelhança morfológica entre *C. urucurana* e as duas outras espécies que fazem parte do complexo, e é por esse motivo que ela está sendo usada como grupo externo. *C. urucurana* pertence à Seção *Cyclostigma*, enquanto as outras amostras são da Seção *Lamprocroton.* Nota-se também que o suporte da união das outras espécies é muito baixo (BS=65).

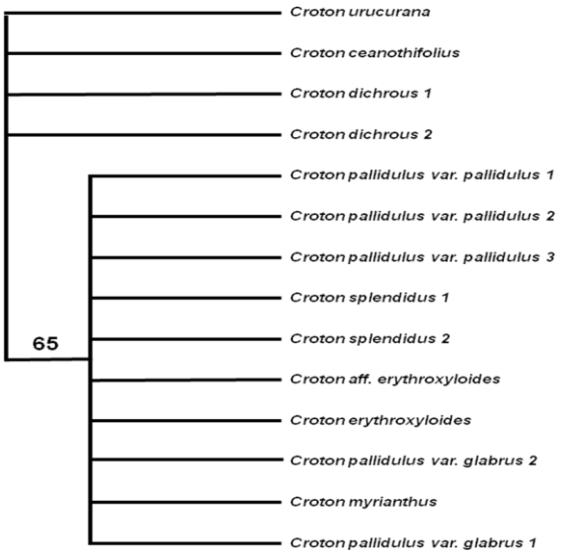


Figura 14. Árvore de consenso estrito baseado em sequências de *rps*16 e máxima parcimônia de espécies de *Croton* que participam do complexo *C. pallidulus.* Os números após as espécies estudadas neste trabalho correspondem às populações analisadas.

Foi realizada uma análise bayesiana para este marcador, porém apenas o grupo externo (*C. urucurana*) separava do resto do Complexo, cujas espécies apareciam todas dentro de uma enorme politomia, sem resolução entre elas. Por esse motivo, essa árvore não foi apresentada neste trabalho.

A Fig. 15 mostra a árvore de consenso estrito da análise combinando as regiões do genoma nuclear *ITS*, e as do cloroplasto *trn*L-F e *rps*16. Observa-se nessa análise que as espécies retiradas do GenBank que pertencem à mesma seção encontram-se, conforme mencionado anteriormente, próximas entre si na árvore.

Os valores encontrados para a combinação dos dados foram de CI 0,7205, RI 0,8211 com 1392 passos para a árvore e 560 caracteres informativos.

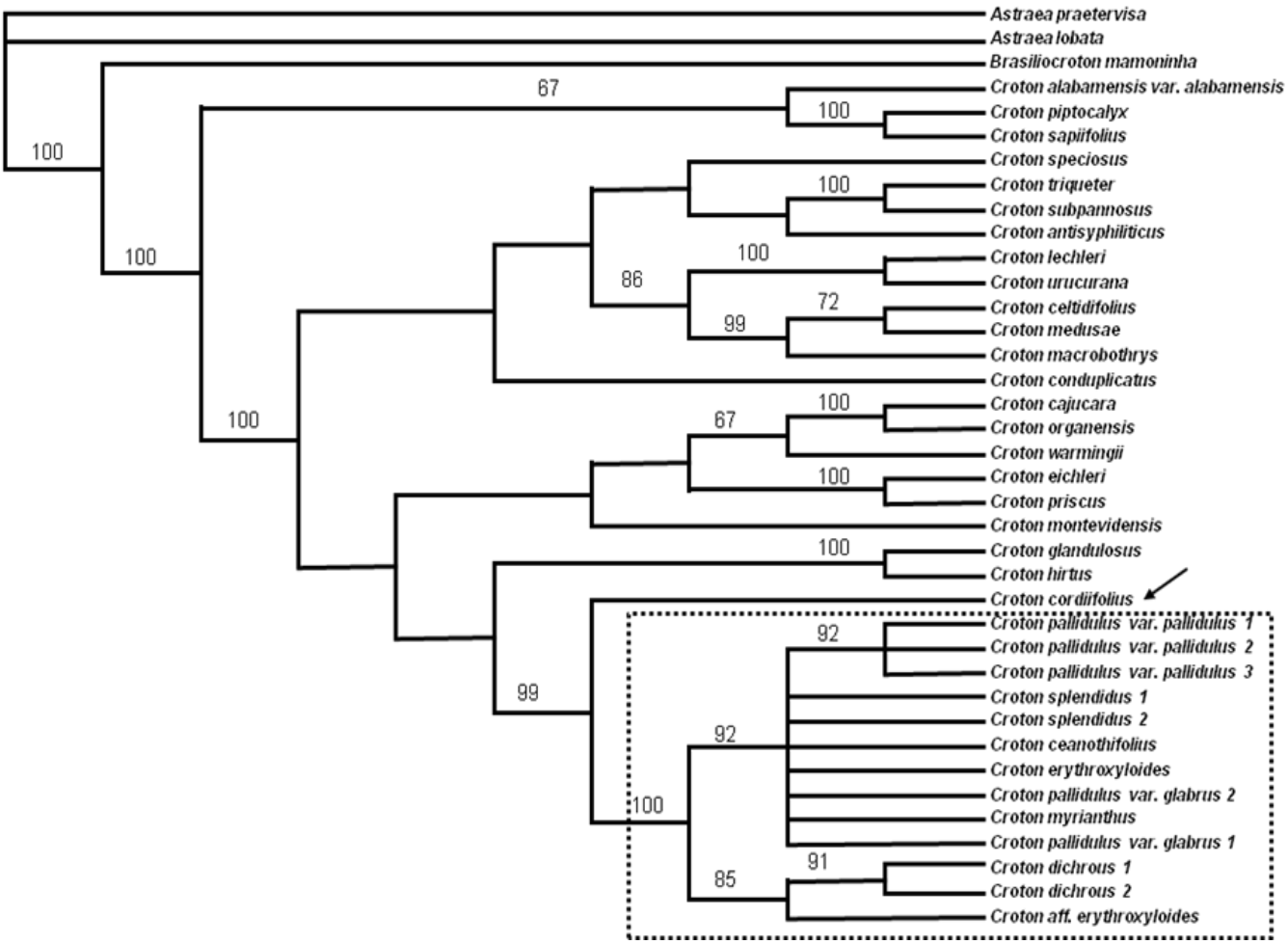


Figura 15. Árvore de consenso estrito, baseadas em sequências de *ITS*, *trn*L-F, *rps*16 e máxima parcimônia. As espécies cercadas por linha pontilhada correspondem àquelas do complexo *C. pallidulus*. As outras foram retiradas do GenBank. Os números após as espécies estudadas neste trabalho correspondem às populações analisadas.

O complexo é bem delimitado (BS=100), assim como a espécie *C. cordiifolius* aparece como espécie-irmã do complexo com índice elevado (BS=99). É interessante observar que a delimitação dos espécimes de *C. pallidulus* var. *pallidulus* aparecem em um clado separado das espécies do complexo, com índice de BS=92. Nota-se que *C. dichrous* (BS=91) continua se separando do resto do clado, juntamente com *C. aff. erythroxyloides* (BS=85), que parece ser mesmo distinta de *C. erythroxyloides*. As outras espécies aparecem, novamente, em uma politomia. As outras espécies retiradas do GenBank comportam-se do mesmo modo como observadas para os marcadores *ITS* e *trn*L-F. O gênero *Croton* mostra-se bem definido, sendo monofilético, com alto índice de BS (100), separando-se das espécies utilizadas como grupo externo (*Astraea*) e também de *Brasiliocroton mamoninha* (Fig. 15).

Foi feita uma análise apenas com as amostras coletadas para o trabalho. Para combinar os marcadores *ITS*, *rps*16 *e* trn*L-F* foi feito o teste IDL, obtendo-se, neste caso, o valor de 0,36, indicando um grau elevado de congruência. Foram utilizadas as sequências de *C. urucurana* como grupo externo. A Fig. 16 mostra a árvore de consenso estrito baseada nas sequências destes marcadores. Observando-se esta figura, pode-se notar que *C. dichrous* e *C. aff. erythroxyloides* emergem num clado, separando-se das outras espécies do complexo, o que já havia sido observado na análise baseada em *ITS* (Fig. 10) e na árvore de consenso unindo os três marcadores (Fig. 15). Essas espécies formam um grupo fortemente suportado (BS=93). O outro clado é formado pelas outras espécies, com o suporte muito baixo (BS=69). Os três espécimes de *C. pallidulus* var. *pallidulus*, pertencentes a populações diferentes, estão juntos nesta análise e são apoiados com um “bootstrap” relativamente alto, BS=82.

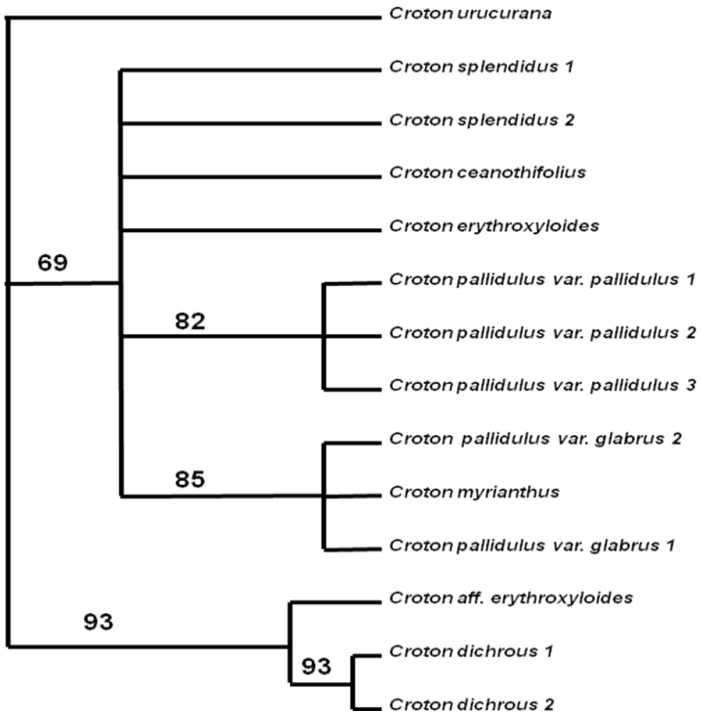


Figura 16. Árvore de consenso estrito, baseadas em sequências de *ITS*, *trn*L-F *e rps*16 e máxima parcimônia de espécies de *Croton* que participam do Complexo. Os números após as espécies estudadas neste trabalho correspondem às populações analisadas.

Nota-se nesta análise que o clado formado por *C.myrianthus* e *C. pallidulus* var. *glabrus* (populações 1 e 2) recebe um bom apoio (BS=85). *C. splendidus* (populações 1 e 2), *C. ceanothifolius* e *C. erythroxyloides* formam uma politomia.

Um teste para a análise bayesiana foi feito também para a combinação dos dados, como pode ser observado na Figura 17.

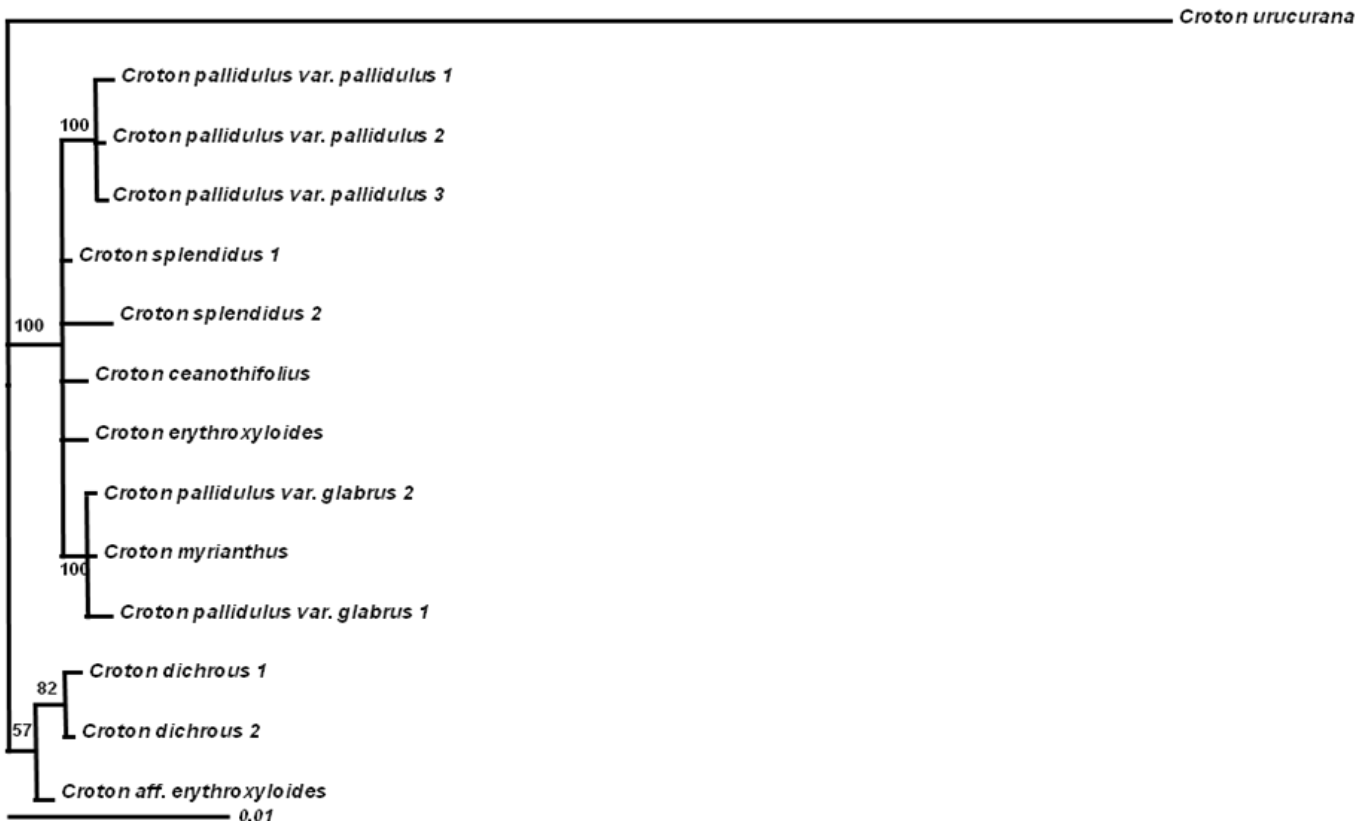


Figura 17. Árvore de consenso de maioria, baseadas em sequências de *ITS, trn*L-F e *rps*16 e análise bayesiana. Os números após as espécies estudadas neste trabalho correspondem às populações analisadas.

Os resultados da Fig.17 apresentam-se tais quais os apresentados para a análise de máxima parcimônia, com *C. dichrous* e *C. aff. erythroxyloides* emergindo em um clado, separando-se das outras espécies do complexo. Porém, diferentemente da análise de máxima parcimônia, o índice apresentado (PP=57) é muito baixo. O outro grupo é formado pelas demais espécies, com forte apoio (PP=100). Os três espécimes de *C. pallidulus* var. *pallidulus*, pertencentes a populações diferentes são apoiados com índice de probabilidades posteriores alto, PP=100. Novamente o outro clado, que compreende *C. pallidulus* var. *glabrus* 2, *C.myrianthus* e *C. pallidulus* var. *glabrus* 1 apresenta elevado suporte (PP=100).

De acordo com todos os resultados apresentados, pode-se observar que os representantes do complexo *C. pallidulus* apresentam forte afinidade entre eles, formando um clado com baixa resolução, em coerência com as semelhanças morfológicas já mencionadas. A análise por sequenciamento utilizando os marcadores *ITS*, *trn*L-F e *rps*16 não ajudaram na resolução do complexo. Apenas *C. dichrous* parece ser uma espécie distinta para análise com *ITS*, com elevado suporte para análise de máxima parcimônia, porém baixa para análise bayesiana.

Isso é corroborado pelo estudo dos caracteres morfológicos. De acordo com Lima & Pirani (2008), *C. dichrous* apresenta caracteres morfológicos que o distingue das outras espécies do complexo e de toda as espécies da Seção *Lamprocroton*, pois suas folhas são discolores, com a face adaxial, em geral, completamente glabra, podendo, entretanto, apresentar na face adaxial, quando jovem, tricomas simples que caem à medida que a folha se desenvolve.

Os resultados, portanto, demonstram que o seqüenciamento feito neste trabalho está de acordo com trabalhos publicados recentemente. Além disso, observa-se claramente que as duas variedades da espécie *C. pallidulus* parecem ter naturezas distintas. Evidencia-se a necessidade de análises com outros tipos de marcadores moleculares que possibilitem evidenciar o relacionamento filogenético entre as espécies componentes do complexo.

4.2. Marcadores AFLP

Para as análises por AFLP foi necessário mudar o protocolo de extração, pois impurezas, como por exemplo, substâncias do metabolismo secundário, dificultam, muitas vezes, o bom desempenho dos procedimentos para obter-se uma boa atuação das enzimas de restrição, e parece ter sido este o caso. A extração de DNA se mostrou muito mais eficiente, devido a pureza do DNA obtido, com a utilização do DNeasy Plant Mini Kit (ver Figs. 7 e 8, para comparação). Os procedimentos de digestão e ligação tornam-se mais eficientes, acarretando uma melhoria na amplificação pré-seletiva, como se pode notar pela documentação dos geis de agarose obtidos por DNAs extraídos pelos dois processos citados (Fig. 18).

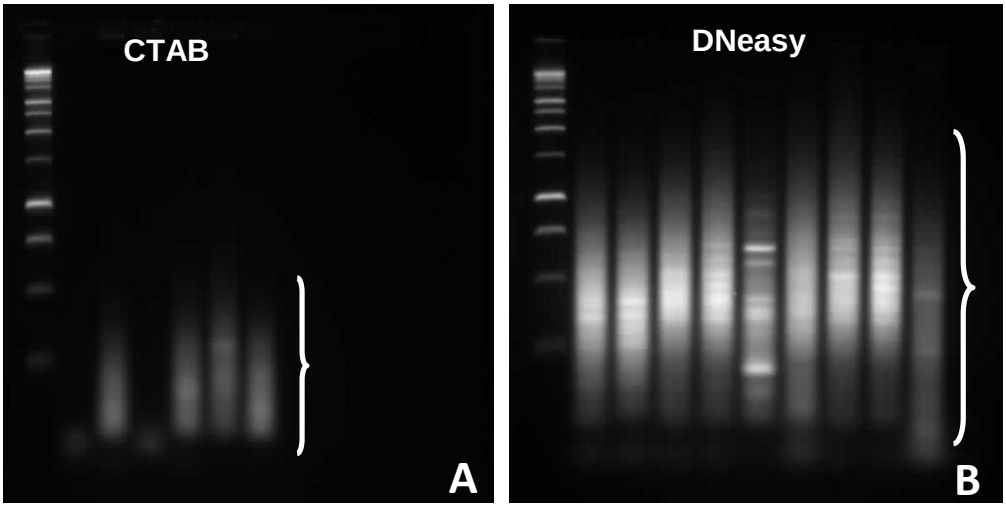


Figura 18. Documentação em gel de agarose 1,5% de amplificações pré-seletivas, de amostras de *Croton*¸ cujos DNAs foram extraídos por CTAB (**A**) e pelo kit DNeasy (QIAGEN) (**B**).

Observa-se na Fig. 18 que pelo método de extração por CTAB, os fragmentos resultantes da amplificação pré-seletiva são muito pequenos, estando na faixa entre 50 e 100 pares de bases (pb). Para comparação dos fragmentos polimórficos em AFLP necessita-se que eles sejam maiores de 75 pb, uma vez que abaixo desse valor eles não são informativos e, portanto, não fornecem caracteres úteis na diferenciação entre espécies, ou mesmo entre diferentes populações da mesma espécie.

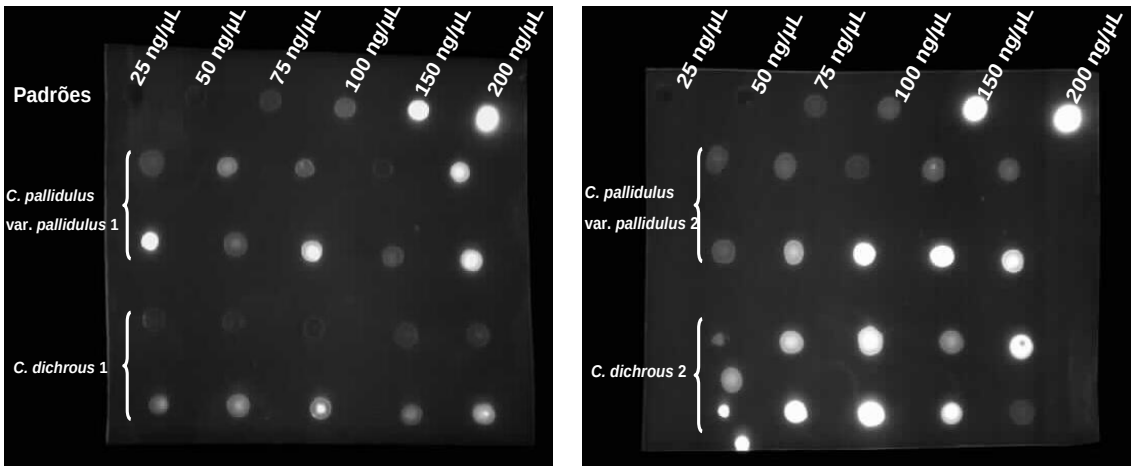


Figura 19. Resultados de quantificação de DNA genômico total observados em transiluminador UV de acordo com os padrões calculados de acordo com o DNA Lambda.

Na Fig. 19 observam-se os resultados da quantificação do DNA dos indivíduos de algumas espécies de *Croton* que foram comparados com os padrões do DNA de fago λ, cujas concentrações foram previamente calculadas. Estes resultados são os observados para a extração realizada pelo método por CTAB modificado. Nota-se que alguns indivíduos, de uma mesma população, apresentam uma intensidade maior na fluorescência do que outros. Isso pode ser devido a contaminantes, que absorvem nos comprimentos de onda da luz UV e que poderiam determinar uma quantificação inexata, o que resultaria na digestão de pouco DNA, por exemplo, e em uma segunda etapa, na reação de digestão e consequentemente em todo o procedimento da análise.

Com o kit de extração, as quantificações foram muito mais homogêneas, com uma pequena diferenciação entre os valores das concentrações obtidas para cada indivíduo extraído. Infelizmente, por um problema técnico, não foi possível a documentação das fluorescências, como foi feito para os DNAs extraídos com CTAB. Entretanto, os resultados, de um modo geral, indicavam a concentração de 50 ng/µL em quase todos os indivíduos, mesmo de populações diferentes.

Para a reação de digestão utilizou-se, portanto, 4 µL do DNA extraído, a fim de que a quantidade de DNA na reação fosse de cerca de 200 ng. Apenas uma vez foi feito um gel de agarose para se verificar a qualidade da digestão, mas a visualização não foi boa, a degradação mostrando-se muito tênue. Isso se deveu ao fato de que o volume colocado no gel foi muito pequeno (3,5 - 4,0 µL), pois os restantes 5,5 µL deveria ser utilizado para a reação de ligação. Optou-se, então, pela verificação da qualidade da reação pré-seletiva por meio de eletroforese com gel de agarose 1,5%,

uma vez que a quantidade aplicada poderia ser de 6 µL, pois na reação seletiva a amostra seria diluída (Fig. 18 B).

Para a reação de amplificação seletiva foram realizadas algumas modificações, principalmente com relação às diluições das amostras. Dependendo da qualidade da amplificação pré-seletiva, visualizada pela intensidade de fluorescência observada no gel, se diluía mais ou menos a amostra para a etapa de amplificação seletiva.

Foram testadas várias combinações, porém sem sucesso. Utilizaram-se, portanto, apenas três combinações. Para as diferentes combinações analisadas, foram obtidos dados para os diferentes indivíduos das 14 populações estudadas no Complexo, como pode ser observado na Tabela 3.

Tabela 3. Quantidade de indivíduos, número de fragmentos e número de populações analisadas para as três combinações utilizadas na amplificação seletiva, na análise de AFLP.

	AGC-CTA (amarelo)	AAG-CAG (verde)	ACA-CTG (azul)
N° indivíduos / N° fragmentos	55/248	54/294	43/190
N° populações	14	14	12

Notam-se, pela análise da Tabela 3, que as duas primeiras combinações foram as que deram resultados melhores, pois o número de fragmentos obtidos foi muito maior do que na terceira combinação. O número de indivíduos que responderam positivamente às duas primeiras combinações foi também maior, o mesmo acontecendo com as populações; em duas delas (*C. urucurana* e *C. myrianthus*), nenhum indivíduo respondeu à combinação ACA-CTG. Em algumas populações, somente um ou dois indivíduos responderam positivamente a uma determinada combinação. Entretanto, para algumas das coletas mais recentes, cujos materiais estavam a menos tempo guardados em sílica, houve um maior número de indivíduos com bons resultados.

Portanto, as combinações AGC-CTA, AAG-CAG e ACA-CTG foram escolhidas, pois apresentam um maior número de fragmentos para todos os indivíduos, isto é, foram as combinações mais polimórficas. Outro ponto importante a ser comentado é a quantidade de fragmentos obtidos para cada indivíduo. Foram selecionados em sua maioria somente os indivíduos que apresentavam no mínimo vinte fragmentos, com bandas bem definidas como os das Figs 20-22. As bandas dos fragmentos deveriam

também ter intensidade (altura) acima de 100 (isto era uma recomendação do software utilizado). Devido a isso, foram desprezadas várias bandas (fragmentos) cujo valor estava abaixo de 100, e dessa maneira, vários indivíduos foram eliminados.

Nas Figs 20–22 podem-se observar alguns exemplos de eletroferogramas obtidos no seqüenciamento dos fragmentos, os quais apresentam fluorescências de cores diferentes dependendo dos iniciadores *Eco*RI empregados na amplificação seletiva; os iniciadores *Mse*I não apresentam fluorescência. As diferentes bandas observadas correspondem aos diversos tamanhos de fragmentos gerados, sendo que os picos observados apresentam valores que variaram de 75pb a 500 pb na análise, e foram comparados pelo software com os dados do padrão adicionado, GeneScan-500 “ROX”, cujos comprimentos são definidos, e nos eletroferogramas aparecem em vermelho. Os resultados podem ser analisados manualmente ou por meio do software Gene Mapper (ver anexo 4), em busca de fragmentos polimórficos que possam ser utilizados para comparação entre os diferentes indivíduos das populações.

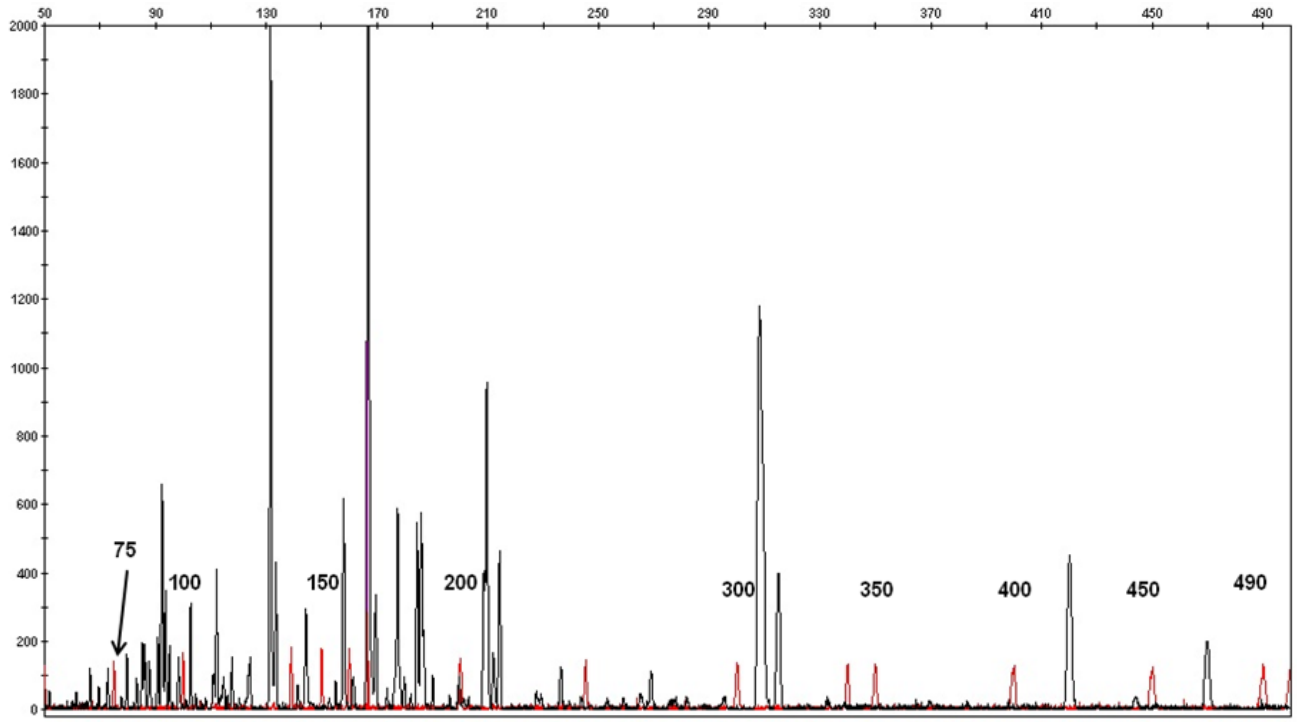


Figura 20. Eletroferograma dos fragmentos obtidos na amplificação seletiva com a combinação AGC–CTA do indivíduo C da população 3 de *Croton pallidulus* var. *pallidulus*. As bandas em vermelho correspondem ao padrão utilizado (GeneScan-500 “Rox”).

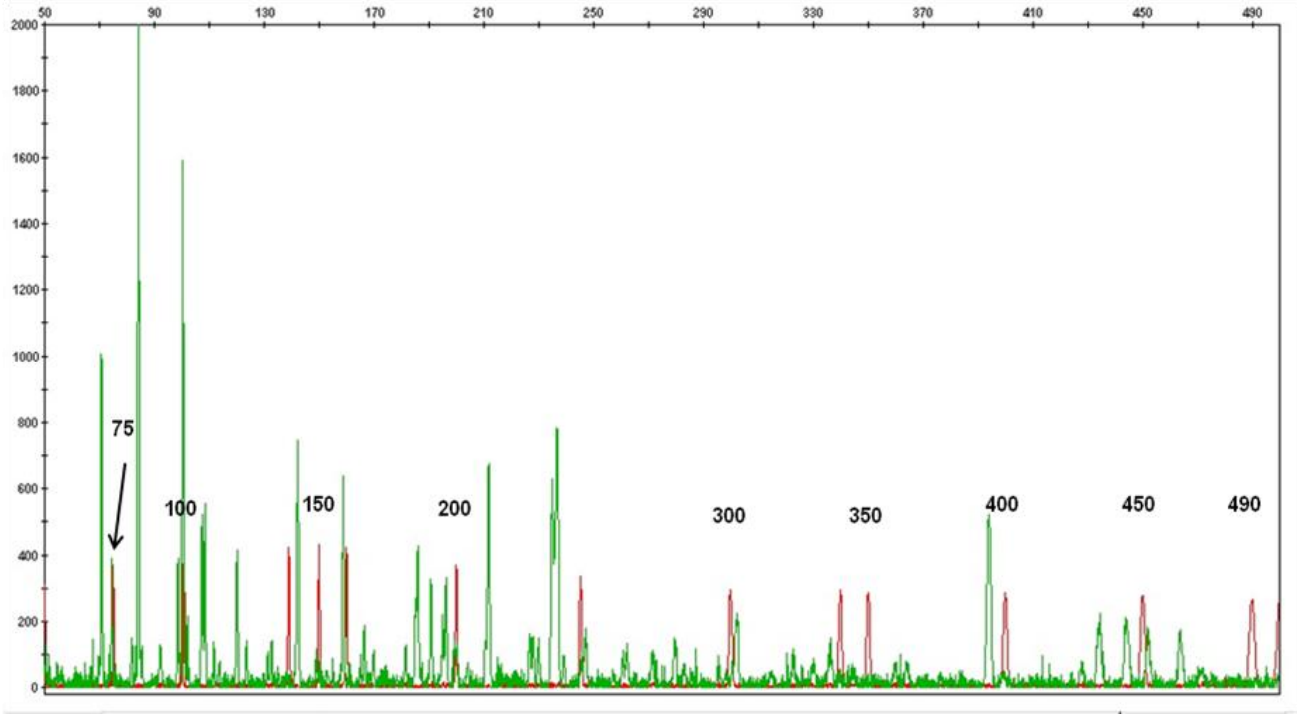


Figura 21. Eletroferograma dos fragmentos obtidos na amplificação seletiva com a combinação AAG–CAG do indivíduo C da população 3 de *Croton pallidulus* var. *pallidulus*. As bandas em vermelho correspondem ao padrão utilizado (GeneScan-500 “Rox”).

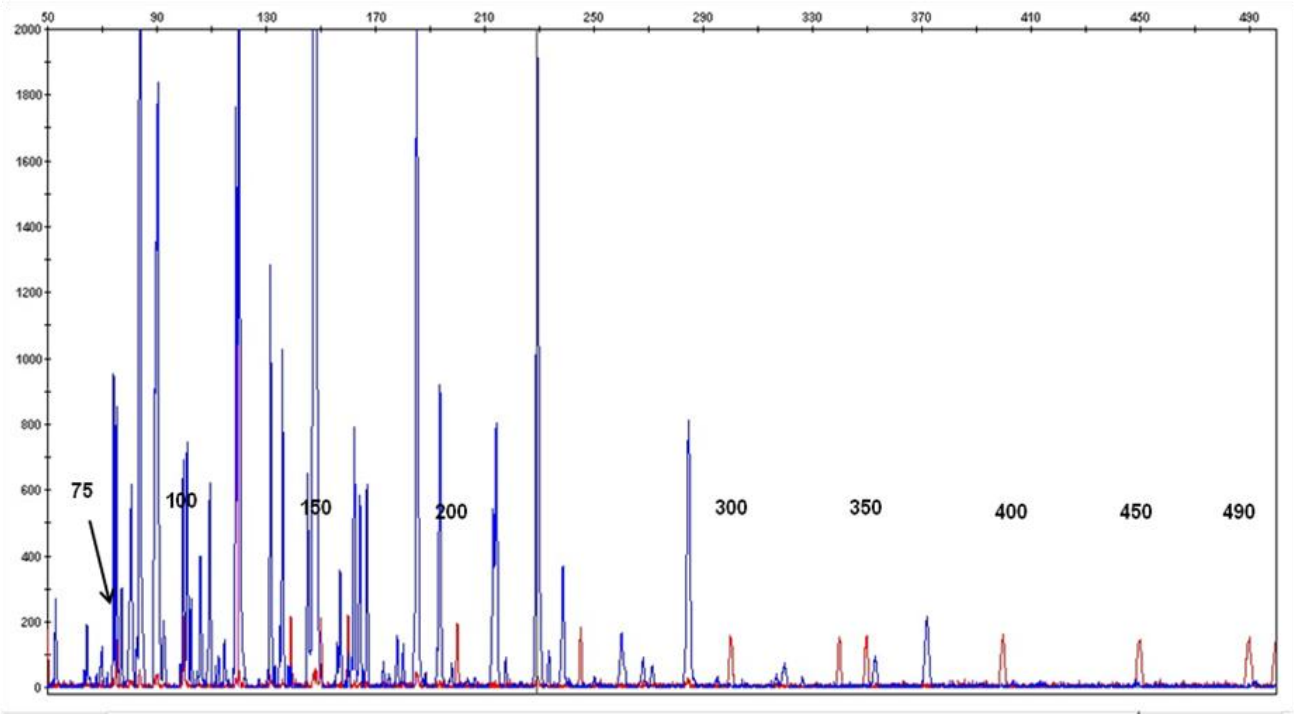


Figura 22. Eletroferograma dos fragmentos obtidos na amplificação seletiva com a combinação ACA–CTG do indivíduo C da população 3 de *Croton pallidulus* var. *pallidulus*. As bandas em vermelho correspondem ao padrão utilizado (GeneScan-500 “Rox”).

As bandas observadas correspondem aos tamanhos dos fragmentos obtidos na amplificação seletiva das amostras. Estes foram analisados primeiramente pelo programa GeneMapper, e posteriormente, foi feito uma correção manual para a confecção da tabela de presença ou ausência de acordo com o tamanho dos fragmentos, para cada indivíduo das populações, nas espécies estudadas.

Estes dados foram então analisados pelo programa PAUP (ver anexo 5), em que foi realizada uma análise de evolução mínima, através do método de distância mínima (Neighbor Joining), sendo que a consistência da monofilia dos ramos foi avaliada através da replicação “bootstrap”, com mil réplicas. Como já foi mencionada, não foi possível a obtenção de dados de todos os indivíduos de todas as populações, tendo sido analisados somente aqueles que apresentavam bons resultados de sequenciamento.

Devido aos problemas mencionados, foi realizada em um primeiro momento uma análise unindo as duas primeiras combinações mais efetivas. O filograma resultante está apresentado na Fig. 23.

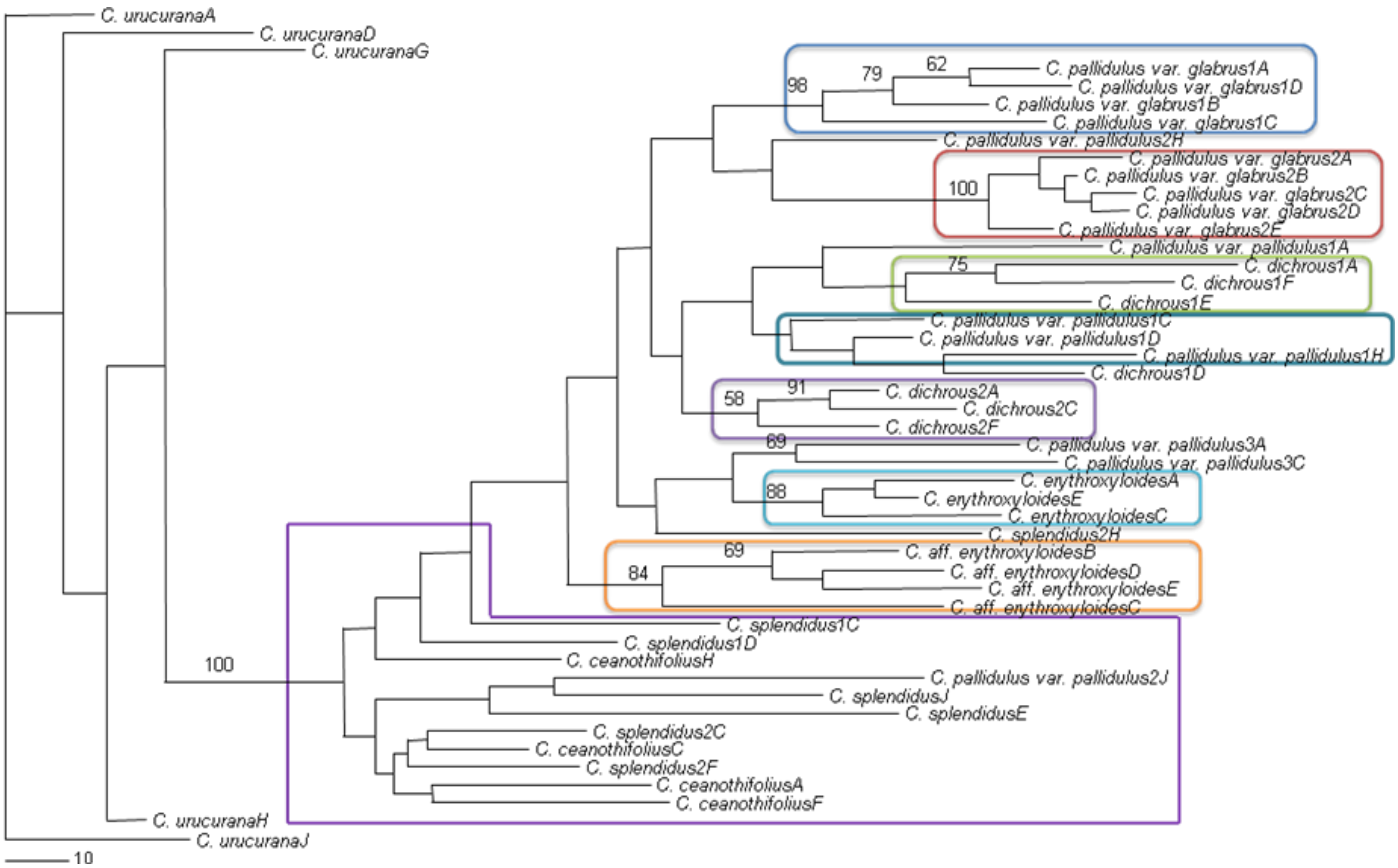


Figura 23. Árvore de relações de afinidade genética entre amostras de espécies de *Croton,* baseadas em fragmentos de restrição amplificados pelas combinações dos AGC-CTA e AAG-CAG e método de agrupamento Neighbor Joining. Os números depois dos nomes das espécies correspondem às populações e as letras, aos indivíduos.

Analisando-se a árvore de afinidades genéticas da Fig. 23 vários resultados interessantes e promissores para o entendimento dos relacionamentos entre membros do Complexo são evidenciados. Pode-se perceber que o Complexo forma um grupo monofilético, com alto índice de sustentação (BS 100), pelo enraizamento da árvore por meio do grupo externo *C. urucurana.* Foram analisados 400 caracteres considerados pelo programa PAUP como informativos (nci).

Os retângulos coloridos delimitam indivíduos da mesma espécie e/ou população. A evidência fornecida pela análise por AFLP é coerente com a obtida por sequências de DNA, no que se refere à distinção entre *C. erythroxyloides* e *C.* aff. *erythroxyloides* (ver Figs. 10 e 15). Ambos os táxons são bem sustentados (BS 88 e 84, respectivamente), porém emergem em clados distintos (Fig. 23).

Levando-se em conta as duas variedades de *C. pallidulus*, a análise por AFLP sugere que ela é um grupo parafilético. Cada população de *C. pallidulus* var. *glabrus* constitui um grupo monofilético com elevado apoio (BS 98 para a população 1 e BS

100 para a população 2). *C. pallidulus* var. *pallidulus* comporta-se como um grupo parafilético, com indivíduos de uma mesma população dispersos entre as outras espécies do Complexo. Por exemplo, indivíduos da população 1 encontram-se no mesmo clado de *C. dichrous*. Dos dois indivíduos da população 2, um está próximo a *C. myrianthus* e o outro, próximo a *C. pallidulus* var. *glabrus.* Indivíduos da população 3 parecem ser grupo irmão de *C. erythroxyloides* (Fig. 23). *C. dichrous*, *C. ceanothifolius* e *C. splendidus*, de acordo com essa análise, também não são grupos monofiléticos.

No agrupamento delimitado no polígono roxo são encontradas várias espécies que apresentam similaridade genética. Dentre elas, *C. myrianthus* é uma espécie bastante semelhante a *C. erythroxyloides* quanto ao hábito e tipo de indumento sendo, muitas vezes, difícil distinguir esses dois táxons, já que em ambos as faces foliares são lepidotas, com indumento formado por tricomas tipicamente lepidotos. Entretanto, em *C. erythroxyloides* as inflorescências são bissexuadas, a face adaxial das sépalas das flores pistiladas e dos estiletes são lepidotos, ao passo que em *C. myrianthus* as inflorescências são unissexuadas, a face adaxial das sépalas das flores pistiladas é tomentosa, coberta por tricomas estrelados, e os estiletes são completamente glabros. A forma e tamanho das folhas são caracteres bastante variáveis nos dois táxons e essa variação é uma característica comum nas diversas espécies de *Croton*, porém em geral, *C. myrianthus* possui as folhas mais estreitas e mais longas quando comparadas às de *C. erythroxyloides* (Lima & Pirani, 2008).

Allem (1978) sinonimizou *C. ceanothifolius* em *C. migrans* Casar (= *C. splendidus* Mart.). Entretanto, esta proposta não é aceita por Lima & Pirani (2008), uma vez que são observadas diferenças consideráveis entre essas duas espécies, que podem ser distinguidas, mesmo no estado vegetativo, devido, principalmente, ao tipo de nervação, forma das folhas e tipo de indumento das duas faces foliares.

Foi realizada também uma análise unindo numa única matriz as três diferentes combinações da amplificação seletiva. Uma vez que os indivíduos de *C. urucurana* não amplificaram para a combinação ACA-CTG, não se pode utilizá-los como grupo externo quando se unem as três combinações. A espécie usada para o enraizamento dessa nova árvore foi *C. ceanothifolius*, por se tratar da espécie com o posicionamento mais “basal” do Complexo, como pode ser visto na Fig. 23. O número de indivíduos que amplificaram para as três combinações foi menor, apenas 33, e estes foram utilizados na matriz dessa árvore, apresentada na Fig. 24.

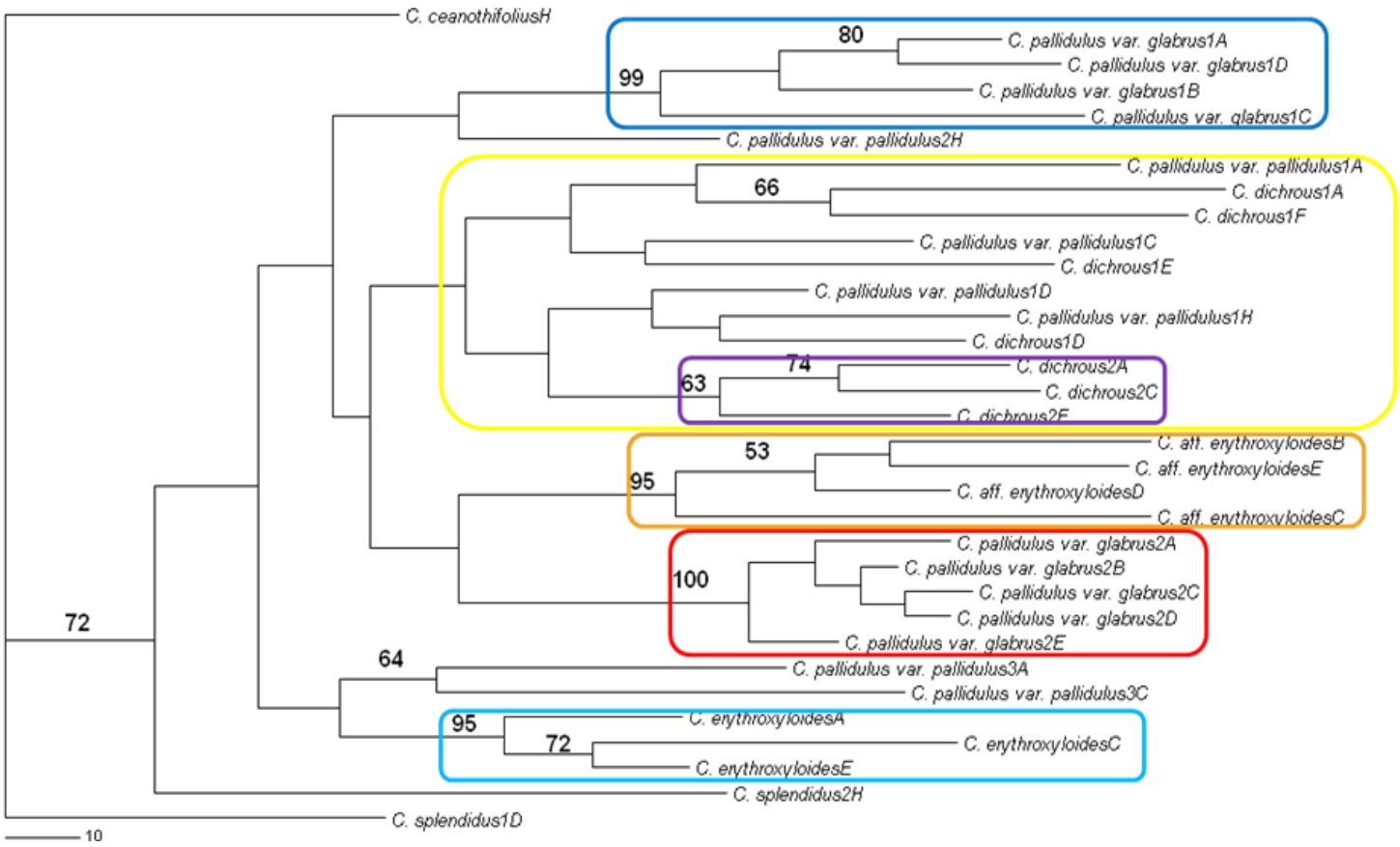


Figura 24. Árvore de relações de afinidade genética entre amostras de espécies de *Croton,* baseadas em fragmentos de restrição amplificados pelas três combinações dos marcadores AFLP e método de agrupamento Neighbor Joining. Os números depois dos nomes das espécies correspondem às populações e as letras, aos indivíduos.

O mesmo padrão da Fig. 23 pode ser observado na árvore da Fig. 24, que reitera a natureza distinta das espécies de *C. erythroxyloides* e *C.* aff. *erythroxyloides*. As duas populações de *C. pallidulus* var. *glabrus* continuam aparecendo em posições distintas também nesta árvore, com bons suportes de boostrap (BS 99 para a população 1 e BS 100 para a população 2). Entretanto, na Fig. 23, as duas populações se encontram no mesmo clado, embora com baixo suporte (<50), enquanto na Fig. 24 elas se distanciaram entre si. A população 1 da variedade *glabrus* tem como grupo irmão um indivíduo de *C. pallidulus* var. *pallidullus* 2 (que aparece como grupo irmão da população 2 de *C. pallidulus* var. *glabrus* na Fig. 23) e a população 2 aparece aqui no mesmo clado de *C.* aff. *erythroxyloides*, embora também aqui, com um BS<50. Foram analisados para este filograma 394 caracteres informativos.

C. myrianthus não foi representada na Fig. 24, por não apresentar nenhum indivíduo que amplificou para essa terceira combinação. Os posicionamentos de *C. splendidus* e *C. ceanothifolius* também não são dignos de confiança pela pouca representação de indivíduos, em relação aos representados na Fig. 23.

Em ambas as árvores (Figs 23 e 24), as duas populações de *C. dichrous* constituem um clado com a população 1 de *C. pallidulus* var. *pallidulus*. Do mesmo modo, em ambas as árvores a população 2 de *C. dichrous* está mais agrupada do que a população 1, estando esta última, muito mais misturada à população 1 de *C. pallidulus* var. *pallidulus*.

É interessante notar o posicionamento de *C. pallidulus* var. *pallidulus*, pois em ambas as árvores as três populações se apresentam em clados diferentes, indicando o parafiletismo desse táxon. Em resumo, a população 1 de *C. pallidulus* var. *pallidulus* está mais próxima de *C. dichrous*, a população 3, mais próxima de *C. erythroxyloides* e a população 2 encontrando-se em várias posições nas duas árvores (Figs 23 e 24). Isso demonstra o grande polimorfismo desta espécie, sugerindo que ela tenha sido possivelmente a origem do complexo estudado.

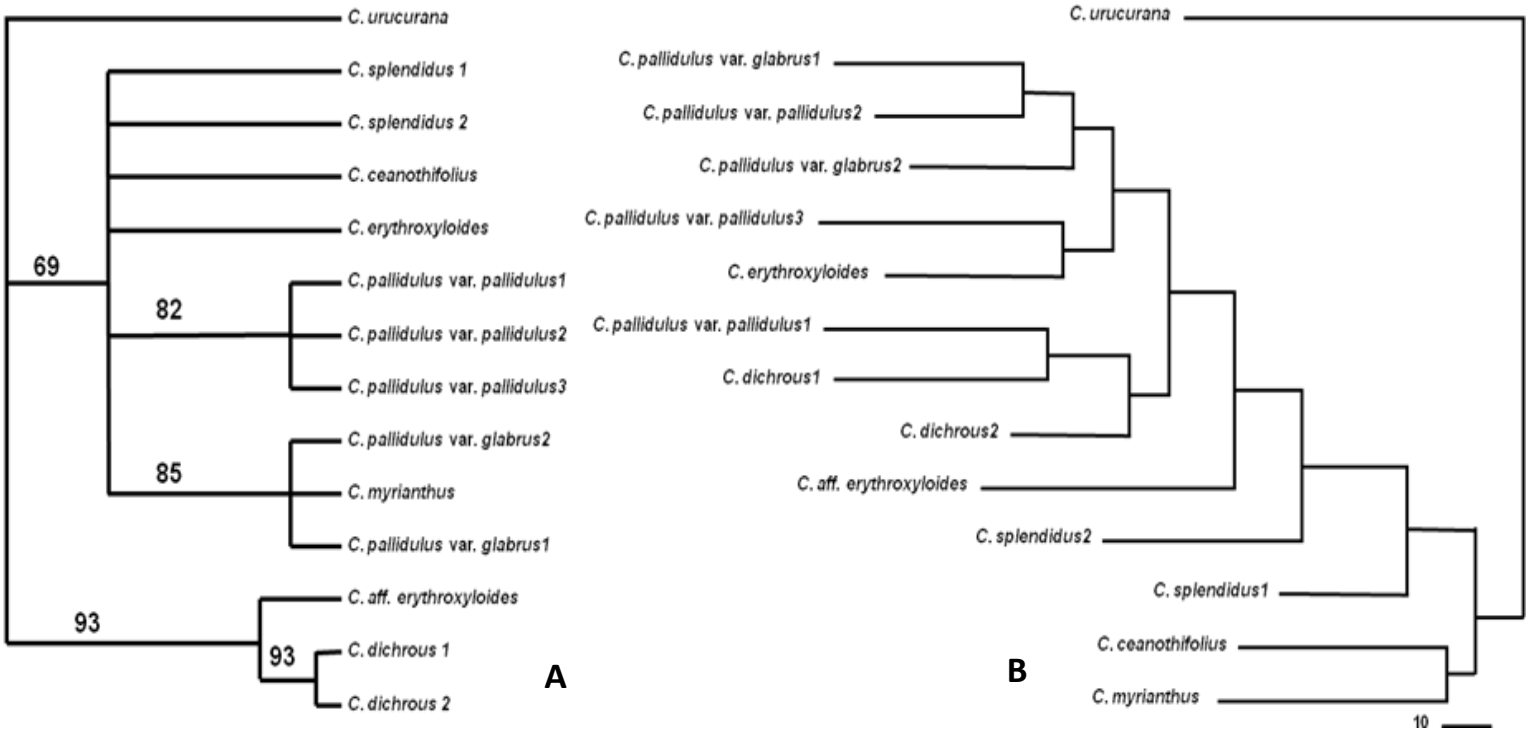


Figura 25. Árvore de máxima parcimônia utilizando sequências de DNA (**A**; resultados retirados da Fig. 16) e árvore de afinidades genéticas utilizando marcadores AFLP (**B**).

Para a elaboração da árvore de afinidades genéticas da Fig. 25 foram feitos todos os procedimentos de AFLP com os DNAs dos mesmos indivíduos utilizados para o sequenciamento. Isso é necessário para que se possam comparar as topologias das duas árvores.

Comparando-se as árvores da Fig. 25 pode-se perceber que os resultados da análise por AFLP diferem dos obtidos pela análise das sequências de DNA, pois nesta

a árvore está pouco resolvida, devido à grande politomia existente, enquanto pela análise por AFLP houve uma melhor resolução da maioria das espécies.

Um importante resultado deste trabalho se revela na diferenciação nítida entre as duas variedades de *C. pallidulus*, principalmente pela análise por AFLP. Na análise de máxima parcimônia, as duas variedades formaram uma politomia com *C. myrianthus* com BS 85. Deve-se lembrar, entretanto, que o número de caracteres informativos na análise de sequenciamento foi de 548, enquanto em AFLP esse número foi de 794; quanto maior o número de caracteres informativos, maior a confiabilidade na análise. Embora os trabalhos morfológicos de Lima & Pirani (2008) indiquem que a diferença entre ambas as variedades ocorre somente em relação ao indumento na face adaxial de *C. pallidulus* var. *glabrus*, os resultados tanto da análise por sequenciamento quanto por marcadores AFLP separam as duas variedades. *C. pallidulus* var. *pallidulus*, como já discutido anteriormente, permeia entre as outras espécies analisadas (Fig. 25B), a população 1 estando próxima da população 1 de *C. dichrous*, a população 2 estando próxima de *C. pallidulus* var. *glabrus*, e a população 3, próxima de *C. erythroxyloides*. Isto pode sugerir que se trata de duas espécies diferentes, e não somente duas variedades de uma mesma espécie.

Pode-se observar pela análise da Fig. 25 que o posicionamento de *C. dichrous* dentro do Complexo foi diferente nas duas análises. Na árvore de máxima parcimônia, *C.* aff*. erythroxyloides* emerge como grupo irmão das duas populações de *C. dichrous,* as quais emergem juntas com um excelente suporte (BS 93), enquanto na árvore de afinidades genéticas aquela espécie é grupo irmão de todo o clado onde se localiza *C. dichrous*. As populações de *C. dichrous*, pela análise de AFLP, saem em diferentes posições; a população 1 de *C. dichrous*, neste caso é mais afim à população 1 de *C. pallidulus* var. *pallidulus*, do que da população 2 de sua própria espécie.

Outra importante contribuição deste trabalho está sendo a diferenciação entre as duas amostras de *C. erythroxyloides*, uma delas denominada aqui como *C.* aff. *erythroxyloides*. De fato, tanto na análise de sequenciamento de regiões genômicas, quanto na de AFLP, essas duas amostras emergem em clados distintos, com elevados índices de suporte, parecendo não ter nenhuma ligação de parentesco entre elas. Aparentemente elas apresentam semelhanças morfológicas, mas talvez precisem ser investigadas com mais cuidado.

C. splendidus, *C. ceanothifolius* e *C. myrianthus* não estão bem resolvidas pela análise de AFLP, as duas primeiras formando uma politomia também pela análise de máxima parcimônia. Pode-se notar, analisando a Fig. 23 que estas três espécies

juntamente com *C. pallidulus* var. *pallidulus* formam um clado, embora com muito pouco suporte (<50). Allem (1978) havia sinonimizado *C. splendidus*, *C. ceanothifolius* e *C. pallidulus* em *C. migrans*, por julgar que estas espécies expressavam apenas variação de hábito de um único táxon.

4.3. Biogeografia do complexo

Os mapas foram obtidos do trabalho de Lima & Pirani (2008), e podem ser observados abaixo (Figuras 26 a 29), para cada espécie em separado.

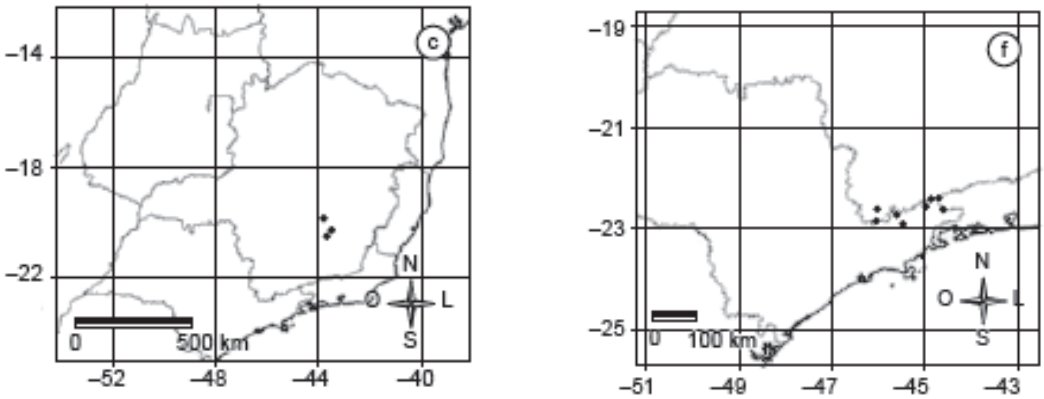


Figura 26. Mapa de distribuição das espécies *C. ceanothifolius* (esquerda e identificada por c) e *C. dichrous* (direita e identificada por f).

Na Fig. 26 observa-se a distribuição restrita ao estado de Minas Gerais de *C. ceanothifolius*, onde é endêmica; ocorrem em poucos municípios, como Ouro Branco, Ouro Preto e Sabará. Desse último há apenas uma coleta, em 1900, sem localidade específica (Lima & Pirani, 2008). Essa distribuição é bastante importante para distingui-la de *Croton pallidulus*, uma vez que ambas compartilham muitos caracteres morfológicos. Esta última, longe de ser endêmica, habita, preferencialmente, altitudes superiores a 1.000 m, em regiões serranas, nos diversos tipos de fisionomias de cerrado, como campo cerrado, campo sujo e cerrado *sensu strictu*, podendo habitar, portanto, diversos tipos de ambientes, como margens de cursos d'água, bordas de capoeiras e matas, campos úmidos, em solo areno-pedregoso. *C. dichrous* aparece somente na região Sudeste do Brasil, ocorrendo nos estados de São Paulo, Minas Gerais e Rio de Janeiro. De acordo com Lima & Pirani (2008) está sempre associada a localidades com altitudes elevadas, superiores aos 1.000 m, podendo ser encontrada até em regiões com elevações superiores aos 2.500 m. Ocorre em diversos tipos de ambientes como campo, campo sujo, campo rupestre e capão de mata, neste caso, especialmente em clareiras. Também foi coletada junto a cursos d'água.

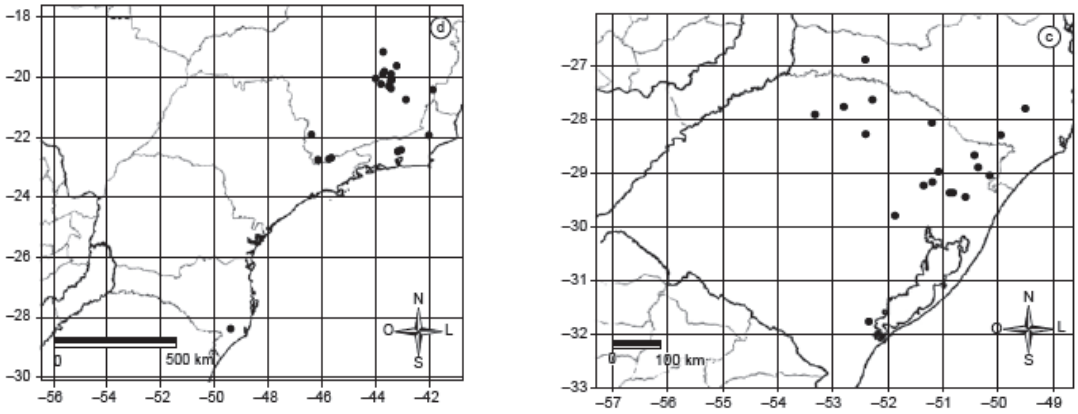


Figura 27. Mapas de distribuição das espécies *C.erythroxyloides* (esquerda e identificada por d) e *C. myrianthus* (direita e identificada por c).

C. erythroxyloides apresenta uma ampla distribuição geográfica, sendo encontrada nos estados de São Paulo (na região de São Bento do Sapucaí, na Pedra do Bauzinho), Rio de Janeiro, Minas Gerais e Santa Catarina (na Serra Geral e na Serra do Rio do Rastro); não ocorrendo, no estado de Paraná (Fig. 27). É encontrada, freqüentemente, em campos de altitude, campos sujos e campos rupestres, e está sempre associada a afloramentos rochosos, em regiões com altitudes superiores a 1.000 m. Isso sem dúvida explica a sua documentação geográfica fragmentada e pontual (Lima & Pirani, 2008). *C. myrianthus* apresenta distribuição restrita à região Sul do Brasil, com ocorrência nos estados de Santa Catarina e Rio Grande do Sul, em altitudes que variam entre 700 e 1.700 m. É encontrada, geralmente, em regiões de campos limpos ou campos sujos, próxima a afloramentos rochosos, na margem de rios, em locais úmidos, em capoeiras e na borda de matas nebulares, intactas ou desmatadas (Fig. 27B) (Lima & Pirani, 2008).

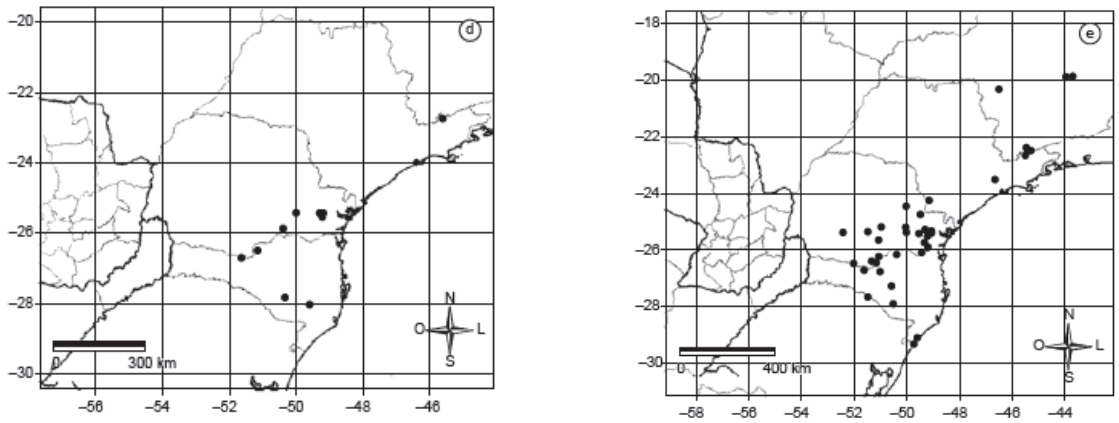


Figura 28. Mapas de distribuição das espécies *C. pallidulus* var. *glabrus* (esquerda e identificada por d) e *C. pallidulus* var. *pallidulus* (direita e identificada por e).

Pode-se observar analisando a Fig. 28 que *C. pallidulus* var. *glabrus* encontra-se somente nos estados de São Paulo, Paraná e Santa Catarina, preferencialmente, em áreas com baixa drenagem, como brejos e margem de rios, em locais com solo arenoso e com altitudes modestas; também foi coletada em borda de mata ciliar. *C. pallidulus* var. *pallidulus* apresenta distribuição principalmente na região Sul do Brasil, nos estados de Paraná, Santa Catarina e Rio Grande do Sul, sendo bastante freqüente no primeiro estado, ocorrendo em vários municípios. Em São Paulo, foi coletada no município de Campos do Jordão, e uma única vez em Itararé e na capital. Ocorre também em alguns municípios de Minas Gerais, em geral, naqueles próximos aos limites com o estado de São Paulo. Há somente uma coleta proveniente do Rio de Janeiro, contudo a ausência de dados na etiqueta da exsicata impossibilita saber em que município do estado a planta teria sido coletada. Habita locais, cujas altitudes variam entre 700 e 2.000 m, e pode ser encontrada nos mais diversos tipos de ambientes como afloramentos rochosos, campo arenoso, cerrado pedregoso, campo úmido, campo sujo, capão de mata, capoeira, campo com palmeiras do gênero *Butia* Becc. ("butiazal") e em borda de mata de galeria. Trata-se também de uma planta invasora, freqüente em regiões de campo degradado (Lima & Pirani, 2008).

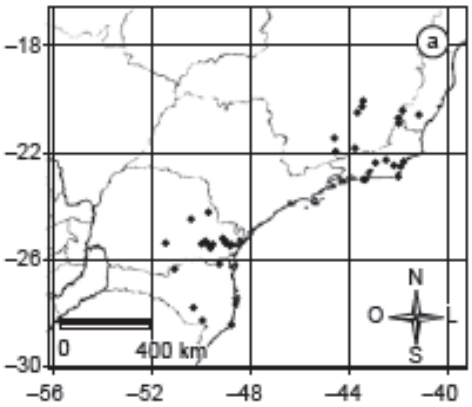


Figura 29. Mapa de distribuição da espécie *C. splendidus*. Retirado de Lima & Pirani (2008).

Observa-se na Fig. 29 que *C. splendidus* apresenta uma área de distribuição compreendendo os estados de Minas Gerais, Espírito Santo, Rio de Janeiro, Paraná e Santa Catarina, e de acordo com Lima & Pirani (2008) desenvolve-se, preferencialmente, como espécie de altitude entre os 1.000 e 1.600 m de altura. Entretanto, ocorre também ao nível do mar, nas restingas; porém, nesse tipo de ambiente, bem como nos campos litorâneos, ocorre esporadicamente como elemento raro e estranho. Trata-se de uma planta heliófila que ocorre nos mais diversos tipos de ambientes como restinga aberta, campo, campo rupestre, campo cerrado e borda de

capão. Também foi coletada em campos de altitude e junto a matas nebulares. Pode ser encontrada tanto em locais com solo arenoso, de restinga e margem de rios, quanto em ambientes com solo pedregoso, como o dos campos rupestres.

A espécie utilizada como grupo externo, *C. urucurana*, apresenta distribuição geográfica nos estados de Rio de Janeiro, Minas Gerais, São Paulo, Paraná, Santa Catarina, Rio Grande do Sul, Bahia e Mato Grosso do Sul (Lorenzi, 1992).

Com os dados disponíveis a respeito da distribuição geográfica das espécies analisadas neste trabalho, pode-se combiná-las com os dados obtidos através da análise por AFLP (ver Figs 23 e 24), técnica que forneceu melhores resultados do que aqueles obtidos pelo sequenciamento das três regiões genômicas (ver Figs 16 e 17).

Os dados extraídos das distribuições geográficas apresentadas acima estão representados graficamente no diagrama de afinidades obtido por AFLP (Fig. 25B) e mostrados na Fig. 30. Uma vez que a maioria das amostras utilizadas neste trabalho está localizada no domínio vegetacional da Mata Atlântica, que corresponderia à província biogeográfica Paranaense, segundo Cabrera & Willink (1973), pode-se utilizar uma mesma metodologia para expressar essa variação.

Estudos visando outras análises biogeográficas não seriam possíveis neste trabalho, devido ao alto grau de similaridade genética observada entre as espécies aqui utilizadas. Isto pode ser observado no filograma da Fig. 17; quanto menores as distâncias relativas entre as diferentes amostras, menores são os graus de diferenças genéticas observadas no sequenciamento. Além disso, o número de espécies utilizadas foi baixo, se comparado com a totalidade de espécies pertencentes a seção *Lamprocroton* (à qual pertencem as espécies estudadas) ou mesmo ao gênero como um todo.

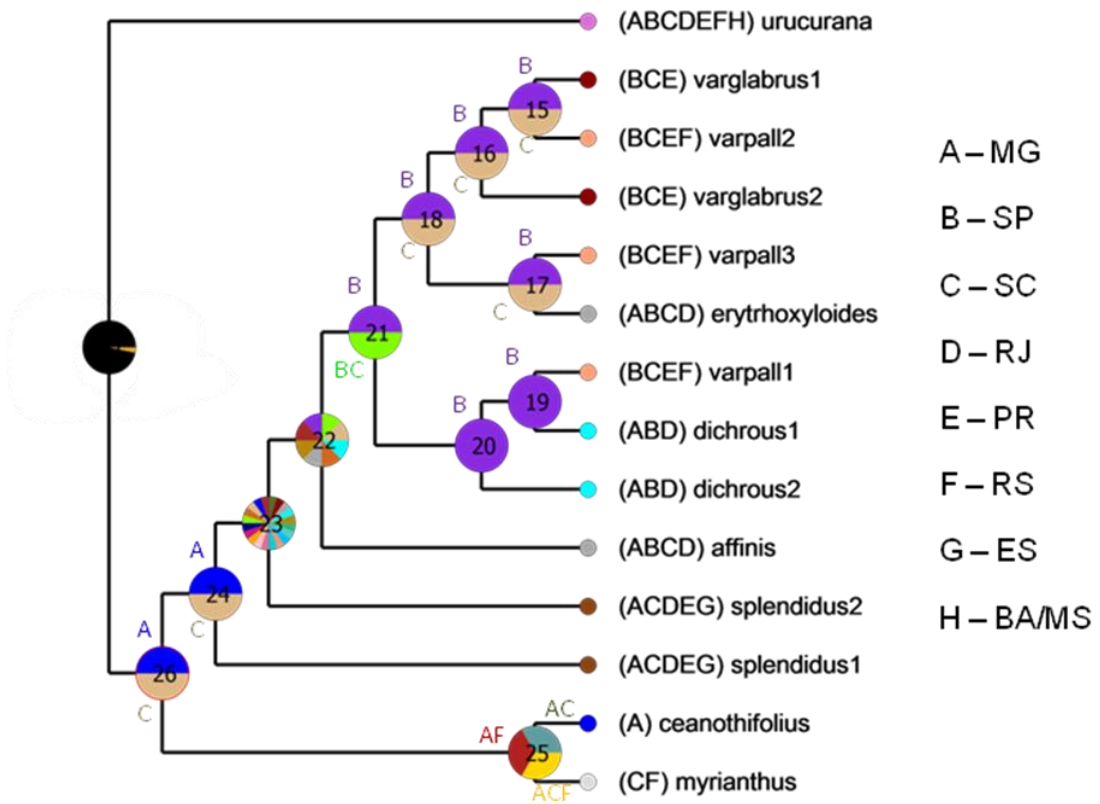


Figura 30. Topologia com as áreas ancestrais otimizadas, indicando áreas reconstruídas como de distribuição ancestral para os diferentes clados.

Os círculos coloridos do gráfico da Fig. 30 representam a probabilidade de distribuições ancestrais em cada nó. Em alguns casos, a otimização resulta em hipóteses alternativas de distribuição ancestral; por exemplo, no nó 25, verifica-se que a reconstrução da área de distribuição ancestral resultou em três hipóteses alternativas AF, AC e ACF. Para o nó 23, a presença de múltiplas cores indica uma reconstrução da área de distribuição ambígua, sem definição correta, pois muitas opções podem ser consideradas como áreas de distribuição ancestral para esse nó.

Os dados da Fig. 30 permitem inferir que um possível ancestral comum às espécies do Complexo, correspondente ao nó 26, provavelmente, ocupava uma área em A (Minas Gerais) ou em C (Santa Catarina), com igual probabilidade para os dois estados, pois as duas colorações dividem o círculo em dois semicírculos iguais. Isso pode ser explicado pelo fato de que a maioria das espécies apresenta distribuição geográfica em apenas um estado, ou nos dois; somente *C. dichrous* não ocorre no estado de Santa Catarina, e as variedades de *C. pallidulus*, bem como *C. myrianthus* não ocorrem no estado de Minas Gerais.

Através desta análise não se pode inferir qual a provável área ancestral comum às duas variedades de *C. pallidulus*, uma vez que há congruência nos dados de

distribuição, havendo uma diferenciação quanto a ocorrência de *C. pallidulus* var. *pallidulus* no estado de Rio Grande do Sul. Portanto, não se pode afirmar que houve uma possível difusão de uma das variedades, isto é, uma expansão gradual da área de distribuição acompanhando colonizações de áreas favoráveis próximas ponto a ponto, ou uma dispersão a longa distância; isto porque elas compartilham até mesmo o hábitat em que ocorrem.

Nos nós 19 e 20, há uma definição quanto a área de distribuição ancestral referente às espécies *C. dichrous* e *C. pallidulus* var. *pallidulus.* O estado de São Paulo parece ser a provável área de distribuição ancestral para este clado, uma vez que a primeira espécie ocorre somente na região Sudeste, e a segunda, na região Sul. Uma possível explicação deve-se à localização "privilegiada" do estado de São Paulo para ambas as espécies, podendo ter sido, dessa maneira, a "nascente" da distribuição de ambas as espécies.

A classificação da distribuição geográfica de *C.* aff*. erythroxyloides* seguiu a distribuição de *C. erythroxyloides*, o que não foi correto, uma vez que se verificou tratar-se de duas amostras com características diferentes.

As reconstruções de áreas ancestrais obtidas com o programa S-DIVA podem informar que o padrão atual de distribuição das espécies do Complexo se deve a um possível ancestral comum a todas elas. Esse ancestral pode ter ocupado uma ampla região, provavelmente, desde Minas Gerais até Santa Catarina, ou seja, entre as regiões Sudeste e Sul do Brasil, indicando uma distribuição ancestral ampla. Porém, devido a algum evento pontual, surgiram especiações, com padrões de distribuição diferenciados para cada espécie. Entretanto, certas características morfológicas estão sendo ainda compartilhadas por várias espécies. Algumas delas, como *C. erythroxyloides* e *C. splendidus* apresentam um padrão de distribuição amplo, enquanto em outras ele é mais específico, até se tornar endêmico, como é o caso de *C. ceanothifolius*.

Um estudo biogeográfico ideal seria possível, por exemplo, com a união de dados de todas as espécies da seção *Lamprocroton*, melhor do que a utilização apenas de dados apenas de um grupo, pois somente dessa forma seria possível considerar todos os viés, e também as relações de parentesco entre as diferentes espécies, não somente aquelas do Complexo. Com isso, poder-se-ia inferir um estudo filogeográfico com resultados mais interessantes e completos.

5. Resumo

O projeto abordou a detecção de polimorfismos genéticos num Complexo formado por algumas espécies de ampla ocorrência no Sudeste-Sul do Brasil, cuja delimitação ainda não está muito bem estabelecida, por apresentar grande polimorfismo morfológico que suscita dúvidas quanto a se tratar de uma única ou de várias espécies. Foram estudadas as seguintes espécies: *C. ceanothifolius*, *C. dichrous*, *C. erythroxyloides*, *C.* aff. *erythroxyloides*, *C. myrianthus*, *C. pallidulus* var. *pallidulus*, *C. pallidulus* var. *glabrus* e *C.* splendidus. A proposta do presente trabalho foi avaliar a consistência das espécies estudadas, por meio de análise filogenética baseada em sequências de *ITS* e duas regiões do DNA do cloroplasto (*trn*L-F e *rps*16), e por meio de marcadores AFLP (Amplified Fragment Length Polymorphism). Como grupo externo foi utilizado *C. urucurana.* Como objetivo adicional procurou-se estudar filogeograficamente as espécies do Complexo, por meio de análises de populações disjuntas, distribuídas desde o Sudeste até o Sul do país. Inferências filogenéticas foram realizadas a partir das seqüências de DNA, realizadas através dos critérios de máxima parcimônia e probabilidades posteriores (análise bayesiana). Os resultados mostraram a grande união da maioria das espécies do Complexo em uma politomia; as populações de *C. dichrous* e *C. aff. erythroxyloides* surgiu separadamente em um clado, assim como duas populações de *C. pallidulus* var. *glabrus* com *C. myrianthus*, além de três populações de *C. pallidulus* var. *pallidulus* e outras espécies do Complexo emergirem em outro clado. Pela técnica de AFLP, obtiveram-se fragmentos polimórficos, utilizados como caracteres nas relações de afinidade genética através da análise Neighbor-Joining (realizada no programa PAUP). Observou-se que as duas variedades de *C. pallidulus* emergem em clados diferentes, enquanto as duas populações de *C. pallidulus* var. *glabrus* emergem em um mesmo clado. *C. pallidulus* var. *pallidulus* apresentou-se como táxon parafilético, na medida em que aparece agrupada com diferentes espécies, resultado este que confirma a natureza distinta das duas variedades. *C. erythroxyloides* e *C.* aff. *erythroxyloides* mostraram, através dos dois métodos de análises moleculares, ser espécies distintas. *C. splendidus*, *C. ceanothifolius* e *C. myrianthus* não estão bem resolvidas, misturando-se, muitas vezes com *C. pallidulus* var. *pallidulus*. Por fim, foi verificada se havia congruência entre as topologias obtidas com a análise baseada na combinação dos dados de sequenciamento com aquela obtida por AFLP. O estudo de biogeografia indicou que a área de distribuição ancestral do grupo provavelmente é ampla, localizando-se entre Minas Gerais e Santa Catarina, uma vez que a distribuição das espécies do complexo se dá preferencialmente nestes estados.

6. Abstract

The project addressed the detection of genetic polymorphisms in a Complex formed by some species of widespread occurrence in southeast of Brazil, whose limits are not yet well established, by presenting great morphological polymorphism that raises doubts as to whether it is a single or various species. The following species were studied: *C. ceanothifolius*, *C. dichrous*, *C. erythroxyloides*, *C. aff. erythroxyloides*, *C. myrianthus*, *C. pallidulus* var. *pallidulus*, *C. pallidulus* var. *glabrus* and *C. splendidus*. The purpose of this study was to evaluate the consistency of the species studied by means of phylogenetic analysis based on sequences of *ITS* and two chloroplast DNA regions (*trn*L-F and *rps*16), and by AFLP markers (Amplified Fragment Length Polymorphism). Was used as outgroup *C. urucurana*. As additional objectives, we tried to study the phylogeography of the species complex, through analysis of disjunct populations distributed from the Southeast to the South. Phylogenetic inferences were made from DNA sequences, performed by the criteria of maximum parsimony and posterior probabilities (Bayesian analysis). The results were a great union of most species of the Complex in a polytomy; the populations of *C. dichrous* and *C. aff. erythroxyloides* emerged separately in a clade, apart from two populations of *C. pallidulus* var. *glabrus* with *C. myrianthus*, besides the three populations of *C. pallidulus* var. *pallidulus* and other species of Complex emerging in another clade. For the AFLP technique, we obtained polymorphic fragments used as characters in the genetic affinity relationships through the neighbor-joining analysis (performed using the PAUP). It was observed that the two varieties of C. *pallidulus* emerge in different clades, while the two populations of *C. pallidulus* var. *glabrus* emerge in the same clade, *C. pallidulus* var. *pallidulus* appeared as a paraphyletic taxon, as it appears grouped with different species, a result that confirms the distinct nature of these two varieties. *C. erythroxyloides* and *C. aff. erythroxyloides* shown by the two methods of molecular analysis, to be distinct species. *C. splendidus*, *C. ceanothifolius* and *C. myrianthus* are not well resolved, blending, often with *C. pallidulus* var. *pallidulus*. Finally, there was congruence between the topologies obtained with the analysis based on combined sequencing data with that obtained by AFLP. The study of biogeography indicated that the distribution area of the ancestral group is probably wide, situated between Minas Gerais and Santa Catarina, since the distribution of species of the complex occurs preferentially in those states.

7. Referências Bibliográficas

Albach D.C., Chase M.W. 2004. Incongruence in Veroniceae (Plantaginaceae): evidence from two plastid and a nuclear ribosomal DNA region. *Molecular Phylogenetics and Evolution* 32:183-197.

Albach D.C., Meudt H.M., Oxelman B. 2005. Piecing together the "new" Plantaginaceae. *American Journal of Botany* 92(2): 297-315.

Allem A.C. 1978. Preliminares para uma abordagem taxonômica do gênero *Croton* L. (Euphorbiaceae) no Rio Grande do Sul, Brasil. Bol. Mus. Bot. Munic. 34:1-33.

Allem A.C. 1979. Quatro novas espécies de *Croton* L. (Euphorbiaceae) do Rio Grande do Sul, Brasil. *Boletin Soc. Argent. Bot*. 18 (3/4): 63-81.

Asmussen C.B., Chase M.W. 2001. Coding and noncoding plastid DNA in Palm systematics. *American Journal of Botany* 88 (6): 1103-1117.

APG (Angiosperm Phylogeny Group) 1998. An ordinal classification for the families of flowering plants. *Annals of Missouri Botanical Garden* 85: 531-553.

APG II. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141: 399-436.

APG III 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105-121.

Baillon H. 1864. Euphorbiacées Américaines. *Adansonia* 4:257-377.

Baldwin B.G., Sanderson M.J., Porter J.M., Wojciechowski M.F., Campbell C.S., Donoghue M.J. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of Missouri Botanical Garden* 82: 247-277.

Barroso G.M. 1991. *Sistemática das angiospermas do Brasil*. Imprensa Universitária. Viçosa, UFV.

Beismann H., Barker J.H.A., Karp A., Speck T. 1997. AFLP analysis sheds light on distribution of two *Salix* species and their hybrid along a natural gradient. *Molecular Ecology*. 6: 989-993.

Bentham G.M. 1880. Notes on Euphorbiaceae. London, Bot. J. Linn. Soc., 37: 185-267.

Berry P.E., Hipp A.L., Wurdack K.J., Ee B.V., Riina R. 2005. Molecular phylogenetics of the giant genus *Croton* and tribe Crotoneae (Euphorbiaceae sensu stricto) using ITS and *trnL-trnF* DNA sequence data. *American Journal of Botany* 92: 1520-1534.

Cabrera A.L., Willink A. 1973. *Biogeografia de America Latina*. Washington: Secretaria General de la Organización de los Estados Americanos.

Carneiro D.S., Cordeiro I., França F. 2002. A Família Euphorbiaceae na flora de *inselbergs* da região de Milagres, Bahia, Brasil. *Bol. Bot. Univ. São Paulo* 20: 31-47.

Chase M.W., Soltis D.E., Olmstead R.G., Morgan D., Les D.H., Mishler B.D., Duvall M.R., Price R.A., Hills H.G., Qiu Y.L., Kron K.A., Rettig J.H., Conti E., Palmer J.D., Manhart J.R., Sytsma K.J., Michaels H.J., Kress W.J., Karol K.G., Clark W.D., Hedren M., Gaut B.S., Jansen R.K., Kim K.J., Wimpee C.F., Smith J.F., Furnier G.R., Strauss S.H., Xiang Q.Y., Plunkett G.M., Soltis P.S., Swensen S.M., Williams S.E., Gadek P.A., Quinn C.J., Eguiarte L.E., Golenberg E., Learn G.H., Graham S.W., Barrett S.C.H., Dayanandan S., Albert V.A. 1993. Phylogenetics of seed plants - An analysis of nucleotide-sequences from the plastid gene *rbc*L. *Annals of the Missouri Botanical Garden* 80 (3): 528-580.

Coley P.G., Bryant J.P., Chapin F.S. 1985. Resource availability and plant antiherbivore defense. *Science* 230: 895-899.

Croizat L. 1944. Additions to the genus *Croton* in South America. *Darwiniana* 6: 442-468.

Cronquist A. 1981. *The evolution and classification of flowering plants*. New York Botanical Garden. Bronx, N.Y., USA

Dahlgren R, Clifford H.T, Yeo P.F. 1985. *The families of the monocotyledons: structure, evolution, and taxonomy*. Berlin; New York: Springer-Verlag.

Doolittle W.F. 1999. Phylogenetic classification and the universal tree (taxonomies based on molecular sequences). *Science* 284: 2124-2126.

Doyle J.J., Doyle J.S. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull*. 19: 11–15.

Emshwiller E., Doyle J.J. 1999. Chloroplast-expressed glutamine synthetase (ncpgs): potential utility for phylogenetic studies with an example from O*xalis* (Oxalidaceae). *Molecular Phylogenetics and Evolution* 12: 310–319.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.

Ferreira M.E., Grattapaglia D. 1996. *Introdução ao uso de marcadores moleculares em análise genética*. 2 ed. Embrapa-Cenargen, Brasília.

Gonçalves E.G. 2002. *Sistemática e evolução da tribo Spathicarpeae (Araceae)*. Tese de doutorado. Universidade de São Paulo, São Paulo, SP.

Gottlieb A., Gilberti G., Poggio L. 2005. Molecular analyses of the genus *Ilex* (Aquifoliaceae) in southern south america, evidence from AFLP and *ITS* sequence data. *American Journal of Botany.* 92(2): 352–369.

Gravendeel B., Chase M.W., Vogel E.F., Roos M.C., Mês T.H.M., Bachmann R. 2001. Molecular phylogeny of *Coelogyne* (Epidendroideae: Orchidaceae) based on plastid RFLPs, *mat*K, and nuclear ribosomal ITS sequences: evidence for polyphyly. *American Journal of Botany* 88 (10): 1915-1927.

Guimarães L.A.C. 2006. *O gênero Croton L. da seção Cyclostigma Griseb. e seção Luntia (Raf.) G. L. Webster (Euphorbiaceae) ocorrentes na Amazônia Brasileira*. Dissertação de Mestrado. Universidade Federal Rural da Amazônia. Belém, AM.

Hall B.G.2001. Phylogenetic trees made easy: a how-to manual for molecular biologists. Sinauer Associates Inc. Massachusetts, USA.

Hahn W.J. 2002. A phylogenetic analysis of the Arecoid line of palms based on plastid DNA sequence data. *Molecular Phylogenetics and Evolution* 23: 189-204.

Hodkinson T.R., Renvoize S.A., Ni Chonghaile G., Stapleton C., Chase M.W. 2000. A comparison of ITS Nuclear rDNA sequence data and AFLP markers for phylogenetic studies in *Phyllostachys* (Bambusoideae - Poaceae). *Journal of Plant Research.* 113: 259-269.

Judd W.S., Campbell C.S., Kellogg E.A., Stevens P.F. 1999. *Plant Systematics – a phylogenetic approach*. Sinauer Associates, Inc. Massachusetts, USA.

Karp A., Seberg O., Buiatti, M. 1996. Molecular techniques in the assessment of botanical diversity. *Annals of Botany* 78: 143-149.

Kardolous J.P., And Eck H.J., Van der Berg R.G. 1998. The potential of AFLPs in biosystematics: a first application in *Solanum* taxonomy (Solanaceae*). Plant Systematic Evolution* 210: 87-103.

Koopman W.J.M. 2005. Phylogenetic signal in AFLP data sets. *Systematic Biology* 54: 197–217.

Lima L.R. 2006. *Estudos Taxonômicos em* Croton *seção Lamprocroton (Müll. Arg.) Pax (Euphorbiaceae).* Tese de doutorado. Universidade de São Paulo, São Paulo, SP.

Lima L.R., Pirani J.R. 2008. Revisão taxonômica de *Croton* sect. *Lamprocroton* (Müll. Arg.) Pax (Euphorbiaceae s.s.). *Biota Neotropica* 8(2): 1-56.

Lohmann L.G. 2006. Untangling the phylogeny of neotropical lianas (Bignonieae, Bignoniaceae). *American Journal of Botany* 93(2): 304-318.

Lorenzi H. 1992. *Árvores brasileiras:* manual de identificação e cultivo de plantas arbóreas nativas do Brasil. Plantarum, Nova Odessa.

Lucena M.F.A. 2000. *Estudos Taxonômicos do gênero Croton L. (Crotonoideae–Euphorbiaceae) nas zonas do Litoral e da Mata do estado de Pernambuco-Brasil*. Dissertação de mestrado. Universidade Federal Rural de Pernambuco, Recife, PE.

Malcomber S.T. 2002. Phylogeny of *Gaertnera* Lam. (Rubiaceae) based on multiple DNA markers: evidence of a rapid radiation in a widespread morphologically diverse genus. *Evolution* 56(1): 42-57.

Mason-Gamer R.J., Kellogg E.A. 1996. Testing for Phylogenetic Conflict Among Molecular Data Sets in the Tribe Triticeae (Gramineae). *Systematic Biology* 45(4): 524-545.

Mello-Silva R. ; Santos D.Y.A.C. ; Salatino M.L.F. ; Mota L. B. ; Cattai M.B. ; Sasaki D. ; Lovo J. ; Pita P. B. ; Rocini C. ; Rodrigues C. D. N. ; Zarrei M. ; Chase M. W. 2011. Five vicariant genera from Gondwana: the Velloziaceae as shown by molecules and morphology. *Annals of Botany* (Print) 108(1): 87-102.

Mueller U.G. & Wolfenbarger L. 1999. AFLP genotyping and fingerprinting. *Trends in Ecology & Evolution* 14(10): 389-394.

Mueller J. 1866. Euphorbiaceae exceto subordo Euphorbieae. In Prodromus systematis universalis regni vegetabilis (A.P. Candolle, ed.). v.15, pars 2, p.189-1286.

Mueller J. 1873. Euphorbiaceae. *In C.F.P. Martius & A.G. Eichler (eds) Flora brasiliensis* 11(2): 1-292, tab. 1-42.

Nei M., Li W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Science of USA.* 76: 5269–5273.

Nowicke J.W. 1994. A palynological study of Crotonoideae (Euphorbiaceae). *Annals of Missouri Botanical Garden* 81: 245-269.

Olmstead R.G., Reeves P.A. 1995. Evidence for the polyphyly of the scrophulariaceae based on chloroplast rbcL and ndhF sequences. *Annals of Missouri Botanical Garden* 82: 176-193.

Oxelman B., Lidén M., Berglund D. 1997. Chloroplast *rps16* intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Systematics and Evolution* 206: 393–410.

Oxelman B., Bremer B. 2000. Discovery of paralogous nuclear gene sequences coding for the second-largest subunit of rna polymerase ii (*rpb2*) and their hylogeneticutility in Gentianales of the Asterids. *Molecular Biology and Evolution* 17(8): 1131-1145.

Pereira A.D., Carbonell S.A., Neto F.R.D., Amaral A.C.F., Barnes R.A. 2002. High-termperature gas chromatography-mass spectrometry with glass capillary columns for the screening of natural products. *Journal of Chromatography A* 947: 255-265.

Pieters L.A., De-Bruyne T. 1995. *In vivo* wound healing activity the drago's blood (Croton spp*). A Tradicional Americans Drug and Constituents Phytomedicine*, 2(1):12-22.

Popp M., Oxelman B. 2001. Inferring the history of the polyploid *Silene aegaea* (Caryophyllaceae) using plastid and homoeologous nuclear DNA sequences. *Molecular Phylogenetics and Evolution* 20(3): 474–481.

Punt W. 1961. Pollen morphology of the Euphorbiaceae with special reference to taxonomy. *Wentia* 7: 1-116.

Riesberg L.H., Soltis D.E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends Plant* 5 (1): 65-84.

Riina R., Berry P.E., Van Ee B.W. 2009. Molecular phylogenetics of the Dragon´s Blood *Croton* section *Cyclostigma* (Euphorbiaceae): A Polyphyletic Assemblage Unraveled. *Systematic Botany* 34(2): 360-374.

Riina R., Cordeiro I., Amorim A. & Berry P.E**.** 2010*. Croton thomasii* Riina & P.E. Berry (Euphorbiaceae), a new species from the Atlantic Forest in the State of Bahia (Brazil) and typification of *Croton sapiifolius* Müll. Arg. *Candollea* 65: 101--107.

Roa A.C., Maya M.M., Duque M.C., Tohme J., Allem A.C., Bonierbale E.M.W. 1997. AFLP analysis of relationships among cassava and other *Manihot* species. *Theoretical and Applied Genetics*. 95: 741-750.

Ronquist F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* 46: 195-203.

Ronsted N., Chae M.W., Albach D.C., Bello M.A. 2002. Phylogenetics relationships within Plantago (Plantaginaceae): evidence from nuclear ribosomal ITS and plastid trnL-F sequence data. *Botanical Journal of the Linnean Society* 139: 323-338.

Sadava D., Heller H.C., Orians G.H., Purves W.K. & Hillis D.M. 2009. *Vida. A ciência da Biologia. Vol. II: Evolução, diversidade e ecologia.* Ed. 8. ARTMED, Porto Alegre.

Sambrook J., Fritsch E.F., Maniatis T. 1989. *Molecular Cloning: a laboratory manual*. 2 ed. Cold Spring Harbor Laboratory Press, New York.

Salatino A., Salatino M.L.F., Mello-Silva R., Sluys M-A.V., Giannasi D.E., Price R.A. 2001. Phylogenetic inference in Velloziaceae using chloroplast *trn*L-F sequences. *Systematic Botany* 26 (1): 92-103.

Salatino A., Salatino MLF, Negri G. 2007. Traditional uses, chemistry and pharmacology of *Croton* species. *Journal of the Brazilian Chemical Society* . 18(1): 11-33.

Saitou N., Nei M. 1987 The neighbour joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology & Evolution*. 4: 406–425.

Schmidt-Lebuhn A.N., Kessler M., Müller J. 2005. Evolution of *Suessenguthia* (Acanthaceae) inferred from morphology, AFLP data, and ITS rDNA sequences. *Organisms Diversity & Evolution*. 5: (1) 1-13.

Schmidt-Lebuhn A.N., Kessler M., Kumar M. 2006. Promiscuity in the Andes: Species Relationships in *Polylepis* (Rosaceae, Sanguisorbeae) Based on AFLP and morphology. *Systematic Botany.* 31(3): 547–559.

Secco R. S. 1992. Notas sobre as lianas do gênero Croton L. (Euphorbiaceae). *Boletim do Museu Paranaense Emílio Goeldi – Botânica*. Pará. 8(2): 265-281.

Secco R. S. 2005. Flora da reserva Ducke, Amazonas, Brasil: Euphorbiaceae – Parte I. *Rodriguésia*. 56(86): 143-168.

Shepherd G.J. 1996. FITOPAC 1. *Manual de usuário*. Departamento de Botânica, UNICAMP.

Smith L.B., Downs R.J., Klein R.M. 1988. Euphorbiaceae – *Croton* L. *Fl. Ilustr. Catarin., Fasc. EUFO*. Pp.62-137.

Steinmann V.W., Porter J.M. 2002. Phylogenetic relationships in Euphorbieae (Euphorbiaceae) based on ITS and *ndh*F sequence data. *Annals of Missouri Botanical Garden* 89 (4): 453-490.

Swofford D.L. 2002. *PAUP*: phylogenetic analysis using parsimony (*other methods), version 4.0*. Sinauer Associates, Sunderland, Massachusetts, USA.

Taberlet P., Gielly L., Pantou G., Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105–1109.

Takhtajan A. 1986. *Floristic regions of the world*. University of California Press, Berkeley.

Takhtajan A. 1991. *Evolutionary trends in flowering plants*. New York: Columbia Univ Press.

Van Ee B.W., Jelinski N., Berry P.E., Hipps A.L. 2006. Phylogeny and biogeography of Croton alabamensis (Euphorbiaceae), a rare shrub from Texas and Alabama, using DNA sequence and AFLP data. *Molecular Ecology* 15: 2735-2751.

Urbatsch L. E., Baldwin B.G., Donoghue M.J. 2000. Phylogeny of the Coneflowers and Relatives (Heliantheae: Asteraceae) Based on Nuclear rDNA Internal Transcribed Spacer (ITS) Sequences and Chlorplast DNA Restriction Site Data. *Systematic Botany* 25(3): 539-565.

Van Ee, B. W., Riina, R., Berry, P.E. 2011. A revised infrageneric classification and molecular phylogeny of New World *Croton* (Euphorbiaceae). *Taxon* 60(3): 791-823.

Veloso H. P., Rangel Filho A. L. R. & Lima J. C. A. 1991. *Classificação da Vegetação brasileira adaptada a um sistema universal.* Rio de Janeiro: IBGE.

Vos P., Hogers R., Bleeker M., Reijans M., Van de Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M., Zabeau M. 1995. AFLP: a new technique for DNA fringerprinting. *Nucleic Acids Research* 23: 4407-4414.

Vos P. and Kuiper M. (1997) AFLP analysis. *DNA Markers:Protocols, Applications, and Overviews* (Caetano-Anollés, G. and Gresshoff, P.M., eds): 115–131, Wiley.

Wallander E., Albert V.A. 2000. Phylogeny and classification of Oleaceae based on rps16 and trnL-F sequence data. *American Journal of Botany* 87 (12): 1827-1841.

Webster G.L. 1993. A provisional synopsis of the sections of the genus *Croton* (Euphorbiaceae). *Taxon* 42: 793-823.

Webster G.L. 1994. Synopsis of the genera and suprageneric taxa of Euphorbiaceae. *Annals of the Missouri Botanical Garden* 81: 33-144.

White T.J., Bruns T., Lee S., Taylor J.W. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315-322 In: *PCR Protocols: A Guide to Methods and Applications*, eds. Innis, M. A., D. H. Gelfand, J. J. Sninsky, and T. J. White. Academic Press, Inc., New York.

Williams J.G.K., Kubelik A.R., Livak J., Rafalski J.A., Tingey S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*. 18: 6531-6535.

Yu,Y. Harris A.J., He X.J. 2010. S-DIVA (Statistical dispersal-vicariance analysis): a tool for inferring biogeographic histories. *Molecular Phylogenetics and Evolution* 56 (2): 848-850.

Yu Y., Harris A.J., He X.J. 2011. RASP (Reconstruct Ancestral State in Phylogenies). Available at http://mnh.scu.edu.cn/soft/blog/RASP. Acesso em: 18 de agosto de 2011.

Zhang L.-B., Comes H.P. & Kadereit J.W. 2001. Phylogeny and quarternary history of the European montane/alpine endemic *Soldanella* (Primulaceae) based on ITS and AFLP variation. *American Journal of Botany.* 88: 2331–2345.

8. ANEXOS

Os anexos abaixo correspondem a análises utilizando a plataforma Windows.

Anexo 1. Análise de máxima parcimônia utilizando o programa PAUP

Para esta análise será necessário ter instalado no computador dois programas:

- BioEdit
- PAUP

1. Preparação para a matriz

No programa BioEdit deve-se alinhar todas as seqüências, editar o nome das espécies (cuidado para não apresentar nomes idênticos, caso haja diversos indivíduos de uma mesma espécie identificá-los de modo diferente) e codificar os eventos de indels caso queira. Com os indels codificados, anote num papel todas as posições de começo e final desses eventos, assim como regiões ambíguas.

Os espaço entre os nomes e identificações não podem ser separados por espaço ( ) ou hífen (-). O ideal é utilizar o "underline" (_). Colocar o terminal que será considerado grupo externo no topo da matriz.

Para montar a base da sua matriz escolha à opção **file → export → tab-delimited text (title.tab.sequence) →** digitar o nome do arquivo _____.txt

2. Construindo sua matriz

Abrir o arquivo ____.txt dentro do programa WordPad ou Bloco de Notas. Pode-se então visualizar (nome – tabulação – seqüência):

nome1
NNNN

NNNN

Para a análise de parcimônia pode-se utilizar o seguinte bloco de comando digitado no WordPad ou Bloco de notas:

```
Begin PAUP;
set autoclose=yes warnreset=no increase=auto;
set criterio=parsimony;
hsearch nreps=1000 nchuck=20 chuckscore=1 addseq=random;
savetrees brlens=yes file= ____hsearch.tre replace=yes;
contree  /strict=yes treefile= ____con.tre percent=50 replace=yes;
log start=yes file= ____.log replace=yes;
describetrees;
bootstrap search= heuristic nreps=1000 conlevel=50;
savetrees from=1 to=1 file= ____boot.tre
savebootP= nodelabels maxdecimals=1 replace=yes;
log stop;

END;
```

Através desse bloco de comando será possível realizar diversas etapas como: busca através do método de parcimônia, salvando as melhores árvores; obter e salvar o consenso estrito; obter a descrição das árvores sobre os caracteres, CI e RI; e obter os valores de “bootstrap” para os ramos.

Neste bloco de comando o critério selecionado foi à parcimônia (linha 3), sendo descrita na linha seguinte os seus parâmetros com: 1000 réplicas, retenção de 20 árvores para cada busca e adição das seqüências de modo aleatório (randômico). A linha 6 mostra o comando para o consenso estrito e o arquivo que será salvo (treefile= _____con.tre). Na posição onde se encontra o “underline” substitua pelo nome desejado para o arquivo.

Na linha 7 será criado um arquivo do tipo log, onde serão guardadas todas as informações das próximas linhas de comando. Na linha 8 serão descritos os dados sobre a árvore. Nas próximas linhas serão gerados os valores de “bootstrap” e sendo salvo um de suas árvores.

4. Executando a matriz no PAUP

Cole os seus arquivos da matriz e do bloco de comando na raiz do programa. Abra o programa PAUP e execute sua matriz da seguinte forma:

execute nomedasuamatriz.txt; <enter>

Verifique se o programa aceitou sua matriz mostrando sua descrição.

Processing of file “nomedoarquivo.txt” begins...
Data read in DNA format
Data matrix has xx taxa, xxxx characters
xxx characters excluded via “Eliminate” command.
Valid character-state symbols: 012..89ACGT
Missing data identified by ‘?’
Gaps identified by ‘-‘
“Equate” macros in effect:

R,r = => {AG}
Y,y = => {CT}
M.m = => {AC}
K,k = => {GT}
S,s = => {CG}
W,w = => {AT}
H,h = => {ACT}
B,b = => {CGT}
V,v = => {ACG}
D,d = => {AGT}
N,n = => {ACGT}

Processing of file “nomedoarquivo.txt” completed.

A seguir execute o bloco de comando da seguinte forma:

exe nomedoblocodecomando.txt; <enter>

Ao terminar todos os comandos digitar **quit;** <enter>

obs: A árvore a ser utilizada é a _____.con porque ela representa o consenso estrito, e os valores de bootstrap a serem plotados para a análise de máxima parcimônia serão obtidos no arquivo log criado na raiz do programa PAUP.

Anexo 2. Análise bayesiana utilizando o programa MrBayes

Para esta análise será necessário ter instalado no computador três programas:

- PAUP
- MrMODELTEST
- MrBAYES.

1. Abrir o programa PAUP.

2. Executar sua matriz com os dados moleculares. Para isso selecionar a opção **FILE → OPEN**. A matriz deve conter um cabeçalho utilizado para o programa PAUP (versão windows) ou digitar a seguinte linha de comando (versão DOS): **execute nomedoarquivo.nex;** . (O arquivo deve estar na mesma pasta que o programa Paup).

3. Copiar o arquivo MrModelblock dentro da pasta do programa MrModeltest2. Colar este arquivo na mesma pasta que o programa Paup. Executar o arquivo Mrmodelblock, digitando a linha de comando: **execute mrmodelblock;**

4. Irão aparecer na tela os diversos testes realizados pelo programa (os modelos evolutivos).

5. Serão criados dois arquivos **mrmodel.scores** e **mrmodelfit**. Estes arquivos podem ser renomeados se achar necessidade.

6. Procurar a pasta onde foi salva o arquivo **mrmodel.scores**, provavelmente na pasta do programa PAUP. Copiar o arquivo.

7. Cole o arquivo na raiz do Windows (C:). Nesse mesmo local também deve estar localizado o programa MrModeltest2.

8. Abrir o **prompt de comando**, para isso clicar em **Iniciar → Todos os programas → Acessórios → Prompt de comando**. A tela de DOS do Windows abrirá.

9. Para chegar até a raiz (C:), dê o comando **cd** e tecle **enter** para chegar à raiz C:/.

10. Digitar a linha de comando **mrmodeltest2 < mrmodel.scores > modelos. out** e tecle **enter**. Aparecerá uma mensagem **Program is done**.

11. Digitar **exit** <enter>.

12. Abrir o arquivo **modelos .out** no bloco de notas. No documento haverá dois blocos de comandos: um para o PAUP para se fazer análise de Máxima verossimilhança e depois o bloco para análise bayesiana **(MrBayes)**. Procurar a

análise referente a hLRT (HIERARCHICAL LIKELIHOOD RATIO TESTS). Observar no final do bloco hLRT a linha de comando referente à análise bayesiana (MrBayes). Copiar desde alinha que começa com a chave (**[!**) até o comando **END;**.

```
[!
MrBayes settings for the best-fit model (xxx+x+x) selected by hLRT in MrModeltest 2.2
]
BEGIN MRBAYES;
        Prset statefreqpr=xxxxxxxxxx;
        Lset  nst=x  rates=xxxxxx;
END;
```

13. Abrir a matriz rodada no PAUP no bloco de notas e colar as linhas de comando no final do arquivo. Deletar a linha de comando **OPTIONS GAPMODE=MISSING;**. Se houver a linha de comando **Eliminate**, substituir pela palavra **Exclude.** Recortar e colar depois do END da matrix e antes das linhas de comando para análise bayesiana. Se houver a linha de comando SYMBOLS "0 1 2 3 ..." deletar. Salvar a matriz num novo arquivo, por exemplo, matrizbay.txt. Fechar o bloco de notas.

14. Abrir o programa MrBayes.

15. Copiar o arquivo **matrizbay.txt** para a pasta do programa MrBayes. Para executar a matriz digite **execute matrizbay.txt;** <enter>.

Irá analisar a matriz mostrando informações a seu respeito.

```
Executing file "matrizbay.txt"
DOS line termination
Longest line length = xxxx
Parsing file
Expecting NEXUS formatted file
Reading data block
        Allocated matrix
        Matrix has xx taxa and xxxx characters
        Data is DNA
        Gaps coded as –
        Setting default partition (does not divide up characters).
        Taxon 1 → nome1
        Taxon 2 → nome2
        …
        Taxon N → nomeN
        Setting output file names to "matrizbay.txt.run<i>.<p/t>"
        Successfully read matrix
Exiting data block
        Excluding character(s)
Reading MrBayes block
        Setting statefreqpr to xxxxxxxxxxxxxx(x,x,x,x)
```

Successfully set prior model parameters
Setting Nst to x
Setting rates to xxxxx
Successfully set likelihood model parameters
Exiting MrBayes block
Reached end of file

16. Caso a matriz não siga até a leitura final é necessário corrigir o erro indicado pelo programa.

17. Para iniciar a análise digitar uma linha de comando:

mcmc ngen=10000000 nruns=2 nchains=4 printfreq=1000 samplefreq=1000;

Esta linha de comando significa:

-mcmc: para iniciar a análise

- ngen: o número de busca que será feita

- nruns: o número de corridas simultâneas, mas independentes

- nchains: o número de cadeias utilizadas

- printfreq: será mostrado na tela a cada 1000 buscas o seu valor

- samplefreq: será salvo os valores no arquivo (matrizbay.run.t) a cada 1000 buscas

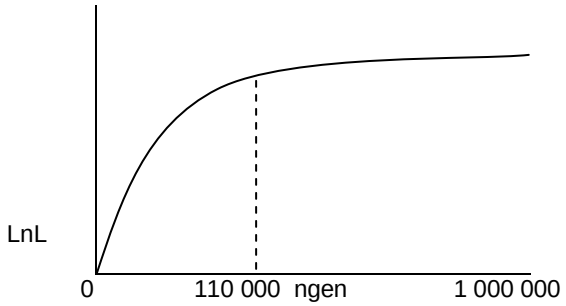
18. Ao executar esse comando aparecerão na tela (a cada 1000 buscas) os menores valores encontrados para as 4 cadeias (os valores entre os 3 (xxx) – cadeias frias e 1 [xxx] – cadeia quente) e para as duas corridas (são separadas pelo *). No final da linha o tempo estimado para o término da análise será indicado (X:XX:XX):

1000 – (-2847.80) (-2850.15) (-2860.13) [-2750.66] * (-2832.80) (-2855.77) [-2800.13] (-2810.66) – X:XX:XX

19. Terminando o número de réplicas proposto, o programa perguntará se haverá prosseguimento na análise. Ao digitar N a análise é terminada. Minimizar o programa MrBayes.

20. Abrir no WorDPad ou bloco de notas os arquivos matrizbay.txt.run1 e matrizbay.txt.run2 (arquivos p).

21. Ao abri-los pode-se observar diversas colunas. Na primeira encontram-se as gerações salvas. Na segunda coluna apresentam-se os comprimentos das árvores (LnL).



Nesta coluna deve-se observar a partir de que geração os valores tornam-se os maiores (uma vez que os valores são negativos). Como no gráfico ao lado, chega-se a um ponto, no qual os valores ficam mais constantes. Este é o valor do “burnin”. O “burnin” é o valor que será utilizado para descartar as árvores antes desse ponto.

Por exemplo: Numa análise com 1 milhão de gerações e salvando uma a cada 1000 gerações [(1milhão/1000) + 1], temos uma total de 100.001 árvores salvas. Observando que o ponto, onde a análise mantém seu valor é de 110.000, o valor do “burnin” é de 110. (logo as 110 primeiras árvores serão descartadas).

22. Voltar para o programa MrBayes.

23. Abrir um bloco de log. (**log start;** <enter>)

24. Para obter as probabilidades posteriores e o consenso de maioria, entrar com a seguinte linha de comando:

sumt burnin=xxx contype=halfcompat showtreeprobs=no nruns=2;

Diversos arquivos serão criados. No arquivo matrizbay.txt.con será salvo uma árvore de consenso de maioria para cada corrida (duas árvores). No arquivo log criado (log.out) estarão as probabilidades posteriores para os ramos.

25. Ao terminar de calcular todas as probabilidades, entrar com a linha de comando para fechar o arquivo log. (**log stop;** <enter>)

26. Para sair do programa digitar: **quit;** <enter>

Anexo 3. Combinação de marcadores e Teste de congruência ILD

1. Combinando os dois marcadores

Para analisar simultaneamente os dados dos dois marcadores, abrir um dos arquivos das matrizes do bloco de notas (A) e considere este arquivo como base. Abrir o outro marcador (B) e copiar somente a matriz. Colar a matriz (B) no arquivo da matriz (A), abaixo desta. Conferir se os terminais estão exatamente na mesma ordem e com os nomes idênticos. No exemplo abaixo, a matriz apresenta dois marcadores. Se quiser, pode-se identificar cada matriz correspondente ao seu marcador, desde que o comentário esteja entre colchetes (por exemplo, [rps16]). Qualquer informação a ser colocada entre colchetes não é lido pelo programa. Verificar se o número de terminais é o mesmo (**ntax**). Se os dois marcadores apresentarem o mesmo número de terminais, e exatamente os mesmos terminais, pode prosseguir a análise. Entretanto, se um dos marcadores não possuir o mesmo número de terminais, este que falta deve ser eliminado da outra matriz. O número total de caracteres será a soma dos dois marcadores. O comando que o programa irá reconhecer que existe mais de uma matriz é o comando “INTERLEAVE”.

```
#Nexus
[ITS, trnL-F, rps16]
Begin data;
Dimensions  ntax= 14  nchar= 1730;
Format datatype= DNA Symbols= “0 1 2 3 4” gap= - INTERLEAVE;
Options gapmode=missing;
Eliminate 36-40 89-93 98-103 1336-1340 1348-1349 1367-1371

Matrix
[ITS]
Croton_urucurana                    TCCGTAGGTGAACCTGCG
Croton_pallidulus_var._pallidulus_1 TCCGTAGGTGAACCTGCG
Croton_pallidulus_var._pallidulus_2 TCCGTAGGTGAACCTGCG
Croton_pallidulus_var._pallidulus_3 TCCGTAGGTGAACCTGCG
Croton_dichrous_1                   TCCGTAGGTGAACCTGCG
Croton_dichrous_2                   TCCGTAGGTGAACCTGCG
Croton_splendidus_1                 TCCGTAGGTGAACCTGCG
Croton_splendidus_2                 TCCGTAGGTGAACCTGCG
Croton_ceanothifolius               TCCGTAGGTGAACCTGCG
```

```
Croton_aff._erythroxyloides          TCCGTAGGTGAACCTGCG
Croton_erythroxyloides               TCCGTAGGTGAACCTGCG
Croton_dusenii                       TCCGTAGGTGAACCTGCG
Croton_myrianthus                    TCCGTAGGTGAACCTGCG
Croton_pallidulus_var._glabrus       TCCGTAGGTGAACCTGCG

[trnL-F]
Croton_urucurana                     ACTAAGTGATAACTTTCAAA
Croton_pallidulus_var._pallidulus_1  ACTAAGTGATAACTTTCAA-
Croton_pallidulus_var._pallidulus_2  ACTAAGTGATAACTTTCAA-
Croton_pallidulus_var._pallidulus_3  ACTAAGTGATAACTTTCAA-
Croton_dichrous_1                    ACTAAGTGATAACTTTCAA-
Croton_dichrous_2                    ACTAAGTGATAACTTTCAA-
Croton_splendidus_1                  ACTAAGTGATAACTTTCAA-
Croton_splendidus_2                  ACTAAGTGATAACTTTCAA-
Croton_ceanothifolius                ACTAAGTGATAACTTTCAA-
Croton_aff._erythroxyloides          ACTAAGTGATAACTTTCAA-
Croton_erythroxyloides               ACTAAGTGATAACTTTCAAA
Croton_dusenii                       ACTAAGTGATAACTTTCAAA
Croton_myrianthus                    ACTAAGTGATAACTTTCAA-
Croton_pallidulus_var._glabrus       ACTAAGTGATAACTTTCAA-

;
end;
```

Um detalhe que deve ser tomado em conta é o comando “eliminate” (mostra quais regiões da sequência tem de ser eliminados). Caso apenas uma matriz apresente regiões eliminadas, é conveniente que esta fique em primeiro, bastando copiar a linha de comando “eliminate” da matriz original e colar. Entretanto se mais de uma matriz apresentar regiões eliminadas, o número de caracteres terá de ser somado a matriz anterior. Vide exemplo:

Matriz 1 – 900 caracteres Eliminate 15-32 45-51

Matriz 2 – 850 caracteres Eliminate 30-35 48-53 (SOMAR 900)

Ao unir todos os dados tem-se que somar os caracteres da matriz anterior, e a linha final de Eliminate será 15-35 45-51 930-935 948-953

Após todas as modificações, deve-se salvar o arquivo da matriz combinada. Com a matriz pronta, pode-se criar o bloco de comando e rodar o programa. Para isso, repetir os passos 3 e 4 do Anexo de Parcimônia.

O teste de congruência ILD é realizado quando dois ou mais marcadores são combinados.

2. Criando um bloco de comando

```
#NEXUS
Begin PAUP;
log file = ildtest.log;
hompart partition= [regioes] nreps=100 /start=stepwise addseq=random nreps=10 savereps=no randomize=addseq rstatus=no hold=1 swap=TBR multrees=yes;
log stop;

End;
```

Através desse bloco de comando será possível realizar o teste ILD para identificar se as diferentes regiões, ou marcadores, podem ser combinados. A terceira linha mostra o comando em que irá salvar no arquivo ildtest.log todas as informações que aparecerem na tela do computador. A seguir, inicia-se a linha de comando para o teste de partição. Os demais comandos já foram comentados no Anexo anterior (Parcimônia) no tópico **3. Gerando um bloco de comando**.

Salvar o arquivo (por exemplo, ildtest.txt). Esta análise será realizada no programa PAUP, portanto, recortar o arquivo e colar na mesma pasta do programa PAUP.

2. Complementando sua matriz

Abrir o arquivo que contém as matrizes combinadas no bloco de notas. O modelo utilizado para cada diferente marcador continua sendo o mesmo para este das combinações, com a diferença de que ao final da matriz, depois de ; End; deve-se acrescentar um comando:

```
;
End; (Final da matriz)
```

```
Begin sets;
charset d1= 1-XX;            (1° bloco de dados)
charset d2= XX-XXX;          (2° bloco de dados)
charpartition [Regioes]= Regiao1:d1, Regiao2:d2;

end;
```

Os dados não preenchidos correspondem a:

- **charset d1**: caracteres do marcador 1 a ser utilizado, por exemplo, se forem utilizados 750 caracteres deve-se anotar 1-750;

- **charset d2**: caracteres do marcador 2 a ser utlizado, por exemplo, se forem utlizados mais 900 caracteres, deve-se anotar 751-1650;

Novamente prestar atenção: se algum dado ou ponto e vírgula faltar na matriz, ao executá-la no programa PAUP este irá mostrar o erro e não será possível continuar o procedimento. Certificar-se de que o nome dado após o comando “charpartition” seja o mesmo do utilizado no bloco de comando na quarta linha (hompart charpartition = **[regiões]**).

3. Executando a matriz no PAUP

Colar os arquivos da matriz e do bloco de comando na raiz do programa. Abrir o programa PAUP e executar sua matriz da seguinte forma:

execute nomedasuamatriz.txt; <enter>

Para iniciar a análise de partição executar o bloco de comando da seguinte forma:

exe ildteste.txt; <enter>

Quando a análise terminar, o programa irá apresentar um valor para P. Como a hipótese nula é que os dados são homogêneos, quando o valor de P for menor ou igual a 0,05 os dados são considerados incongruentes, caso contrário, os dados podem ser utilizados e combinados.

Anexo 4. Utilização do software GeneMapper Version 4.1

1. Ao abrir o programa, o primeiro passo é clicar em **File>New Project** e escolher AFLP, posteriormente adicionar amostras ao projeto em **File>Add Samples to Project**, selecionar a aba **Files** e navegar pelas pastas até encontrar o arquivo ou pasta em que foram salvos os dados de seqüenciamento, depois de feita esta etapa clicar em **Add to List** e posteriomente em **Add**;

2. Observar se existe um método de análise ou se deverá ser criado um (consultar o AFLP System Analysis Getting Started Guide na pagina 15);

3. Após estas etapas observar o método de análise a ser empregado para o projeto, e também o **Size Standard** a ser utilizado (em geral utiliza-se o GS500(-250) ROX);

4. Clicar em **Analysis>Analyse** ou simplesmente na seta verde acima do painel de amostras, e escolher um nome a ser salvo para a sua análise;

5. As amostras a serem analisadas serão aquelas em que aparecerem um quadrado verde na **coluna SQ** indicando que o sequenciamento foi realizado de modo correto, ou que sua amostra está adequada para ser analisada. Se não aparecer este quadrado verde em SQ sua amostra não poderá ser analisada;

6. Depois de realizada esta etapa, clicar em **Analysis>Display Plots** para mostrar o eletroferograma para cada amostra que se quer analisar, nesta etapa é importante observar cada pico, pois muitas vezes o software considera picos que estão muito próximos como dois diferentes picos, o que pode interferir na análise posteriormente;

7. Na tela aberta do **Samples Plot**, observar que existem três diferentes colorações indicando qual marcador de fluorescência foi utilizado (o vermelho corresponde ao padrão ROX utilizado), mais ao lado podem-se observar três caixas à direita, indicando: **No Table**, **Sizing Table** e **Genotypes Table**, clicar na segunda e aparecerá uma lista abaixo do eletroferograma indicando quais os picos que foram selecionados de acordo com o método de analise escolhido, sendo importante observar o tamanho do pico (que vai variar se desejar uma intensidade alta ou baixa, por exemplo, a altura acima de 100 que pode ser considerado um pico interessante para a análise), e nesta etapa, se desejar modificar o tamanho do pico a ser escolhido, terá que modificar o método de analise, em **Analysis>Analysis Method Editor**;

8. Aparecerá uma tela com diversas abas, em que a correspondente a **Allele** diversos componentes podem ser observados, e modificados de acordo com a necessidade, sendo que os principais são **Analyse Dyes** (em que se pode modificar o marcador de fluorescência utilizado) e em **Panel>Bin width (bp)** o normal é 1.0 (equivalente a agrupar tamanhos dos picos de um em um, por exemplo de 100 a 101, de 101 a 102 e assim por diante), porém pode-se utilizar tamanhos maiores, por exemplo 2.0 (equivaleria a agrupar os tamanhos dos picos de 2 em 2). Na aba **Peak Detector**, pode-se modificar em **Ranges>Sizing** o **Start Size** (a partir de qual tamanho se quer que a analise comece, por exemplo, no caso deste projeto a partir de 75pb) e o **Stop Size** (geralmente 500pb);

9. Após as modificações, procede-se novamente a observação dos picos, como na etapa seis. Todas as amostras da corrida devem ser observadas, pois como fora dito, muitas vezes o software pode considerar um pico muito próximo do outro como diferentes, sendo que podem ser os mesmos, ao menos para a análise. Feita a observação, procede-se a etapa de exportar os dados de presença ou ausência dos picos para cada amostra em um arquivo do tipo Excel ou Bloco de Notas.

10. Para tanto, observa-se uma aba **Genotypes** no painel inicial, em que diversos resultados podem ser analisados, sendo que o importante a ser observado é a quantidade de alelos obtidos para a análise. Clicar em **Tools>Gene Mapper Manager**, e abrir a aba **Report Settings**, clicar em **New** para abrir o editor. Adicionar o nome que se deseja (por exemplo, o numero da combinação a ser testado), e abaixo em número de alelos acrescentar a quantidade de alelos observado anteriormente. Na aba **Genotype**, adicionar as colunas que desejar, podendo utilizar as teclas **Ctrl** ou **Shift** para selecionar **Sample File**, ou **Sample Name**, e todos os alelos. Salvar em **OK,** e depois clicar em **Done** para fechar o Gene Mapper Manager;

11. Selecionar todas as amostras do projeto em **Edit>Select All**, clicar em **Analysis>Report Manager**, e selecionar em **Report Setting** o nome digitado. Clicar em **File>Export Table**, escolher a pasta em que se deseja salvar o arquivo, nomeá-lo e clicar em **Save**. Existem algumas opções de como exportar o arquivo, uma em cima **Export File as** em que se pode salvar em arquivo **.txt** ou **.csv** (o primeiro se refere a arquivo de texto, podendo ser utilizado para o Bloco de Notas posteriormente, e o outro a arquivo do Excel), e o abaixo existem basicamente as mesmas opções (as etapas 10 e 11 deste protocolo podem ser melhor detalhadas nas paginas 81 e 82 do guia mencionado anteriormente na etapa 2);

12. Fechar o Report Manager e posteriormente o programa em **File>Exit**.

Anexo 5. Análise de Neighbour Joining utilizando o programa PAUP

1*. Construindo sua matriz*

Abrir o arquivo ____.txt obtido no anexo anterior, do programa GeneMapper, dentro do programa WordPad ou Bloco de Notas. Pode-se então visualizar (nome – tabulação – seqüência) (NNNNN indica os 0 e 1 – ausência ou presença dos picos):

nome1
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
nome2
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
nome3
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
...
nomeN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

Antes da matriz, adicionar algumas linhas. A matriz tem de seguir o seguinte modelo:

#Nexus
Begin data;
Dimensions ntax= xx nchar= xx;
Format datatype= STANDARD gap= - ;
Options gapmode=missing;

Matrix [Começa sua matriz de dados]
nome1 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
nome2 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
nome3 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
...
nomeN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
;
End;

Os dados não preenchidos correspondem a:

- **ntax**: número de terminais da sua análise;

- **nchar**: número de caracteres totais da sua análise;

Obs: como a matriz trata de presença ou ausência dos caracteres (no caso de AFLP os picos, indicando a presença ou não de fragmentos) o formato a ser executado é STANDARD e não DNA, que corresponde às bases A,T,C e G.

Preste atenção: se algum dado ou ponto e vírgula faltar na matriz, ao executá-la no programa PAUP este irá mostrar o erro e não será possível continuar o procedimento. **Importante lembrar também que a espécie a ser considerada como grupo-externo deve ser posicionada como a primeira da análise, pois somente deste modo o programa vai enraizar nesta espécie.**

Obs: se necessário combinar dados de diferentes combinações, não se esquecer de posicionar os nomes de cada combinação na mesma ordem da primeira, que será considerada a original. Além disso, deve ser acrescido um comando INTERLEAVE após gap:- e antes de ;.

2. Executando a matriz no PAUP

Cole o seu arquivo da matriz na raiz do programa. Abra o programa PAUP e execute sua matriz da seguinte forma:

execute nomedasuamatriz.txt; <enter>

Verifique se o programa aceitou sua matriz mostrando sua descrição.

Processing of file “nomedoarquivo.txt” begins...

Data matrix has xx taxa, xxxx characters
Valid character-state symbols: 01
Missing data identified by ‘?’
Gaps identified by ‘-‘
Processing of file “nomedoarquivo.txt” completed.

A seguir execute seguinte comando:**NJ;** <enter> Aparecerá uma árvore de Neighbour Joining. Posteriormente a este comando deve-se salvar a árvore com o seguinte comando: **savetrees file=_____.tre brlens=yes;** Desta forma, será salva a árvore, com as distâncias relativas apresentadas como no programa, na forma de um filograma, que poderão ser analisadas posteriormente em um programa específico

para visualização de árvores filogenéticas. (Obs: o programa Treeview não aceita a visualização, indicando um erro de sintaxe, porém na versão mais atualizada TreeviewX é possível a observação e análise).

Se necessário para uma análise mais apurada, deve-se buscar através do modo heurístico réplicas de bootstrap para verificação dos índices de confiabilidade para os diferentes ramos e clados apresentados na árvore. Para tanto, deve-se digitar um comando **log start;** que indicará um arquivo em que será salvo as informações da sua análise (se estiver de modo correto aparecerá Logging output to file" _____.log".). Após feita esta etapa deve-se então digitar **bootstrap nreps=1000;** (se necessário aumentar o número de réplicas modificar este número). Em seguida aparecerá na tela as seguintes informações:

```
Bootstrap method with heuristic search:
  Number of bootstrap replicates = 1000
  Starting seed = ____________
  Optimality criterion = parsimony
    Character-status summary:
      Of ____ total characters:
        All characters are of type 'unord'
        All characters have equal weight
        __ characters are constant
        ___ variable characters are parsimony-uninformative
        Number of parsimony-informative characters = ____
    Gaps are treated as "missing"
  Starting tree(s) obtained via stepwise addition
  Addition sequence: simple (reference taxon = ______)
  Number of trees held at each step during stepwise addition = 1
  Branch-swapping algorithm: tree-bisection-reconnection (TBR)
  Steepest descent option not in effect
  Initial 'MaxTrees' setting = 100
  Branches collapsed (creating polytomies) if maximum branch length is zero
  'MulTrees' option in effect
  Topological constraints not enforced
  Trees are unrooted

The limit of 100 trees (= 'MaxTrees') has been reached.
Do you want to increase 'MaxTrees'? (Y/n) paup> n
```

Antes de terminar o número de réplicas proposto, o programa perguntará se haverá prosseguimento na análise, pois quer saber se deseja aumentar o número máximo de árvores a serem salvas. Ao digitar N a análise continua.

Ao executar este comando aparecerão na tela os valores encontrados para bootstrap. No começo da linha o tempo corrido para a análise será indicado (X:XX:XX). Ao final das 1000 réplicas propostas, aparecerá a seguinte informação:

1000 bootstrap replicates completed
  Note: Effectiveness of search may have been diminished due to tree-buffer
    overflow.
  Time used = XX:XX:XX.X

Bootstrap 50% majority-rule consensus tree

Depois desta linha aparecerá uma árvore de consenso de maioria, com os índices de confiabilidade de bootstrap a serem empregados posteriormente na confecção da árvore a ser apresentada. Outras informações abaixo serão passadas, sendo que ao final aparecerá:

XXXX groups at (relative) frequency less than 5% not shown.

Ao terminar todos os comandos digitar **quit;** <enter> As informações (índices de bootstrap) a serem observadas e posteriormente plotadas na árvore de Neighbour Joining devem ser lidas no arquivo log criado, que estará na raiz do programa PAUP.