

Universidade de Brasília
Departamento de Biologia Celular
Curso de Pós-Graduação em Biologia Molecular

**DESENVOLVIMENTO, CARACTERIZAÇÃO E
MAPEAMENTO DE MICROSSATÉLITES
DE TETRA E PENTANUCLEOTÍDEOS
EM *Eucalyptus spp.***

Dissertação apresentada ao programa de Pós-Graduação em Biologia Molecular do Instituto de Ciências Biológicas da Universidade de Brasília como requisito parcial para a obtenção do Título de Mestre.

**Aluna: Carolina Paola Sansaloni
Orientador: Dr. Dario Grattapaglia**

Brasília - DF
Fevereiro - 2008

Universidade de Brasília
Departamento de Biologia Celular
Curso de Pós-Graduação em Biologia Molecular

DESENVOLVIMENTO, CARACTERIZAÇÃO E MAPEAMENTO DE
MICROSSATÉLITES DE TETRA E PENTANUCLEOTÍDEOS
EM *Eucalyptus spp.*

CAROLINA PAOLA SANSALONI

APROVADA POR:

.....
Dario Grattapaglia, PhD (Universidade de Brasília)
CPF: 286.974.301-78 (Orientador) e-mail: dario@cenargen.embrapa.br

.....
Marcio Elias Ferreira, PhD (Universidade de Brasília)
CPF: 317.541.981-04 (Membro Interno) e-mail: ferreira@cenargen.embrapa.br

.....
Rinaldo Wellerson Pereira, PhD (Universidade Católica de Brasília)
CPF: 844.342.796-53 (Membro Externo) e-mail: rinaldo@pos.ucb.br

A César,
meu companheiro de vida...

DEDICO

AGRADECIMENTOS

À Universidade de Brasília (UnB), e especialmente ao programa de pós-graduação em Biologia Molecular, representados por todos os professores e funcionários, pela oportunidade de realização deste curso.

Ao CNPq pelo apoio financeiro que tornou possível a realização desse trabalho.

À EMBRAPA Recursos Genéticos e Biotecnologia, pela infra-estrutura de trabalho.

Ao Projeto Genolyptus e a todas as empresas e Instituições que a compõem, por facilitar as amostras e os recursos necessários para a realização deste trabalho.

Ao meu orientador Dario Grattapaglia, por ter acreditado em mim desde o começo, pelos ensinamentos transmitidos, sua confiança, dedicação, paciência, oportunidade de crescimento e principalmente pela sua amizade.

Ao Marcio Elias Ferreira e Dario Grattapaglia, pela oportunidade de aprendizagem e crescimento profissional no laboratório de Hereditas.

À Eva, pelos ensinamentos transmitidos com paciência e carinho, por estar sempre disposta a tender uma mão para ajudar em tudo e pela grande amizade brindada. Gracias Eva!!!!

À Marilia e a Dany pela ajuda outorgada sempre, pelos ensinamentos e pela amizade.

Aos colegas e amigos do laboratório, Mariana, Laís, Leo, e Juliana, pela amizade, companheirismo e troca de conhecimentos.

Aos meus amigos do “cafezinho”, Adri, Thaiza, Bruna, Andréia, Karen, Dany, Marília, Bete, Túlio e Marco, pela grande amizade, apoio, conselhos e boas risadas compartilhadas. Adoro vocês!!!!

Ao Marcão, o mestre dos géis de poliacrilamida, pela ajuda dispensada.

Ao pessoal da Hereditas, Mônica, Nathalia, Maria, Zilma, Claudinha, Sandra, Militizi, Polyanna, André, Saulo, Ronaldo, Tânia e Elen, pela grande amizade, companheirismo, confiança e ajuda incondicional. Eu encontrei uma verdadeira família em vocês.

Aos meninos e meninas do LGA (Laboratório de Genética Animal), pelo carinho e amizade.

Às minhas amigas argentinas, Mariana, Sory, Macarena, Manu, Ale e Gaby por nunca esquecer dos momentos vividos juntas. Mesmo estando na distancia as levo sempre no meu coração.

Principalmente, aos meus pais, Norma e Juan, pelos seus ensinamentos, apoio incondicional, alento a continuar crescendo cada dia como pessoa e profissional. Todo o que eu sou eu devo a vocês. Os amo !!

A meu irmão Adrián e minha cunhada Sole, pelo apoio, amor e principalmente por ter me presenteado com uma das coisas mais lindas da vida, minha afilhada Valentina, a luz dos meus olhos e a minha sobrinha Florência que espero ansiosamente o nascimento o dia 12 de fevereiro.

A meus irmãos do coração, Mariana e Tomas, pela amizade, carinho, confiança, e por tantos momentos compartilhados, a Viole, por fazer esquecer qualquer problema com um simples sorriso e ao sobrinho que chegara!

E finalmente, ao meu grande amor, César, pelos momentos vividos juntos que fortalece nosso amor a cada dia. Pela paciência e compreensão nos momentos de estresse e principalmente por ser meu companheiro nesta vida. Te Amo Gordo!!!

Ao Brasil, ao seu povo maravilhoso e sua cultura cativante.

E a todos, familiares, tios, primos, amigos e colegas, que me apoiaram sempre e fizeram possível a realização deste trabalho.

Muito obrigada!!!

ÍNDICE

	Página
Lista de Figuras	ix
Lista de Tabelas	x
Resumo	1
Abstract	2
1. Introdução	4
2. Revisão bibliográfica	7
2.1 <i>Eucalyptus</i> : florestas plantadas e melhoramento	7
2.2 Seleção de clones de eucalipto	10
2.3 Proteção varietal de clones de eucalipto	12
2.4 Marcadores Moleculares	14
2.5 Sistemas multiplex de co-amplificação de microssatélites	16
2.6 Microssatélites derivados de tetra e pentanucleotídeos em plantas	20
3. Objetivos	22
4. Material e métodos	22
4.1 Material genético	22
4.2 Mineração de microssatélites derivados de tetra e pentanucleotídeos	24
4.3 Triagem de marcadores microssatélites	24
4.4 Caracterização genética dos microssatélites de tetra e pentanucleotídeos	26
4.5 Herança, segregação e mapeamento de microssatélites tetra e pentanucleotídeos selecionados	27
4.6 Montagem e aplicação de sistemas multiplex de co-amplificação de microssatélites	28
5. Resultados	29
5.1 Desenvolvimento de microssatélites derivados de tetra e pentanucleotídeos	29
5.2 Caracterização genética dos microssatélites selecionados em três espécies de <i>Eucalyptus</i> .	40
5.3 Mapeamento dos locos tetra e pentanucleotídeos	49
5.4 Otimização e aplicação de sistemas multiplex de co-amplificação de microssatélites	58
6. Discussão	66
7. Conclusões	74

8. Anexos	78
Anexo 1: Freqüências alélicas dos locos tetranucleotídeos nas seis procedências das espécies estudadas.	78
Anexo 2: Freqüências alélicas dos locos pentanucleotídeos nas seis procedências das espécies estudadas.	80
Anexo 3: Histogramas de freqüências alélicas dos 22 locos tetranucleotídeos selecionados para caracterização genética	83
Anexo 4: Histogramas de freqüências alélicas dos 10 locos pentanucleotídeos selecionados para caracterização genética.	89
Anexo 5: Freqüências alélicas os locos tetranucleotídeos nas duas procedências de cada uma das espécies <i>E. grandis</i> (A), <i>E. globulus</i> (B) e <i>E. urophylla</i> (C).	91
Anexo 6: Caracterização de locos pentanucleotídeos de duas procedências de cada uma das espécies <i>E. grandis</i> (A), <i>E. globulus</i> (B) e <i>E. urophylla</i> (C).	93
9. Referências bibliográficas	95

LISTA DE FIGURAS

	Página
Figura 1: Descritores morfológicos de <i>Eucalyptus</i> .	13
Figura 2: Exemplos de resultados da etapa 1 de triagem de locos tetra e pentanucleotídeos em gel de agarose.	30
Figura 3: Exemplos de resultados da etapa 2 de triagem em gel de poliacrilamida	34
Figura 4: Distribuição de freqüência das diferentes categorias de locos tetra e pentanucleotídeos após a segunda etapa de triagem.	37
Figura 5: Distribuição de freqüência de locos por número de alelos para tetra e pentanucleotídeos	38
Figura 6: Freqüência de microssatélites tetra e pentanucleotídeos	39
Figura 7: Eletroferogramas de locos polimórficos	42
Figura 8: Configurações de segregação observadas para os locos tetra e pentanucleotídeos	51
Figura 9: Mapeamento genético de locos EMBRA tetra e pentanucleotídeos nos grupos de ligação de cada genitor (♀ <i>E. urophylla</i> / ♂ <i>E. grandis</i>) da família IP	54
Figura 10: Mapeamento genético de locos EMBRA tetra e pentanucleotídeos nos grupos de ligação de cada genitor (♀ <i>E. urophylla</i> / ♂ <i>E. grandis</i>) da família VCP	56
Figura 11: Eletroferogramas ilustrativos dos sistemas multiplex 1 (A) e multiplex 2 (B)	61
Figura 12: Fenograma representando a discriminação de 46 indivíduos utilizando 12 locos tetra e pentanucleotídeos que corresponde ao multiplex 1 e 2.	62
Figura 13: Fenograma representando a discriminação de 46 indivíduos utilizando 7 locos (tetra e pentanucleotídeos) que correspondem ao multiplex 1.	63
Figura 14: Fenograma representando a discriminação de 46 indivíduos utilizando 7 locos (tetra e pentanucleotídeos) que correspondem ao multiplex 2.	64
Figura 15: Fenograma representando a discriminação de 46 indivíduos utilizando 9 locos dinucleotídeos.	65

LISTA DE TABELAS

	Página
Tabela 1: Caracterização genética de microssatélites tetranucleotídeos	46
Tabela 2: Caracterização genética de microssatélites pentanucleotídeos.	47
Tabela 3: Alelos privados dos locos tetranucleotídeos nas espécies <i>E. grandis</i> , <i>E. globulus</i> e <i>E. urophylla</i> .	48
Tabela 4: Alelos privados dos locos pentanucleotídeos nas espécies <i>E. grandis</i> , <i>E. globulus</i> e <i>E. urophylla</i> .	49
Tabela 5: Sistemas <i>multiplex</i> de co-amplificação de PCR testados com sucesso neste trabalho.	60

RESUMO

Marcadores microssatélites permitem o gerenciamento de variabilidade genética, proteção varietal e identificação individual em populações de melhoramento e produção de *Eucalyptus*. A análise de locos microssatélites em sistemas “multiplex” (i.e. vários locos analisados simultaneamente) vem sendo realizada tradicionalmente com microssatélites dinucleotídeos cujos alelos diferem em apenas dois pares de bases. Entretanto o padrão de qualidade recomendado internacionalmente para a investigação genética forense estabeleceu a utilização exclusiva de microssatélites baseados em repetições de tetra ou pentanucleotídeos. A maior diferença de tamanho entre os alelos e a minimização de artefatos de amplificação permitem uma interpretação significativamente mais robusta de genótipos em comparação a dinucleotídeos e conseqüentemente uma fácil e confiável comparação de perfis genotípicos entre laboratórios. O objetivo deste trabalho foi, portanto, o desenvolvimento, caracterização e mapeamento genético de uma nova categoria de microssatélites para espécies de *Eucalyptus*, baseados em repetições de tetra e pentanucleotídeos. Um banco de dados de 18.510 mil seqüências genômicas derivadas de clones de cromossomos artificiais de bactérias (BAC) de *E. grandis* totalizando 9,6 mega pares de bases foram mineradas para repetições de tetra e pentanucleotídeos e pares de iniciadores desenhados automaticamente para flanquear repetições totalizando dez ou mais pb. Dos 304 e 196 pares de iniciadores desenhados para a amplificação de locos tetra e pentanucleotídeos perfeitos, respectivamente, foram selecionados 80 locos de tetra e 30 de pentanucleotídeos para estudos detalhados. Após uma triagem em eletroforese de alta resolução foram selecionados 22 locos tetra e 10 pentanucleotídeos os quais foram geneticamente mapeados e caracterizados para número de alelos, heterozigosidade esperada (He) e observada (Ho), probabilidade de identidade (PI), probabilidade de exclusão de paternidade (PE) e conteúdo de informação polimórfica (PIC) em amostras populacionais de seis procedências das três principais espécies plantadas de *Eucalyptus* no mundo (*E. grandis*, *E. globulus* e *E. urophylla*). Os locos tetra e

pentanucleotídeos selecionados revelaram um conteúdo informativo inferior àquele de dinucleotídeos, embora adequado para discriminar indivíduos e determinar parentesco com elevada precisão. Ao avaliar a capacidade efetiva de discriminação individual em um grupo de 46 clones elite plantados comercialmente 5 locos tetra ou pentanucleotídeos foram suficientes para identificar unicamente todos os clones. A herança mendeliana de todos os 32 locos selecionados foi confirmada. Destes, 23 locos apresentaram uma configuração de segregação que permitiu o seu mapeamento genético em 8 dos 11 grupos de ligação com base em mapas de referência de duas famílias de *E. grandis* x *E. urophylla*. Dois sistemas multiplex, cada um com 7 locos selecionados foram testados com sucesso disponibilizando assim os primeiros sistemas de análise genética de alta precisão para espécies de *Eucalyptus*, passíveis de adoção em escala internacional para a análise genética de clones e populações.

ABSTRACT

Microsatellite markers allow the management of genetic variation, varietal protection and individual identification in breeding and production populations of *Eucalyptus*. The analysis of microsatellite markers in multiplexed systems has been traditionally accomplished using dinucleotide repeat based loci whose alleles differ by only two base pairs. However the golden standard recommended for individual identification in forensic genetics requires the exclusive use of microsatellites based on equal or higher than tetranucleotide repeats. The larger difference in allele size and the reduction in amplification stuttering artifacts allows a significantly more robust genotype calling and thus an easier and more reliable comparison of multilocus genotypes across laboratories. The objective of this work was, therefore, the development, characterization and mapping of a novel category of microsatellites for *Eucalyptus* species, based on tetra and pentanucleotide repeats. A database of 18,510 sequences

derived from Bacterial Artificial Chromosomes (BAC) clones of *E. grandis* totalling 9.6 mega basepairs of high quality sequence was mined for tetra and pentanucleotide repeats and primers designed automatically to flank repeats equal or larger than 10 bp. Out of the 304 and 196 primer pairs designed for perfect tetra and pentanucleotide repeats respectively, 80 tetranucleotide and 30 pentanucleotide based loci were selected for detailed studies. Following a screening step for polymorphism in high resolution PAGE, 22 tetra and 10 pentanucleotide repeat based loci were selected and genetically mapped and characterized for numbers of alleles, observed and expected heterozygosity, probability of identity (PI), probability of paternity exclusion (PE) and polymorphism information content in population samples of six provenances of the three most widely planted species of the genus (*E. grandis*, *E. globulus* e *E. urophylla*). The tetra and pentanucleotide repeat loci showed a lower information content when compared to dinucleotide repeat markers however fully adequate to discriminate individuals and determine parentage with confidence. When assessing the actual discrimination power on a group of 46 commercially planted Eucalyptus clones, five tetra or pentanucleotide markers were sufficient to uniquely identify all the clones. The mendelian inheritance of all 32 selected loci was confirmed. Twenty three displayed a segregation configuration that allowed genetically mapping them on 8 of the 11 linkage groups using two reference maps of *E. grandis* x *E. urophylla*. Two multiplex systems, with 7 loci each were successfully tested introducing the first high precision systems of genetic analysis for species of *Eucalyptus*, likely to be internationally adopted for the genetic analysis of clones and populations.

1. INTRODUÇÃO

Plantios intensivos de florestas de espécies do gênero *Eucalyptus* de rápido crescimento suprem, de modo racional e eficaz, a demanda por biomassa lenhosa de alta qualidade que de outro modo viria de florestas tropicais nativas. Nas décadas por vir, a expansão destas "fazendas de fibras" será cada vez mais limitada pelo crescimento do plantio de culturas agrícolas e pela pressão ambientalista. Produtividades florestais crescentes e refinamentos na qualidade dos produtos de madeira por meio de melhoramento genético tornaram-se cada vez mais estratégicos para a indústria florestal. Ferramentas moleculares baseadas na identificação de polimorfismos no DNA, envolvidos no controle genético de fenótipos de interesse, prometem fornecer novas oportunidades para a seleção de características de crescimento, forma e propriedades da madeira de árvores cultivadas (Grattapaglia, 2000).

Enquanto que a aplicação de ferramentas moleculares para a seleção direcional de genótipos superiores de eucalipto ainda representa um desafio que depende do entendimento das relações complexas entre genótipo e fenótipo, marcadores moleculares podem ser usados imediatamente e de forma eficiente para resolver várias questões relacionadas ao gerenciamento da variabilidade genética e identificação individual em populações de melhoramento e produção. Marcadores de DNA são muito eficientes para a resolução de questões de identidade clonal e determinação de parentesco em espécies de *Eucalyptus*. Embora marcadores RAPD e AFLP forneçam um bom poder de discriminação para a identificação clonal (Keil e Griffin 1994; Costa e Silva e Grattapaglia 1997; Nesbitt *et al.* 1997), análise de variabilidade em bancos de germoplasma (Nesbitt *et al.* 1995), pomares de sementes (Marcucci *et al.* 2003) bem como quantificação da taxa de fecundação cruzada em populações (Gaiotto *et al.* 1997), seu modo de herança dominante não é adequado para investigações mais refinadas de relacionamento genético entre clones, ou ainda para o estabelecimento de um *fingerprint* genético robusto para fins de registro de cultivares ou disputa legal. Nestes casos torna-se necessário o uso de

marcadores moleculares co-dominantes, robustos e com maior conteúdo informativo tais como marcadores baseados em microssatélites (Brondani *et al.* 1998; Kirst *et al.* 2005).

A tecnologia de detecção de polimorfismos de microssatélites também chamados de STR ("Short Tandem Repeat") ou SSR (Simple Sequence Repeats) via PCR é hoje o padrão internacional para a identificação individual e investigação de vínculo genético em seres humanos, animais domésticos e plantas cultivadas (Butler 2005a). Marcadores microssatélites são baseados na amplificação de seqüências de DNA compostas por um número variável de nucleotídeos, normalmente de 1 a 7 pares de bases, repetidos várias vezes e encontradas no genoma de eucariotos (Hamada *et al.* 1982; Tautz e Renz 1984; Litt e Luty, 1989; Weber e May 1989), embora tenham sido catalogados também em procariotos (Sreenu *et al.* 2003). Em plantas, uma proporção menor do genoma é constituída de microssatélites em comparação a vertebrados e uma maior proporção em relação a invertebrados e fungos. Por exemplo, em *Arabidopsis* 0,85% do genoma é constituído de seqüências simples repetidas, em milho 0,37%. Em peixes a proporção sobe para 2,12% no gênero *Fugu* e no cromossomo 22 de humanos, por exemplo, é de 1,07%. Em *C. elegans* é de 0,21% e em *Saccaromyces* é de 0,30% (Morgante *et al.* 2002). Em plantas, microssatélites estão preferencialmente associados a regiões transcritas do genoma especialmente nas regiões não traduzidas (Morgante *et al.* 2002).

O alto grau de multialelismo e a herança mendeliana co-dominante simples e clara dos microssatélites fornecem um sistema potente para a identificação única de árvores de *Eucalyptus* principalmente quando existe a possibilidade de existir ancestralidade em comum entre os indivíduos. Marcadores microssatélites permitem, por exemplo, determinar com alta precisão a composição paterna das sementes geradas em pomares de sementes e identificar possíveis contaminações por pólen externo ao pomar bem como polinizadores de maior sucesso na geração de descendentes superiores (Grattapaglia *et al.* 2004). A ampla variabilidade genética encontrada em espécies de *Eucalyptus* e o a

locos microssatélites permitiu discriminar todos os 192 indivíduos de uma população de melhoramento de *E. grandis* com apenas 3 marcadores (Kirst *et al.* 2005). A análise de locos microssatélites em sistemas “multiplex” (vários locos analisados simultaneamente) com múltiplas fluorescências e genotipagem em seqüenciador automático vem sendo utilizada (Grattapaglia *et al.* 2003a; Ottewell *et al.* 2005; Missiaggia *et al.* 2005; Mamani *et al.* 2004). Entretanto ainda não existem sistemas otimizados de co-amplificação de vários marcadores caracterizados geneticamente o que permitiria uniformizar e comparar perfis genotípicos entre laboratórios e plataformas de análise.

Este trabalho teve por objetivo a identificação de microssatélites baseados em tetra e pentanucleotídeos e o desenvolvimento, caracterização e mapeamento genético de uma bateria de marcadores para as principais espécies plantadas de *Eucalyptus*. Foi realizada inicialmente uma mineração de seqüências genômicas derivadas de pontas de clones BAC de *E. grandis* seqüenciadas no âmbito do projeto Genolyptus, buscando SSR de tetra e pentanucleotídeos. Experimentos de amplificação e otimização, transferibilidade interespecífica e avaliação do nível de polimorfismo foram realizados. Os novos marcadores microssatélites desenvolvidos foram posicionados geneticamente em mapas genéticos referência de *E. grandis* e *E. urophylla* e caracterizados pela análise de variabilidade em populações de árvores geneticamente não relacionadas das três espécies mais plantadas no mundo e uma coleção de clones elite plantados no Brasil. Os dados permitiram ainda uma comparação do poder de discriminação genética do conjunto de locos tetra e pentanucleotídeos selecionados com locos baseados em dinucleotídeos atualmente utilizados para fins de identificação em *Eucalyptus*. Dois sistemas multiplex de co-amplificação envolvendo 12 locos foram, por fim, testados com sucesso constituindo os primeiros sistemas de alta precisão para a análise genética de clones e populações de *Eucalyptus*.

2. REVISÃO BIBLIOGRÁFICA

2.1 *Eucalyptus*: florestas plantadas e melhoramento

O gênero *Eucalyptus* L'Herit pertence à família *Myrtaceae*. Possui mais de 700 espécies distribuídos em 8 subgêneros, sendo o principal deles o *Symphyomyrtus*, com mais de 300 espécies, dentre as quais estão as espécies mais plantadas para fins comerciais, *E.grandis*, *E. globulus*, *E. urophylla*, *E. camaldulensis*, *E. saligna* e *E. tereticornis* (Eldridge, 1994). O gênero é originário da Austrália, com exceção de *E. urophylla* e *E. deglupta*, que são naturais do Timor e Papua Nova Guiné, respectivamente. Florestas plantadas de *Eucalyptus* são conhecidas por seu rápido crescimento, forma reta, qualidade superior da madeira para múltiplas aplicações, ampla adaptabilidade a solos e climas e facilidade de manejo por plantio direto e rebrota (Eldridge *et al.* 1994; Potts 2004). Espécies de *Eucalyptus* são atualmente plantadas em mais de 90 países onde são utilizadas para diversos produtos florestais tais como madeira sólida, postes, energia, celulose, carvão vegetal, óleos essenciais, mel e tanino bem como para sombra em parques e jardins (Doughty 2000).

Um relatório da FAO estimou um total de 17,9 milhões de hectares plantados de *Eucalyptus* em todo o mundo com a Índia como maior plantador com mais de 8 milhões de hectares, seguida pelo Brasil com 3 milhões (FAO 2000). A grande maioria das plantações é formada com algumas poucas espécies e híbridos. As mais importantes são *E. grandis*, *E. globulus*, *E. urophylla* e *E. camaldulensis*, que, juntamente com seus híbridos perfazem cerca de 80% da área plantada, seguidas por *E. nitens*, *E. saligna*, *E. deglupta*, *E. pilularis*, *Corymbia citriodora* e *E. tereticornis* (Eldridge *et al.* 1994; Waugh 2004).

Embora o melhoramento de *Eucalyptus* seja atualmente uma operação muito dinâmica e tecnicamente avançada, executada principalmente pelas empresas privadas de base florestal, o eucalipto ainda se encontra na infância da sua domesticação quando comparado com culturas anuais. A maioria dos programas de melhoramento se encontra hoje na segunda ou terceira geração de seleção a partir de populações naturais. Entretanto, a combinação de uma ampla variabilidade genética natural seja dentro como entre espécies, juntamente com a possibilidade de hibridização inter-específica e clonagem de genótipos elite, os eucaliptos rapidamente se tornaram os materiais genéticos florestais mais avançados da atualidade no que se refere à combinação matéria prima-produto final, principalmente em países como o Brasil, África do Sul, Chile e Portugal que adotaram o eucalipto para o suprimento de matéria prima para grandes indústrias de base florestal na área de papel, celulose e siderurgia. Uma revisão recente sobre o status do melhoramento do eucalipto no mundo foi publicada por Potts (2004).

A cultura do *Eucalyptus* para exploração comercial da madeira teve início no Brasil por volta do início do século passado, por iniciativa da Companhia Paulista de Estradas de Ferro (Mora e Garcia 2000). O interesse pelo eucalipto, na época, surgiu da necessidade de carvão para abastecer as locomotivas e de dormentes para suas ferrovias. Florestas industriais de *Eucalyptus* estão distribuídas por todo o território nacional, principalmente no Sul, Sudeste e regiões meridionais do Nordeste. Os principais produtos obtidos a partir da madeira do *Eucalyptus* são a celulose e o carvão vegetal para indústria siderúrgica, além de outros derivados da madeira, para móveis e construção, papelão, óleos e outros (Mora e Garcia 2000). A maior parte dos três milhões de hectares cobertos por *Eucalyptus* pertence a empresas produtoras de papel e celulose, sendo que as árvores para produção presentes nestas áreas são, em sua maioria, clones ou híbridos inter-específicos superiores provenientes de programas de melhoramento genético. Graças a estes programas e ao manejo florestal, nas últimas três décadas a produtividade desta espécie foi elevada, a área de plantação foi triplicada através de melhoramento

visando adaptação a diferentes regiões do Brasil e a qualidade da madeira foi melhorada de forma a atender à demanda das indústrias de papel e de celulose.

Plantios intensivos de *Eucalyptus*, entretanto, ainda não suprem integralmente a demanda nacional. A produtividade e as áreas plantadas anualmente têm sido consideradas insuficientes para suprir a demanda, no chamado apagão florestal em discussão desde 2003. Um dos maiores reflexos da escassez de madeira são os valores de comercialização. Em 2003, o metro cúbico de madeira de eucalipto estava sendo negociado a R\$ 7 e no final de 2005 custava R\$ 25, acumulando crescimento de 257% em 24 meses. Juntos, a produção de celulose, siderúrgica e de móveis em Minas Gerais, por exemplo, consome mais de 30 milhões de metros cúbicos de madeira reflorestada por ano. O estado produz somente 58% do que é consumido, com os 42% restantes comprados na Bahia, Mato Grosso do Sul, São Paulo e Goiás, entre outros. O déficit de madeira calculado para o Brasil como um todo é de 225 milhões de metros cúbicos de estéreos por ano (www.ipef.br, nov. 2005). O melhoramento genético buscando aumento de qualidade e produtividade na mesma área plantada constitui, portanto, uma estratégia essencial para a sustentabilidade de empreendimentos de base florestal.

Nos últimos anos, novas metodologias e estratégias de seleção direcional têm sido incorporadas aos programas de melhoramento genético de eucalipto. Dentre as novas tecnologias que vem sendo gradativamente incorporadas, destaca-se a utilização de marcadores moleculares para fins de gerenciamento de variabilidade, proteção varietal e controle de qualidade, e, ainda, em menor escala, o melhoramento molecular envolvendo o mapeamento genético e a transgenia. O objetivo, em geral, gira em torno da melhoria da qualidade da madeira, especificamente quanto às características ligadas a produção industrial buscando otimizar o rendimento das florestas em termos de celulose por hectare ou carvão por hectare. Do ponto de vista genômico, o eucalipto tem sido objeto de estudo nos últimos 10 anos, principalmente com o desenvolvimento de mapas genéticos, a localização de QTL (Quantitative

Trait Loci) e genes e mais recentemente a geração de bancos de ESTs (Expressed Sequence Tags) e a construção de bibliotecas de clones BAC para manipulações físicas do genoma. O genoma de *Eucalyptus* tem entre 500 e 650 Mbp, distribuídos em 11 pares de cromossomos (Grattapaglia e Bradshaw 1994). As espécies *Eucalyptus grandis*, *E. globulus* e *E. urophylla* estão entre as mais cultivadas e também entre as mais estudadas do ponto de vista molecular. Revisões recentes sobre o status da pesquisa genômica, seleção assistida por marcadores e perspectivas da integração destas tecnologias no melhoramento operacional de espécies de *Eucalyptus* tem sido publicadas (Grattapaglia 2004; Myburg *et al.* 2006; Grattapaglia 2006).

2.2 Seleção de clones de eucalipto

O clone designa um indivíduo ou grupo de indivíduos geneticamente idênticos, derivados de um ancestral comum pela multiplicação assexuada. Em outras palavras, no processo de clonagem o genótipo de interesse é integralmente transmitido à descendência permitindo capturar toda a superioridade daquela árvore, seja devida a efeitos gênicos aditivos transmissíveis entre gerações, bem como efeitos não aditivos, resultantes de interações entre alelos ao mesmo loco (dominância) bem como entre alelos de diferentes locos (epistasia). Portanto, as características silviculturais e de qualidade da madeira que diferenciam um clone, podem ser perpetuadas por diferentes gerações, sem que ocorra a recombinação e segregação observada na propagação sexuada.

Tipicamente, para se produzir um clone superior de eucalipto parte-se de uma combinação genotípica superior obtida pelo intercruzamento de árvores elite da mesma espécie ou de espécies diferentes buscando a exploração de complementaridade de características e possivelmente efeitos heteróticos. Estas árvores elite parentais são em geral obtidas via seleção recorrente, processo que envolve a

avaliação, seleção e recombinação (intercruzamento) de indivíduos superiores por sucessivas gerações.

Em cada ciclo de seleção recorrente são obtidos clones potenciais, a partir da propagação assexuada dos indivíduos selecionados quanto aos caracteres de interesse. O método mais comum de propagação vegetativa é o enraizamento de estacas, embora a micropropagação também seja empregada nesta etapa. Os clones potenciais são avaliados em ensaios de competição de materiais genéticos, em delineamentos experimentais apropriados, nas diferentes condições ambientais em questão. As características silviculturais e de qualidade da madeira de interesse são novamente avaliadas nesta etapa, em idade mínima para a qual a seleção precoce seja eficiente. A indicação para implantação em testes piloto ou em plantios comerciais é feita a partir da análise estatística destes dados.

A clonagem de árvores superiores de eucalipto é a uma estratégia de curto prazo para a obtenção de ganhos em produtividade, qualidade e adaptabilidade nos programas de melhoramento genético conduzidos nas empresas florestais. Cada empresa possui hoje clones superiores utilizados nos programas comerciais de plantio e nas estratégias de melhoramento. Atualmente, a maior parte dos plantios comerciais de eucalipto no Brasil é estabelecida com clones altamente produtivos, propagados vegetativamente por miniestaquia em viveiros florestais. A área plantada com clones corresponde a cerca de 60% do total dos quase 4 milhões de hectares plantados com eucalipto no Brasil, com tendência à expansão, já que a maioria das empresas empregam clones em mais de 80% da área plantada anualmente (de Assis 2000).

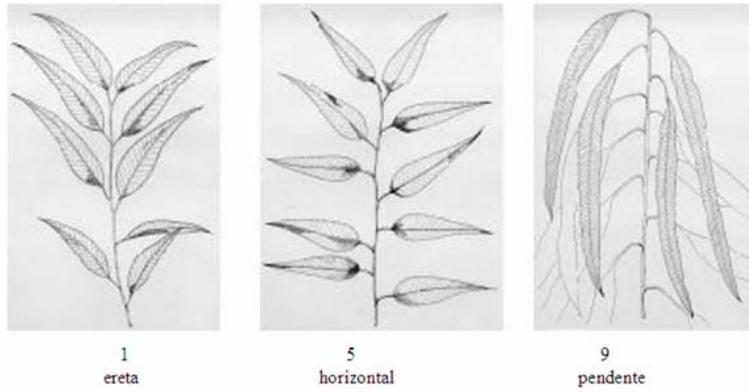
2.3 Proteção varietal de clones de eucalipto

A lei de proteção de cultivares N.º 9.456, sancionada no Brasil em 25 de abril de 1997, representou um importante avanço no sentido de proteger e estimular o contínuo desenvolvimento de materiais genéticos superiores de culturas agrícolas, frutíferas e florestais. Com a publicação das instruções para a execução dos ensaios de distinguibilidade, homogeneidade e estabilidade (DHE) de cultivares de eucalipto em fevereiro de 2002, espécies do gênero *Eucalyptus* foram incluídas na lista daquelas cujos cultivares passam a ser protegidos pelo SNPC (Serviço Nacional de Proteção de Cultivares) do Ministério da Agricultura e Pecuária. A identificação correta de clones de eucalipto tem implicações importantes em diversos procedimentos de melhoramento genético, gerenciamento e controle de qualidade de pomares de sementes e plantios operacionais, intercâmbio e licenciamento de clones e execução de contratos de terceirização de produção de mudas.

Os principais usuários deste processo de proteção têm sido as empresas de base florestal, notadamente aquelas que produzem papel e celulose. O principal objeto de proteção de cultivares de eucalipto tem sido o clone elite, desenvolvido segundo um rigoroso processo de seleção genética direcional a partir dos programas de melhoramento florestal. Clones superiores permitem a obtenção de ganhos elevados e em curto prazo, dentro dos programas de melhoramento genético e constituem um importante componente da vantagem competitiva de uma empresa em um mercado internacional de concorrência intensa.

A proteção de clones superiores de eucalipto envolve hoje a utilização de um conjunto selecionado e validado de marcadores morfológicos principalmente em folhas, flores e casca como se observa na Figura 1. Estes descritores foram selecionados por atender os requisitos básicos de estabilidade e pouca influência ambiental.

A



B

19. Característica 35. Casca na base: textura (planta com 5 anos de idade).



C

15. Característica 31. Fruto: forma

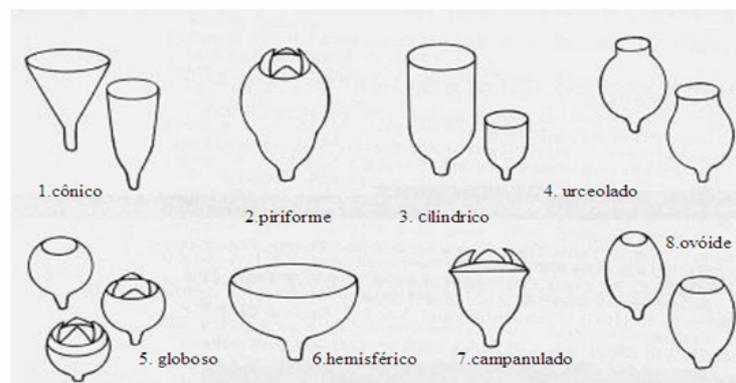


Figura 1: Descritores morfológicos de *Eucalyptus*. A: folha, B: casca, C: fruto.

Entretanto, além das dificuldades de caracterização morfológica às vezes encontrada em função da mudança de fase fisiológica entre o estado juvenil e adulto, clones elite de eucalipto muitas vezes apresentam co-ancestralidade. É comum, por exemplo, que clones elite sejam selecionados dentro da mesma família de irmãos completos ou meio-irmãos. A baixa variabilidade observada para características morfológicas dificulta assim uma clara distinção entre os indivíduos tornando-se necessária a utilização de outros descritores de maior poder de resolução.

Em vista das potenciais limitações inerentes ao processo de descrição morfológica, já na instrução normativa que regulamentou a proteção de cultivares de eucalipto, marcadores moleculares foram incluídos como descritores facultativos. Interessante observar que, apesar de serem apenas facultativos, os perfis genéticos de marcadores moleculares têm sido rotineiramente incluídos nos pedidos de proteção de cultivares de eucalipto. Isto se deve por um lado pela percepção que os melhoristas têm do valor de um clone elite para o negócio florestal e por outro pela posição avançada que o Brasil ocupa internacionalmente na pesquisa e desenvolvimento em genética, melhoramento e análise molecular do eucalipto (Grattapaglia *et al.* 2003b).

2.4 Marcadores Moleculares

O marcador molecular pode ser definido como sendo qualquer fenótipo molecular oriundo de um gene expresso, como no caso das isoenzimas, ou de um segmento específico de DNA correspondente a regiões expressas, ou não, do genoma (Ferreira & Gattapaglia, 1998). Marcadores moleculares que apresentam um comportamento Mendeliano simples podem, por sua vez, ser empregados como marcadores genéticos. Outra característica importante dos marcadores moleculares é o fato de sua herança não ser influenciada pelo meio ambiente e deles serem, em geral, neutros.

Uma das técnicas de utilização de marcadores moleculares depende da amplificação de seqüências de DNA. A amplificação do DNA tornou-se possível com a invenção da tecnologia PCR (*Polymerase Chain Reaction* – Reação em Cadeia da Polimerase), concebida por Kary Mullis em meados da década de oitenta. Desde sua concepção, esta tecnologia causou uma verdadeira revolução na biologia. A facilidade, rapidez, versatilidade e sensibilidade da PCR, a torna particularmente poderosa para estudos genéticos moleculares envolvendo grande número de indivíduos de qualquer organismo vivo (Ferreira e Grattapágia, 1998; Nicholas, 1999). Esta técnica promove a síntese enzimática *in vitro* de milhões de cópias de um segmento específico de DNA (Ferreira & Grattapaglia, 1998; Nicholas, 1999; Ramalho, Santos e Pinto, 2000), requerendo pequenas quantidades de DNA mesmo sem grande pureza (Ramalho, Santos e Pinto, 2000). A duplicação do DNA *in vitro* ocorre de forma semelhante ao que ocorre dentro da célula, exigindo basicamente os mesmos componentes: DNA molde, desoxirribonucleotídeos trifosfatados (dATP, dCTP, dGTP, dTTP), DNA polimerase, e iniciadores oligonucleotídeos (*primers*). A PCR inclui, em geral, de 30 a 45 ciclos repetidos de amplificação, com três etapas em cada ciclo: inicialmente a desnaturação do DNA a temperaturas que podem variar de 90 a 95°C, em seguida o anelamento dos iniciadores (*primers*) a uma temperatura que depende do *primer* utilizado e, por último o alongamento da cadeia pela enzima a 72°C.

Os marcadores moleculares internacionalmente adotados para a identificação individual de plantas, animais e seres humanos são baseados na amplificação de segmentos curtos de DNA de 1 a 7 pares de bases repetidos em tandem, comumente conhecidos como microssatélites. Seqüências simples repetidas ou microssatélites foram originalmente identificados como cadeias de polipirimidinas, quando dos primeiros trabalhos que sequenciaram segmentos de DNA de eucariotos (Birnboim e Straus 1975; Hamada *et. al.* 1982) e mais tarde formalmente descritos como fontes de marcadores hipervariáveis no genoma humano e outros genomas de mamíferos (Tautz e Renz 1984; Tautz 1989; Legercrantz *et al* 1993). Logo em seguida foram também identificados e isolados no genoma de várias espécies de

plantas (Akkaya *et al.* 1992; Morgante e Olivieri 1993; Wang *et al.* 1994; Cho *et al.* 2000). Estas seqüências repetitivas são distribuídas por todo o genoma. O polimorfismo decorrente da variação no número de unidades repetitivas é revelado pela análise de produtos da PCR gerados com iniciadores específicos que flanqueiam o microssatélite. Para a resolução de diferenças de 2 pares de bases entre estes produtos, que correspondem aos alelos, a análise é realizada preferencialmente em géis de eletroforese de alta resolução com detecção automatizada de fluorescências em seqüenciador automático de DNA. Os alelos são identificados pelo seu tamanho em pares de bases de DNA. Este tipo de marcador molecular é amplamente utilizado para estudos de caracterização, identificação individual, testes de paternidade, construção de mapas genéticos, estudos de genética de populações (Ferreira & Grattapaglia, 1998; Cañon *et al.*, 2000; Bjornstad, Gumby & Roed, 2000; Cunningham *et al.*, 2001) e experimentos de mapeamento genético visando identificar regiões controladoras da variação em fenótipos de interesse (Collins *et al.*, 1997).

2.5 Sistemas multiplex de co-amplificação de microssatélites

Nos últimos 12 anos, com o avanço da automação e reagentes para análise molecular, o uso de marcadores moleculares baseados em microssatélites para fins de genotipagem individual evoluiu para o desenvolvimento de sistemas de análise multiloco semi-automatizada (Kimpton *et al.* 1993). Estes sistemas foram desenvolvidos inicialmente para estudos de genética humana (Fregeau *et al.* 1993; Urquhart *et al.* 1995). Eles baseiam-se na amplificação via PCR de um ou mais locos microssatélites, simultaneamente, utilizando iniciadores marcados com fluorocromos, e separação e visualização dos fragmentos em máquinas que permitam a detecção de fluorescência.

Sistemas semelhantes têm sido desenvolvidos para análise de espécies animais de importância econômica. O objetivo tem sido estabelecer sistemas universais e altamente robustos de identificação

genética para fins de exames de vínculo genético, estudos de variabilidade em populações e estimativas de distância genética. Entre as vantagens destes sistemas estão: (1) rapidez e acurácia na geração de dados; (2) avaliação simultânea de vários locos microssatélites em uma única pista do gel, reduzindo o número de análises necessárias em relação às técnicas convencionais; (3) aquisição e armazenamento dos dados (computadorizados) em tempo real, reduzindo a manipulação manual da informação; (4) possibilidade de exportação direta dos dados para programas de análise genética e (5) transferibilidade imediata do sistema para outros laboratórios, permitindo a padronização dos locos do genoma amostrados e a comparação e compartilhamento de resultados.

O desenvolvimento de sistemas *multiplex* com um maior número de marcadores não é, entretanto, uma tarefa trivial. Alguns trabalhos têm sido publicados procurando fornecer procedimentos sistemáticos para isso (Shuber *et al.* 1995; Henegariu *et al.* 1997; Butler 2005b), e *softwares* tem sido desenvolvidos para facilitar a avaliação da complementaridade entre os vários *primers* (Vallone e Butler 2004). Na prática, entretanto, este procedimento ainda envolve uma boa dose de tentativa e erro. Os principais requerimentos para que microssatélites sejam candidatos a fazerem parte de um sistema *multiplex* robusto incluem: (1) os marcadores devem ser geneticamente independentes de forma que seja possível aplicar a regra do produto; para isso é necessário conhecer a posição de mapa dos locos; (2) os marcadores devem ter um número elevado de alelos com freqüências o mais homogêneas possível de forma a maximizar o conteúdo informativo do loco; (3) os marcadores preferivelmente não devem ter alelos microvariantes de forma a facilitar a análise a qual seguirá uma série alélica constante de diferença de pares de base; (4) ter seqüência genômica disponível dos dois lados dos *primers* inicialmente desenhados de forma a permitir o redesenho de *primers* para amplificação de segmentos de tamanhos menores ou maiores e com isso permitir o melhor encaixe no *multiplex*; (5) os vários *primers* que serão utilizados conjuntamente na mesma PCR devem ter seqüências compatíveis, ou seja, com possibilidade zero ou mínima de anelamento entre si.

No caso de plantas cultivadas, entretanto, embora muitos artigos tenham sido publicados relatando a utilização de microssatélites para as mais variadas aplicações, muito pouco tem sido feito no sentido de otimizar sistemas *multiplex* de microssatélites para a geração de perfis multiloco. Em milho, Matsuoka *et al.* (2002) genotiparam 46 locos em 17 *multiplexes* com em média 3 marcadores por PCR. Em *Eucalyptus leucoxyton* oito microssatélites foram otimizados em três PCRs distintas (Ottewell *et al.* 2005). Embora a eficiência seja maior do que a análise de locos individualmente, a simples combinação de três locos de cada vez ainda representa uma exploração muito limitada do potencial de *multiplexagem* de microssatélites. Em girassol, por exemplo, um conjunto de 13 sistemas *multiplex* de 6 locos cada foi desenvolvido para cobrir grande parte do genoma para fins de mapeamento (Tang *et al.* 2003). Em humanos e animais sistemas com 10 a 20 microssatélites co-amplificados na mesma PCR tem sido desenvolvidos (Butler 2005a,b) representando hoje o paradigma em termos de otimização de sistemas de genotipagem de microssatélites.

No caso de espécies de *Eucalyptus*, cerca de 370 microssatélites já foram publicados, sendo 300 desenvolvidos a partir de *E. grandis* e *E. urophylla* (Brondani *et al.* 1998; 2002; 2006) e algumas dezenas a partir de outras espécies como *E. globulus*, *E. nitens*, *E. sieberi* e *E. leucoxyton* (Steane *et al.* 2001; Byrne *et al.* 1996; Glaubitz *et al.* 2001; <http://www.ffp.csiro.au/tigr/molecular/eucmsps.html>, acessado em nov. 2007; Ottewell *et al.* 2005). Embora alguns trabalhos relataram o uso de microssatélites genotipados em sistemas *multiplex*, estes se limitaram a combinar 3 a no máximo 6 locos em cada PCR e nenhum investimento mais específico foi feito no sentido de otimizar sistemas com maior número de locos (Grattapaglia *et al.* 2003a; Mamani *et al.* 2004).

Com a utilização crescente de microssatélites para fins de identificação clonal e proteção de cultivares, combinada com a prática cada vez mais freqüente de intercâmbio de material genético elite na forma de clones entre empresas nacionalmente e internacionalmente, tem se tornado cada vez mais

premente a necessidade de um conjunto otimizado de microssatélites co-amplificado para *Eucalyptus*. O fato que o melhoramento genético de *Eucalyptus* nos trópicos se baseia em grande parte na geração e utilização de híbridos interespecíficos, é fundamental que os microssatélites utilizados tenham ampla e robusta transferibilidade interespecífica ao mesmo tempo em que apresentem níveis adequados de multialelismo na grande maioria das principais espécies cultivadas.

Além dos aspectos de transferibilidade interespecífica e elevado nível de polimorfismo, um aspecto fundamental ao se pensar na utilização de marcadores microssatélites para fins de identificação individual e comparação interlaboratorial de genótipos é a facilidade e robustez com a qual são corretamente declarados os alelos a cada loco. A quase totalidade dos microssatélites desenvolvidos e publicados em plantas cultivadas são derivados de seqüências repetidas de dinucleotídeos e uma pequena proporção derivada de trinucleotídeos. É consenso na comunidade científica que utiliza microssatélites que a interpretação robusta de dinucleotídeos é muitas vezes complexa em função da pequena diferença esperada de apenas 2 pares de bases no tamanho dos alelos, ou mesmo de apenas um par de bases no caso de alelos microvariantes. Além disso, dificuldades de interpretação surgem com (1) a ocorrência generalizada do fenômeno do gaguejamento (*stuttering*) da DNA polimerase que gera produtos que diferem do produto correto em dois ou mais pares de bases gerando muitas vezes múltiplos picos em eletroferogramas; (2) a presença de produtos de PCR aos quais não foi incorporada a base Adenina gerando assim os chamados produtos-A que diferem em um par de bases; e (3) migração e visualização das duas fitas de DNA amplificado nos casos da detecção dos alelos serem realizada por calorimetria e não por marcação fluorescente de um dos *primers*. Na utilização de análise de DNA para fins forenses humanos já foi estabelecido há mais de 10 anos por diferentes grupos de trabalho, que todos os marcadores microssatélites devem ser derivados de seqüências de tetra e pentanucleotídeos de forma de evitar ou pelo menos minimizar significativamente

estes artefatos analíticos que podem gerar interpretações errôneas de genótipos principalmente nos casos de análise de amostras misturadas.

2.6 Microssatélites derivados de tetra e pentanucleotídeos em plantas

Atualmente todos os sistemas padronizados de microssatélites para identificação individual em seres humanos são baseados em tetranucleotídeos. No caso de cães (Halverson e Basten 2005) e gatos (Menotti-Raymond 2005), da mesma forma, sistemas otimizados de tetranucleotídeos foram desenvolvidos para identificação individual. No caso de eqüinos, bovinos e ovinos, entretanto, os microssatélites recomendados até hoje pela ISAG (International Society for Animal Genetics) ainda são derivados de dinucleotídeos (ISAG International Society for Animal Genetics <http://www.isag.org.uk>, acessado em nov. 2007). Isto provavelmente se deve à baixa frequência de seqüências de tetranucleotídeos no genoma destas espécies (Stone *et al.* 1995) o que explica a ausência de microssatélites de tetranucleotídeos nos bancos de dados de eqüinos e bovinos mantidos pelo Roslin Institute (<http://www.thearkdb.org>, acessado em nov. 2007).

Em plantas, praticamente nada tem sido feito no sentido de se desenvolver marcadores microssatélites baseados em tetra e/ou pentanucleotídeos para fins de identificação individual. A maioria dos trabalhos apenas cita a simples observação de microssatélites de tetranucleotídeos em diferentes espécies de plantas, porém somente alguns relatam avaliações experimentais limitadas do desenvolvimento de marcadores moleculares a partir destas seqüências. Baterias de microssatélites derivados de trinucleotídeos vêm sendo, entretanto, desenvolvidas em algumas das principais espécies. Em soja, por exemplo, várias dezenas de microssatélites baseados em trinucleotídeos foram desenvolvidos e mapeados (Song *et al.* 2004). Em *Picea abies*, uma conífera, um trabalho de desenvolvimento de

microssatélites de trinucleotídeos revelou um nível mais baixo de polimorfismo do que dinucleotídeos (Scotti *et al.* 2002).

Pouco se sabe sobre o comportamento de marcadores microssatélites derivados de tetra e menos ainda pentanucleotídeos em plantas do ponto de vista de polimorfismo e robustez analítica e nada existe ainda sobre este tipo de marcadores em espécies de *Eucalyptus*. Embora esta observação sugira que a frequência de tetra e pentanucleotídeos em genomas de plantas seja baixa, isso não parece ser verdade. Com os recentes esforços de seqüenciamento genômico parcial ou total de diversas espécies de plantas, bem como a construção de bancos de seqüências EST, uma grande quantidade de seqüências se tornou disponível o que tem permitido a localização de um número significativo de seqüências simples repetidas de tetra e pentanucleotídeos. Morgante *et al.* (2002) mineiraram uma grande quantidade de seqüências genômicas e de EST disponíveis em bancos públicos e privados de cinco espécies de plantas (*Arabidopsis*, milho, arroz, soja e trigo) buscando SSR com tamanho mínimo total de 12 bases e pelo menos três repetições de motivos de 1 a 5 pares de bases. Eles identificaram uma densidade de tetranucleotídeos perfeitos variando de um mínimo de 45/Mpb para milho a um máximo de 86/Mpb para *Arabidopsis* e arroz enquanto que dinucleotídeos foram observados com frequência de 27/Mpb em milho a um máximo de 77/Mpb em *Arabidopsis*. Por outro lado pentanucleotídeos efetivamente são mais raros com um mínimo de 8/Mpb em trigo a um máximo de 30/Mpb em *Arabidopsis*. No genoma de *Eucalyptus* nada se sabe até o momento sobre a frequência de repetições de tetra e pentanucleotídeos e marcadores baseados nestas seqüências não existem.

3. OBJETIVOS

1. Desenvolver um conjunto de marcadores microssatélites de amplificação robusta baseados em tetra e pentanucleotídeos derivados de seqüências genômicas de extremidades de clones BAC.
2. Caracterizar a variabilidade alélica, desempenho forense e transferibilidade de uma bateria de locos selecionados após etapas de triagem para amplificação, transferibilidade interespecífica e hipervariabilidade nas três principais espécies de *Eucalyptus* plantadas no mundo.
3. Mapear os locos polimórficos derivados de tetra e pentanucleotídeos em mapas genéticos referência de *Eucalyptus*.
4. Para um conjunto de locos selecionados, comparar a variabilidade alélica e poder de discriminação individual de tetra e pentanucleotídeos e dinucleotídeos em um grupo de 46 clones elite de *Eucalyptus*.
5. Otimizar as condições de co-amplificação de sistemas *multiplex* preliminares baseados na co-amplificação de locos tetra e pentanucleotídeos.

4. MATERIAL E MÉTODOS

4.1 Material genético

Para a realização dos estudos foram utilizados os seguintes materiais genéticos:

(a) Clones elite: Foi estudado um grupo de 46 clones elite públicos, representativo de material elite utilizado operacionalmente em plantios clonais de eucalipto no Brasil. Dos 46 clones elites, 33 são clones híbridos envolvendo principalmente *E. grandis* e em menor escala *E. urophylla* e *E. camaldulensis* plantados em um teste clonal no município de Cristalina, e os demais 13 clones de *E. globulus* pertencentes à Mundial Forestación no Uruguay. Estes clones são resultado de uma elevada intensidade de seleção para crescimento volumétrico (1.200.000 m³ em média) praticada em plantios comerciais entre os anos 70 e 80 para os híbridos de *E. grandis* e anos 90 para *E. globulus*.

(b) Populações naturais: Um conjunto de 48 plantas de cada uma de três espécies de *Eucalyptus* (*E. grandis*, *E. globulus* e *E. urophylla*) foi amostrado a partir de coleções de germoplasma estabelecidas com sementes oriundas da Austrália. Para cada espécie foram estudadas 24 plantas de cada uma das duas principais procedências plantadas no Brasil e no mundo. No caso de *E. globulus* foram estudadas as procedências de Flinder Island (ilhas ao sul do continente Australiano paralelo 40° sul) e de Jeerlang (Victoria, paralelo 37° sul), ambas de especial importância para o melhoramento devido à sua excelente performance e que além disso apresentaram características distintas de variabilidade em estudos recentes (Steane *et al.* 2006). Para *E. grandis* foram estudadas as procedências de Atherton (Queensland, norte da Austrália paralelo 17° sul) e Coffs Harbor (New South Wales, paralelo 30° sul), as duas principais procedências comerciais utilizadas nos programas de melhoramento do Brasil. No caso de *E. urophylla* foram analisadas árvores das procedências de Timor (Timor-leste paralelo 9° sul) e da ilha de Flores (Indonésia, paralelo 8° sul). É interessante notar ainda que a distribuição geográfica das espécies é contrastante: enquanto *E. globulus* ocorre em manchas ao sul da Austrália, *E. grandis* tem distribuição contínua ao longo da costa de New South Wales e Queensland e *E. urophylla* ocorre fora do continente australiano, nas ilhas ao norte e Indonésia. Cada árvore amostrada pertence a uma família de meios-irmãos de polinização aberta diferente, de modo a minimizar a probabilidade de elas

serem geneticamente relacionadas. Estas árvores estão plantadas em coleções de germoplasma nas áreas da Aracruz Celulose nos municípios de Guaíba (RS) e Aracruz (ES). As amostras foram enviadas na forma de folhas, devidamente embaladas e identificadas. A extração do DNA de todos os materiais foi realizada a partir de tecido foliar fresco utilizando o protocolo de descrito por Grattapaglia e Sederoff (1994).

4.2 Mineração de microssatélites derivados de tetra e pentanucleotídeos

Esta etapa do trabalho envolveu a mineração de um banco de dados composto por 18.510 seqüências totalizando 9.635.099 bases de alta qualidade derivadas do seqüenciamento das extremidades de clones BAC (cromossomos artificiais de bactérias) de uma biblioteca da árvore G21 de *E. grandis* construída no âmbito do projeto Genolyptus entre 2003 e 2005 por Sergio Brommonschenkel. A mineração foi realizada buscando microssatélites perfeitos, ou seja, sem interrupções, com um total de pelo menos 10 bases de seqüência simples repetitiva. Para os microssatélites de tetra e pentanucleotídeos foram desenhados pares de iniciadores flanqueando os microssatélites buscando condições padrão de qualidade, ou seja, ausência de dímeros e complementaridade entre os iniciadores, iniciadores com 20 a 25 bases e uma temperatura de anelamento homogênea entre 50 e 60 °C. A mineração e desenho de iniciadores foram realizadas por Georgios Pappas utilizando um “pipeline” do Sistema Genoma.

4.3 Triagem de marcadores microssatélites

Visando o desenvolvimento de marcadores microssatélites operacionais foram realizadas duas etapas sucessivas de triagem de um conjunto de pares de iniciadores desenhados após a mineração.

Etapa de triagem 1: Foi realizada uma triagem de 110 pares de *primers* selecionados a partir da mineração, privilegiando microssatélites mais longos, ou seja, com maior número de unidades repetitivas. Na triagem foi utilizado um painel de 6 plantas sendo duas de cada uma de três espécies (*E. grandis*, *E. urophylla* e *E. globulus* – representantes das três principais seções do subgênero *Symphyomyrtus*, *Transversaria*, *Exertaria* e *Maidenaria* respectivamente) para análise de transferibilidade, sucesso e robustez de amplificação e sugestão de polimorfismo em gel de agarose. Como controle positivo foi utilizado DNA da árvore G21 de *E. grandis* da qual foi originalmente construída a biblioteca de BAC e, portanto, cuja seqüência foi utilizada para o desenho dos *primers*, e como controle negativo foi utilizada água. A amplificação do DNA foi realizada mediante reação de PCR (Reação em Cadeia da Polimerase) utilizando 4 ng. de DNA, 1X de Tampão (Phoneutria) (contendo em sua composição 1,5 mM de MgCl₂) 0,4 µM de dNTP (Invitrogen), 0,6 unit Taq DNA polymerase Phoneutria, 0,3 µM de cada iniciador, 0,25 mg/ml de BSA (Albumina de Soro Bovina) e H₂O para completar 12 µl de volume final. As reações foram realizadas em um termociclador Applied Biosystem 9600 thermal controller (Applied Biosystems, Foster City, CA) sob as seguintes condições: 94°C por 5 min (1 ciclo), 94°C por 1 min, 57 °C por 1 min., 72 °C por 1 min. (30 ciclos), 72 °C por 40 min. (1 ciclo). Os produtos de amplificação foram visualizados em minigéis de agarose 2,5% em tampão SB 1X (Sodium Boric Acid, Brody e Kem, 2004) a 80V. por 50 min.

Etapa de triagem 2: Considerando que a declaração de presença ou ausência de polimorfismo em géis de agarose seria apenas aproximada - podendo vir a descartar locos que efetivamente seriam polimórficos - todos os mesmos 110 pares de iniciadores foram novamente avaliados em um painel aumentado com 24 plantas (oito de cada espécie) utilizando a mesma reação e programa de amplificação utilizado na triagem 1, mas com detecção em géis de poliacrilamida 5% a 90V., por 90 min. em condição desnaturante. A visualização foi realizada com coloração com nitrato de prata

visando uma avaliação do potencial de revelação de polimorfismo, seja dentro como entre espécies, e uma estimativa preliminar do número, frequência e faixa de tamanho dos principais alelos observados. Após a segunda triagem foram selecionados os microssatélites que apresentaram transferibilidade interespecífica, amplificação robusta e fácil interpretação para síntese do iniciador direto com marcação fluorescente no terminal 5'. Uma análise detalhada foi realizada para selecionar a melhor combinação de fluorescências para cada loco com base na faixa de tamanho de alelos observada na etapa 2 de triagem visando: (a) separar os locos por faixa de tamanho aproximada (3 faixas); (b) dentro de cada faixa marcar um loco com cada uma de três cores diferentes, FAM (marcação azul), HEX (marcação verde) e NED (marcação amarela).

4.4 Caracterização genética dos microssatélites de tetra e pentanucleotídeos

Para uma bateria selecionada de locos após as etapas de triagem e eliminação de locos duplicados ou não inteiramente transferíveis entre espécies, foi analisado um conjunto de 144 árvores de populações naturais das três espécies alvo (veja item 4.1 para detalhamento) visando uma caracterização detalhada dos locos e do desempenho forense dos mesmos individualmente e em conjunto. A amplificação dos microssatélites com o iniciador marcado com fluorescência foi realizada em sistemas multiplex de co-amplificação utilizando o Kit QIAGEN® Multiplex PCR 2X sob as seguintes condições: 95°C por 15 min (1 ciclo), 94°C por 30 seg. para desnaturação, 57°C por 90 seg. para anelamento, 72°C por 1 min para extensão (30 ciclos) e 60°C por 30 min para extensão final (1 ciclo). Os produtos amplificados foram eletroinjetados em um seqüenciador automático de DNA (ABI PRISM 3100, Applied Biosystems) e o filtro D foi utilizado para a leitura das fluorescências. A análise dos fragmentos foi realizada com o programa GeneScan Analysis 2.1 (*Applied Biosystems*) e a genotipagem com o programa Genotyper 2.1 (*Applied Biosystems*). Os dados genotípicos em pares de bases com duas casas decimais de precisão foram ajustados para classes alélicas definidas utilizando o *software*

Allelobin que implementa um método de minimização baseada em quadrados mínimos descrita por Idury & Cardon (1997). Foram estimadas a heterozigidade esperada (H_e) e observada (H_o) o número de alelos e freqüências alélicas para cada loco a partir dos dados genotípicos observados. Estas análises foram realizadas utilizando o programa GENALEX (Peakall e Smouse 2006). O desempenho forense esperado de cada loco e de baterias específicas combinadas para análises genéticas forenses nas três espécies foi avaliado através das seguintes estimativas de parâmetros: Probabilidade de identidade (PI); Poder de exclusão de paternidade (PE) e PIC – (Polymorphism Information Content) Conteúdo de informação polimórfica. A planilha Powerstats (Tereba 1999) foi utilizada para automatizar estas análises. Estas estimativas foram obtidas para cada procedência individualmente, para cada espécie consolidando as duas procedências e a média das três espécies. Estimativas de média e variância de PI e PE foram obtidas visando classificar os locos quanto ao desempenho para utilização ampla no gênero.

4.5 Herança, segregação e mapeamento de microssatélites tetra e pentanucleotídeos selecionados

Todos os microssatélites desenvolvidos foram avaliados quanto ao comportamento mendeliano de herança e segregação. Este trabalho envolveu duas etapas. Inicialmente foi realizada uma etapa de verificação de herança dos marcadores de parentais para descendentes e segregação mendeliana em duas famílias de *E. grandis* x *E. urophylla* (família IP International Paper e família VCP Votorantim Celulose e Papel) para as quais haviam sido construídos no âmbito do projeto Genolyptus dois mapas genéticos de referência com cerca de 200 microssatélites dinucleotídeos. Foram utilizadas duas famílias com o objetivo de validar a herança e segregação em dois pedigrees distintos e não relacionados e ao mesmo tempo maximizar a probabilidade de identificar uma configuração de segregação informativa que permitisse o mapeamento do maior número possível dos novos

microsatélites desenvolvidos. A triagem foi realizada utilizando os dois pais e 6 descendentes para cada família. Com base na detecção de polimorfismo e do nível de informação genética da configuração de segregação observada foi organizada a segunda etapa de mapeamento definindo quais famílias seriam usadas para mapear quais microsatélites, priorizando configurações totalmente informativas que permitissem o mapeamento nos mapas de ambos os parentais.

Na etapa de mapeamento foram genotipados os locos microsatélites em 96 descendentes por família, número suficiente para estimar a sua frequência de recombinação em relação aos demais marcadores já mapeados bem como ordená-los com elevada confiabilidade estatística utilizando um suporte de LOD > 3,0. A análise de mapeamento foi realizada com o programa MapMaker utilizando um LOD mínimo de 4,0 e uma fração de recombinação de 0,4 analisando conjuntamente os dados de segregação dos novos microsatélites juntamente com o conjunto de dados existente para cada uma das duas famílias.

4.6 Montagem e aplicação de sistemas multiplex de co-amplificação de microsatélites

Um conjunto de locos microsatélites de tetra ou pentanucleotídeos com maior conteúdo informativo foram reunidos em sistemas *multiplex* de PCR procurando a melhor combinação entre a cor da marcação com fluorescência disponível, a faixa de tamanho dos alelos de cada marcador, a compatibilidade dos iniciadores avaliada com o software Multiplexer (Coelho A., não publicado) e os parâmetros de maior média de PE, menor média de PI e menor variância interespecífica destes dois parâmetros. A meta foi utilizar 12 locos em dois sistemas multiplex de seis locos cada. Os locos foram ainda selecionados por estar em grupos de ligação distintos e, caso no mesmo grupo, a distâncias grandes de forma a segregar de forma independente. Para um maior controle da amplificação e

certificação de uma provável troca de amostras entre multiplex, dois locos foram replicados entre os dois *multiplexes*, totalizando assim 7 locos por multiplex.

Os dois sistemas multiplex foram utilizados para genotipar uma coleção de 46 clones elite comerciais. Estes mesmos 46 clones foram também genotipados com 9 locos microssatélites dinucleotídeos utilizados em rotina para identificação em *Eucalyptus* visando fazer uma análise comparativa da capacidade de discriminação individual entre di e tetra/pentanucleotídeos. Os resultados foram representados em fenogramas ilustrativos baseados no compartilhamento de alelos construídos com o software *Share Allele Distance* (<http://www2.biology.ualberta.ca/jbrzusto/sharedst.php>).

5. RESULTADOS

5.1 Desenvolvimento de microssatélites derivados de tetra e pentanucleotídeos

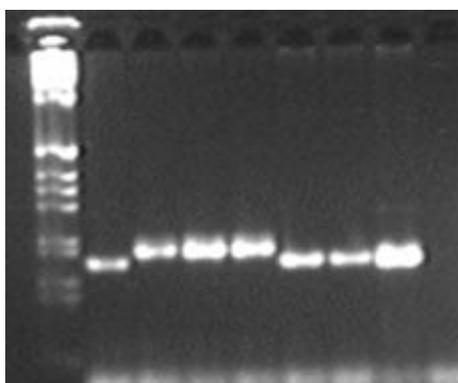
A mineração das 18.510 seqüências genômicas de clones BAC para microssatélites perfeitos identificou 2.406 seqüências contendo um ou mais microssatélites de duas ou mais bases e um total de 3.034 microssatélites perfeitos com pelo menos 10 bases de seqüência simples repetitiva (ex. 5 dinucleotídeos ou 4 trinucleotídeos etc.). Os 3.034 microssatélites foram encontrados ao se analisar 9.635.099 bases correspondendo assim a uma densidade média de 1 microssatélite dinucleotídeo ou maior perfeito a cada 3.175 pb no genoma de *E. grandis*. Para estes 3.034 microssatélites foi possível desenhar pares de *primers* flanqueando 1.896 microssatélites nas condições de Tm dos iniciadores entre 59 e 60 °C, tamanho de 20 a 25 bases e produto de tamanho mínimo de 80 pb. e máximo de 550 pb. Para os demais 1.138 microssatélites não foi possível desenhar iniciadores nas condições estabelecidas. Nestas condições de seleção, foram desenhados iniciadores para 728 dinucleotídeos, 495 trinucleotídeos, 304 tetranucleotídeos, 196 pentanucleotídeos e 107 hexanucleotídeos, além de

mais 66 seqüências repetitivas com 7 ou mais bases repetidas das quais um minissatélite com a seqüência núcleo de 71 pb repetida 3 vezes.

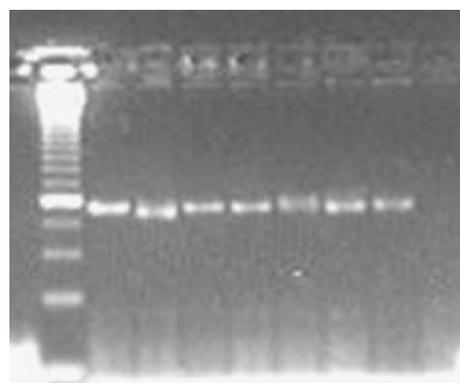
Considerando que, em geral, microsatélites mais longos tendem a ser mais hipervariáveis, para o desenvolvimento de marcadores foi realizada uma seleção de tetra e pentanucleotídeos priorizando aqueles que apresentassem quatro ou mais unidades repetitivas, ou seja, 16 pb para tetra ou 20 pb para pentanucleotídeos com tamanho de produto abaixo de 450 pb. Esta seleção resultou em um conjunto de 110 pares de iniciadores utilizados neste trabalho, sendo 80 tetranucleotídeos e 30 pentanucleotídeos.

Na primeira etapa da triagem, todos os pares de iniciadores (*primers*) foram testados utilizando 2 indivíduos de cada uma das espécies (*E. grandis*, *E. globulus* e *E. urophylla*). Nesta triagem preliminar foi observada transferibilidade de 100% em 51 locos tetranucleotídeos e 26 locos pentanucleotídeos para as três espécies. Dentro deste grupo de locos 100% transferíveis, em 27 tetras (53%) e 6 pentas (23%) foi possível detectar polimorfismo com resolução em agarose (Figura 2). Entretanto neste nível de resolução foi difícil declarar com certeza a não transferibilidade e/ou a inexistência de polimorfismo o que demandou a segunda etapa de triagem para uma avaliação mais precisa do potencial de cada par de iniciadores.

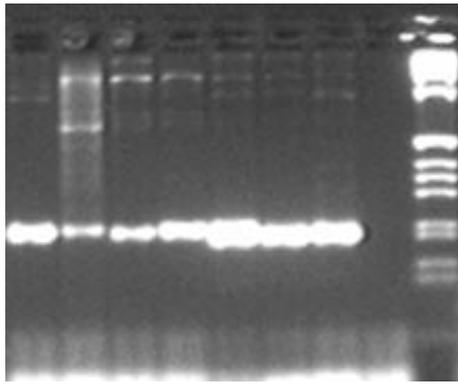
A



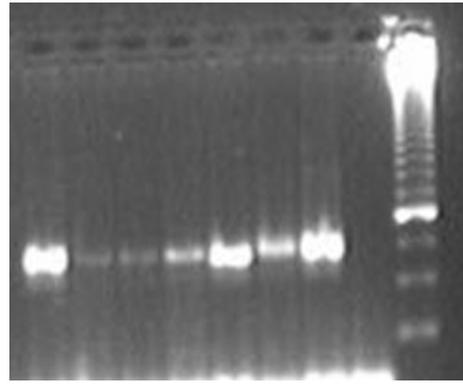
EMBRA 2549



EMBRA 3275

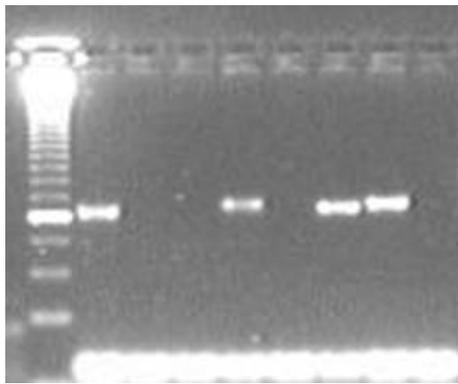


EMBRA 2059

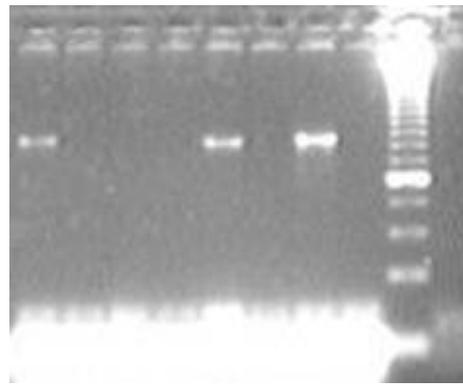


EMBRA 2134

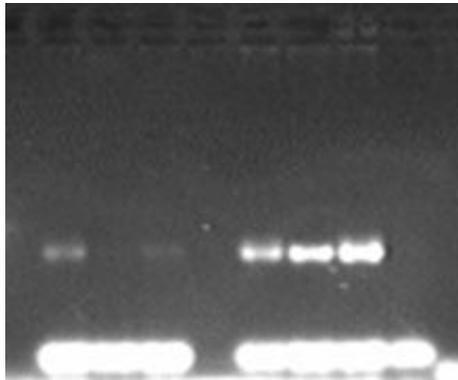
B



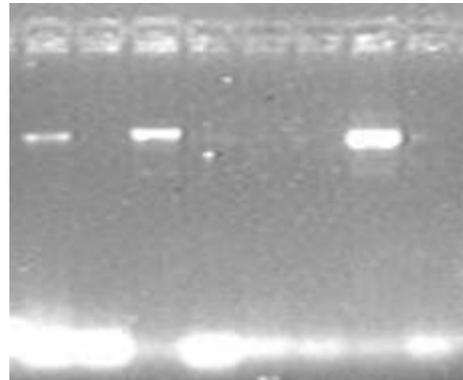
EMBRA 3372



EMBRA 3819

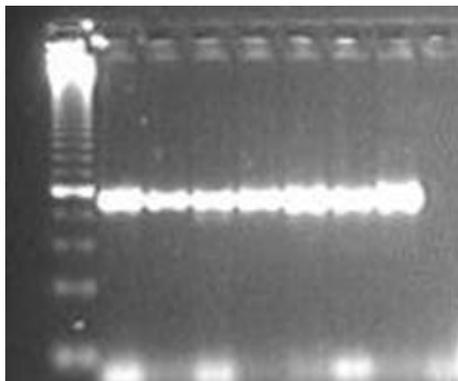


EMBRA 2086

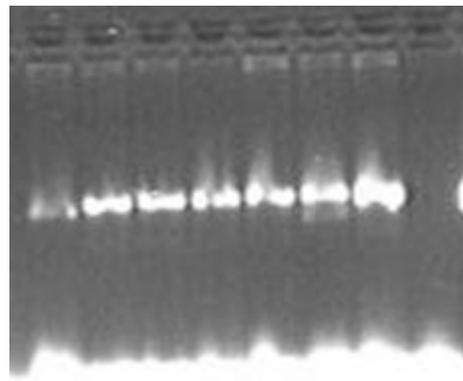


EMBRA 3086

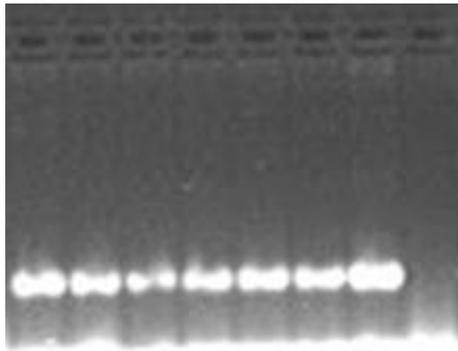
C



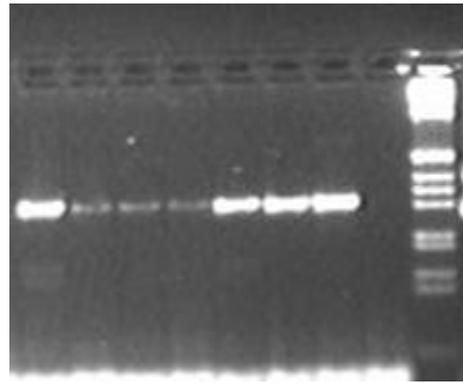
EMBRA 2052



EMBRA 2321

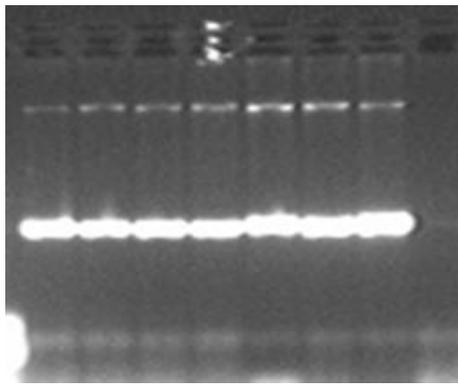


EMBRA 2108

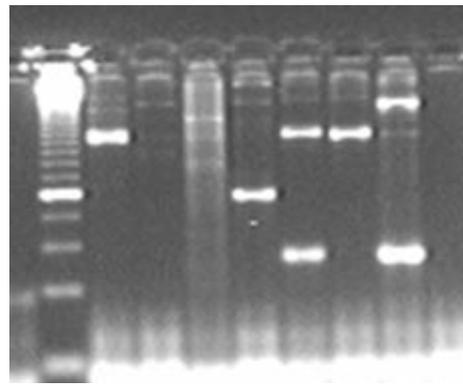


EMBRA 3205

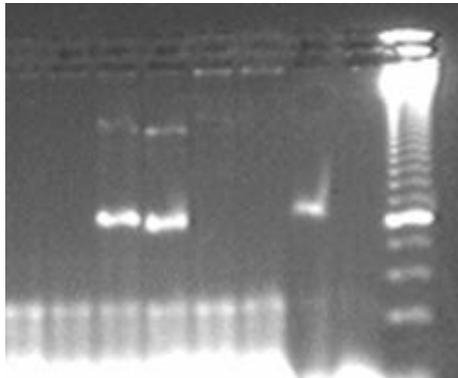
D



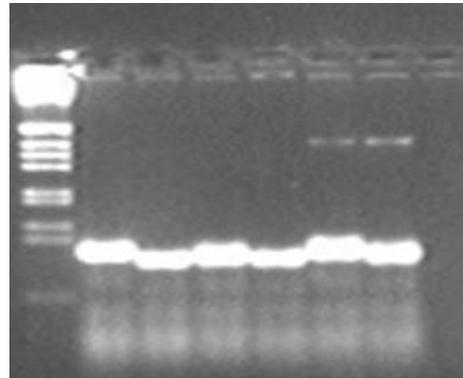
EMBRA 3457



EMBRA 2827

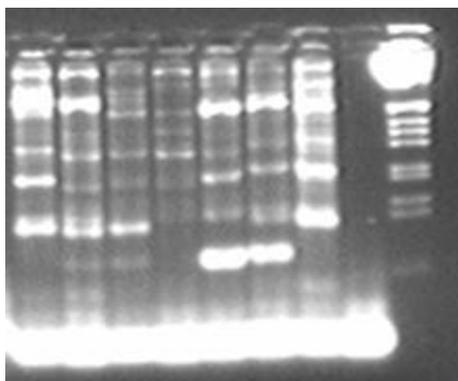


EMBRA 2918

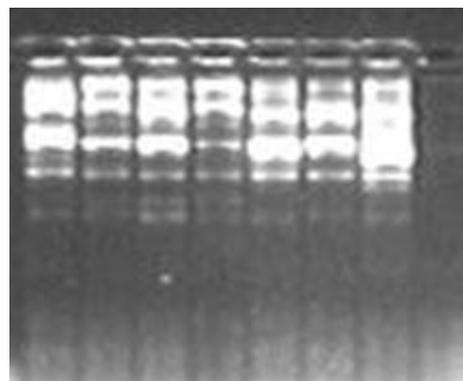


EMBRA 2420

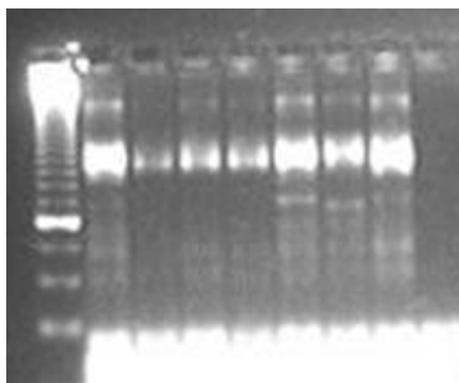
E



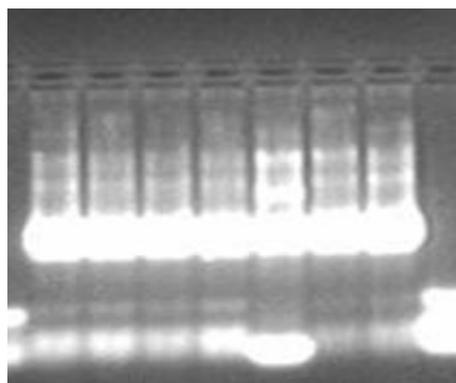
EMBRA 3465



EMBRA 3047



EMBRA 2967



EMBRA 2214

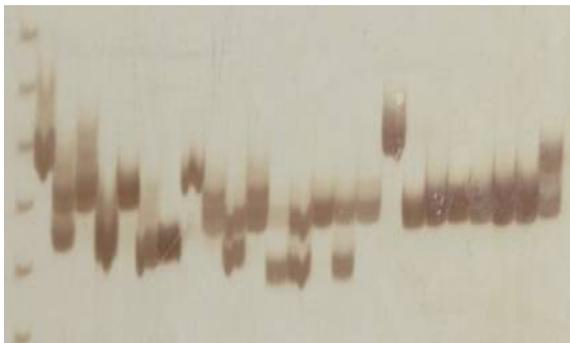
Figura 2: Exemplos de resultados da etapa 1 de triagem de locos tetra e pentanucleotídeos em gel de agarose. Em cada painel, da esquerda para direita, são analisados dois indivíduos de cada espécie (*E. grandis*, *E. globulus* e *E. urophylla*), controle positivo (*E. grandis* indivíduo G21) e controle negativo (água). Na coluna da esquerda são apresentados locos tetranucleotídeos e na coluna da direita os locos pentanucleotídeos. Grupo A: locos com indicação de polimorfismo e 100% de transferibilidade interespecífica; Grupo B: locos com < 100% de transferibilidade; Grupo C: locos monomórficos; Grupo D: locos duplicados; Grupo E: locos com amplificação de produtos de forma inespecífica.

Além das dificuldades de se declarar o efetivo nível de transferibilidade interespecífica ou mesmo o sucesso de amplificação, a resolução de fragmentos em géis de agarose 2,5% não permite distinguir heterozigotos com precisão ou alelos próximos em tamanho. Desta forma para poder averiguar a efetiva ocorrência de polimorfismo nestes locos, todos os locos da primeira triagem, independentemente do resultado da primeira triagem, foram novamente analisados na segunda etapa de triagem em géis de poliacrilamida (PAGE) em condições desnaturantes, metodologia esta que permite uma melhor declaração do sucesso de amplificação, transferibilidade e distinção de alelos com 2pb de diferença o que corresponderia a microvariantes no caso de tetra e pentanucleotídeos.

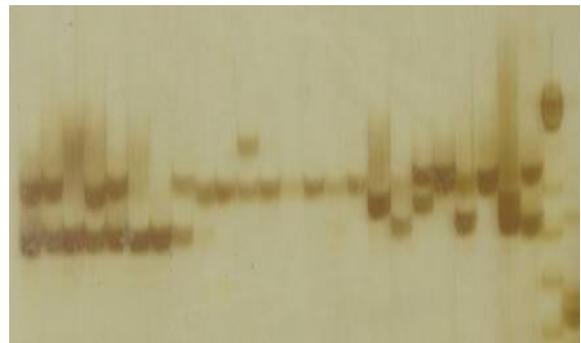
A visualização realizada em PAGE constituiu a segunda etapa de triagem dos marcadores. Após a segunda etapa de triagem os locos de cada tipo (tetra e pentanucleotídeos) foram classificados em

cinco categorias, resultantes da combinação dos seguintes critérios: qualidade de amplificação, presença de polimorfismo alélico e presença de duplicação do loco. Na figura 3, são mostrados alguns exemplos de resultados da análise em géis de poliacrilamida de microssatélites tetra e pentanucleotídeos dos locos durante a segunda etapa de triagem.

A



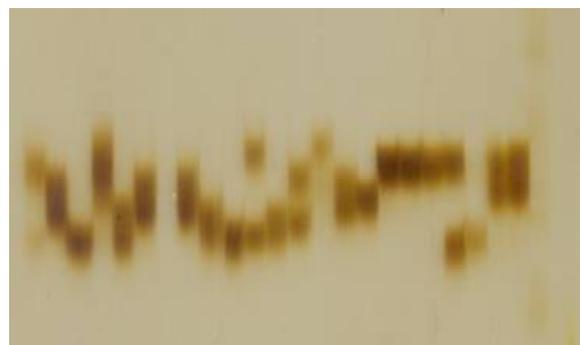
EMBRA 2052



EMBRA 3808



EMBRA 3769



EMBRA 3894

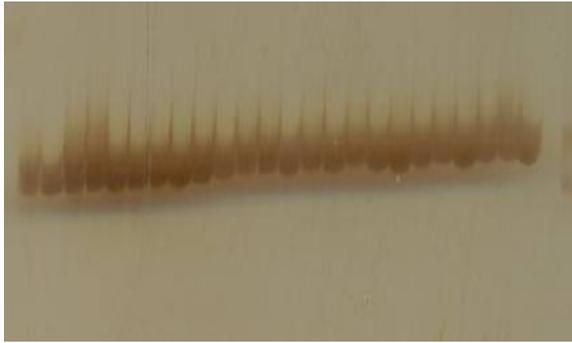
B



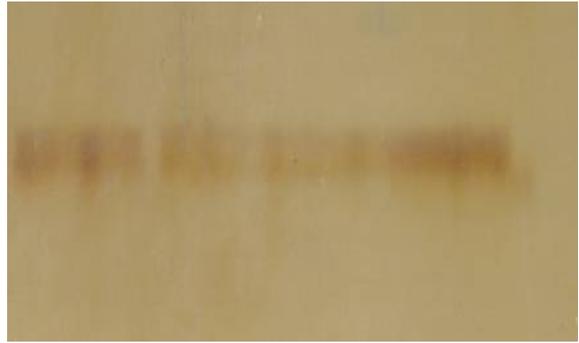
EMBRA 3058



EMBRA 3205



EMBRA 3312



EMBRA 2671

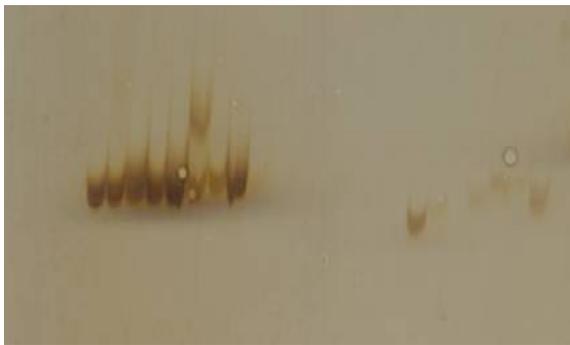
C



EMBRA 3086



EMBRA 2910

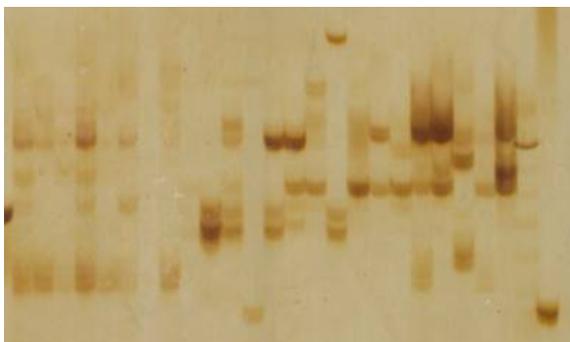


EMBRA 3490

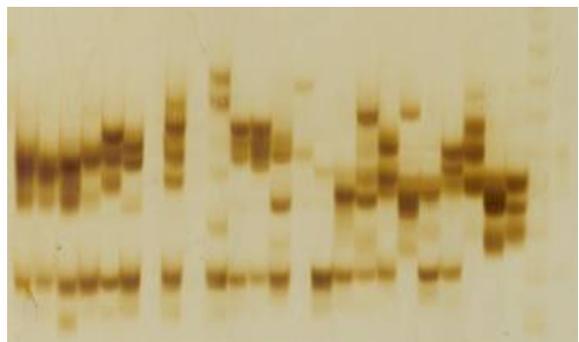


EMBRA 3386

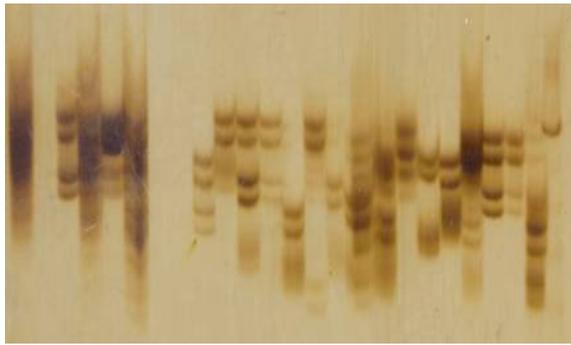
D



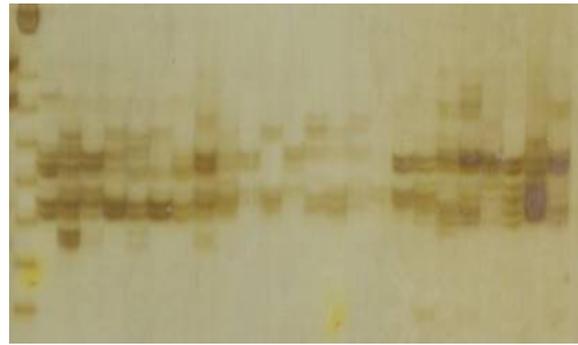
EMBRA 2322



EMBRA 2462



EMBRA 2011



EMBRA 3865

Figura 3: Exemplos de resultados da etapa 2 de triagem em gel de poliacrilamida desnaturante com 24 indivíduos (8 indivíduos de cada uma de três espécies *E. grandis*, *E. globulus* e *E. urophylla*). A coluna da esquerda apresenta locos tetranucleotídeos e a coluna da direita, locos pentanucleotídeos. Grupo A: locos polimórficos com 100% de transferibilidade; Grupo B: locos monomórficos; Grupo C: locos com <100% de transferibilidade; Grupo D: locos com amplificação inespecífica e/ou duplicações do loco.

Após a segunda etapa de triagem foi verificado que 30 pares de iniciadores, sendo 27 tetranucleotídeos e 3 pentanucleotídeos efetivamente não amplificaram produtos. Estes foram descartados neste ponto do desenvolvimento. Os locos que amplificaram produtos com sucesso em pelo menos um indivíduo de cada uma das três espécies foram classificados de acordo com uma estimativa preliminar de heterozigidade esperada, resultante do número e frequências relativas de alelos observado na amostra de 24 indivíduos e presença de duplicações do loco amplificado. As categorias estabelecidas foram as seguintes: locos polimórficos com heterozigidade esperada H_e maior do que 45%, locos polimórficos com H_e menor do que 45%, locos duplicados, locos monomórficos e locos de difícil interpretação devido às amplificações inespecíficas (Figura 4).

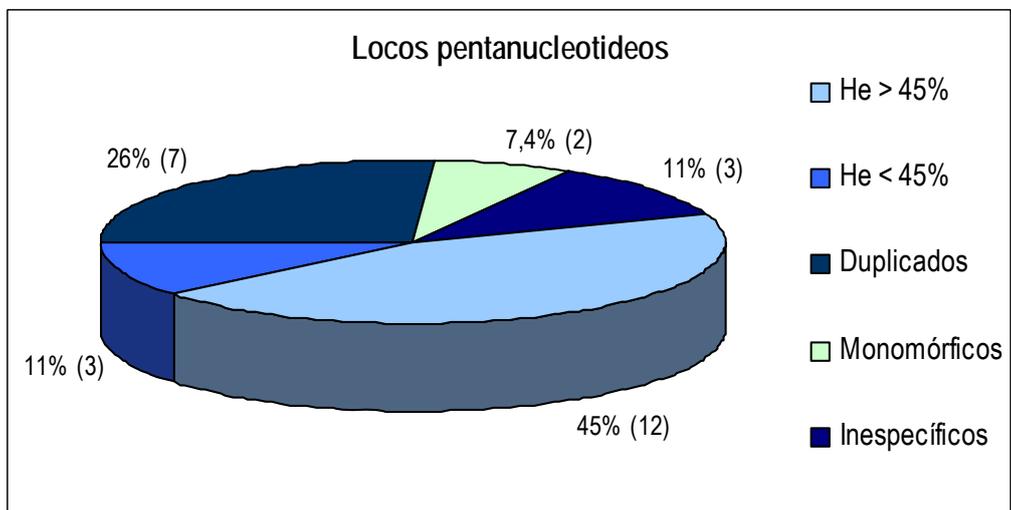
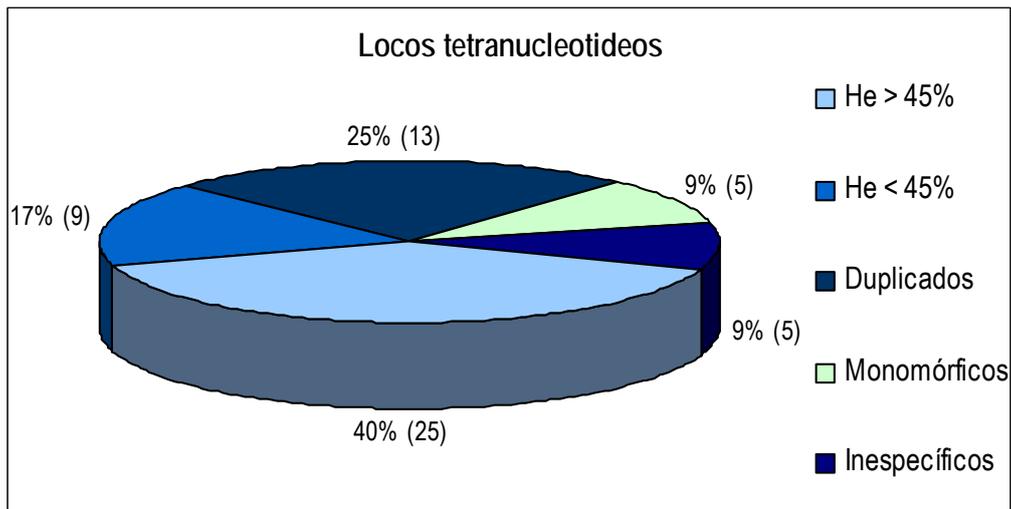


Figura 4: Distribuição de frequência das diferentes categorias de locos tetra (acima) e pentanucleotídeos (embaixo) após a segunda etapa de triagem.

Um número considerável de microssatélites apresentou a amplificação de locos duplicados. No entanto, vários destes locos apresentaram robustez de amplificação, transferibilidade interespecífica e polimorfismo. Estas características permitiriam incluí-los no grupo de locos selecionados para avaliação posterior com detecção fluorescente. Embora vários locos duplicados apresentaram alelos em faixas alélicas próximas, foi possível observar dois locos tetranucleotídeos e três pentanucleotídeos que apresentaram faixas alélicas de tamanho distinto na etapa de PAGE o que permitiu diferenciar os alelos a ambos os locos.

Já nesta etapa preliminar de triagem com apenas 24 indivíduos analisados, diferenças no número total de alelos foram observadas entre os locos tetra e pentanucleotídeos. A proporção de locos com 4 ou 5 alelos na amostra de indivíduos foi maior para os tetranucleotídeos, enquanto que maior proporção de pentanucleotídeos apresentou 3 alelos. Locos com 6 e 7 alelos só foram observados em pentanucleotídeos e apenas um loco tetranucleotídeo com 9 alelos (Figura 5).

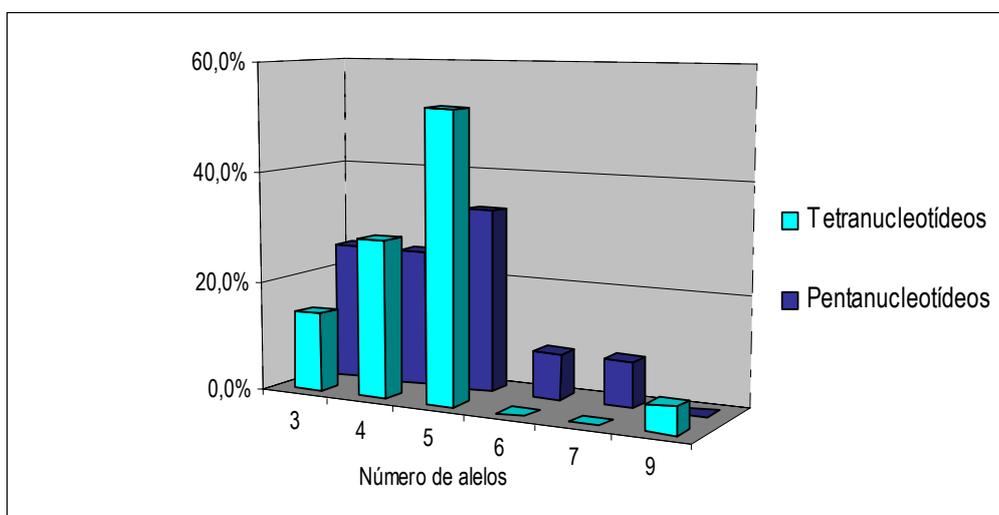


Figura 5: Distribuição de frequência de locos por número de alelos para tetra e pentanucleotídeos

A análise da transferibilidade de amplificação entre as três espécies estudadas, de *E. grandis* (a espécie na qual os microssatélites foram descobertos) para *E. urophylla* (mesma seção e subgênero) e *E. globulus* (outra seção do mesmo subgênero) mostrou que para os tetranucleotídeos, a transferibilidade de *E. grandis* para *E. urophylla* foi de 100 % e para *E. globulus* de 91% conforme esperado pela proximidade taxonômica. Para os locos pentanucleotídeos, todos os locos que amplificaram regiões polimórficas em *E. grandis* amplificaram também nas outras duas espécies. Ainda na análise de transferibilidade, para alguns locos, embora tenham amplificado nas espécies *E. globulus* e *E. urophylla*, não houve uniformidade de desempenho entre os indivíduos avaliados. Estas diferenças podem ser observadas na Figura 6, onde é mostrada uma análise comparativa entre o número de indivíduos para os quais a amplificação foi bem sucedida dentro de cada espécie. Foram estabelecidas

quatro classes (8 indivíduos, 7-6 indivíduos, ≤ 5 indivíduos e 0 indivíduos), e a porcentagem de locos amplificados para cada espécie dentro de cada classe. Em *E. urophylla* foi observada uma maior falha de amplificação entre os indivíduos em relação a *E. globulus* e *E. grandis*, sugerindo um elevado nível de polimorfismo de seqüência nas regiões de anelamento dos *primers*.

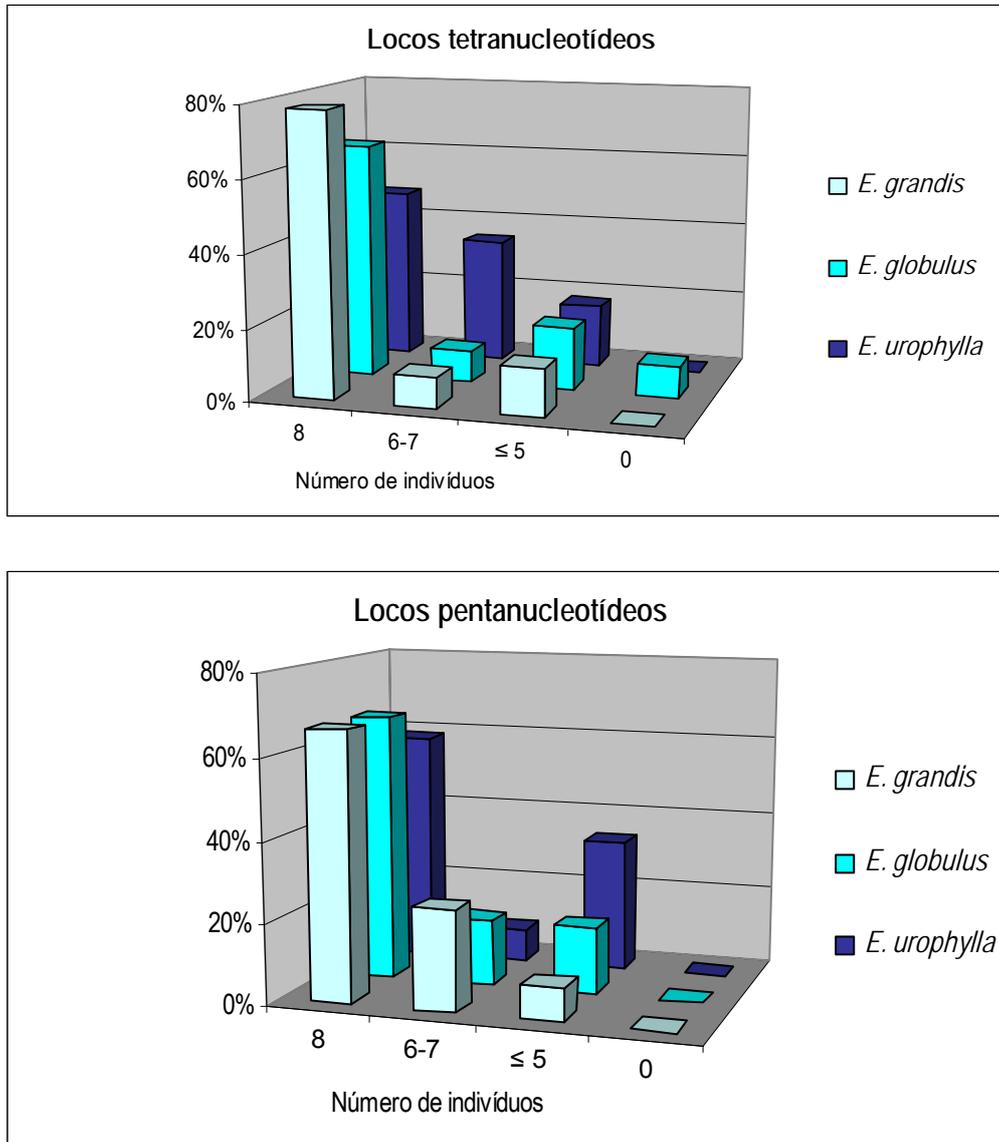


Figura 6: Frequência de microssatélites tetra e pentanucleotídeos com diferentes níveis de transferibilidade entre *E. grandis* (a espécie fonte das seqüências) e *E. globulus* e *E. urophylla* avaliados com base no número de indivíduos nos quais foi observada amplificação de produtos de PCR.

Com base em todas as etapas de seleção descritas anteriormente foi possível selecionar 45 locos microssatélites, 41 % do conjunto inicial, sendo 30 tetranucleotídeos e 15 pentanucleotídeos. Estes marcadores reuniram as características desejadas para a continuidade do desenvolvimento de uma bateria de microssatélites para utilização operacional

5.2 Caracterização genética dos microssatélites selecionados em três espécies de Eucalyptus.

Com base nos resultados da segunda etapa de triagem, os iniciadores diretos dos 45 locos selecionados foram novamente sintetizados agora com fluorocromo de três tipos distintos (FAM, HEX, NED), o que permitiu a montagem de multiplexes. A detecção fluorescente permitiu uma melhor discriminação de alelos em relação à detecção em PAGE com nitrato de prata, incluindo microvariantes, aumentando a precisão na caracterização genética destes locos microssatélites.

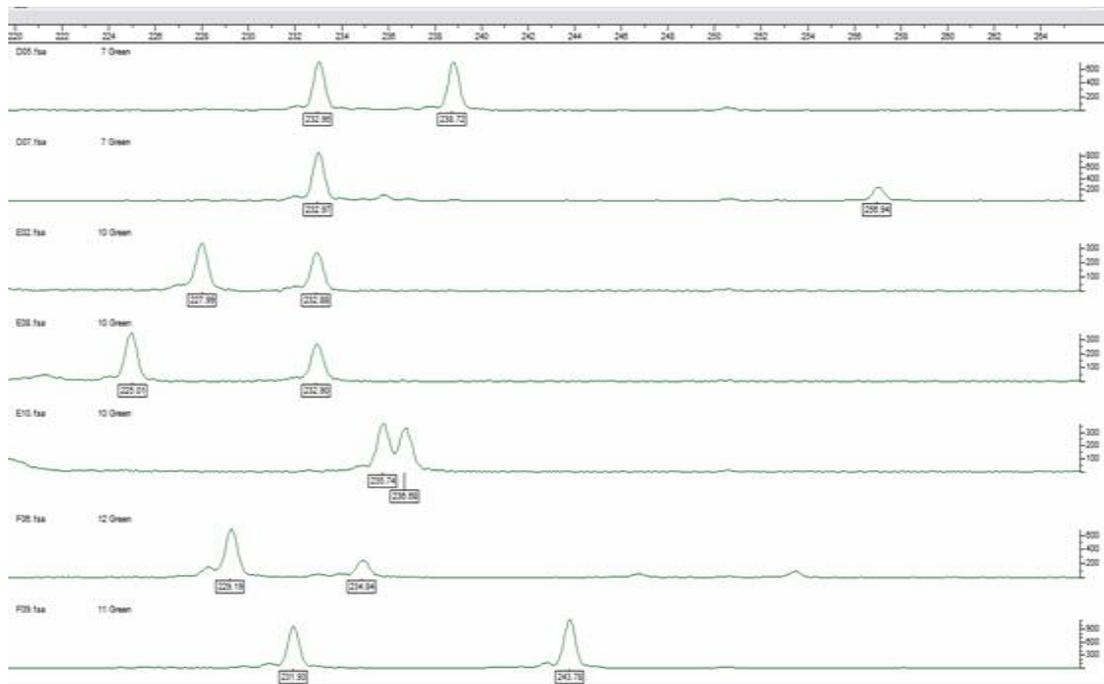
Como teste inicial para identificar o poder de multiplexagem dos microssatélites, os 45 locos foram agrupados em 15 triplexes pelo programa Multiplexer desenvolvido no âmbito do projeto Genolyptus (A. Coelho não publicado). Este programa utiliza para a montagem dos multiplexes, os seguintes critérios: tipo de fluorocromo, temperatura de anelamento (TM), autopareamento (*loops*) do *primer* e a possível formação de dímeros entre eles. Dos 15 multiplexes testados, em 9 deles todos os locos amplificaram com sucesso e em 6 houve falha de amplificação em 1 ou 2 locos. Os locos que não amplificaram nestes sistemas foram amplificados individualmente com sucesso.

Para a caracterização foi realizada uma análise intra e interespecífica em amostras de populações naturais de *E. grandis*, *E. globulus* e *E. urophylla*. Dos 30 locos tetranucleotídeos, 8 apresentaram duplicações, enquanto que dos 15 pentanucleotídeos, foram 5 os que apresentaram esta característica. Os locos duplicados foram eliminados do procedimento de caracterização. Finalmente, 22 locos

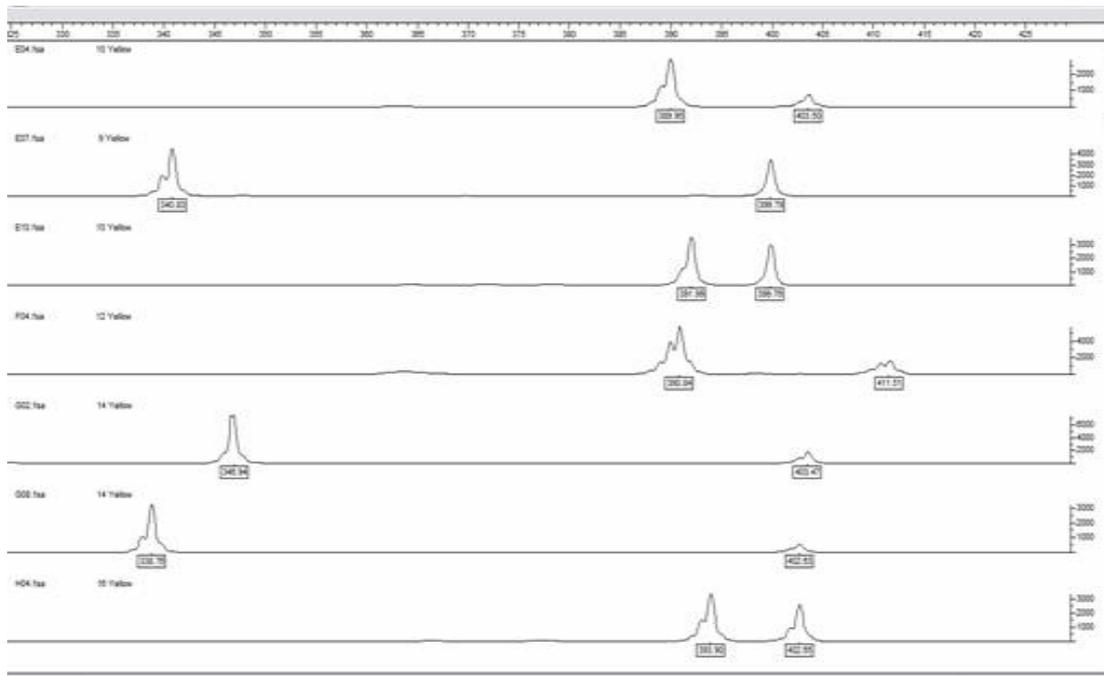
tetranucleotídeos e 10 locos pentanucleotídeos foram caracterizados geneticamente nas três espécies. Nas tabelas 1 e 2 são apresentados os resultados da caracterização genética dos locos tetra e pentanucleotídeos por espécie, e consolidados para o gênero, com estimativas médias e variâncias para o poder de exclusão de paternidade (PE) e probabilidade de identidade (PI). O número médio de alelos entre as três espécies foi similar nos locos tetra e pentanucleotídeos com um valor de 4,7 e 4,4 respectivamente. Cabe destacar que nos locos tetranucleotídeos foram observados os maiores valores médios entre as espécies estudadas nos locos EMBRA 3769 (8,33), EMBRA 2059 (7,66) e EMBRA 2052 (7,33), e nos pentanucleotídeos os maiores valores foram observados nos locos EMBRA 3275 e 2343 (5,66). Em tetranucleotídeos as heterozigosidades médias observadas (H_o) e esperadas (H_e) foram de 0,30 e 0,50, respectivamente, já nos pentanucleotídeos, estes valores foram de 0,34 e 0,46. Entretanto, foram identificados alguns marcadores tetranucleotídeos interessantes que apresentaram uma $H_o > 0,45$ e $H_e > 0,55$ (EMBRA 2052, EMBRA 2086, EMBRA 3595, EMBRA 3769, EMBRA 2764, EMBRA 2339, EMBRA 2059 e EMBRA 3034). Também nos pentanucleotídeos foram registradas $H_o > 0,45$ e $H_e > 0,5$ (EMBRA 3275, EMBRA 3894, EMBRA 2134 e EMBRA 2343). Na figura 7 são apresentados os eletroferogramas de alguns locos tetra e pentanucleotídeos. Os locos possuem perfis de alta qualidade com a ocorrência rara de gaguejo (*stuttering*) ou presença de artefatos +A-A. Os locos EMBRA 2241 e EMBRA 3645 são exemplos de locos triplicados, mas que apresentam perfis de alta qualidade, passíveis de utilização para identificação individual de clones sem necessidade de uma determinação de relação alélica entre os alelos observados. A maioria dos locos, tanto tetra, como pentanucleotídeos apresentaram pelo menos um alelo privativo por espécie (Tabelas 3 e 4), embora de forma geral os alelos privativos apareçam em baixa frequência, com exceção do alelo 213 pb do pentanucleotídeo EMBRA 3808 em *E. urophylla*, cuja frequência estimada foi de 0,94 (Tabela 4). Observaram-se ainda locos que revelaram uma elevada diferenciação entre espécies, segundo as estimativas do coeficiente de endocruzamento devido ao processo de deriva (F_{st}) (ex. EMBRA 2420,

EMBRA 2134, EMBRA 3808, EMBRA 3543 e EMBRA 2578), em função da distribuição diferenciada de frequências alélicas entre as três espécies.

A

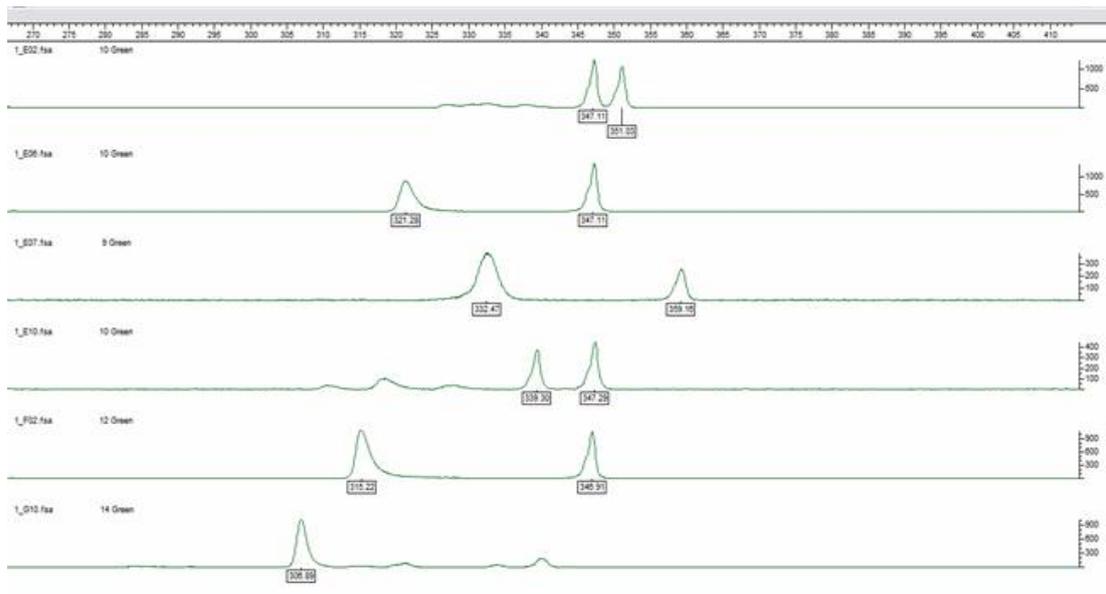


EMBRA 2052

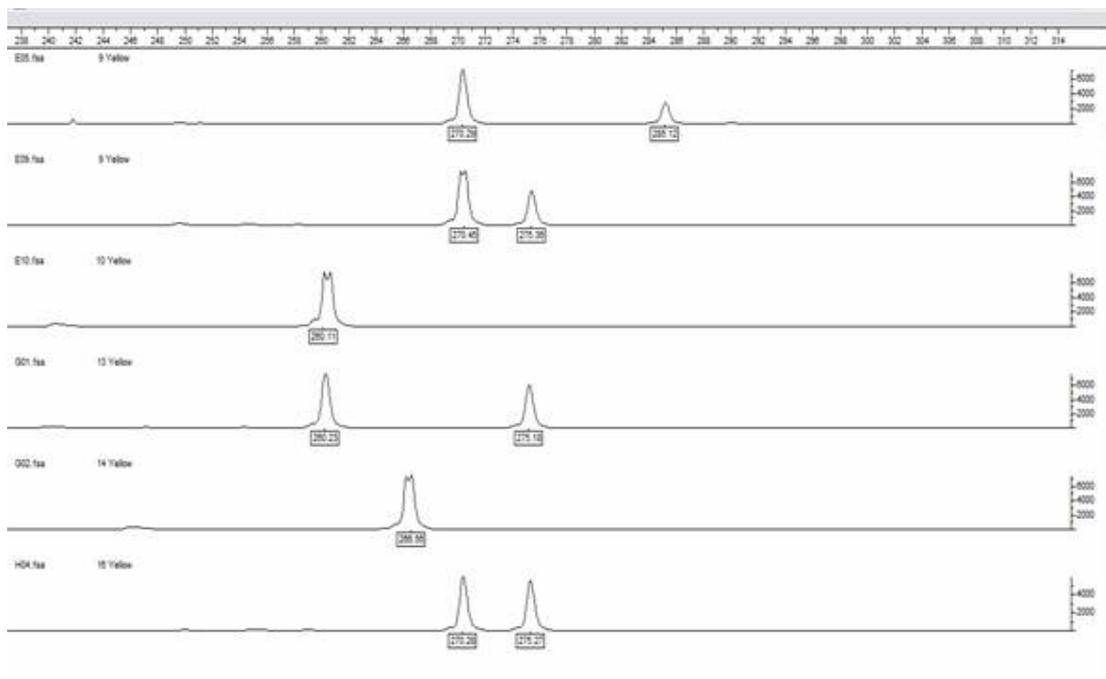


EMBRA 3769

B

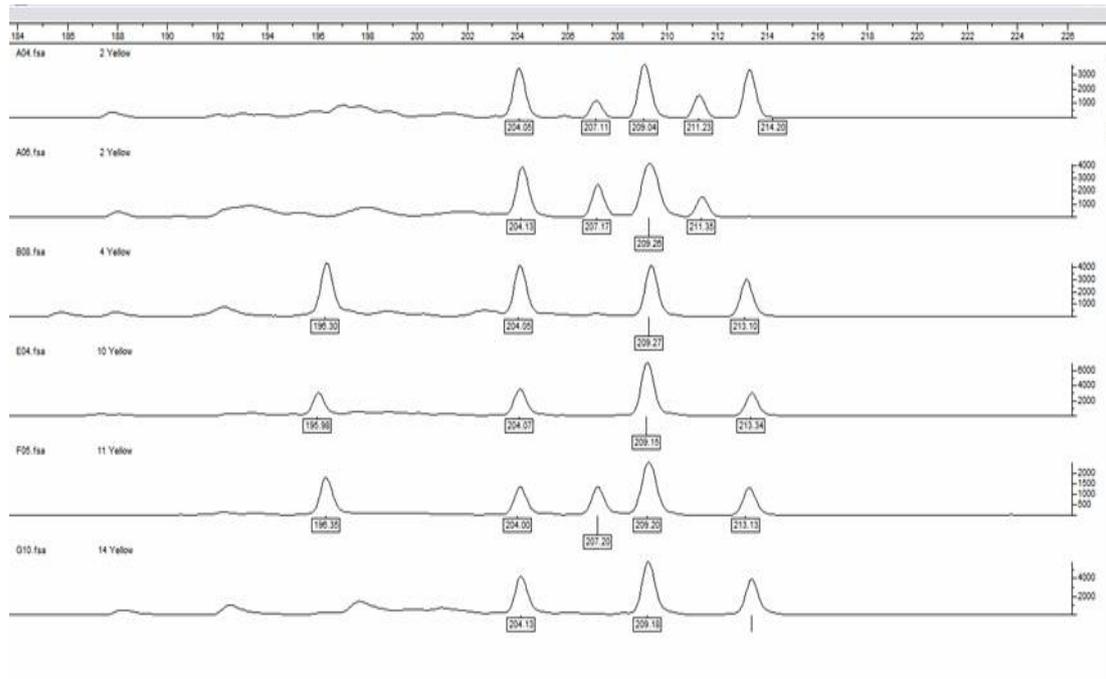


EMBRA 3275



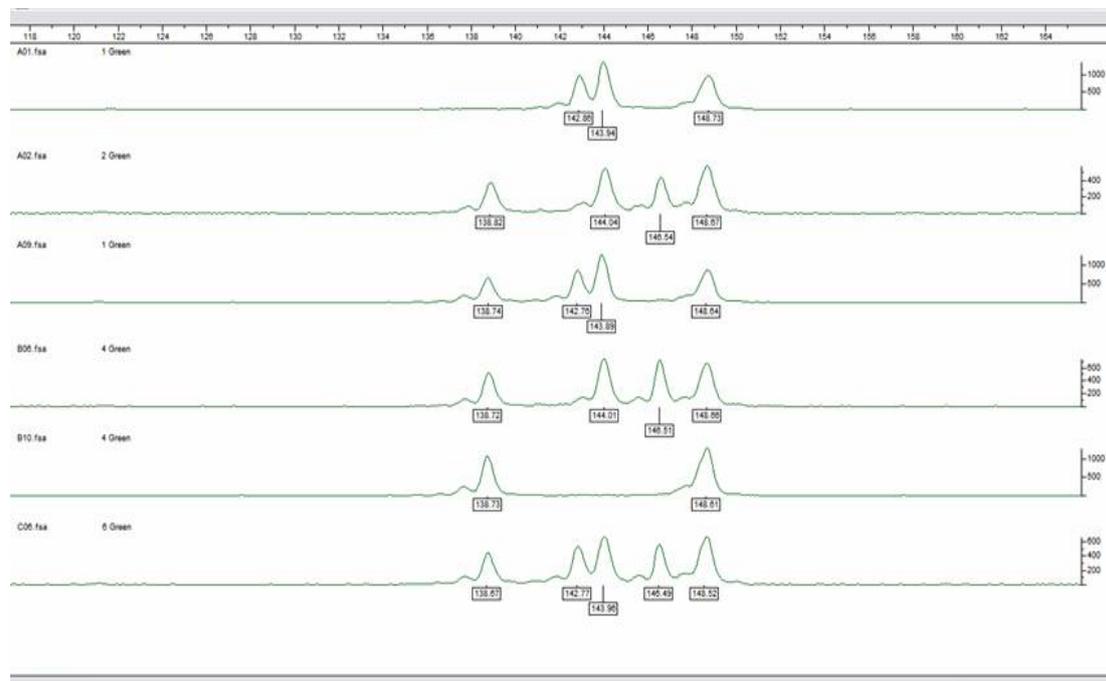
EMBRA 2321

C



EMBRA 2241

D



EMBRA 3645

Figura 7: Eletroferogramas de locos polimórficos. Grupo A: Locos tetranucleotídeos; Grupo B: Locos pentanucleotídeos; C: Loco tetranucleotídeo duplicado; D: Loco pentanucleotídeo duplicado.

Foram estimadas as probabilidade de identidade (PI) de cada loco, isto é, a probabilidade de dois indivíduos apresentarem o mesmo genótipo, com um valor médio de 0,42 nos locos tetranucleotídeos e 0,44 nos pentanucleotídeos. Observando a estimativa de conteúdo de informação polimórfica (PIC), parâmetro relevante para fins de mapeamento genético, foram estimados valores similares nas três espécies analisadas com uma média de 0,42 nos locos tetranucleotídeos e de 0,41 nos pentanucleotídeos.

Um poder de exclusão de paternidade (PE) moderado foi observado nos locos tetranucleotídeos com um valor médio das três espécies de 0,19, sendo o maior valor apresentado pelo loco EMBRA 3769 (0,34). Nos locos pentanucleotídeos foi estimado um valor de PE médio entre as espécies de 0,12, com o maior valor no loco EMBRA 2134 (0,38).

Com base em todas as estimativas teóricas de desempenho esperado para identificação individual em *Eucalyptus* com prioridade para o poder de exclusão de paternidade (PE), por ser este o mais limitante para a investigação genética, foram identificados 12 locos para a montagem de sistemas multiplex. Um aspecto importante também levado em consideração foi a variância das estimativas de PE e PI visando locos com menor valor para este parâmetro, o que indicaria um desempenho equilibrado nas três espécies. Os 12 locos também foram selecionados com base nos resultados empíricos de sucesso de co-amplificação em sistemas multiplex, avaliado preliminarmente pelo *software* Multiplexer.

Tabela 1: Caracterização genética de microssatélites tetranucleotídeos. N: Número de indivíduos analisados; Na: Número de alelos; Ho: Heterozigosidade observada; He: heterozigosidade esperada; PIC: Conteúdo de informação polimórfica; PI: probabilidade de identidade; PE: Poder de exclusão; Médias das espécies; Variância de PI e PE; Coeficiente de endocruzamento devido ao processo de diferenciação entre espécies (Fst).

Loco	<i>E.grandis</i>							<i>E.globulus</i>							<i>E.urophylla</i>							Média das espécies						Variância		Fst
	N	Na	Ho	He	P.I.C.	P.I.	P.E.	N	Na	Ho	He	P.I.C.	P.I.	P.E.	N	Na	Ho	He	P.I.C.	P.I.	P.E.	Na	Ho	He	P.I.C.	P.I.	P.E.	PI	PE	
Embra 2113	45	5	0,56	0,50	0,47	0,29	0,24	47	5	0,30	0,42	0,39	0,41	0,06	34	5	0,18	0,31	0,30	0,55	0,02	5,00	0,34	0,41	0,38	0,42	0,11	0,018	0,013	0,02
Embra 2675	45	3	0,73	0,51	0,40	0,52	0,48	43	4	0,49	0,57	0,50	0,27	0,18	28	4	0,07	0,60	0,53	0,35	0,00	3,66	0,43	0,56	0,47	0,38	0,22	0,016	0,058	0,03
Embra 3513	24	3	0,08	0,66	0,58	0,30	0,01	27	4	0,19	0,66	0,61	0,24	0,03	31	3	0,06	0,18	0,17	0,76	0,00	3,33	0,11	0,50	0,45	0,44	0,01	0,082	0,000	0,23
Embra 3086	40	3	0,30	0,51	0,44	0,36	0,06	45	4	0,56	0,56	0,46	0,35	0,24	36	6	0,33	0,68	0,64	0,22	0,08	4,33	0,40	0,58	0,51	0,31	0,13	0,006	0,010	0,07
Embra 2549	42	4	0,36	0,31	0,29	0,48	0,09	46	3	0,37	0,33	0,28	0,49	0,10	48	1	0,00	0,00	0,00	1,00	0,00	2,66	0,24	0,21	0,19	0,66	0,06	0,088	0,003	0,07
Embra 2052	46	9	0,67	0,73	0,69	0,14	0,39	47	7	0,55	0,61	0,54	0,20	0,24	37	6	0,32	0,42	0,39	0,42	0,07	7,33	0,52	0,59	0,54	0,25	0,23	0,022	0,025	0,09
Embra 2086	48	4	0,54	0,61	0,55	0,21	0,23	48	5	0,31	0,65	0,59	0,21	0,07	46	4	0,50	0,45	0,38	0,38	0,19	4,33	0,45	0,57	0,5	0,27	0,16	0,010	0,007	0,13
Embra3490	47	6	0,30	0,50	0,48	0,36	0,06	46	7	0,17	0,37	0,36	0,53	0,02	30	8	0,17	0,69	0,67	0,25	0,02	7,00	0,21	0,52	0,5	0,38	0,04	0,020	0,001	0,04
Embra 3595	48	5	0,73	0,64	0,57	0,25	0,47	47	6	0,43	0,57	0,52	0,26	0,13	44	6	0,45	0,55	0,52	0,27	0,15	5,66	0,54	0,59	0,53	0,26	0,25	0,000	0,037	0,12
Embra3196	43	2	0,02	0,02	0,02	0,95	0,00	32	2	0,13	0,17	0,16	0,73	0,01	34	5	0,24	0,22	0,21	0,60	0,04	3,00	0,13	0,14	0,13	0,76	0,02	0,032	0,000	0,04
Embra 3435	48	1	0,00	0,00	0,00	1,00	0,00	47	2	0,02	0,02	0,02	0,96	0,00	45	2	0,04	0,04	0,04	0,92	0,00	1,66	0,02	0,02	0,02	0,96	0,00	0,002	0,000	0,01
Embra 3769	42	8	0,67	0,81	0,78	0,10	0,38	42	12	0,74	0,79	0,76	0,09	0,49	40	5	0,45	0,66	0,61	0,19	0,15	8,33	0,62	0,75	0,72	0,13	0,34	0,003	0,030	0,08
Embra 2764	46	3	0,59	0,45	0,37	0,47	0,28	46	4	0,72	0,54	0,48	0,37	0,46	48	4	0,15	0,63	0,57	0,31	0,02	3,66	0,48	0,54	0,47	0,38	0,25	0,007	0,049	0,20
Embra 2339	47	4	0,62	0,52	0,42	0,38	0,31	43	7	0,49	0,50	0,47	0,30	0,18	43	6	0,58	0,66	0,61	0,16	0,27	5,66	0,56	0,56	0,5	0,28	0,25	0,012	0,005	0,14
Embra 2965	47	5	0,30	0,37	0,34	0,46	0,06	44	4	0,11	0,39	0,36	0,58	0,01	47	5	0,26	0,66	0,60	0,22	0,05	4,66	0,22	0,47	0,43	0,42	0,04	0,034	0,001	0,19
Embra 2081	47	3	0,13	0,12	0,12	0,77	0,01	44	7	0,48	0,53	0,50	0,32	0,17	44	6	0,20	0,69	0,66	0,22	0,03	5,33	0,27	0,45	0,42	0,44	0,07	0,088	0,007	0,09
Embra 3185	45	3	0,29	0,33	0,29	0,50	0,06	44	3	0,16	0,36	0,33	0,54	0,02	43	4	0,40	0,61	0,56	0,21	0,11	3,33	0,28	0,44	0,39	0,42	0,06	0,032	0,002	0,21
Embra 2527	39	9	0,44	0,59	0,56	0,25	0,14	32	5	0,19	0,50	0,42	0,39	0,03	40	7	0,33	0,35	0,34	0,44	0,07	7,00	0,32	0,48	0,43	0,36	0,08	0,010	0,003	0,20
Embra 2059	47	6	0,57	0,68	0,63	0,15	0,26	48	9	0,48	0,60	0,57	0,23	0,17	42	8	0,55	0,74	0,70	0,12	0,23	7,66	0,53	0,68	0,63	0,17	0,22	0,003	0,002	0,12
Embra 3034	48	4	0,33	0,40	0,36	0,41	0,08	46	6	0,43	0,51	0,47	0,28	0,14	31	11	0,71	0,82	0,80	0,08	0,44	7,00	0,49	0,58	0,54	0,26	0,22	0,028	0,038	0,09
Embra 3543	47	1	0,00	0,00	0,00	1,00	0,00	41	4	0,39	0,52	0,42	0,32	0,11	27	5	0,30	0,35	0,33	0,47	0,06	3,33	0,23	0,29	0,25	0,60	0,06	0,128	0,003	0,47
Embra 2578	40	3	0,10	0,36	0,33	0,58	0,01	36	2	0,06	0,28	0,24	0,67	0,00	18	4	0,11	0,29	0,28	0,62	0,01	3,00	0,09	0,31	0,28	0,62	0,01	0,002	0,000	0,41
Média	44	4	0,38	0,44	0,40	0,45	0,16	42,8	5,09	0,35	0,48	0,43	0,40	0,13	38,0	5,23	0,29	0,48	0,45	0,40	0,09	4,92	0,34	0,46	0,42	0,42	0,13	0,001	0,001	0,14

Tabela 2: Caracterização genética de microssatélites pentanucleotídeos. N: Número de indivíduos analisados; Na: Número de alelos; Ho: Heterozigidade observada; He: heterozigidade esperada; PIC: Conteúdo de informação polimórfica; PI: probabilidade de identidade; PE: Poder de exclusão; Médias das espécies; Variância de PI e PE; Coeficiente de endocruzamento devido ao processo de diferenciação entre espécies (Fst).

Loco	<i>E.grandis</i>							<i>E.globulus</i>							<i>E.urophylla</i>							Média das espécies						Variância		Fst
	N	Na	Ho	He	P.I.C.	P.I.	P.E.	N	Na	Ho	He	P.I.C.	P.I.	P.E.	N	Na	Ho	He	P.I.C.	P.I.	P.E.	Na	Ho	He	P.I.C.	P.I.	P.E.	PI	PE	
Embra 3275	41	6	0,54	0,53	0,50	0,26	0,22	45	4	0,62	0,55	0,50	0,26	0,32	44	7	0,27	0,42	0,39	0,44	0,05	5,66	0,48	0,50	0,46	0,32	0,20	0,010	0,018	0,02
Embra 3894	47	4	0,43	0,66	0,60	0,19	0,13	48	4	0,48	0,59	0,52	0,24	0,17	43	6	0,44	0,61	0,53	0,23	0,14	4,66	0,45	0,62	0,55	0,22	0,15	0,001	0,000	0,14
Embra 2420	47	3	0,15	0,28	0,26	0,60	0,02	47	3	0,21	0,41	0,37	0,49	0,03	28	5	0,14	0,68	0,62	0,29	0,02	3,66	0,17	0,46	0,42	0,46	0,02	0,025	0,000	0,29
Embra 3297	48	4	0,60	0,67	0,61	0,17	0,30	48	3	0,10	0,32	0,28	0,59	0,01	45	5	0,51	0,54	0,48	0,28	0,20	4,00	0,41	0,51	0,46	0,35	0,17	0,047	0,021	0,15
Embra 2134	48	3	1,00	0,55	0,46	0,78	1,00	48	4	0,25	0,38	0,36	0,56	0,04	47	6	0,34	0,64	0,60	0,32	0,08	4,33	0,53	0,53	0,47	0,56	0,38	0,052	0,293	0,21
Embra 2321	40	6	0,13	0,55	0,52	0,39	0,01	45	5	0,24	0,49	0,46	0,36	0,04	38	3	0,03	0,17	0,16	0,80	0,00	4,66	0,13	0,40	0,38	0,52	0,02	0,062	0,000	0,05
Embra 2570	47	4	0,21	0,43	0,38	0,42	0,03	45	5	0,31	0,35	0,33	0,45	0,07	33	4	0,03	0,25	0,24	0,73	0,00	4,33	0,18	0,34	0,31	0,53	0,03	0,028	0,001	0,03
Embra 2587	43	4	0,21	0,27	0,26	0,58	0,03	34	5	0,21	0,26	0,25	0,61	0,03	31	5	0,10	0,21	0,20	0,71	0,01	4,66	0,17	0,25	0,23	0,63	0,02	0,005	0,000	0,01
Embra 3808	46	2	0,17	0,31	0,27	0,56	0,02	42	5	0,19	0,33	0,30	0,54	0,03	9	2	0,11	0,10	0,10	0,80	0,01	3,00	0,16	0,25	0,23	0,63	0,02	0,021	0,000	0,59
Embra 2343	43	6	0,53	0,72	0,68	0,17	0,22	31	4	0,23	0,57	0,48	0,30	0,04	31	7	0,55	0,65	0,62	0,17	0,23	5,66	0,44	0,65	0,59	0,21	0,16	0,005	0,012	0,12
<i>Média</i>	<i>45,0</i>	<i>4,20</i>	<i>0,40</i>	<i>0,50</i>	<i>0,45</i>	<i>0,41</i>	<i>0,20</i>	<i>43,3</i>	<i>4,20</i>	<i>0,28</i>	<i>0,43</i>	<i>0,39</i>	<i>0,44</i>	<i>0,08</i>	<i>34,9</i>	<i>5,00</i>	<i>0,25</i>	<i>0,43</i>	<i>0,39</i>	<i>0,48</i>	<i>0,07</i>	<i>4,46</i>	<i>0,31</i>	<i>0,45</i>	<i>0,41</i>	<i>0,44</i>	<i>0,12</i>	<i>0,001</i>	<i>0,005</i>	<i>0,16</i>

Tabela 3: Estimativas de freqüências alélicas de alelos privativos aos locos tetranucleotídeos nas espécies *E. grandis*, *E. globulus* e *E. urophylla*.

		<i>E. grandis</i>	<i>E. globulus</i>	<i>E. urophylla</i>
Embra 2113	alelo (pb)	168		144
	freqüência	0,03		0,01
Embra 2675	alelo (pb)		359	411
	freqüência		0,03	0,04
Embra 3513	alelo (pb)		293	301
	freqüência		0,07	0,08
Embra 3086	alelo (pb)			251 263
	freqüência			0,04 0,07
Embra 2549	alelo (pb)	196		
	freqüência	0,01		
Embra 2052	alelo (pb)	220 248 256	224	
	freqüência	0,01 0,05 0,01	0,03	
Embra 3490	alelo (pb)	398 410 414	402 418	374 406 430
	freqüência	0,03 0,07 0,10	0,01 0,02	0,12 0,03 0,03
Embra 3196	alelo (pb)		365	341 357 361
	freqüência		0,09	0,01 0,06 0,03
Embra 3769	alelo (pb)	416	336 340 344 348 396	
	freqüência	0,05	0,01 0,04 0,02 0,05 0,06	
Embra 2764	alelo (pb)		80	84
	freqüência		0,10	0,31
Embra 2339	alelo (pb)		218	206
	freqüência		0,01	0,05
Embra 2965	alelo (pb)	230	226	234
	freqüência	0,02	0,02	0,01
Embra 2081	alelo (pb)		97	
	freqüência		0,01	
Embra 3185	alelo (pb)			268
	freqüência			0,05
Embra 2527	alelo (pb)	285		317
	freqüência	0,01		0,04
Embra 2059	alelo (pb)		203 231	191
	freqüência		0,02 0,01	0,02
Embra 3034	alelo (pb)			307 315 319 327 359
	freqüência			0,03 0,03 0,06 0,03 0,02
Embra 3543	alelo (pb)			215
	freqüência			0,02
Embra 2578	alelo (pb)			300
	freqüência			0,06

Tabela 4: Estimativas de frequências alélicas de alelos privativos aos locos pentanucleotídeos nas espécies *E. grandis*, *E. globulus* e *E. urophylla*.

		<i>E. grandis</i>			<i>E. globulus</i>			<i>E. urophylla</i>		
Embra 3275	Alelo (pb)	224						214	244	
	Frequência	0,11						0,01	0,01	
Embra 3894	Alelo (pb)							197	217	227
	Frequência							0,02	0,02	0,02
Embra 2420	Alelo (pb)							89	264	
	Frequência							0,05	0,04	
Embra 3297	Alelo (pb)							186		
	Frequência							0,01		
Embra 2134	Alelo (pb)							227	232	
	Frequência							0,04	0,01	
Embra 2321	Alelo (pb)	255	280	295	285					
	Frequência	0,11	0,1	0,05	0,01					
Embra 2570	Alelo (pb)							125		
	Frequência							0,08		
Embra 2587	Alelo (pb)				100			130	135	
	Frequência				0,01			0,02	0,03	
Embra 3808	Alelo (pb)				283	288	303	213	248	
	Frequência				0,02	0,01	0,02	0,94	0,06	
Embra 2343	Alelo (pb)	156			116			141		
	Frequência	0,02			0,05			0,03		

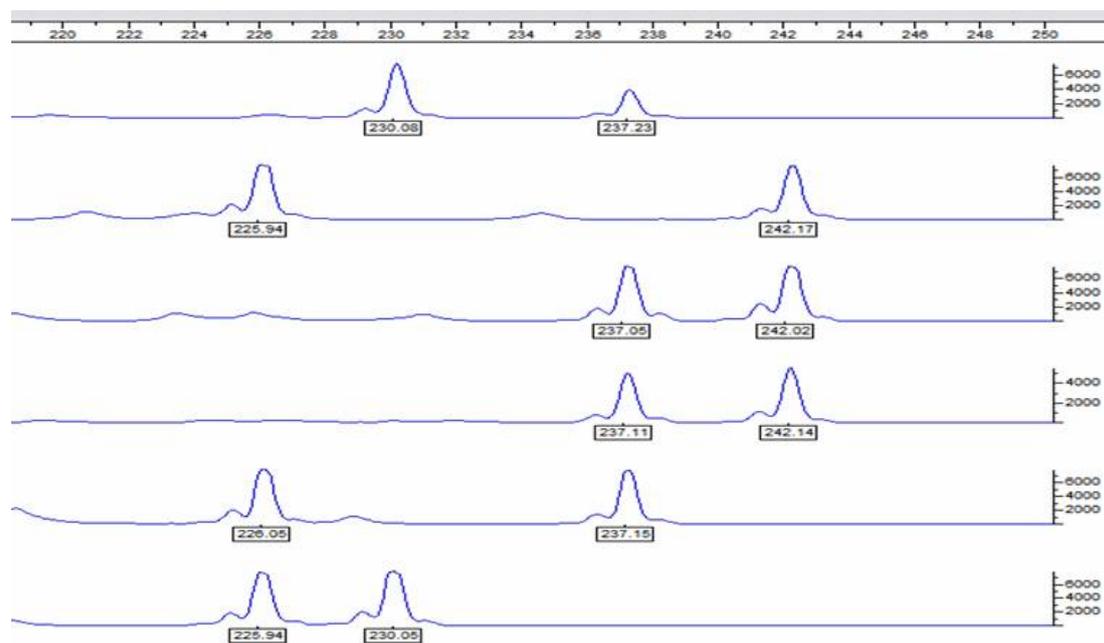
5.3 Mapeamento dos locos tetra e pentanucleotídeos

Os 45 locos selecionados foram avaliados quanto à herança e configuração de segregação visando o mapeamento. Na triagem com a família segregante IP (International Paper), 25% segregaram em configuração totalmente informativa com 3 ou 4 alelos distintos (Figura 8), 20% segregaram a partir de um dos genitores, 5% apresentaram segregação tipo F2 (1:2:1), 15% apresentaram alelo nulo, 5% foram informativos, mas apresentaram alelos microvariantes, 17,5% foram de difícil interpretação devido a duplicações ou “*stutter*” e apenas 12,5% foram monomórficos. Em vista da configuração de segregação observada na família IP, foi possível posicionar 8 locos, dos quais 7 são tetranucleotídeos e um pentanucleotídeo (Figura 9). Dos 8 locos, 4 segregaram de forma totalmente informativa e

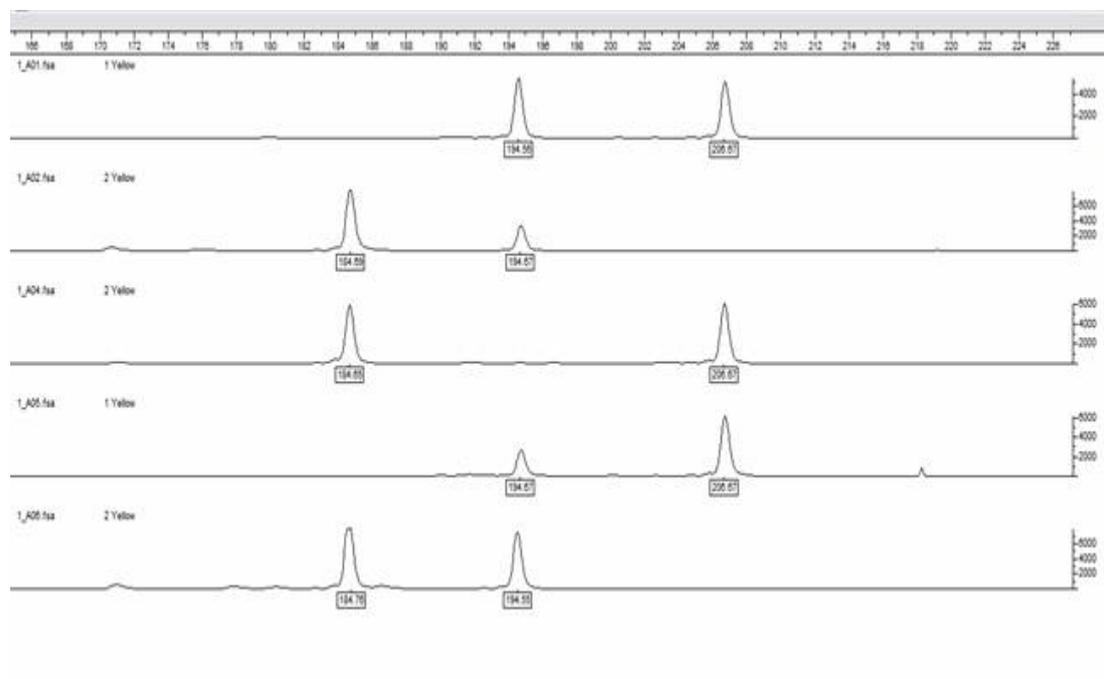
mapearam nos mapas de ambos os parentais (EMBRA 3372, EMBRA 3595, EMBRA 3543 e EMBRA 2339), dois locos se posicionaram unicamente no mapa da mãe (EMBRA 2965 e EMBRA 2587) e dois no mapa do pai (EMBRA 3185 e EMBRA 3086). Estes locos se localizaram em 5 grupos de ligação dos 11 grupos existentes para *Eucalyptus*.

Na triagem realizada com a família segregante VCP (Votorantim Celulose e Papel) 22,3% dos locos apresentaram segregação totalmente informativa, 15,6 % segregaram a partir de um dos genitores, 22,3 % apresentaram segregação tipo F2, 8,8 % apresentaram alelo nulo, 2,2 % foram informativos, mas apresentaram alelos microvariantes, 24,3 % foram de difícil interpretação e 4,5% foram monomórficos. Do total de marcadores microssatélites tetra e pentanucleotídeos foram possíveis mapear 16 locos no mapa de referência de *Eucalyptus* da família VCP (Novaes, dados não publicados) (Figura 10), dos quais 6 locos foram posicionados no mapa de referência de ambos os parentais (EMBRA 3297, EMBRA 2059, EMBRA 3769, EMBRA 3275, EMBRA 3086 e EMBRA 3894), 5 locos mapearam unicamente no mapa da mãe (EMBRA 2462, EMBRA 2134, EMBRA 2086, EMBRA 2113 e EMBRA 2560) e 5 locos exclusivamente no mapa do pai (EMBRA 3513, EMBRA 2549, EMBRA 2011, EMBRA 2527 e EMBRA 2764). Os 16 locos mapeados na família VCP se localizaram em 7 grupos de ligação dos 11 grupos de *Eucalyptus*. Somando os resultados obtidos com as duas famílias segregantes, foram mapeados 23 locos, 8 na família IP e 16 na família VCP, sendo que um loco (EMBRA 3086) foi mapeado nas duas famílias. Conforme esperado, este loco foi mapeado corretamente no mesmo grupo de ligação em ambos os mapas.

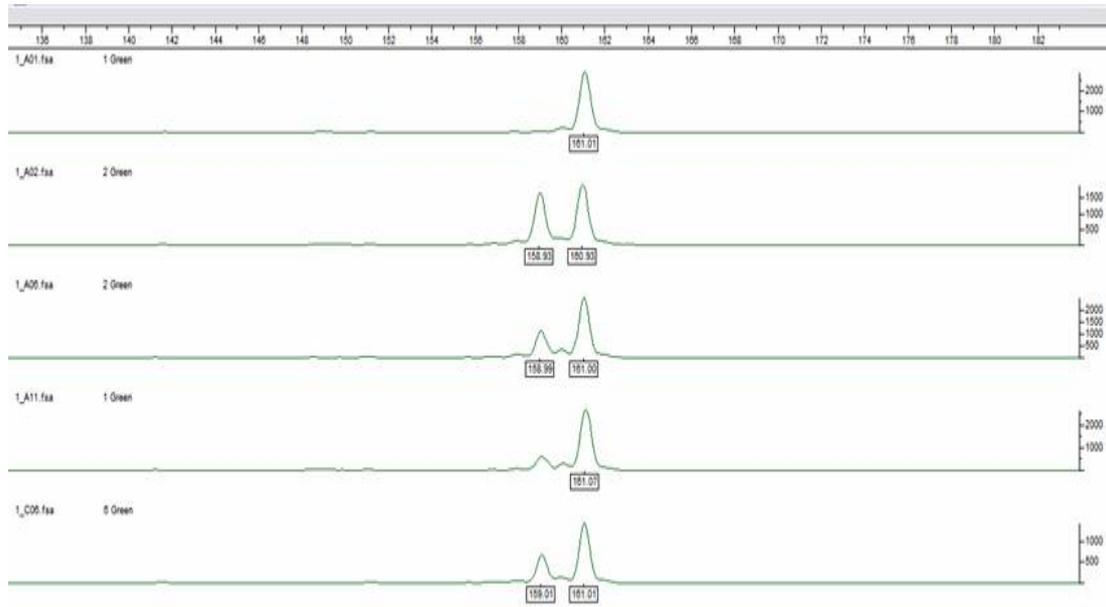
A



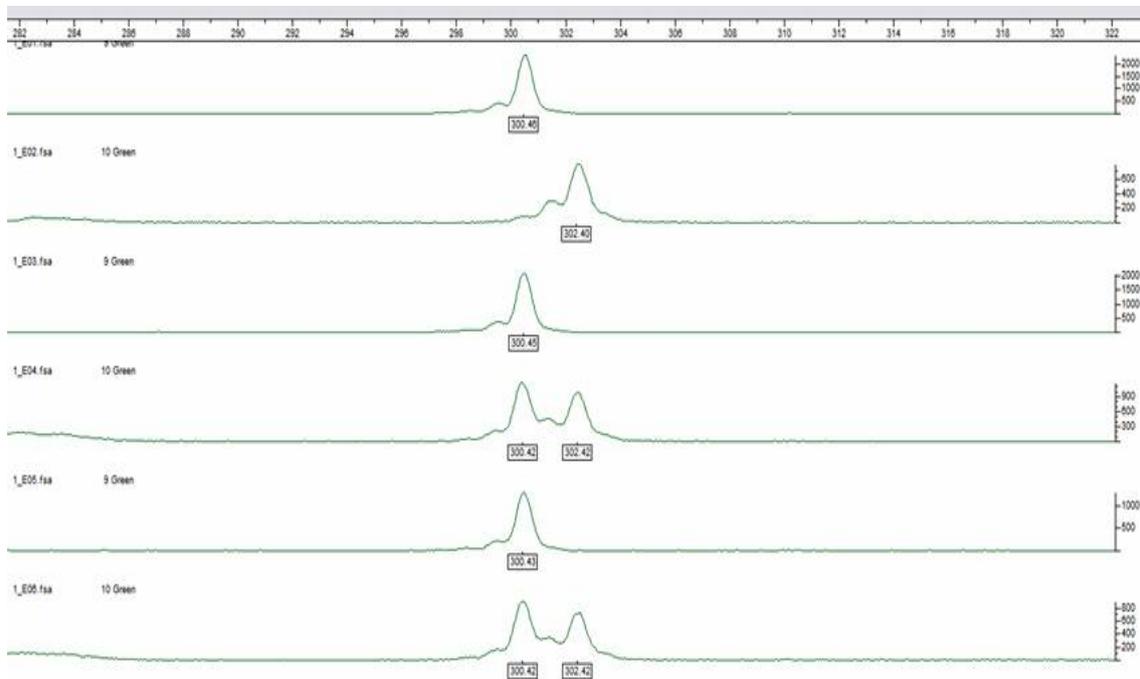
B



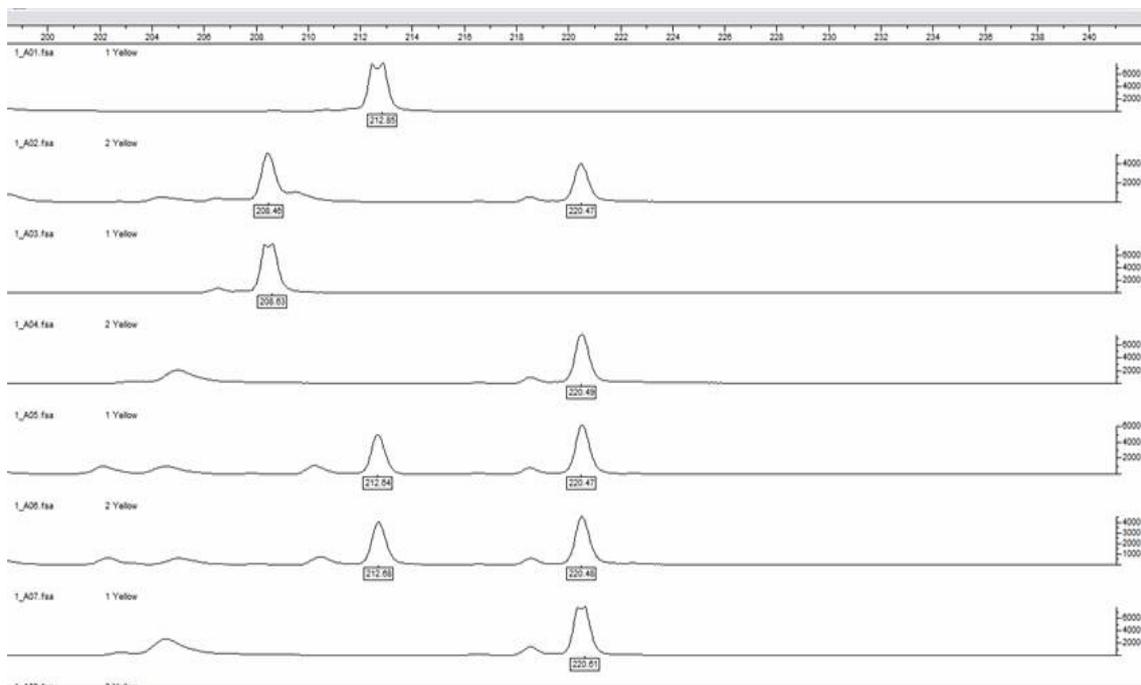
C



D



E



F

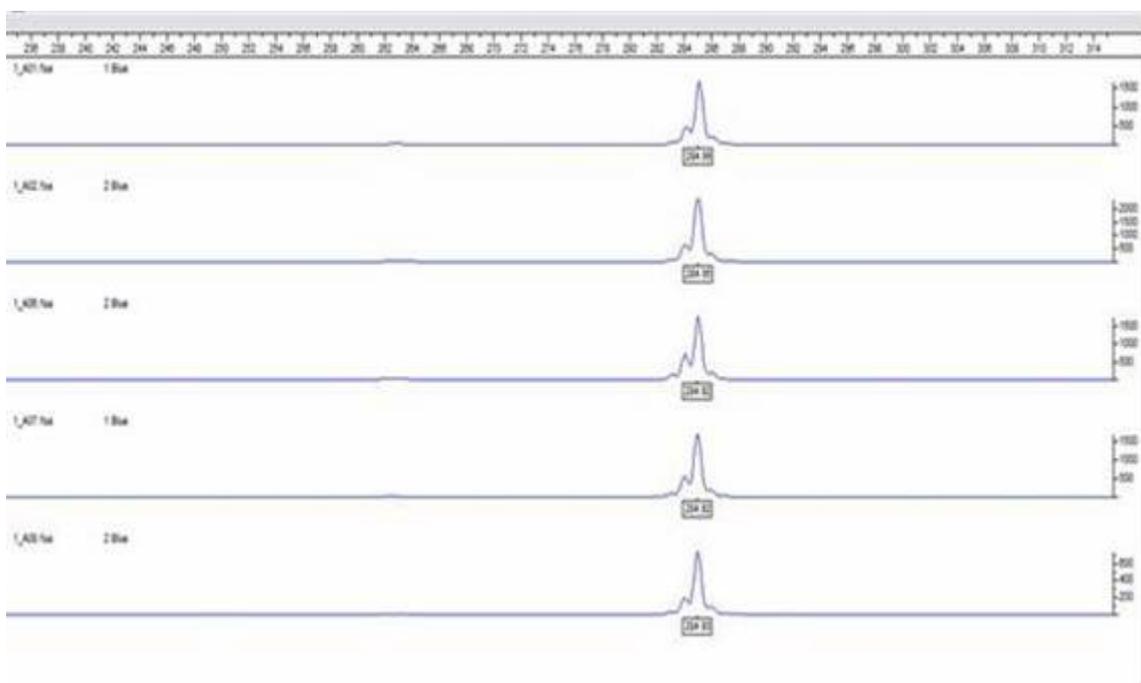
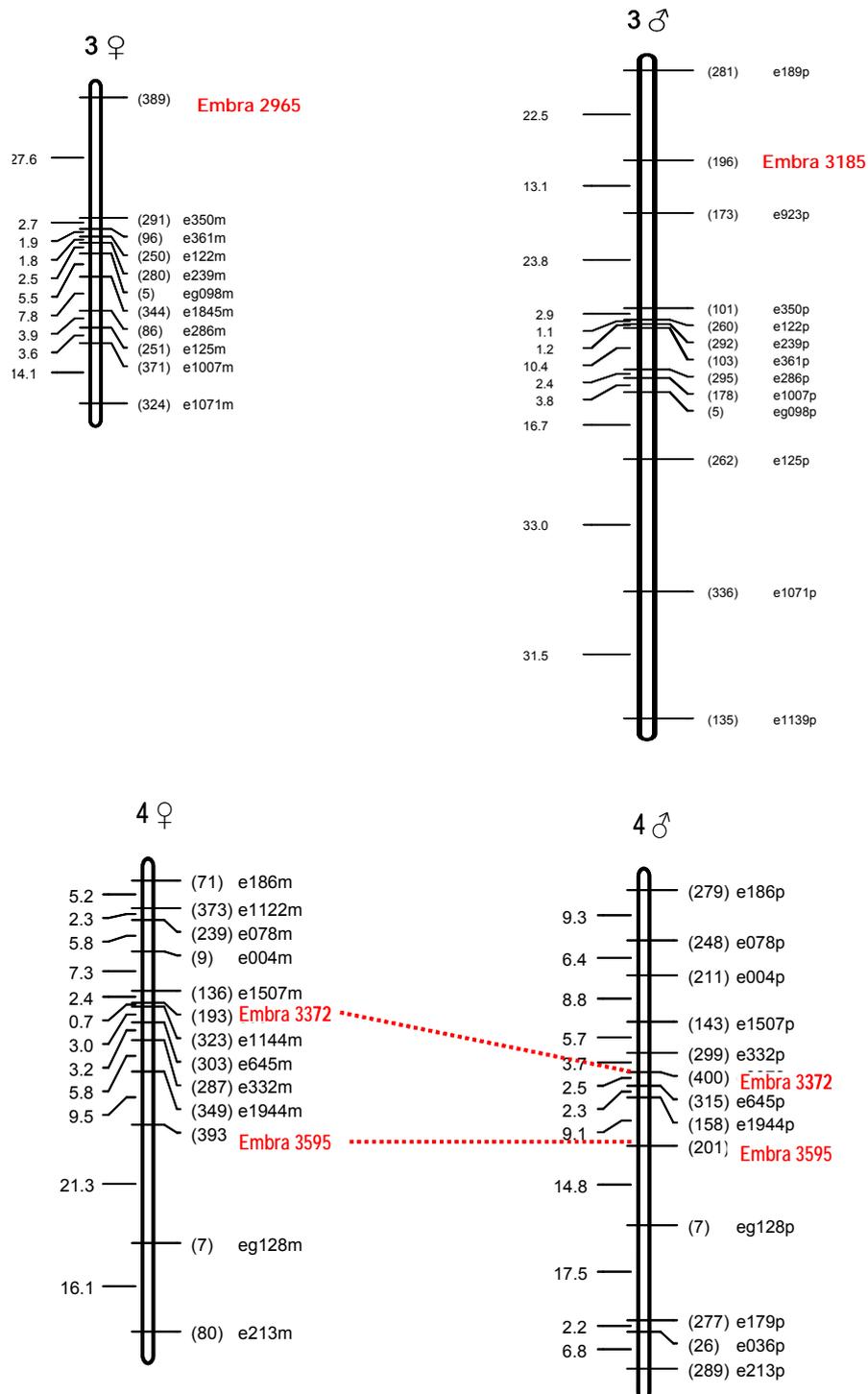


Figura 8: Configurações de segregação observadas para os locos tetra e pentanucleotídeos. Para cada loco, em ordem descendente encontram-se, os parentais (Mãe e Pai) e os filhos. A) e B) segregação totalmente informativa; C) segregação de um dos parentais; D) presença de alelos microvariantes com diferença de 2pbe alelo nulo para um parental; E) presença de alelo nulo para um parental e F) configuração não informativa.



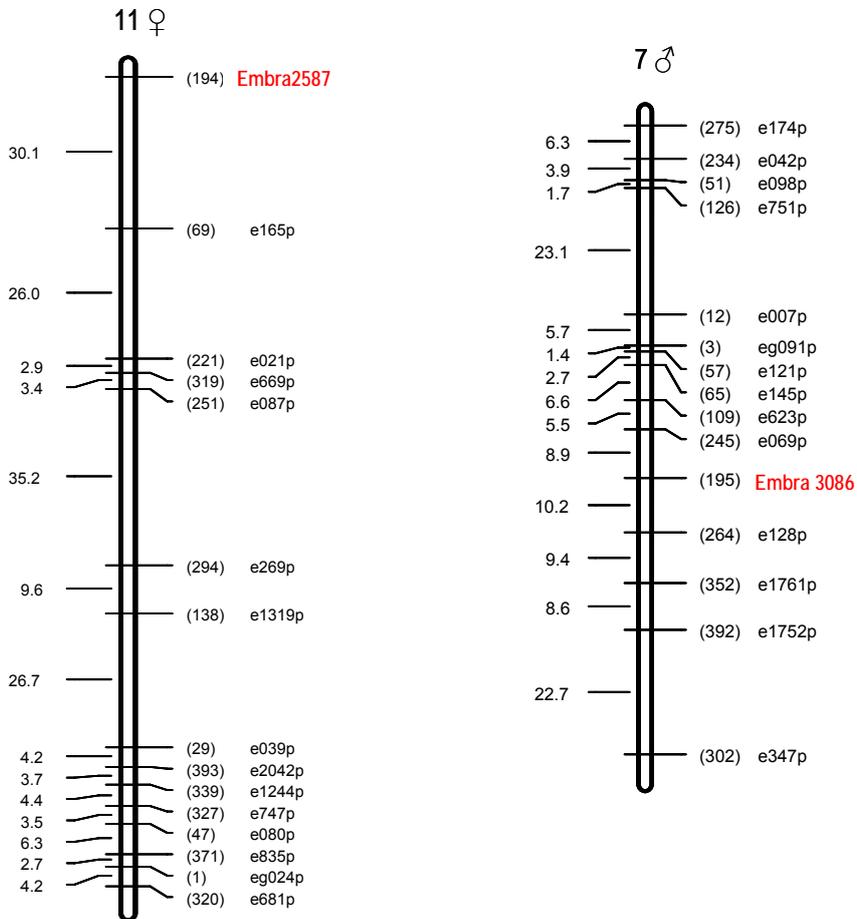
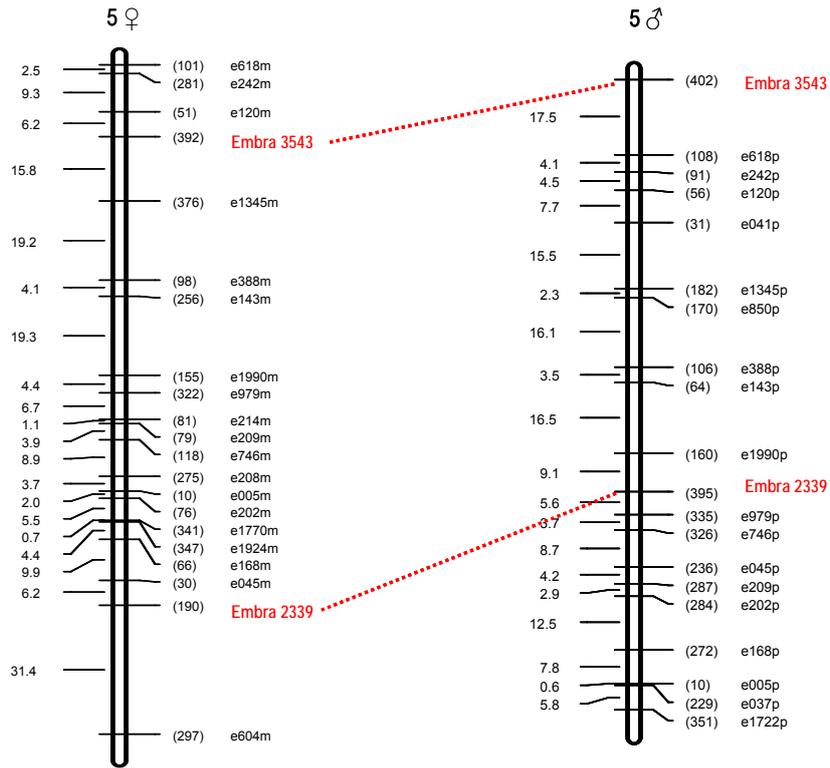
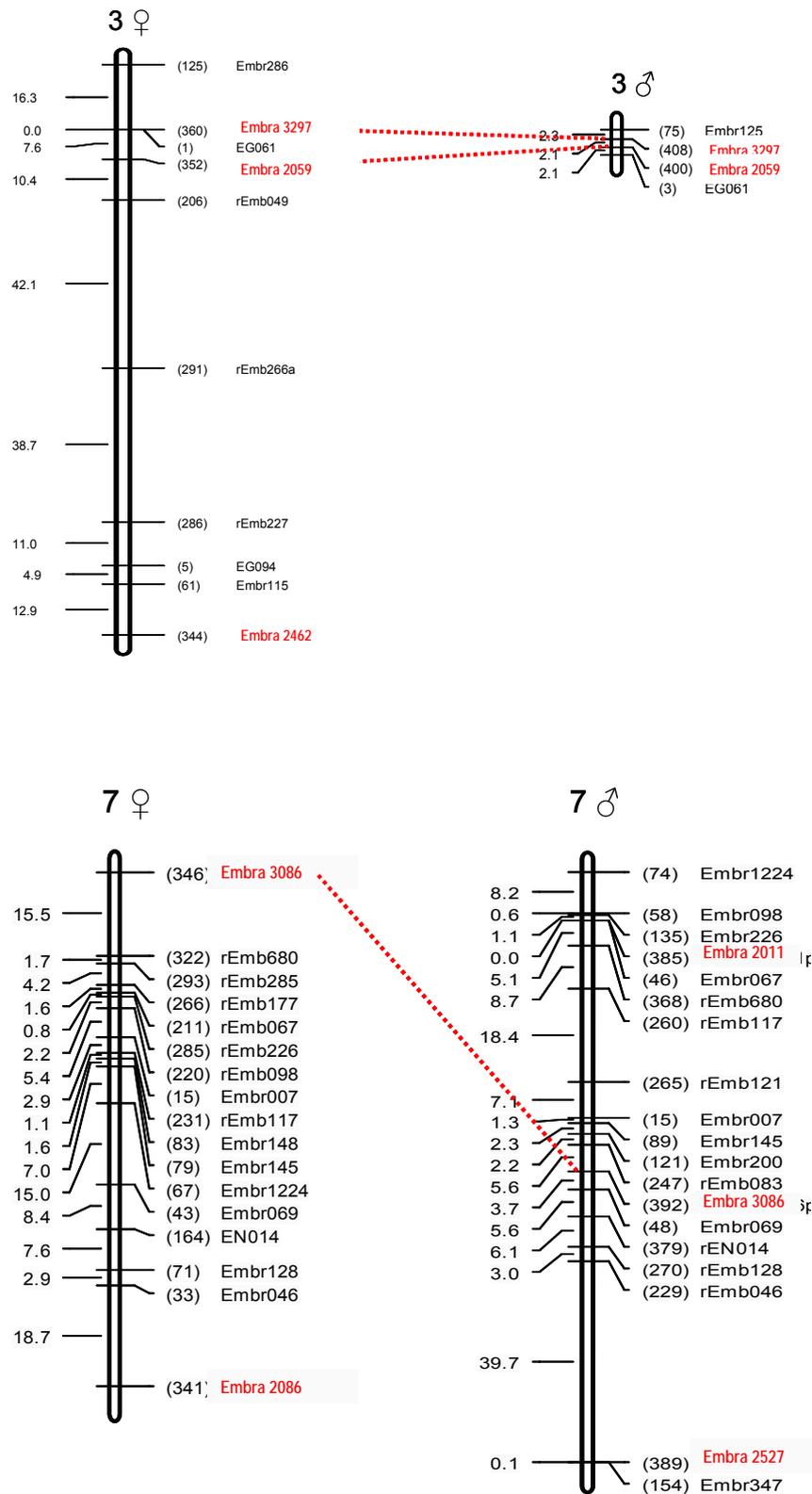
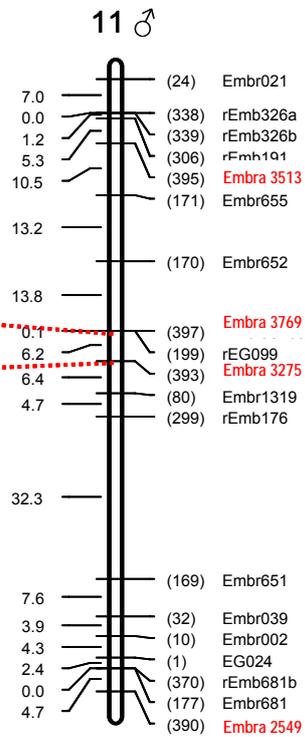
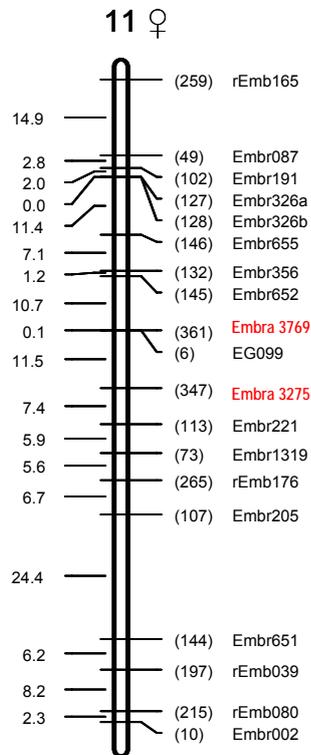
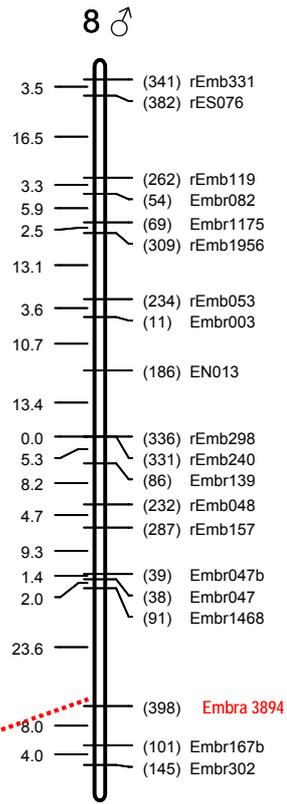
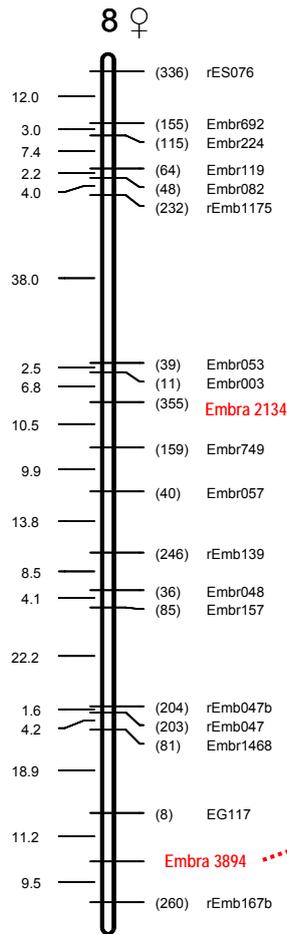


Figura 9: Mapeamento genético de locos EMBRA tetra e pentanucleotídeos nos grupos de ligação de cada genitor (♀ *E. urophylla* | ♂ *E. grandis*) da família IP O mapa genético base utilizado foi o de Mamani *et al.* (não publicado)





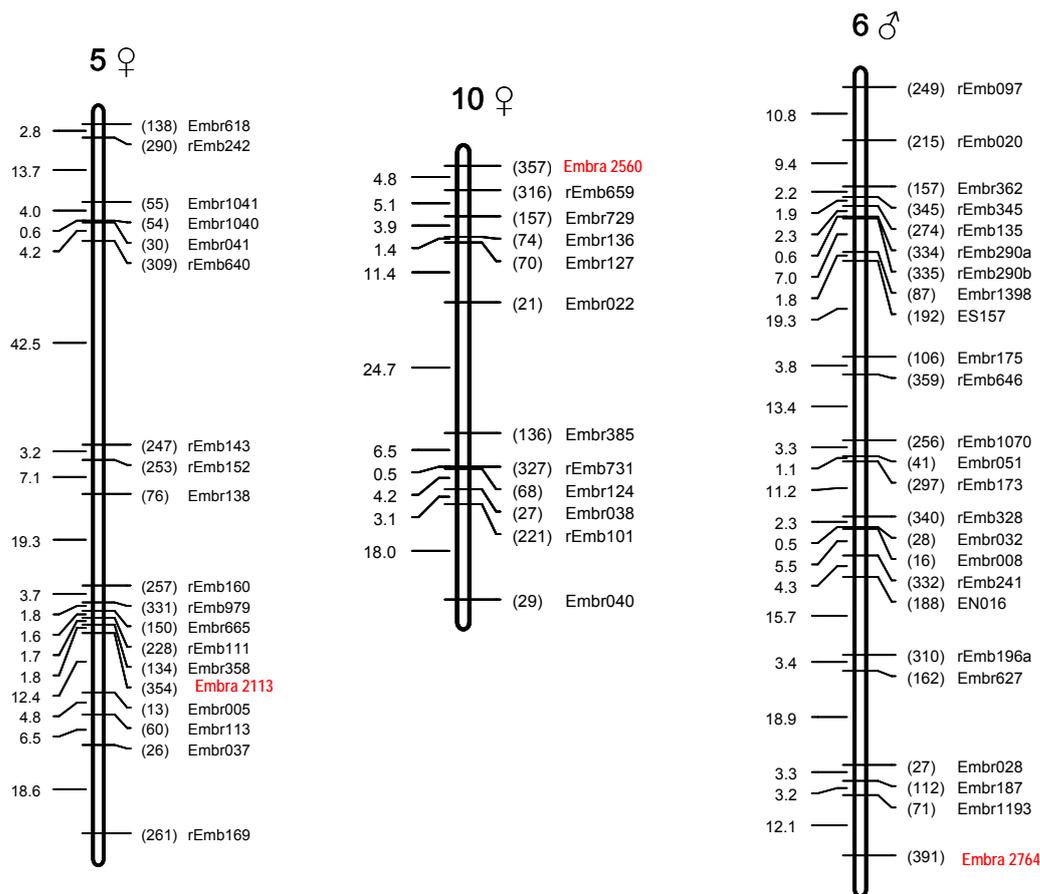


Figura 10: Mapeamento genético de locos EMBRA tetra e pentanucleotídeos nos grupos de ligação de cada genitor (♀ *E. urophylla* | ♂ *E. grandis*) da família VCP. O mapa genético base utilizado foi o de Novaes *et al.* (não publicado)

5.4 Otimização e aplicação de sistemas multiplex de co-amplificação de microssatélites

Considerando os dados de caracterização dos locos tetra e pentanucleotídeos foi possível selecionar um conjunto de locos altamente informativos e montar com eles dois sistemas multiplex de co-amplificação de PCR levando em consideração a cor da marcação com fluorescência FAM (azul), HEX (verde) e NED (amarelo), e a faixa de tamanhos dos alelos esperados de cada marcador. Cada multiplex montado incluiu uma bateria de 7 locos (Tabela 5), sendo que dois locos são repetidos em cada multiplex como sistema de controle da qualidade da declaração de genótipos e possíveis erros de

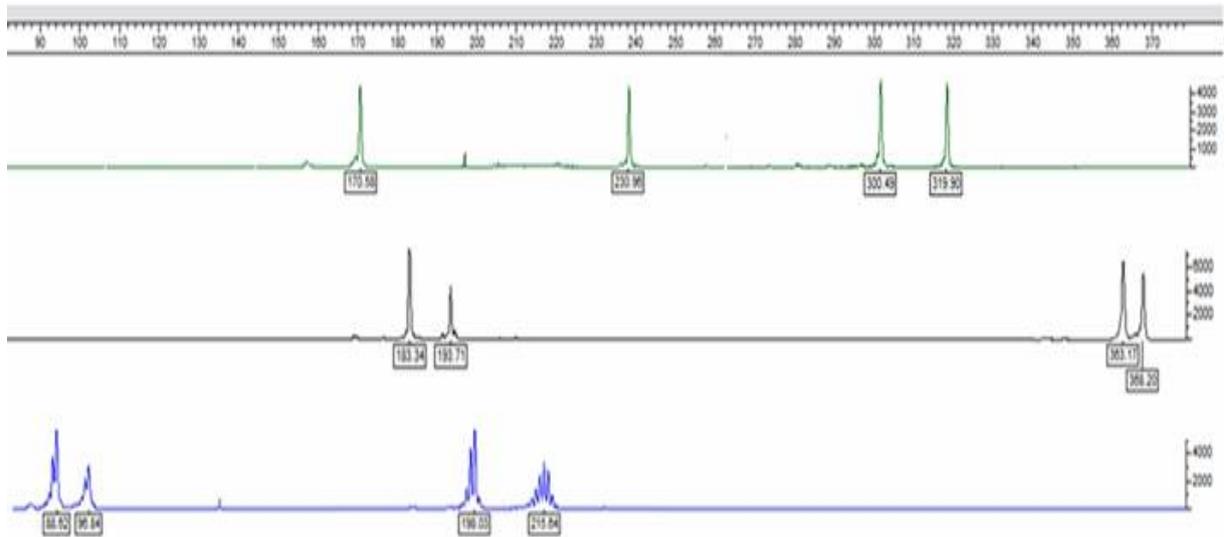
troca de amostras ao longo das análises. Com estes dois sistemas amostras de *Eucalyptus* das três espécies podem ser genotipadas com 12 locos com duas reações de PCR (Figura 11). Os 12 locos combinados fornecem uma probabilidade média nas três espécies de exclusão de paternidade (PE) de 93,92% e uma probabilidade de identidade, parâmetro relevante para trabalhos de identificação de clones, de $1,635 \times 10^{-7}$. Para fins de identificação individual de clones, cada multiplex individualmente já fornece uma probabilidade de identidade suficientemente baixa, da ordem de 10^{-4} o que permite a utilização de apenas um dos multiplex para a grande maioria das aplicações de rotina.

Foram construídos fenogramas representando a discriminação de 46 indivíduos baseados no compartilhamento de alelos utilizando o *software Shared Allele Distance*. Na figura 12 o fenograma foi produzido com os multiplexes 1 e 2, totalizando 12 locos, nas figuras 13 e 14 são apresentados os fenogramas construídos a partir dos multiplex 1 e 2 respectivamente com 7 locos cada um, e finalmente na figura 15 é mostrado o fenograma montado com 9 locos dinucleotídeos. Em todos os casos puderam ser discriminados todos os 46 indivíduos.

Tabela 5: Sistemas *multiplex* de co-amplificação de PCR testados com sucesso neste trabalho. Fluor: marcação FAM (azul), HEX (verde) e NED (amarela); Alelo (pb) faixa de tamanho esperado de alelos; Na: Numero médio de alelos; Ho: heterozigosidade media observada; He: heterozigosidade media esperada; P.I.C.: conteúdo de informação polimórfica; P.I.: Probabilidade de identidade; P.E.: poder de exclusão de paternidade. Estimativas de PI e PE combinadas de cada multiplex e dos 12 locos combinados são também apresentadas.

Multiplex 1								
Loco	Fluor.	Alelo (pb)	Na	Ho	He	P.I.C.	P.I.	P.E.
Embra 3894	FAM	190-230	4,66	0,45	0,62	0,55	0,22	0,15
Embra 2086	HEX	145-185	4,33	0,45	0,57	0,5	0,27	0,16
Embra 2052	HEX	215-260	7,33	0,52	0,59	0,54	0,25	0,23
Embra 2339	NED	160-215	5,66	0,56	0,56	0,5	0,28	0,25
Embra 2675	NED	265-380	3,66	0,43	0,56	0,47	0,38	0,22
Embra 2764	FAM	80-105	3,66	0,48	0,54	0,47	0,38	0,25
Embra 2527	HEX	290-330	7,00	0,32	0,48	0,43	0,36	0,08
							0,000220071	0,77900
Multiplex 2								
Loco	Fluor.	Alelo (pb)	Na	Ho	He	P.I.C.	P.I.	P.E.
Embra 3769	NED	335-420	8,33	0,62	0,75	0,72	0,13	0,34
Embra 3595	FAM	215-240	5,66	0,54	0,59	0,53	0,26	0,25
Embra 2113	HEX	140-175	5,00	0,34	0,41	0,38	0,42	0,11
Embra 3275	HEX	215-260	5,66	0,48	0,50	0,46	0,32	0,20
Embra 2059	NED	195-220	7,66	0,53	0,68	0,63	0,17	0,22
Embra 2764	FAM	80-105	3,66	0,48	0,54	0,47	0,38	0,25
Embra 2527	HEX	290-330	7,00	0,32	0,48	0,43	0,36	0,08
							0,000103307	0,8096
Poder combinado dos 12 locos nos dois multiplex:							1,635E-07	0,9392

A



B

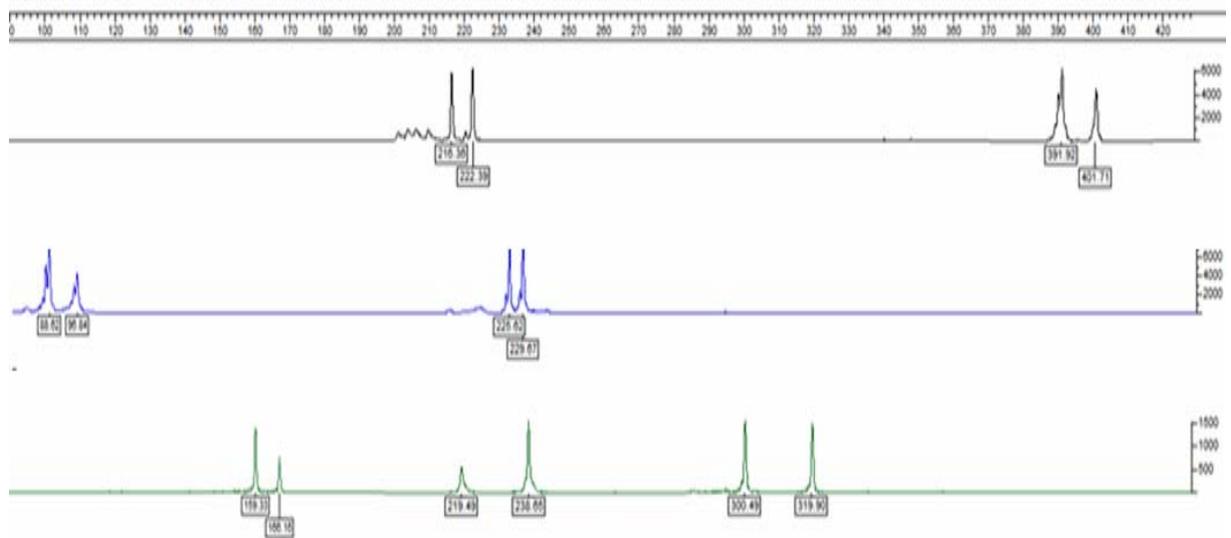


Figura 11: Eletroferogramas ilustrativos dos sistemas multiplex 1 (A) e multiplex 2 (B)

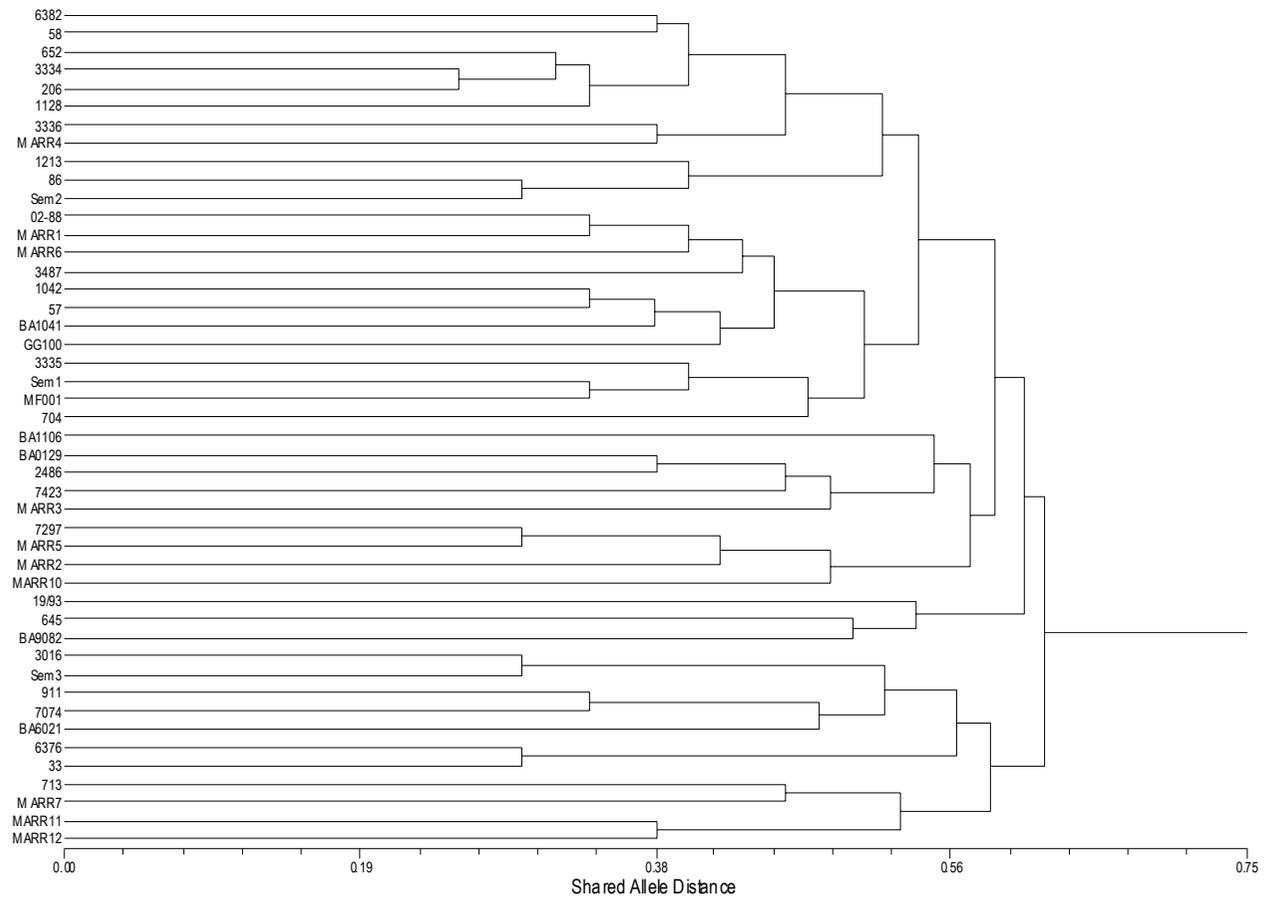


Figura 12: Fenograma representando a discriminação de 46 indivíduos utilizando 12 locos tetra e pentanucleotídeos que corresponde ao multiplex 1 e 2.

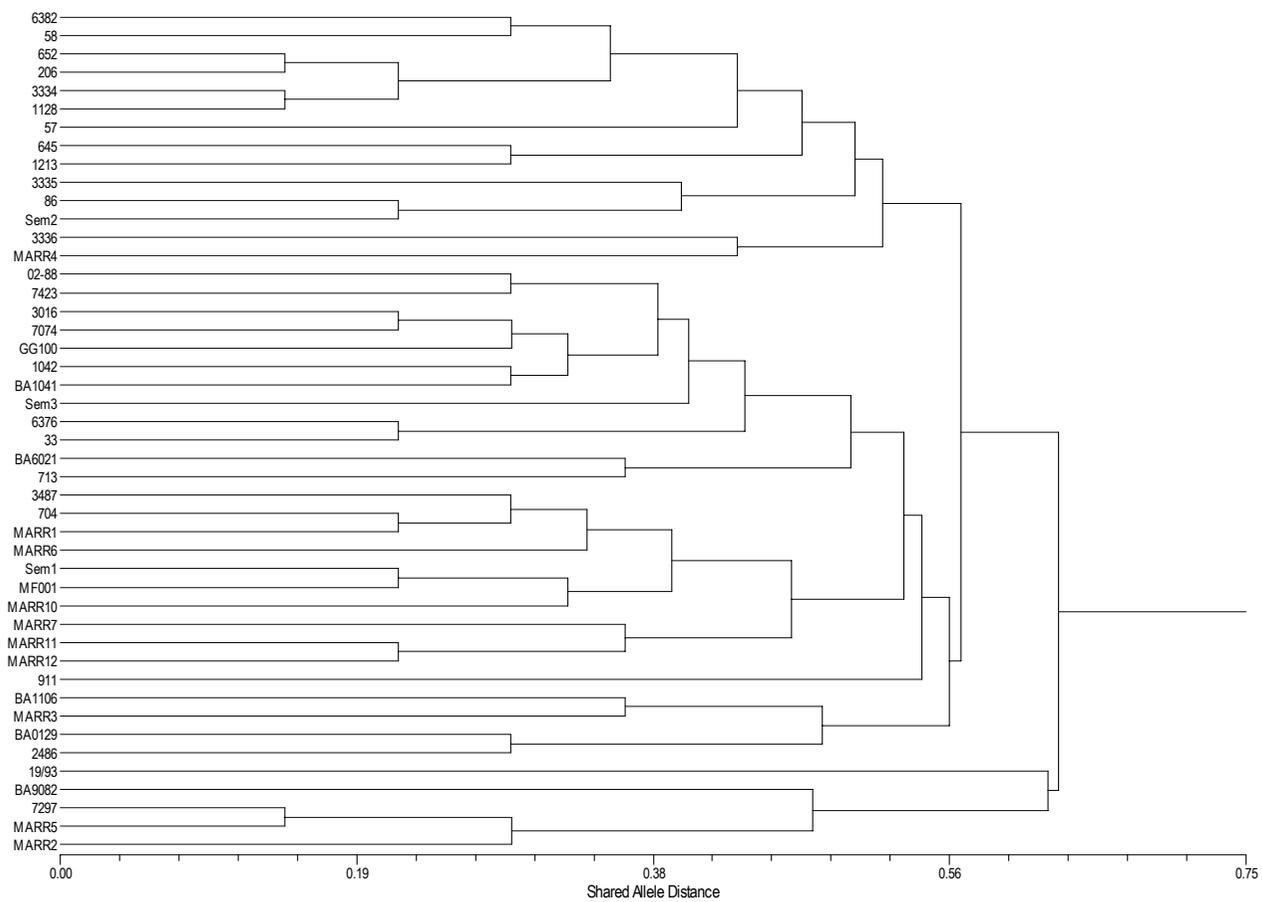


Figura 13: Fenograma representando a discriminação de 46 indivíduos utilizando 7 locos (tetra e pentanucleotídeos) que correspondem ao multiplex 1.

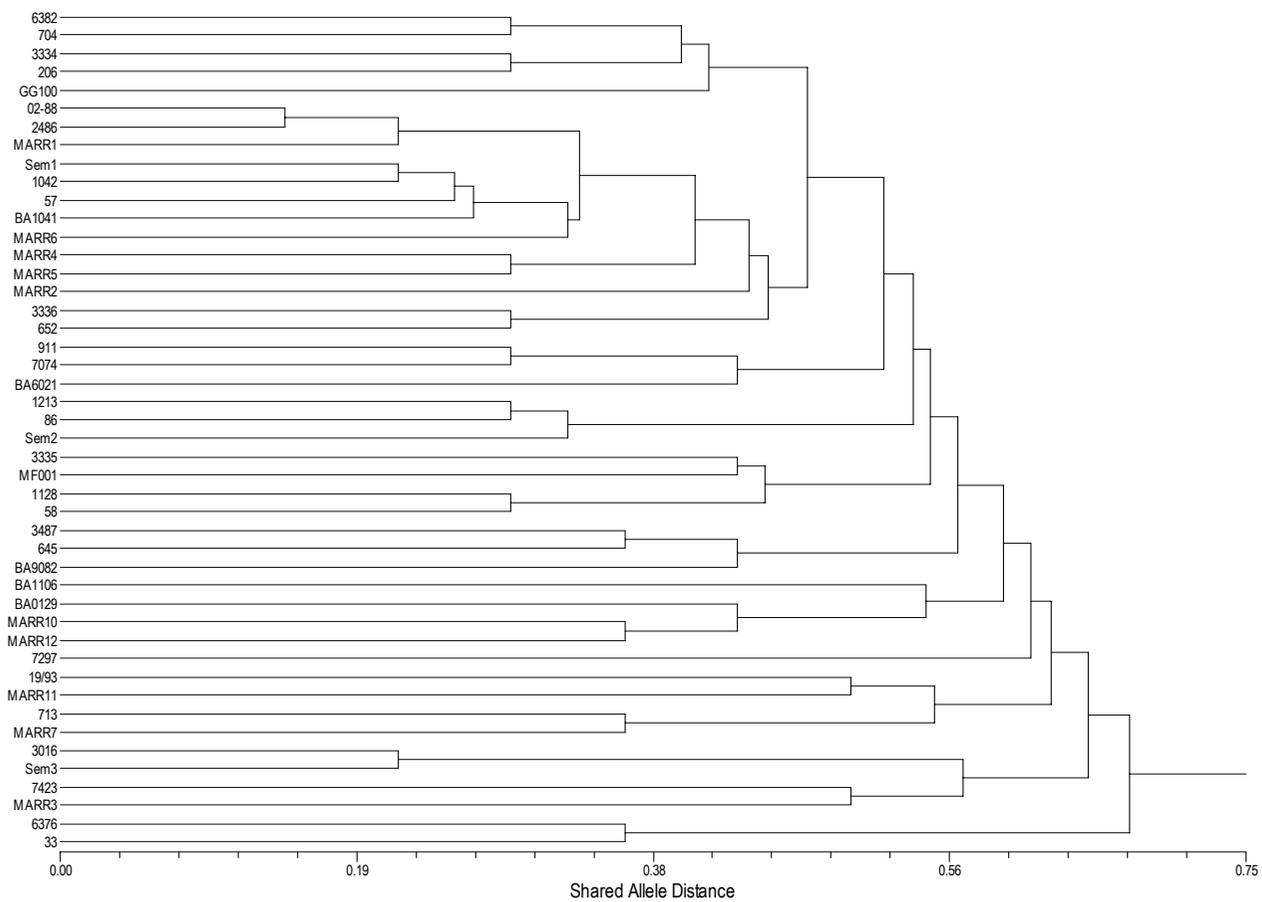


Figura 14: Fenograma representando a discriminação de 46 indivíduos utilizando 7 locos (tetra e pentanucleotídeos) que correspondem ao multiplex 2.

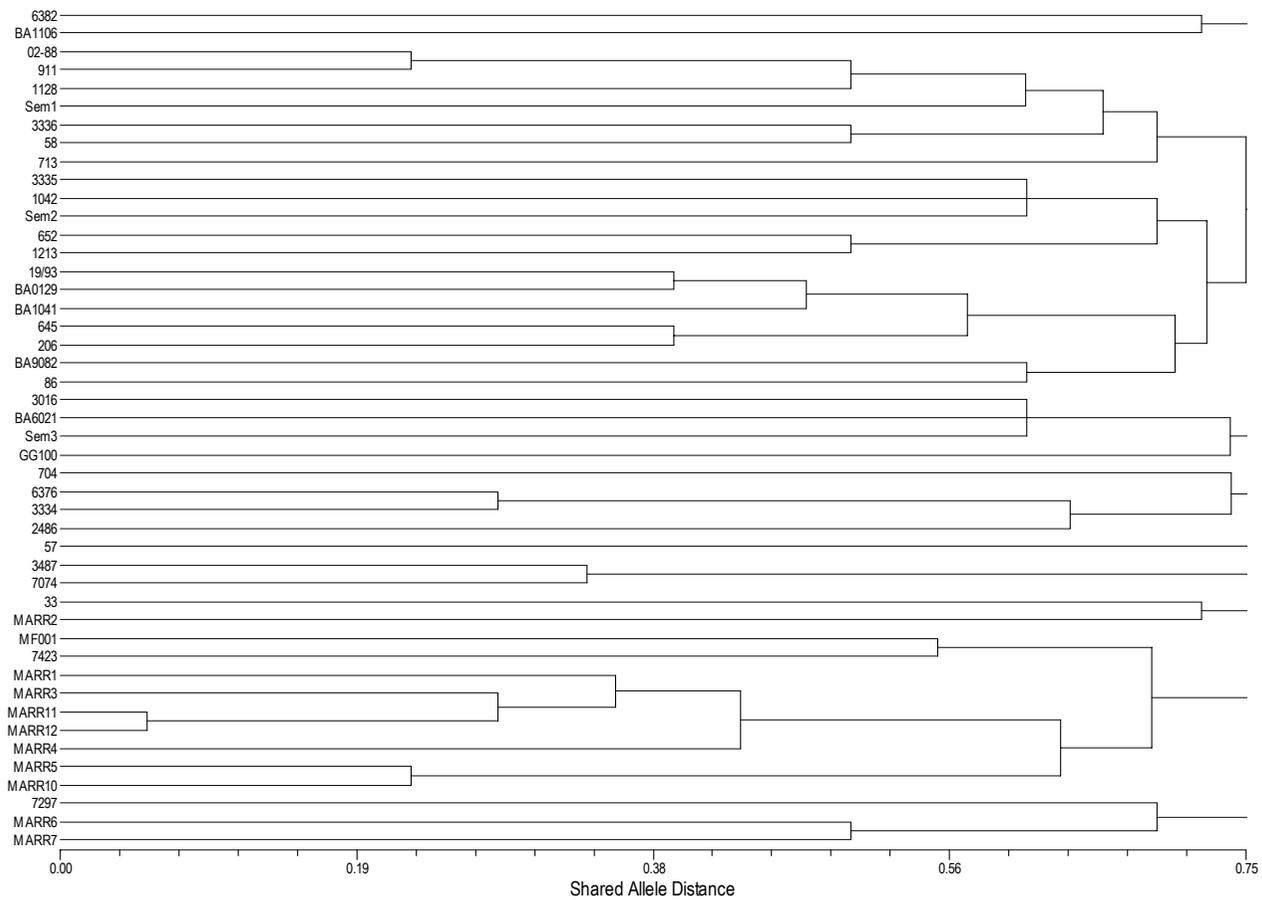


Figura 15: Fenograma representando a discriminação de 46 indivíduos utilizando 9 locos dinucleotídeos.

6. DISCUSSÃO

No presente trabalho foi realizada uma busca de microssatélites perfeitos partindo da mineração de 18.510 seqüências genômicas de clones BAC, identificando-se 2.406 seqüências contendo um ou mais microssatélites de duas ou mais bases. A partir dessas seqüências foram identificados 3.034 microssatélites perfeitos para os quais foi possível desenhar pares de iniciadores para 1896 microssatélites, segundo as condições de temperatura (T_m entre 59-60°C) e tamanho do produto esperado (entre 80 e 550 pb.). O insucesso no desenho de iniciadores para cerca de 38% dos microssatélites ocorreu em função da ausência de seqüência flanqueante ao microssatélite ou da baixa qualidade da seqüência disponível. Mesmo assim o número de potenciais marcadores derivados do banco de seqüências genômicas disponíveis é elevado e certamente suficiente para o desenvolvimento adicional de um grande número de marcadores. Foram desenhados iniciadores para 728 dinucleotídeos, 495 trinucleotídeos, 304 tetranucleotídeos, 196 pentanucleotídeos e 107 hexanucleotídeos. Do total de marcadores, foram selecionados para síntese de iniciadores os microssatélites com repetições em tandem de 4 e 5 *motifs* (tetra e pentanucleotídeos respectivamente) com tamanho de produto de amplificação esperado menor a 450pb, sendo estes, 110 potenciais marcadores microssatélites. Neste trabalho, portanto, somente foram explorados uma pequena parte dos tetra e pentanucleotídeos disponíveis, cerca de 25% dos tetra e 15% dos pentanucleotídeos, restando um grande número de microssatélites a serem testados. As triagens realizadas, tanto com resolução em agarose, como em PAGE permitiram avaliar com detalhamento crescente a natureza e o comportamento dos marcadores em diferentes espécies de *Eucalyptus*. Isso permitiu selecionar os locos mais robustos, com ampla transferibilidade entre espécies e com potencial de multiplexagem em função da faixa de tamanho de alelos observada. Dos 80 locos tetranucleotídeos após as etapas de triagem foram obtidos 48 locos (60%) de amplificação

robusta, 5 locos (6,3%) de amplificação inespecífica e 27 pares de iniciadores (33,7%) que não resultaram em amplificação de produtos de PCR. Dos locos que tiveram amplificação robusta, 62,5% (30 locos) foram polimórficos, 27% (13 locos) apresentaram duplicações e 10,4% (5 locos) foram monomórficos. Dos 30 locos pentanucleotídeos, 80% (24) apresentaram amplificação robusta, 10% amplificação de produtos inespecíficos e 10% não amplificaram. Dos locos pentanucleotídeos que apresentaram amplificação robusta 62,5% (15 locos) foram polimórficos, 29% (7 locos) mostraram duplicações e 8,4% (2 locos) foram monomórficos. Foram selecionados por fim 45 locos que consolidaram todos os atributos objetivados, sendo 30 tetra e 15 pentanucleotídeos, obtendo assim uma eficiência de resgate de marcadores tetra e pentanucleotídeos de 37,5% e 50% respectivamente, com potencial de se tornarem marcadores operacionais. Estes foram analisados com detecção fluorescente para a identificação mais precisa e análise detalhada em populações e clones elite. No momento, está sendo desenvolvida uma bateria adicional de locos tetra e pentanucleotídeos (Santos L.P. com. pess.) com o objetivo de aumentar o número de locos com alto conteúdo informativo. Resultados similares de eficiência de resgate foram observados no trabalho de Brondani *et. al.* (1998), no qual foram desenvolvidos marcadores microssatélites dinucleotídeos em *Eucalyptus grandis* e *E. urophylla*, de 180 marcadores, 35% (63 locos) tiveram uma amplificação robusta e de fácil interpretação; 37% (67 locos) apresentaram bandas inespecíficas, e 28% (50 locos) não amplificaram. Naquele trabalho uma eficiência final de 35 % de resgate do grupo inicial de marcadores foi obtida. Num trabalho realizado em arroz por Temnykh *et. at.* (2001), de 42 marcadores tetranucleotídeos identificados, observou-se um sucesso de amplificação em 71,4 % dos locos, dos quais 60% apresentaram polimorfismo, ou seja, uma eficiência final de aproximadamente 42%. Em outro estudo realizado em girassol, Yu *et. al.* (2002), desenvolveram 11 marcadores tetranucleotídeos dos quais 7 amplificaram com sucesso (63,6%), desses 7 somente 4 (57,1%) foram polimórficos. No trabalho de Yi *et. al.* (2006), realizado em pimenta, foram desenvolvidos 1201 marcadores,

sendo 216 di, 793 tri, 93 tetra e 99 pentanucleotídeos, dos 93 SSR tetranucleotídeos, foram desenhados 65 pares de iniciadores dos quais 33 amplificaram com sucesso (50,8%) e somente 14 deles apresentaram polimorfismo (42,4%). Dos 99 marcadores pentanucleotídeos foram desenhados 60 pares de iniciadores, 45 deles amplificaram (75%) e somente 6 marcadores foram polimórficos (13,3%).

No trabalho de Yu *et al.* (2002), o número médio de alelos observado foi de 3,7 em dinucleotídeos, 3,6 em trinucleotídeos e 9,5 em tetranucleotídeos, demonstrando surpreendentemente que os locos tetranucleotídeos possuíam uma maior variabilidade alélica que os di e trinucleotídeos, além de parecer uma das classes mais polimórficas de marcadores de DNA em girassol. Em *Eucalyptus* o número médio de alelos observado foi de 4,7 e 4,4 alelos nos locos tetra e pentanucleotídeos respectivamente, significativamente menor do que uma média de 9 a 26 alelos observados em dinucleotídeos (Brondani *et al.* 1998; 2002) o que está de acordo com o esperado. Dinucleotídeos tendem a evoluir muito mais rapidamente do que tetra e pentanucleotídeos pela maior facilidade da ocorrência do fenômeno de escorregamento (*slippage*) da DNA polimerase durante a replicação. Entretanto foi observado um loco com um número de alelos quase o dobro da média, o loco tetranucleotídeo EMBRA 3769 com um número médio de alelos de 8,33 e locos pentanucleotídeos também com mais alelos do que a média, exemplos EMBRA 3275 e EMBRA 2343 com 5,66. Este resultado sugere que existe a possibilidade de se identificar locos tetra ou pentanucleotídeos significativamente mais polimórficos possivelmente por serem mais antigos e terem assim acumulado um maior número de mutações no tamanho do SSR com freqüências suficientemente altas para se tornarem marcadores hipervariáveis equivalentes aos dinucleotídeos. Assumindo que a freqüência de locos tetranucleotídeos mais hipervariáveis é da ordem de 1 a cada 80 testados, conforme observado neste trabalho existiria a possibilidade de descobrir mais três marcadores com estas

características a partir dos 304 tetranucleotídeos descobertos na mineração das pontas de BAC. Com base neste mesmo raciocínio, seriam 4 tetranucleotídeos mais hipervariáveis para os 9.6 mega pb amostrados neste trabalho, o que forneceria cerca de 252 tetranucleotídeos altamente hipervariáveis para os 630 Mpb do genoma de *E. grandis*.

Em seres humanos os locos tetra e pentanucleotídeos são amplamente utilizados para identificação de indivíduos e estudos de diversidade genética (Redd *et al.*, 2002; Sarkar *et al.* 2002; e Tandon *et al.*, 2002). Para fins de identificação forense somente são utilizados tetra e pentanucleotídeos por questões de precisão e intercâmbio de resultados entre laboratórios. Na análise realizada por Sarkar *et al.* (2002), foram utilizados 15 locos (Penta E, D18S51, D21S11, THO1, D3S1358, FGA, TPOX, D8S1179, VWA, Amelogenina, Penta D, CSF1PO, D16S539, D7S820, D13S317 e D5S818), dos quais dois são penta e 13 tetranucleotídeos. Várias populações humanas foram estudadas e em todas elas o número de alelos foi bastante homogêneo, da ordem de 9 a 15 alelos, apenas com pequenas diferenças de frequência entre as populações. O fato que o melhoramento genético de *Eucalyptus* envolve em geral múltiplas espécies demandou que o desenvolvimento e avaliação dos locos neste trabalho fosse feita em três espécies. Isso obviamente torna o trabalho mais complexo uma vez que foi necessário buscar locos que fossem igualmente polimórficos nas três espécies. Ao se analisar os locos por espécie isoladamente, este trabalho revelou locos com um elevado número de alelos, bastante próximo daquele tipicamente observado em tetranucleotídeos na espécie humana ou em dinucleotídeos de *Eucalyptus*, como por exemplo, o loco Embra 3769 com 12 alelos em *E. globulus* e o loco Embra 3034 em *E. urophylla* com 11 alelos. Este resultado sugere que o desempenho forense dos sistemas multiplex poderá ser significativamente melhor do que o teórico em situações operacionais a depender da espécie estudada.

Os locos tetra e pentanucleotídeos desenvolvidos apresentaram uma amplificação robusta, de fácil interpretação e alta transferibilidade (100%) entre as três espécies de *Eucalyptus* (*E. grandis*, *E. globulus* e *E. urophylla*) utilizadas em programas de melhoramento e plantios no mundo. Num estudo realizado por Chagné *et al.* (2004), foi comparada a transferibilidade de marcadores microssatélites di, tri e tetranucleotídeos entre várias espécies de *Pinus*. Observou-se que os marcadores mais freqüentes e com maior transferibilidade foram os trinucleotídeos (63% em *Pinus tadea* e 51% em *Pinus pinaster*), seguido pelos marcadores dinucleotídeos (36% em *P. tadea* e 45% em *P. pinaster*) e finalmente os marcadores tetranucleotídeos que foram praticamente ausentes no descobrimento feito a partir de um banco de EST de *Pinus* (1% em *P. taeda* e 3% em *P. pinaster*). Esta alta taxa de transferibilidade observada em *Eucalyptus* permite maximizar seu uso em vários tipos de estudos em diferentes espécies e possivelmente nas várias espécies filogeneticamente relacionadas. Estes estudos incluem a construção de mapas genéticos, a análise de diversidade em populações naturais e de melhoramento e o *fingerprinting* para identificação individual em casos de proteção varietal.

O grande diferencial destes novos marcadores em relação à ampla bateria de locos dinucleotídeos desenvolvidos e mapeados até hoje (Brondani *et al.* 1998; 2002; 2006) é a facilidade significativamente maior de interpretação de alelos e declaração e comparação interlaboratorial de genótipos principalmente em sistemas multiplex de co-amplificação de múltiplos locos. A fácil identificação é devida ao fato dos picos alélicos poderem ser visualizados com precisão por estarem separados por um tamanho maior (4 a 5 pb) comparando com locos di ou trinucleotídeos (2 a 3 pb). Além disso, estes alelos raramente sofrem do fenômeno de gaguejo (*stutter bands*), muito comum em marcadores dinucleotídeos. A maior separação de alelos facilita ainda a identificação de microvariantes, encontrados frequentemente em microssatélites, porém dificilmente discriminados em di e trinucleotídeos enquanto que em tetra e

penta estes podem ser identificados com maior facilidade e alta precisão. Estas características tomadas conjuntamente permitirão uma comparação imediata e de alta confiabilidade de resultados entre laboratórios. A possibilidade de discriminação precisa de alelos permite ainda a definição de intervalos (*bins*) alélicos definidos que podem ser organizados em escadas alélicas controle nos mesmos moldes do que vem sendo praticado em genética humana. Para isso seria recomendado o seqüenciamento dos alelos observados a cada loco o que permitiria ainda a definição precisa de alelos microvariantes.

As estimativas de parâmetros forenses, conteúdo de informação polimórfica (PIC) probabilidade de identidade (PI) e poder de exclusão (PE), assim como as Heterozigosidades observada (H_o) e esperada (H_e) permitiram adotar critérios objetivos para a seleção de locos candidatos a compor sistemas multiplex para utilização em rotina para identificação individual de *Eucalyptus*. Entre todos os parâmetros forenses estimados o mais relevante para utilização em *Eucalyptus* é a probabilidade de identidade (PI) tendo em vista que a maior demanda por análises genéticas tem sido e continuará sendo a discriminação precisa de clones e análises de identidade. Ao analisar a estimativa de probabilidade de identidade foi possível identificar marcadores informativos, como por exemplo, o loco tetranucleotídeo EMBRA 3769 (0,13) e o pentanucleotídeo EMBRA 2343 (0,21). O poder combinado dos 12 locos selecionados para os sistemas multiplex alcançou um PI da ordem de $1,64 \times 10^{-7}$ o que em geral é plenamente suficiente para a correta discriminação de clones de *Eucalyptus*. Vale ressaltar, entretanto, que uma bateria de apenas 6 locos dinucleotídeos selecionados com PI individual variando do mais informativo com $PI = 0,008$ ao menos informativo com $PI = 0,168$ forneceu um PI combinado da ordem de 2×10^{-9} ou seja duas ordens de magnitude mais potente com metade do número de locos (Kirst *et al.* 2005). O poder de exclusão de paternidade também constitui um parâmetro interessante e relevante para as situações nas quais é demandada a investigação de parentesco

de árvores selecionadas necessária na estratégia de investigação retrospectiva de parentesco proposta por Grattapaglia *et al.* (2004). Em função da menor hipervariabilidade observada e presença de alelos dominantes em frequência, o poder de exclusão de paternidade observado para os locos tetra e pentanucleotídeos desenvolvidos é relativamente limitado com um valor médio de 0,13 nos locos tetranucleotídeos, e 0,12 nos pentanucleotídeos. O maior valor identificado nos locos tetranucleotídeos foi novamente no loco EMBRA 3769 (0,34) e nos pentanucleotídeos no loco 2134 (0,38). Locos dinucleotídeos certamente são significativamente mais potentes para esta aplicação. Os seis locos selecionados e avaliados por Kirst *et al.* (2005) variaram em PE desde um mínimo de 0,434 a um máximo de 0,941 e um PE combinado da ordem de 99,999% enquanto que os 2 locos tetra e pentanucleotídeos selecionados para os dois multiplexes neste estudo alcançaram uma PE combinada de apenas 93,9%. Embora o PE combinado permita a exclusão de apenas cerca de 94% do falso parental de um indivíduo, na presença de um deles, para a maioria das situações este poder será suficiente. Entretanto em situações operacionais, ao longo do processo seqüencial de exclusão de paternidade poderá ser necessária a utilização de uma bateria adicional de dinucleotídeos principalmente em situações nas quais os possíveis parentais não excluídos apresentem algum grau de coancestralidade.

A análise de herança e segregação dos locos tetra e pentanucleotídeos nas famílias IP e VCP identificou 28 locos passíveis de serem mapeados na família IP (62%) e 22 na família VCP (49%). Ou seja, em média 55,5%. Esta estimativa obtida é bastante próxima da estimativa teórica média de PIC, 59%. Do total de locos possíveis de mapear, foram por fim mapeados 23 marcadores, sendo 8 pertencentes ao mapa da família IP e 16 ao mapa referência da família VCP, um dos marcadores foi localizado nos dois mapas. Esta eficiência de mapeamento é significativamente inferior àquela observada tipicamente para locos dinucleotídeos com mais de 90% dos locos passíveis de mapeamento em qualquer família segregante analisada (Brondani

et. al., 1998; 2002; 2006). Além disso, para dinucleotídeos entre 55 e 60% dos microsatélites segregam em configuração totalmente informativa enquanto que somente 22 a 25% dos tetra e pentanucleotídeos apresentaram este comportamento. Claramente locos tetra e pentanucleotídeos não serão os mais adequados para fins de mapeamento genético de QTL, por exemplo, pois apenas uma pequena proporção apresentará segregação informativa em uma dada família. Vale observar ainda que um número relativamente alto de locos tetranucleotídeos, 10 em 45 na família VCP, segregou em configuração tipo F2, ou seja, segregação 1:2:1 o que não possibilitou o seu mapeamento em função da ausência de informação de fase de ligação.

Com base nas estimativas individuais de desempenho forense dos vários locos obtidas neste trabalho foi possível propor um sistema de dois sistemas multiplexes envolvendo 7 locos cada um, 2 deles replicados nos dois multiplexes para atuar como controle totalizando assim 12 locos diferentes. Os dois sistemas multiplex foram utilizados para genotipar uma coleção de 46 clones elite. Foi possível discriminar com precisão todos os 46 clones com pelo menos 3 locos apresentando genótipos distintos em todas as comparações tomadas duas a duas. Um fenograma ilustrativo baseado na estimativa do compartilhamento de alelos (Shared Allele Distance) foi construído indicando que todos os clones foram discriminados com uma distância média de $0,58 \pm 0,11$. Os mesmos 46 clones também foram genotipados com uma bateria de nove locos dinucleotídeos o que também permitiu a discriminação precisa de todos os clones. Conforme esperado, uma comparação preliminar dos resultados obtidos com os dois sistemas indicou que já a partir de 3 a 4 locos dinucleotídeos foi possível discriminar todos os clones com pelo menos duas diferenças genotípicas, enquanto que com os sistemas de tetra e pentanucleotídeos isso somente foi alcançado a partir de 5 a 6 locos. Mesmo assim, do ponto de vista prático, esta diferença não é relevante uma vez que em situações operacionais não seriam utilizados menos do que seis locos microsatélites sejam eles di ou tetranucleotídeos, sugerindo

assim que os sistemas multiplex propostos neste trabalho constituem ferramentas efetivamente operacionais para a identificação individual de clones de *Eucalyptus*.

Vale por fim ressaltar que várias outras combinações de locos tetra e/ou pentanucleotídeos poderiam ser sugeridas para os sistemas multiplex e poderão vir a ser testadas em trabalhos futuros. Desenhos alternativos de iniciadores flanqueando os microsatélites visando aumentar a compatibilidade das co-amplificações de locos e marcações alternativas com os diferentes fluorocromos poderão ser planejadas visando aumentar o número de locos co-amplificados e analisados na eletroforese. Em conclusão, este trabalho, revelou com sucesso a possibilidade de identificar locos tetra e pentanucleotídeos polimórficos e transferíveis entre espécies de *Eucalyptus* e com características genéticas adequadas para a utilização operacional. Os experimentos realizados lançaram as bases para desenvolvimento adicionais seja de um maior número de locos, seja da identificação de locos mais hipervariáveis e de combinações de sistemas multiplex eficientes e de alta precisão para espécies de *Eucalyptus*, passíveis de adoção em escala internacional para a análise genética de clones e populações.

7. Conclusões

1. O trabalho de descobrimento e posterior teste de conversão de pares de iniciadores em marcadores operacionais resultou em uma eficiência de resgate de marcadores tetra e pentanucleotídeos da ordem de 37,5% e 50% respectivamente, com potencial de se tornarem marcadores operacionais. Esta eficiência foi semelhante àquela obtida para o desenvolvimento de marcadores baseados em dinucleotídeos a partir de bibliotecas enriquecidas.

2. O número médio de alelos observado foi de 4,7 e 4,4 alelos nos locos tetra e pentanucleotídeos respectivamente, significativamente menor do que uma média de 9 a 26 alelos observados em dinucleotídeos. Entretanto foram identificados seja tetra como pentanucleotídeos com um número de alelos muito acima da média sugerindo que é possível identificar locos tetra ou pentanucleotídeos com hipervariabilidade equivalente àquela de dinucleotídeos.

3. Assumindo que a frequência de locos tetranucleotídeos perfeitos mais hipervariáveis é da ordem de 1 a cada 80 testados conforme observado neste trabalho, existiria a possibilidade de descobrir mais três marcadores com estas características a partir dos 304 tetranucleotídeos descobertos na mineração das pontas de BAC. Com base neste mesmo raciocínio, seriam 4 tetranucleotídeos mais hipervariáveis para os 9.6 mega pb amostrados neste trabalho, o que forneceria cerca de 252 tetranucleotídeos altamente hipervariáveis para os 630 Mpb do genoma de *E. grandis*.

4. Os locos tetra e pentanucleotídeos desenvolvidos apresentaram uma amplificação robusta, de fácil interpretação e alta transferibilidade (100%) entre as três espécies de *Eucalyptus* (*E. grandis*, *E. globulus* e *E. urophylla*) utilizadas em programas de melhoramento e plantios no mundo. Esta alta taxa de transferibilidade observada em *Eucalyptus* permite maximizar seu uso em vários tipos de estudos em diferentes espécies e possivelmente nas várias espécies filogeneticamente relacionadas. Estes estudos incluem a construção de mapas genéticos, a análise de diversidade em populações naturais e de melhoramento e o *fingerprinting* para identificação individual em casos de proteção varietal.

5. O grande diferencial destes novos marcadores em relação à ampla bateria de locos dinucleotídeos disponíveis hoje é a elevada qualidade analítica e conseqüente facilidade significativamente maior de interpretação de alelos e declaração e comparação interlaboratorial de genótipos incluindo a identificação de microvariantes. Estas características tomadas conjuntamente permitirão uma comparação imediata e de alta confiabilidade de resultados entre laboratórios.
6. O poder combinado dos 12 locos selecionados para os sistemas multiplex propostos neste trabalho alcançou uma probabilidade de identidade (PI) da ordem de $1,64 \times 10^{-7}$ o que em geral é plenamente suficiente para a correta discriminação de clones de *Eucalyptus* embora duas ordens de magnitude abaixo daquele estimado para uma bateria de apenas 6 dinucleotídeos.
7. Em função da menor hipervariabilidade observada e presença de alelos dominantes em freqüência, o poder de exclusão de paternidade (PE) estimado para os locos tetra e pentanucleotídeos é relativamente limitado com uma PE combinada de apenas 93,9% para os 12 locos selecionados embora suficiente para a maioria das situações. Entretanto em situações operacionais, ao longo do processo seqüencial de exclusão de paternidade poderá ser necessária a utilização de uma bateria adicional de dinucleotídeos principalmente em situações nas quais os possíveis parentais não excluídos apresentem algum grau de coancestralidade.
8. A eficiência de mapeamento de locos tetra e pentanucleotídeos, da ordem de 55 a 59% é significativamente inferior àquela observada tipicamente para locos dinucleotídeos com mais de 90% dos locos passíveis de mapeamento em qualquer família segregante analisada. Além disso, para dinucleotídeos entre 55 e 60% dos microsatélites

segregam em configuração totalmente informativa enquanto que somente 22 a 25% dos tetra e pentanucleotídeos apresentaram este comportamento. Claramente locos tetra e pentanucleotídeos não serão os mais adequados para fins de mapeamento genético de QTL, por exemplo, pois apenas uma pequena proporção apresentará segregação informativa em uma dada família segregante.

9. Uma comparação preliminar da capacidade de discriminação individual observada em uma coleção de 46 clones elite indicou que já a partir de 3 a 4 locos dinucleotídeos é possível discriminar todos os clones com pelo menos duas diferenças genotípicas, enquanto que com os sistemas de tetra e pentanucleotídeos isso somente foi alcançado a partir de 5 a 6 locos. Entretanto, do ponto de vista prático os sistemas multiplex propostos neste trabalho constituem ferramentas efetivamente operacionais para a identificação individual de clones de *Eucalyptus*.
10. Vale por fim ressaltar que várias outras combinações de locos tetra e/ou pentanucleotídeos poderiam ser sugeridas para os sistemas multiplex e poderão vir a ser testadas em trabalhos futuros. Desenhos alternativos de iniciadores flanqueando os microssatélites visando aumentar a compatibilidade das co-amplificações de locos e marcações alternativas com os diferentes fluorocromos poderão ser planejados visando aumentar o número de locos co-amplificados e analisados na eletroforese.
11. Em conclusão, este trabalho, revelou com sucesso a possibilidade de identificar locos tetra e pentanucleotídeos polimórficos e transferíveis entre espécies de *Eucalyptus* e com características genéticas adequadas para a utilização operacional. Os experimentos realizados lançaram as bases para desenvolvimento adicionais seja de um maior número de locos, seja da identificação de locos mais hipervariáveis e de combinações

de sistemas multiplex eficientes e de alta precisão para espécies de *Eucalyptus*, passíveis de adoção em escala internacional para a análise genética de clones e populações.

8. ANEXOS

Anexo 1: Frequências alélicas dos locos tetranucleotídeos nas seis procedências das espécies estudadas. A: *E. grandis* de Coff's Harbor; B: *E. grandis* de Atherton Queensland; C: *E. globulus* de Jeerlang; D: *E. globulus* de Flinder; E: *E. urophylla* de Timor; F: *E. urophylla* de Flores.

Locos	Alelo	A	B	C	D	E	F
EMBRA 3275	214	0,00	0,00	0,00	0,00	0,02	0,00
	219	0,03	0,00	0,00	0,00	0,02	0,00
	224	0,23	0,00	0,00	0,00	0,00	0,00
	229	0,00	0,02	0,07	0,05	0,00	0,10
	234	0,13	0,17	0,11	0,25	0,02	0,00
	239	0,50	0,81	0,72	0,55	0,80	0,67
	244	0,00	0,00	0,00	0,00	0,02	0,00
	249	0,13	0,00	0,11	0,16	0,11	0,24
EMBRA 3894	197	0,00	0,00	0,00	0,00	0,00	0,05
	201	0,04	0,04	0,19	0,00	0,48	0,43
	206	0,39	0,06	0,35	0,77	0,05	0,05
	211	0,22	0,69	0,35	0,23	0,48	0,39
	216	0,35	0,21	0,10	0,00	0,00	0,00
	217	0,00	0,00	0,00	0,00	0,00	0,05
	227	0,00	0,00	0,00	0,00	0,00	0,05
EMBRA 2420	89	0,00	0,00	0,00	0,00	0,00	0,11
	104	0,15	0,07	0,52	0,96	0,04	0,32
	109	0,08	0,02	0,33	0,00	0,18	0,36

	114	0,77	0,91	0,15	0,04	0,71	0,21
	264	0,00	0,00	0,00	0,00	0,07	0,00
EMBRA 3297	166	0,15	0,13	0,00	0,00	0,00	0,09
	171	0,33	0,54	0,35	0,00	0,57	0,66
	176	0,38	0,29	0,65	0,96	0,35	0,20
	181	0,15	0,04	0,00	0,04	0,09	0,02
	186	0,00	0,00	0,00	0,00	0,00	0,02
EMBRA 2134	197	0,50	0,50	0,19	0,02	0,17	0,17
	202	0,48	0,40	0,56	0,98	0,17	0,20
	207	0,02	0,10	0,21	0,00	0,52	0,57
	212	0,00	0,00	0,04	0,00	0,08	0,02
	227	0,00	0,00	0,00	0,00	0,04	0,04
	232	0,00	0,00	0,00	0,00	0,02	0,00
EMBRA 2321	255	0,00	0,23	0,00	0,00	0,00	0,00
	260	0,05	0,08	0,05	0,11	0,00	0,00
	265	0,05	0,00	0,00	0,20	0,00	0,11
	270	0,60	0,70	0,89	0,50	0,95	0,87
	275	0,00	0,00	0,05	0,20	0,05	0,03
	280	0,20	0,00	0,00	0,00	0,00	0,00
	285	0,00	0,00	0,02	0,00	0,00	0,00
	295	0,10	0,00	0,00	0,00	0,00	0,00
EMBRA 2570	125	0,00	0,00	0,00	0,00	0,00	0,17
	130	0,65	0,79	0,93	0,69	0,89	0,83
	135	0,02	0,00	0,00	0,04	0,00	0,00
	140	0,00	0,00	0,02	0,10	0,06	0,00
	150	0,22	0,19	0,05	0,06	0,06	0,00
	175	0,11	0,02	0,00	0,10	0,00	0,00
EMBRA 2587	100	0,00	0,00	0,04	0,00	0,00	0,00
	105	0,06	0,08	0,11	0,05	0,00	0,00
	110	0,00	0,13	0,04	0,05	0,00	0,04

	115	0,90	0,79	0,82	0,88	0,83	0,96
	120	0,04	0,00	0,00	0,03	0,08	0,00
	130	0,00	0,00	0,00	0,00	0,03	0,00
	135	0,00	0,00	0,00	0,00	0,06	0,00
EMBRA 3808	213	0,00	0,00	0,00	0,00	0,00	0,94
	248	0,00	0,00	0,00	0,00	0,00	0,06
	283	0,00	0,00	0,00	0,05	0,00	0,00
	288	0,00	0,00	0,00	0,02	0,00	0,00
	298	0,79	0,82	0,73	0,89	0,00	0,00
	303	0,00	0,00	0,05	0,00	0,00	0,00
	308	0,21	0,18	0,23	0,05	0,00	0,00
EMBRA 2343	111	0,14	0,11	0,00	0,00	0,21	0,00
	116	0,00	0,00	0,00	0,09	0,00	0,00
	121	0,29	0,25	0,37	0,53	0,46	0,61
	126	0,00	0,00	0,57	0,38	0,04	0,11
	131	0,05	0,00	0,00	0,00	0,00	0,03
	136	0,19	0,11	0,00	0,00	0,13	0,05
	141	0,00	0,00	0,00	0,00	0,00	0,05
	146	0,29	0,52	0,07	0,00	0,17	0,16
	156	0,05	0,00	0,00	0,00	0,00	0,00

Anexo 2 Frequências alélicas dos locos pentanucleotídeos nas seis procedências das espécies estudadas. A: *E. grandis* de Coff's Harbor; B: *E. grandis* de Atherton Queensland; C: *E. globulus* de Jeerlang; D: *E. globulus* de Flinder; E: *E. urophylla* de Timor; F: *E. urophylla* de Flores.

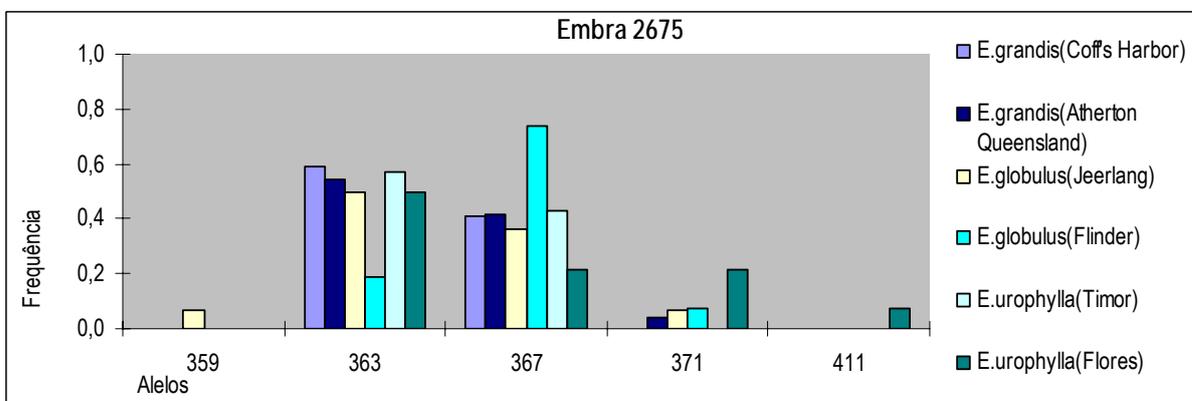
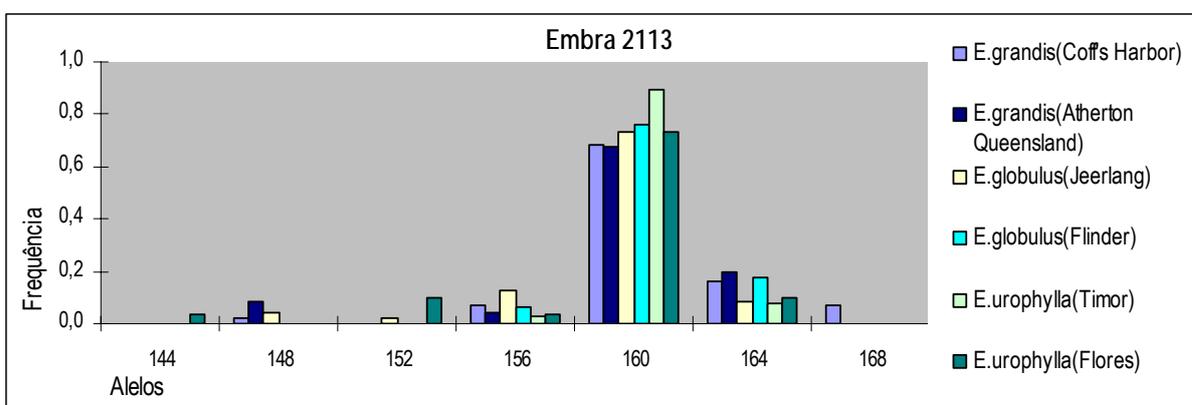
Locos	Alelo	A	B	C	D	E	F
EMBRA 3275	214	0,00	0,00	0,00	0,00	0,02	0,00
	219	0,03	0,00	0,00	0,00	0,02	0,00
	224	0,23	0,00	0,00	0,00	0,00	0,00

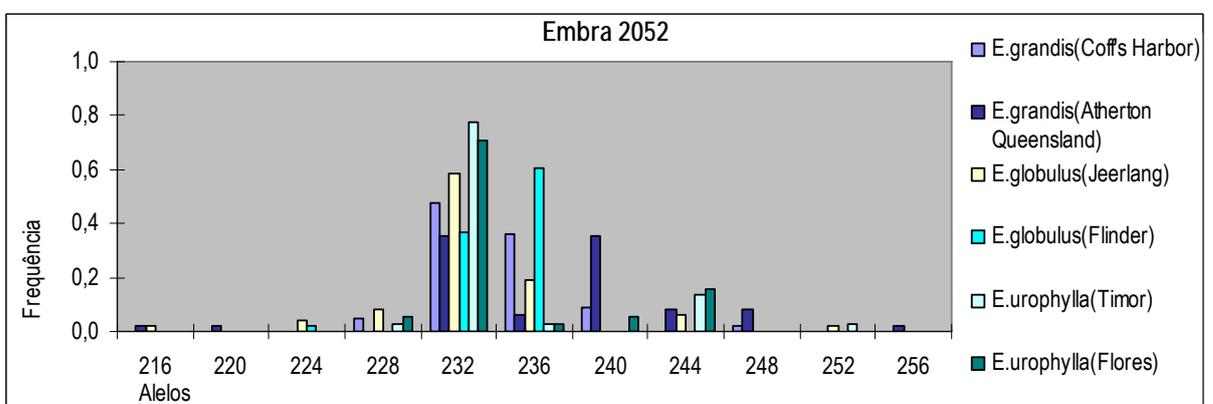
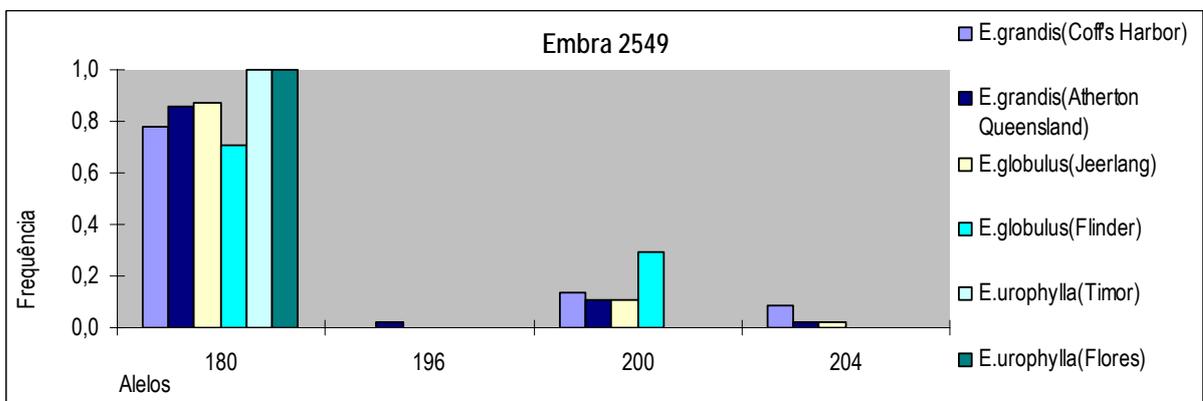
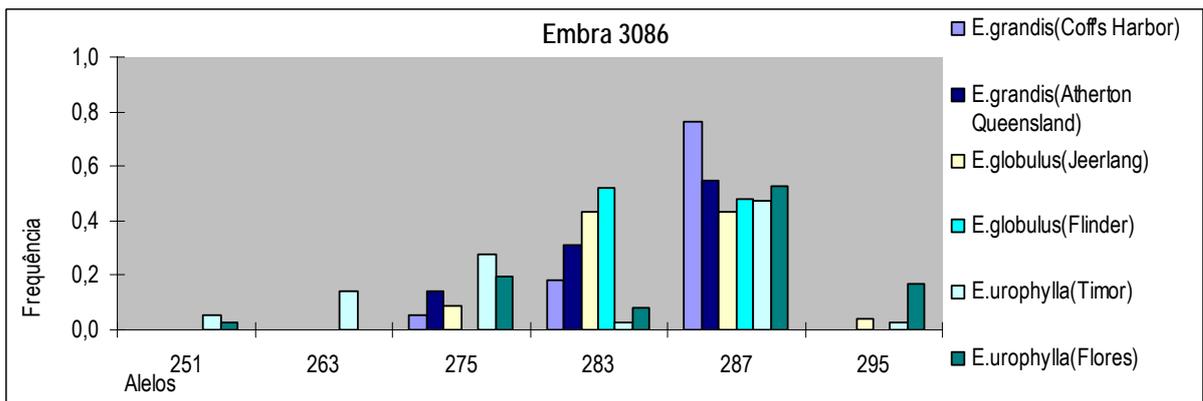
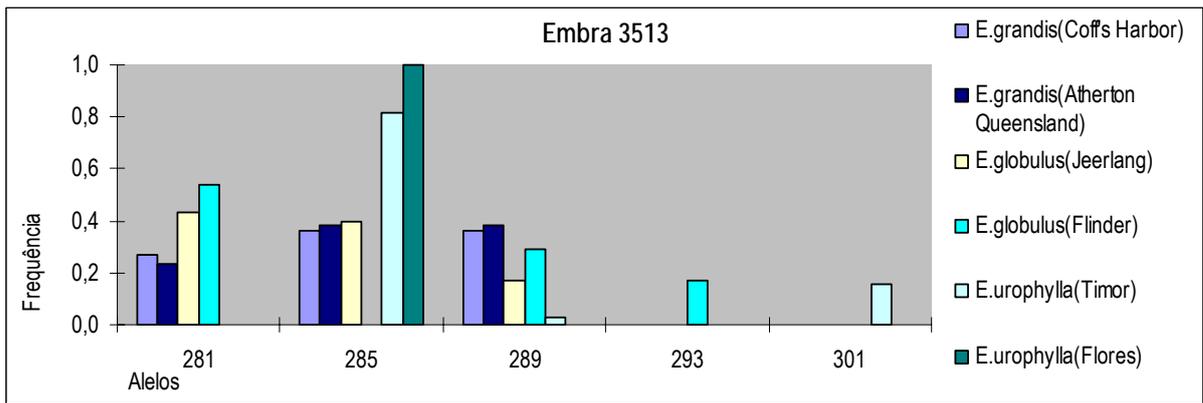
	229	0,00	0,02	0,07	0,05	0,00	0,10
	234	0,13	0,17	0,11	0,25	0,02	0,00
	239	0,50	0,81	0,72	0,55	0,80	0,67
	244	0,00	0,00	0,00	0,00	0,02	0,00
	249	0,13	0,00	0,11	0,16	0,11	0,24
EMBRA 3894	197	0,00	0,00	0,00	0,00	0,00	0,05
	201	0,04	0,04	0,19	0,00	0,48	0,43
	206	0,39	0,06	0,35	0,77	0,05	0,05
	211	0,22	0,69	0,35	0,23	0,48	0,39
	216	0,35	0,21	0,10	0,00	0,00	0,00
	217	0,00	0,00	0,00	0,00	0,00	0,05
	227	0,00	0,00	0,00	0,00	0,00	0,05
EMBRA 2420	89	0,00	0,00	0,00	0,00	0,00	0,11
	104	0,15	0,07	0,52	0,96	0,04	0,32
	109	0,08	0,02	0,33	0,00	0,18	0,36
	114	0,77	0,91	0,15	0,04	0,71	0,21
	264	0,00	0,00	0,00	0,00	0,07	0,00
EMBRA 3297	166	0,15	0,13	0,00	0,00	0,00	0,09
	171	0,33	0,54	0,35	0,00	0,57	0,66
	176	0,38	0,29	0,65	0,96	0,35	0,20
	181	0,15	0,04	0,00	0,04	0,09	0,02
	186	0,00	0,00	0,00	0,00	0,00	0,02
EMBRA 2134	197	0,50	0,50	0,19	0,02	0,17	0,17
	202	0,48	0,40	0,56	0,98	0,17	0,20
	207	0,02	0,10	0,21	0,00	0,52	0,57
	212	0,00	0,00	0,04	0,00	0,08	0,02
	227	0,00	0,00	0,00	0,00	0,04	0,04
	232	0,00	0,00	0,00	0,00	0,02	0,00
EMBRA 2321	255	0,00	0,23	0,00	0,00	0,00	0,00
	260	0,05	0,08	0,05	0,11	0,00	0,00

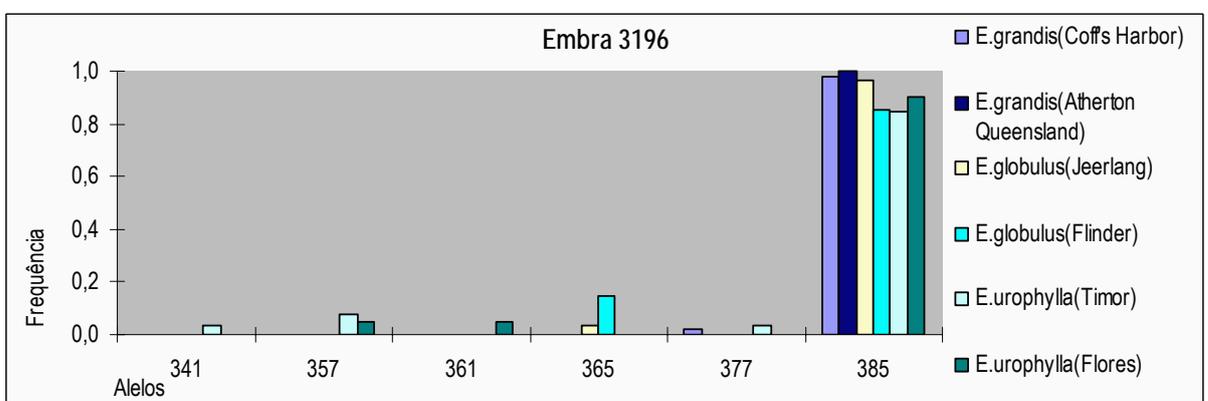
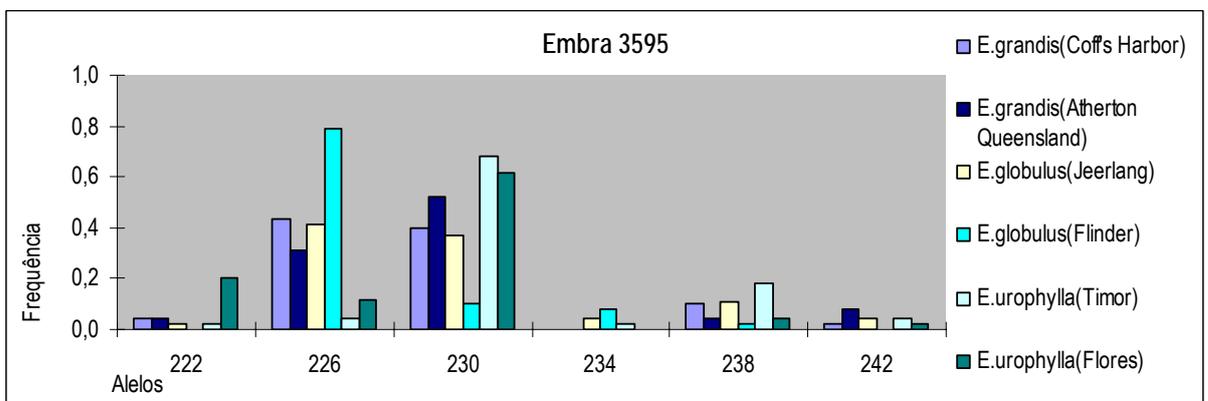
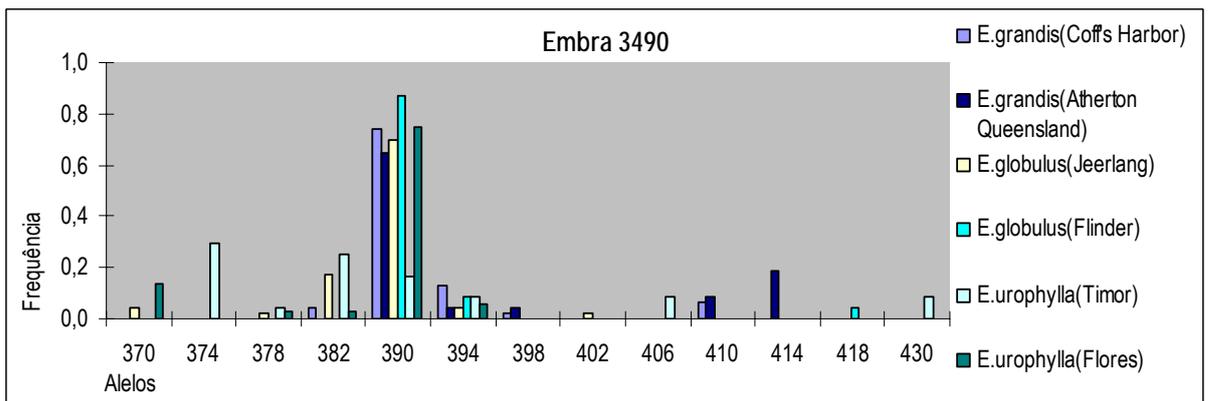
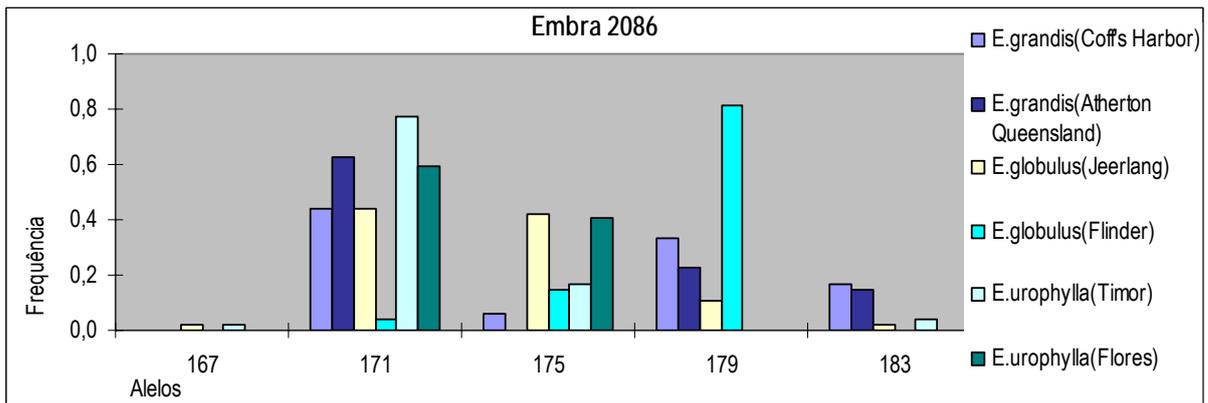
265	0,05	0,00	0,00	0,20	0,00	0,11	
270	0,60	0,70	0,89	0,50	0,95	0,87	
275	0,00	0,00	0,05	0,20	0,05	0,03	
280	0,20	0,00	0,00	0,00	0,00	0,00	
285	0,00	0,00	0,02	0,00	0,00	0,00	
295	0,10	0,00	0,00	0,00	0,00	0,00	
<hr/>							
EMBRA 2570	125	0,00	0,00	0,00	0,00	0,00	0,17
	130	0,65	0,79	0,93	0,69	0,89	0,83
	135	0,02	0,00	0,00	0,04	0,00	0,00
	140	0,00	0,00	0,02	0,10	0,06	0,00
	150	0,22	0,19	0,05	0,06	0,06	0,00
	175	0,11	0,02	0,00	0,10	0,00	0,00
<hr/>							
EMBRA 2587	100	0,00	0,00	0,04	0,00	0,00	0,00
	105	0,06	0,08	0,11	0,05	0,00	0,00
	110	0,00	0,13	0,04	0,05	0,00	0,04
	115	0,90	0,79	0,82	0,88	0,83	0,96
	120	0,04	0,00	0,00	0,03	0,08	0,00
	130	0,00	0,00	0,00	0,00	0,03	0,00
	135	0,00	0,00	0,00	0,00	0,06	0,00
<hr/>							
EMBRA 3808	213	0,00	0,00	0,00	0,00	0,00	0,94
	248	0,00	0,00	0,00	0,00	0,00	0,06
	283	0,00	0,00	0,00	0,05	0,00	0,00
	288	0,00	0,00	0,00	0,02	0,00	0,00
	298	0,79	0,82	0,73	0,89	0,00	0,00
	303	0,00	0,00	0,05	0,00	0,00	0,00
	308	0,21	0,18	0,23	0,05	0,00	0,00
<hr/>							
EMBRA 2343	111	0,14	0,11	0,00	0,00	0,21	0,00
	116	0,00	0,00	0,00	0,09	0,00	0,00
	121	0,29	0,25	0,37	0,53	0,46	0,61
	126	0,00	0,00	0,57	0,38	0,04	0,11

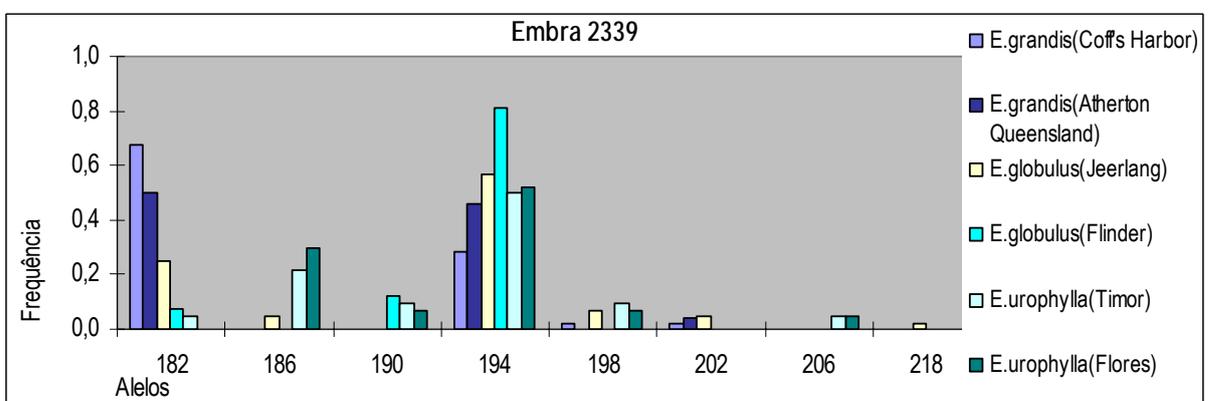
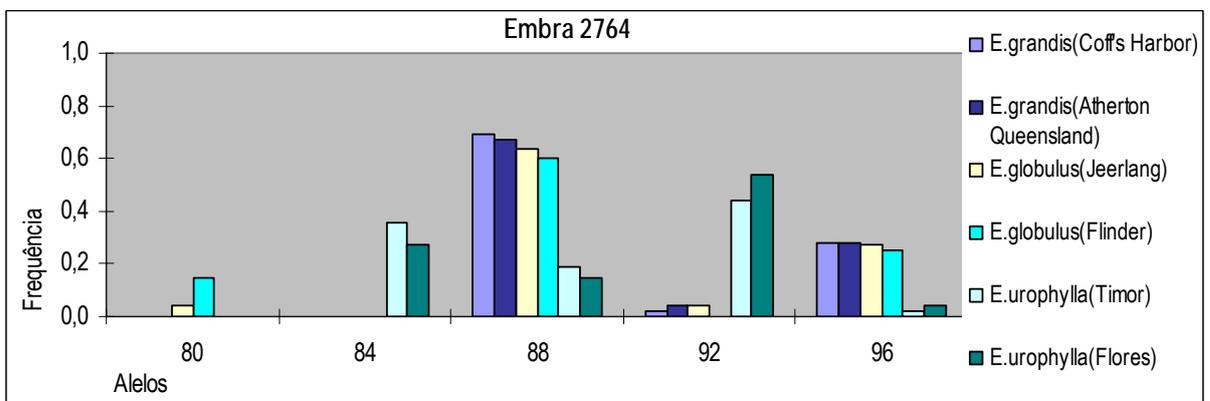
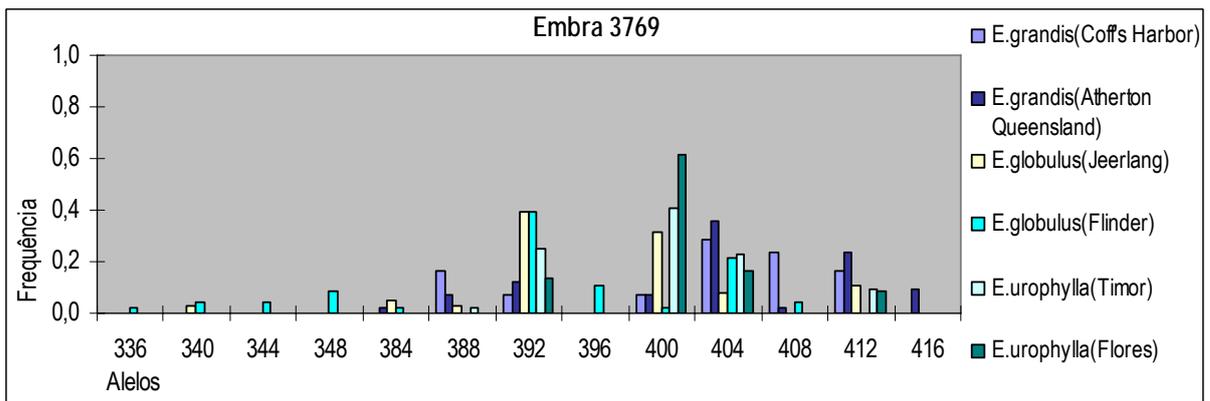
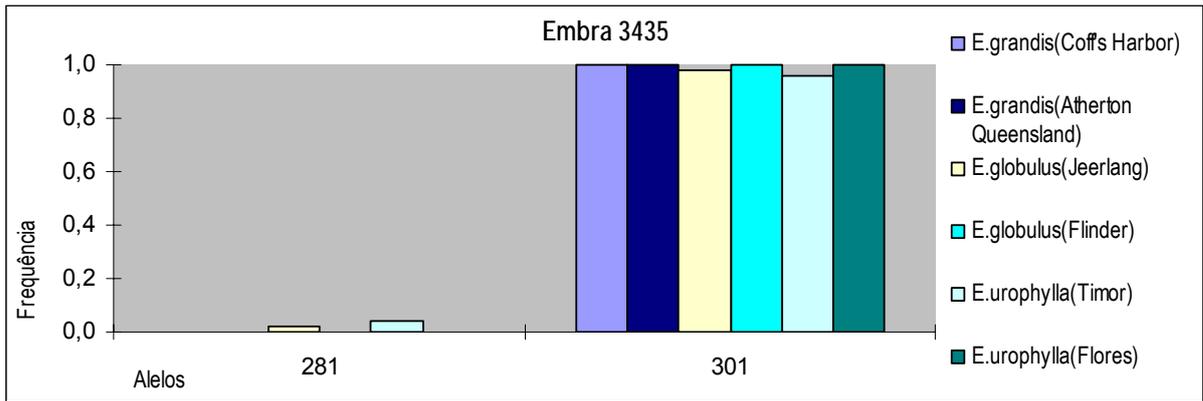
131	0,05	0,00	0,00	0,00	0,00	0,03
136	0,19	0,11	0,00	0,00	0,13	0,05
141	0,00	0,00	0,00	0,00	0,00	0,05
146	0,29	0,52	0,07	0,00	0,17	0,16
156	0,05	0,00	0,00	0,00	0,00	0,00

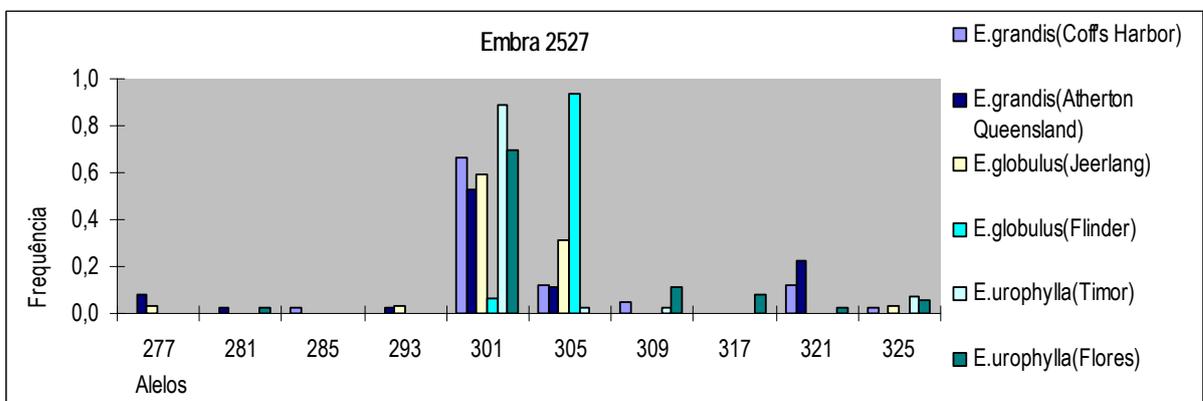
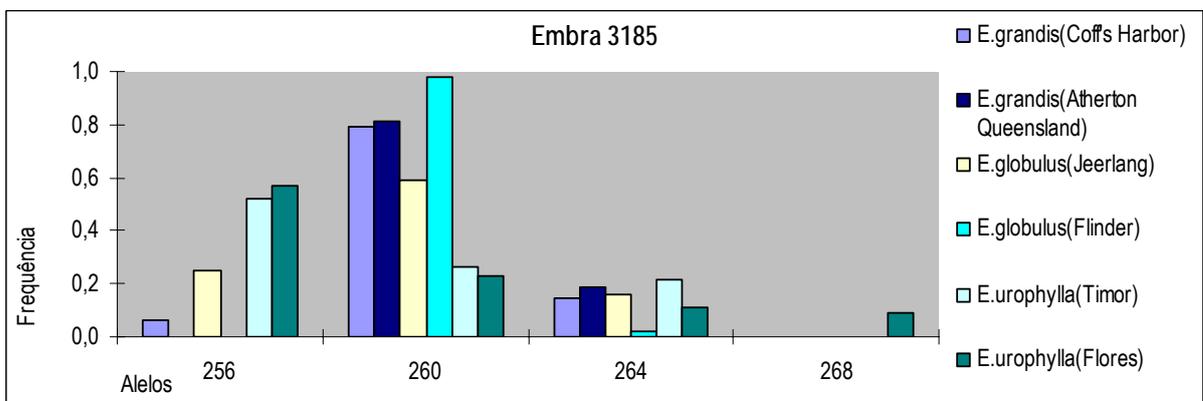
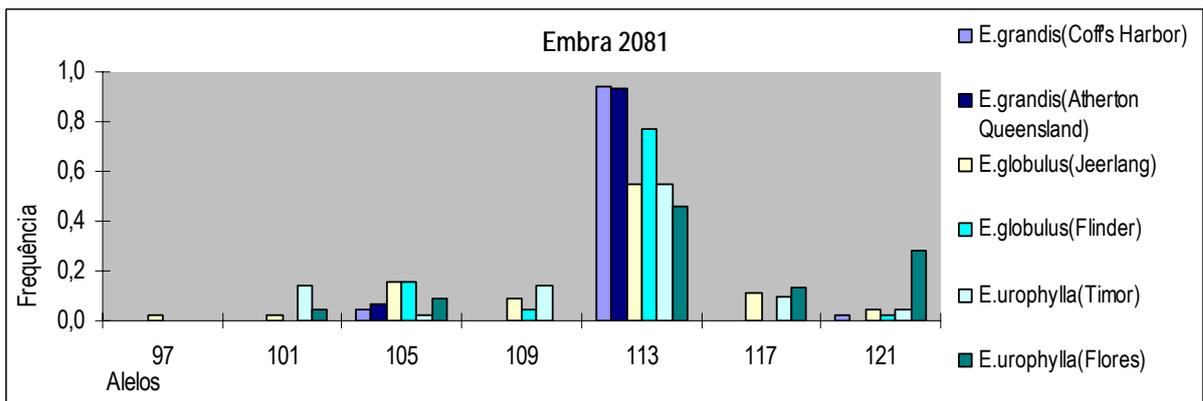
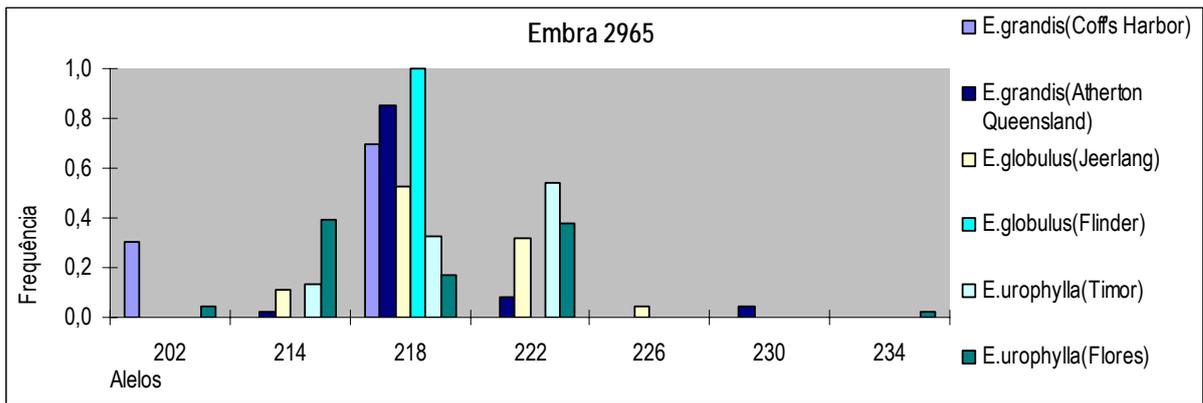
Anexo 3. Histogramas de freqüências alélicas dos 22 locos tetranucleotídeos selecionados para caracterização genética.

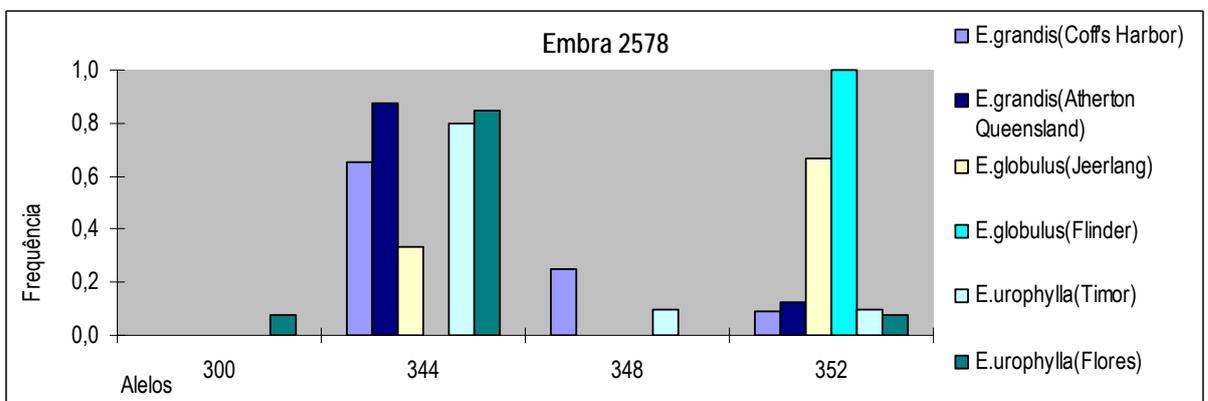
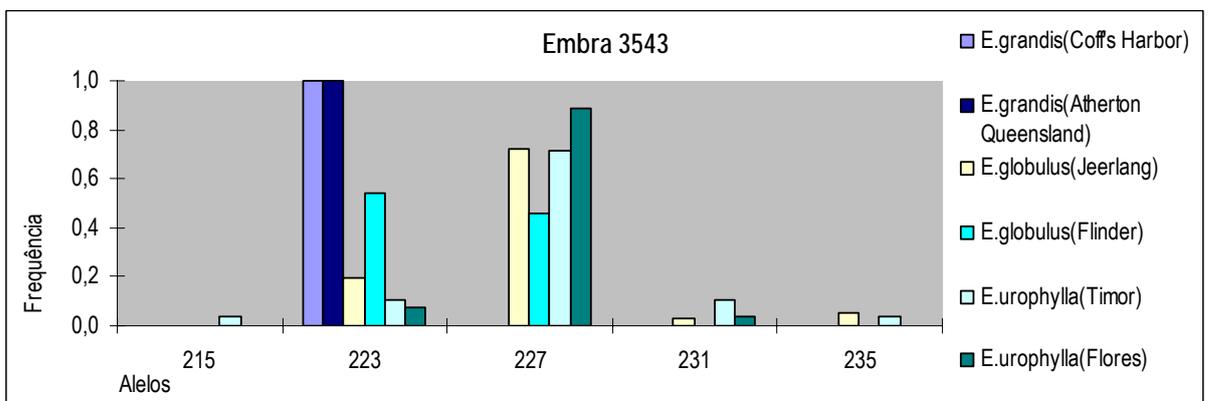
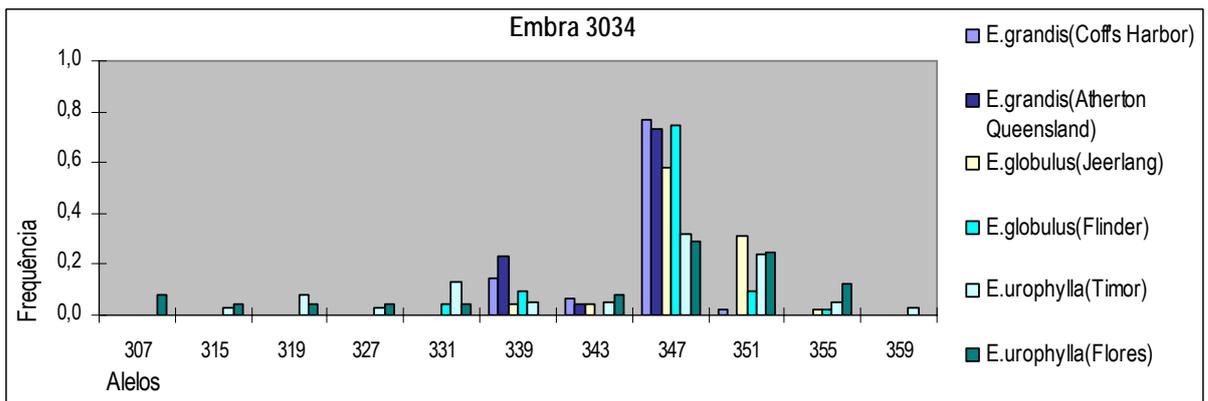
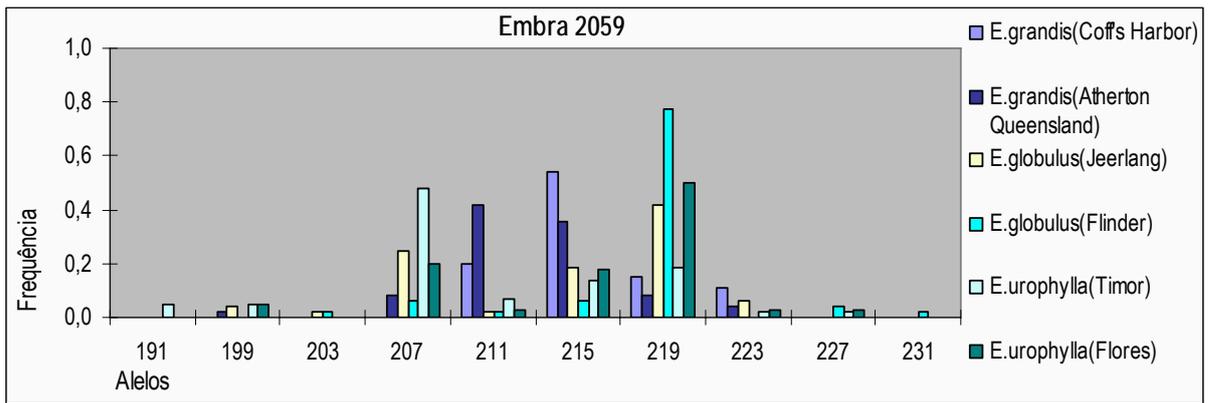




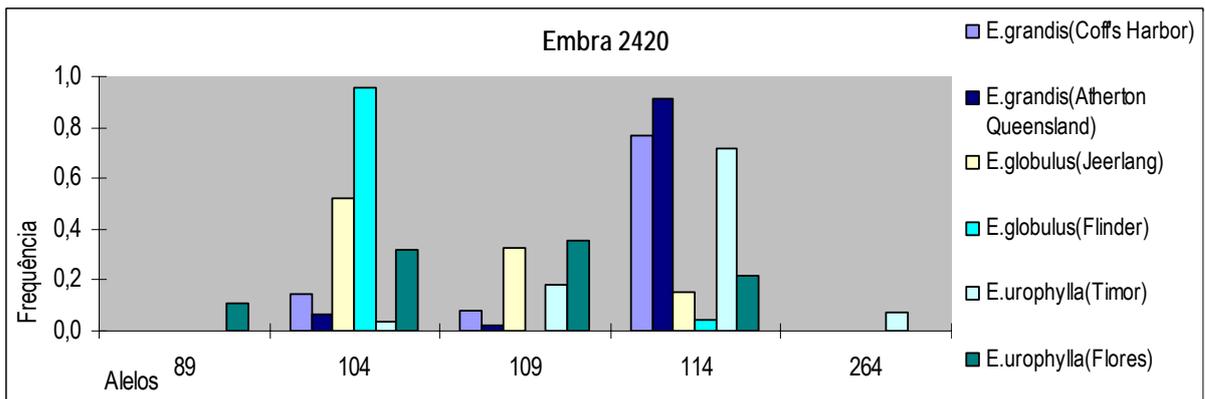
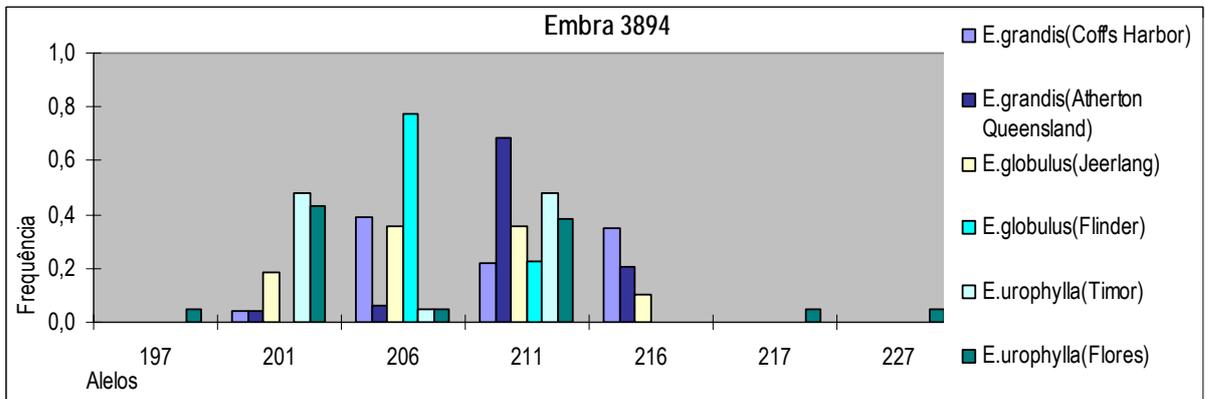
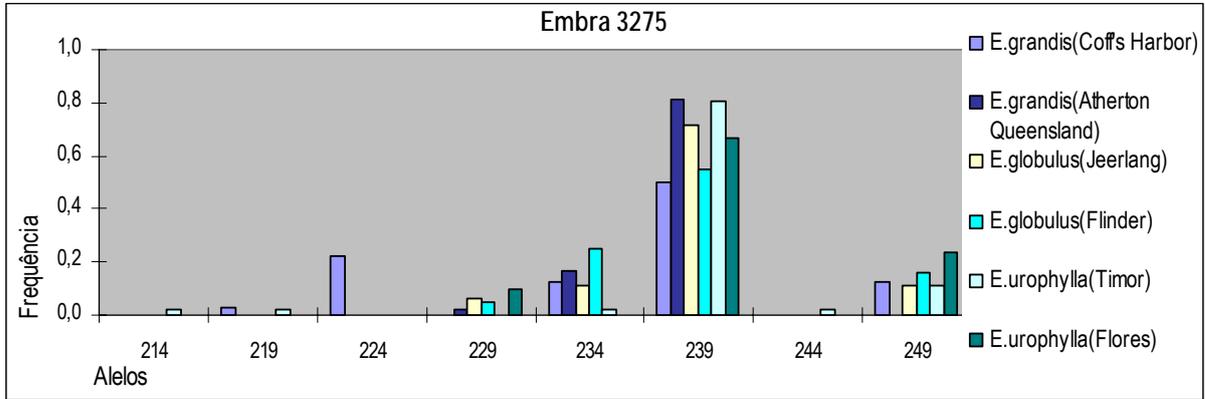


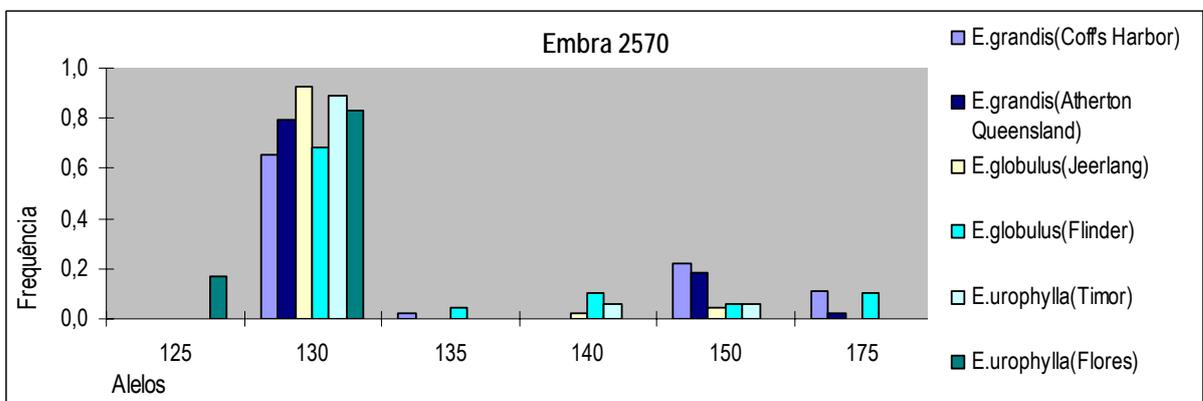
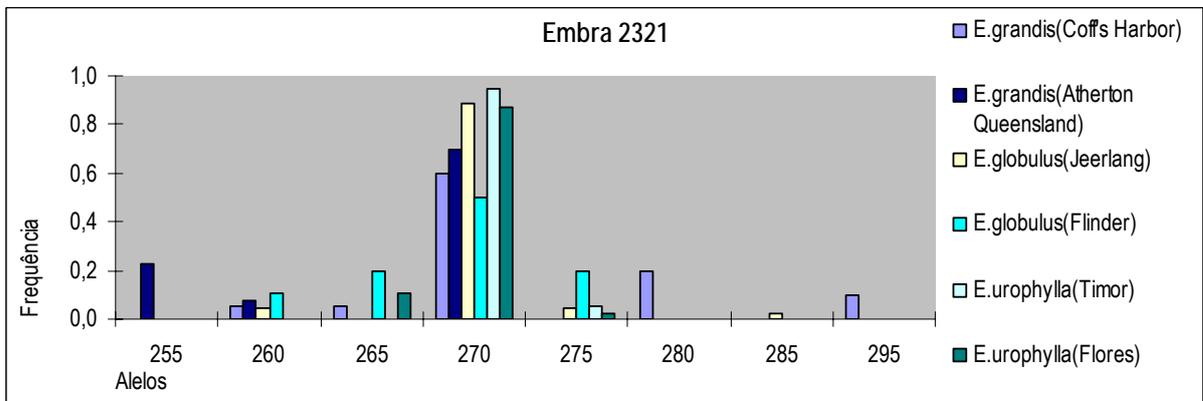
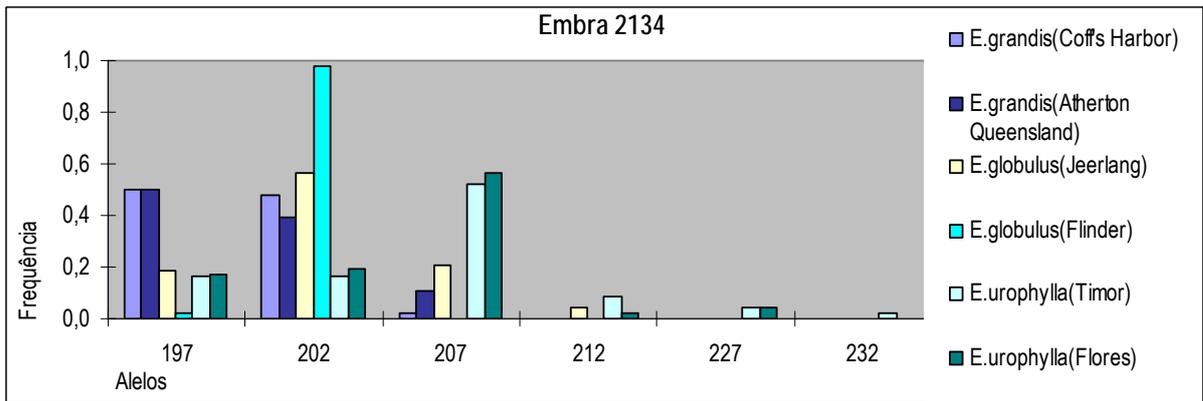
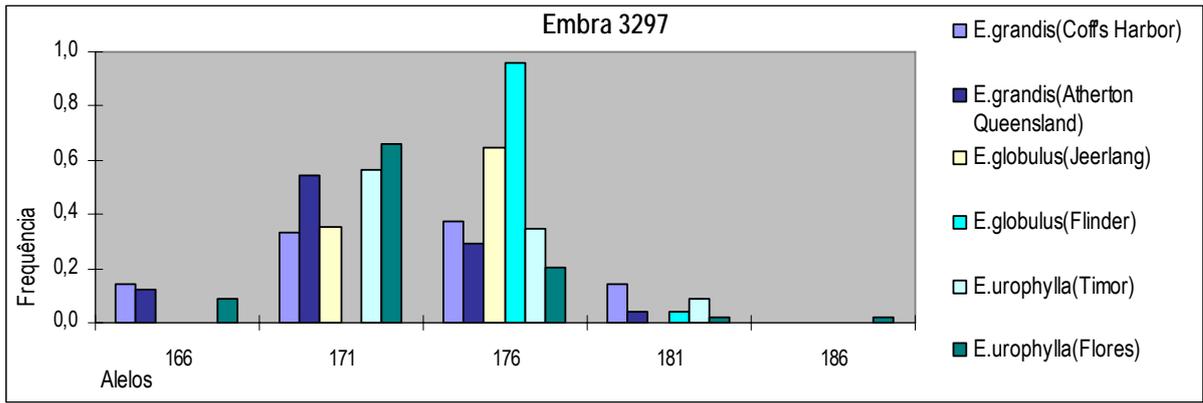


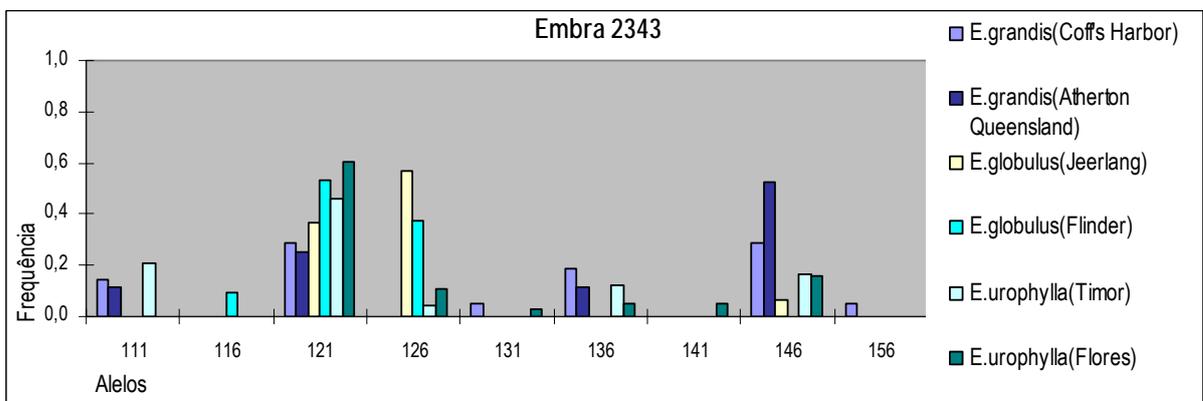
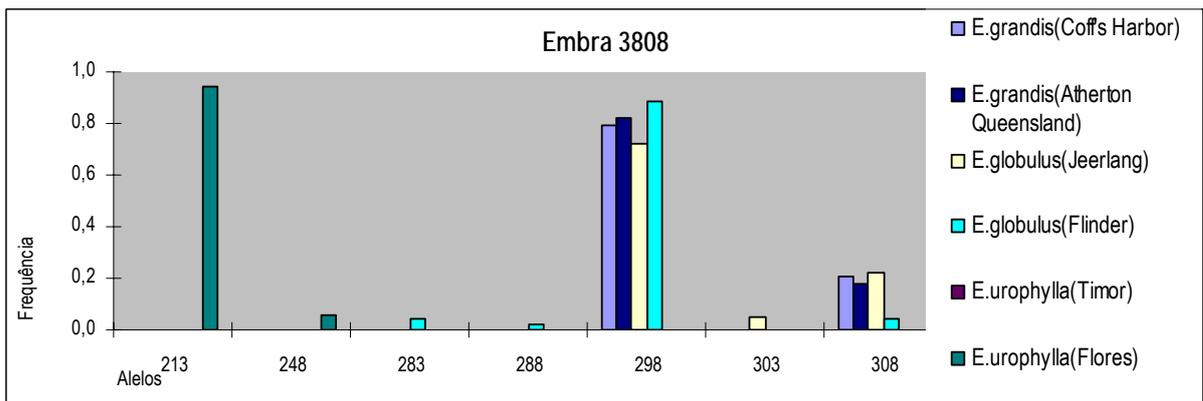
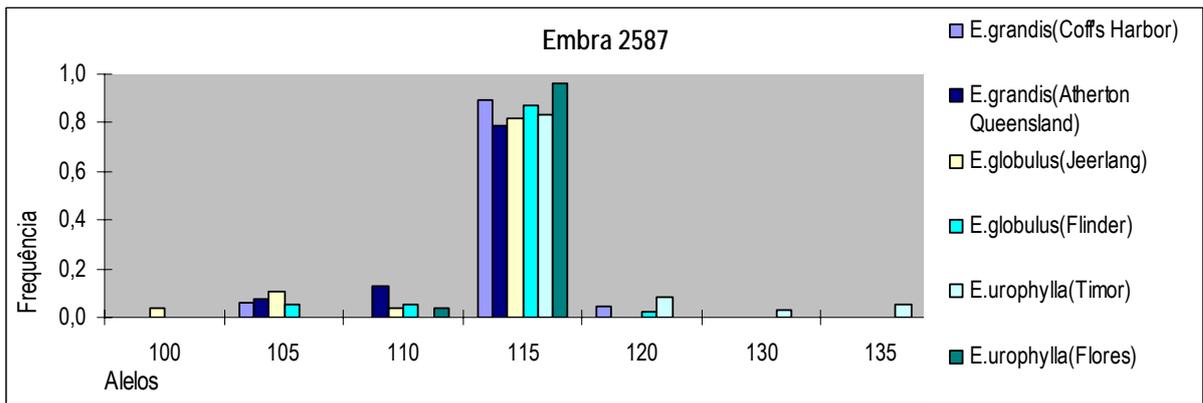




Anexo 4. Histogramas de frequências alélicas dos 10 locos pentanucleotídeos selecionados para caracterização genética.







Anexo 5: Frequências alélicas os locos tetranucleotídeos nas duas procedências de cada uma das espécies *E. grandis* (A), *E. globulus* (B) e *E. urophylla* (C). N: Numero de indivíduos analisados; Na: número de alelos; Ho: Heterozigotidade observada; He: heterozigotidade esperada, p valor EHW: p valor do teste de aderência ao equilíbrio de Hardy Weinberg

A

<i>E. grandis</i>										
Coffs Harbor						Atherton				
Loco	N	Na	Ho	He	p valor EHW	N	Na	Ho	He	p valor EHW
embra 2113	22	5	0,50	0,50	0,6707	23	4	0,61	0,50	0,8224
embra 2675	22	2	0,73	0,48	0,0180	23	3	0,74	0,53	0,1194
embra 3513	11	3	0,00	0,66	0,0001	13	3	0,15	0,65	0,0012
embra 3086	19	3	0,26	0,38	0,0002	21	3	0,33	0,58	0,0100
embra 2549	18	3	0,44	0,37	0,6894	24	4	0,29	0,26	0,9945
embra 2052	22	5	0,59	0,63	0,7158	24	8	0,75	0,73	0,0907
embra 2086	24	4	0,58	0,67	0,8343	24	3	0,50	0,54	0,6404
embra 3490	23	5	0,17	0,43	0,0000	24	5	0,42	0,54	0,0001
embra 3595	24	5	0,71	0,64	0,1938	24	5	0,75	0,62	0,4202
embra 3196	23	2	0,04	0,04	0,9151	20	1	0,00	0,00	monomórfico
embra 3435	24	1	0,00	0,00	monomórfico	24	1	0,00	0,00	monomórfico
embra 3769	21	6	0,62	0,80	0,0797	21	8	0,71	0,78	0,0061
embra 2764	23	3	0,61	0,44	0,2212	23	3	0,57	0,46	0,0000
embra 2339	23	4	0,57	0,47	0,9124	24	3	0,67	0,54	0,2474
embra 2965	23	2	0,43	0,42	0,8978	24	4	0,17	0,26	0,0993
embra 2081	24	3	0,13	0,12	0,9910	23	2	0,13	0,12	0,7379
embra 3185	24	3	0,38	0,35	0,6998	21	2	0,19	0,31	0,0797
embra 2527	21	6	0,38	0,52	0,0012	18	6	0,50	0,65	0,0306
embra 2059	23	4	0,35	0,63	0,0056	24	6	0,79	0,68	0,8437
embra 3034	24	4	0,29	0,38	0,0959	24	3	0,38	0,41	0,7364
embra 3543	23	1	0,00	0,00	monomórfico	24	1	0,00	0,00	monomórfico
embra 2578	16	3	0,13	0,50	0,0011	24	2	0,08	0,22	0,0024

B

<i>E. globulus</i>										
Jeerlang						Flinder				
Loco	N	Na	Ho	He	p valor EHW	N	Na	Ho	He	p valor EHW
embra 2113	24	5	0,21	0,44	0,0000	23	3	0,39	0,39	0,0125
embra 2675	22	4	0,55	0,61	0,1082	21	3	0,43	0,41	0,0169
embra 3513	15	3	0,27	0,62	0,0264	12	3	0,08	0,59	0,0002
embra 3086	23	4	0,52	0,61	0,0003	22	2	0,59	0,50	0,3874
embra 2549	24	3	0,25	0,22	0,9211	22	2	0,50	0,42	0,3458
embra 2052	24	7	0,58	0,61	0,9863	23	3	0,52	0,49	0,8705
embra 2086	24	5	0,33	0,62	0,0000	24	3	0,29	0,32	0,8075
embra 3490	23	6	0,17	0,48	0,0001	23	3	0,17	0,23	0,0000
embra 3595	23	6	0,57	0,68	0,0119	24	4	0,29	0,36	0,6260
embra 3196	15	2	0,07	0,06	0,8938	17	2	0,18	0,25	0,2214
embra 3435	23	2	0,04	0,04	0,9151	24	1	0,00	0,00	monomórfico
embra 3769	19	7	0,58	0,72	0,0064	23	10	0,87	0,77	0,0529
embra 2764	22	4	0,73	0,52	0,3042	24	3	0,71	0,55	0,0494
embra 2339	22	6	0,59	0,61	0,2374	21	3	0,38	0,33	0,7620
embra 2965	22	4	0,23	0,61	0,0015	22	1	0,00	0,00	monomórfico
embra 2081	22	7	0,59	0,65	0,0747	22	4	0,36	0,38	0,0007
embra 3185	22	3	0,27	0,56	0,0098	22	2	0,05	0,04	0,9131
embra 2527	16	5	0,25	0,55	0,2640	16	2	0,13	0,12	0,7897
embra 2059	24	7	0,63	0,72	0,7646	24	7	0,33	0,39	0,0127
embra 3034	24	5	0,42	0,56	0,0022	22	5	0,45	0,42	0,8577
embra 3543	18	4	0,17	0,44	0,0000	23	2	0,57	0,50	0,5049
embra 2578	18	2	0,11	0,44	0,0015	18	1	0,00	0,00	monomórfico

<i>E. urophylla</i>										
Timor						Flores				
Loco	N	Na	Ho	He	p valor EHW	N	Na	Ho	He	p valor EHW
embra 2113	19	3	0,11	0,19	0,0511	15	5	0,27	0,44	0,2843
embra 2675	14	2	0,14	0,49	0,0080	14	4	0,00	0,65	0,0000
embra 3513	16	3	0,13	0,31	0,0250	15	1	0,00	0,00	monomórfico
embra 3086	18	6	0,50	0,68	0,1139	18	5	0,17	0,65	0,0001
embra 2549	24	1	0,00	0,00	monomórfico	24	1	0,00	0,00	monomórfico
embra 2052	18	5	0,39	0,37	0,0001	19	5	0,26	0,46	0,0000
embra 2086	24	4	0,46	0,38	0,9082	22	2	0,55	0,48	0,5476
embra 3490	12	7	0,25	0,80	0,0011	18	5	0,11	0,41	0,0001
embra 3595	22	6	0,50	0,50	0,6918	22	5	0,41	0,57	0,7500
embra 3196	13	4	0,31	0,28	0,9986	21	3	0,19	0,18	0,9721
embra 3435	23	2	0,09	0,08	0,8274	22	1	0,00	0,00	monomórfico
embra 3769	22	5	0,45	0,71	0,0743	18	4	0,44	0,57	0,0024
embra 2764	24	4	0,13	0,65	0,0000	24	4	0,17	0,61	0,0000
embra 2339	21	6	0,52	0,68	0,4251	22	5	0,64	0,63	0,3141
embra 2965	23	3	0,35	0,58	0,0740	24	5	0,17	0,67	0,0000
embra 2081	21	6	0,24	0,65	0,0000	23	5	0,17	0,69	0,0000
embra 3185	21	3	0,33	0,61	0,0209	22	4	0,45	0,60	0,1542
embra 2527	22	4	0,14	0,21	0,1650	18	6	0,56	0,49	0,9993
embra 2059	22	8	0,55	0,71	0,0186	20	7	0,55	0,68	0,8313
embra 3034	19	10	0,68	0,81	0,2619	12	9	0,75	0,82	0,4585
embra 3543	14	5	0,36	0,46	0,1619	13	3	0,23	0,21	0,9741
embra 2578	5	3	0,40	0,34	0,9577	13	3	0,00	0,27	0,0000

Anexo 6. Caracterização de locos pentanucleotídeos de duas procedências de cada uma das espécies *E. grandis* (A), *E. globulus* (B) e *E. urophylla* (C). N: Numero de indivíduos analisados; Na: número de alelos; Ho: Heterozigosidade observada; He: heterozigosidade esperada, p valor EHW: p valor do teste de aderência ao equilíbrio de Hardy Weinberg

A										
<i>E. grandis</i>										
Coffs Harbor						Atherton				
Loco	N	Na	Ho	He	p valor EHW	N	Na	Ho	He	p valor EHW
embra 3275	20	5	0,70	0,67	0,4312	21	3	0,38	0,32	0,7620
embra 3894	23	4	0,52	0,68	0,3473	24	4	0,33	0,48	0,0724
embra 2420	24	3	0,21	0,38	0,0131	23	3	0,09	0,16	0,0229
embra 3297	24	4	0,79	0,71	0,0741	24	4	0,42	0,60	0,0002
embra 2134	24	3	1,00	0,52	0,0000	24	3	1,00	0,58	0,0000
embra 2321	20	5	0,00	0,59	0,0000	20	3	0,25	0,45	0,0549
embra 2570	23	4	0,26	0,52	0,0201	24	3	0,17	0,34	0,0383
embra 2587	24	3	0,13	0,19	0,0000	19	3	0,32	0,35	0,5164
embra 3808	24	2	0,25	0,33	0,2356	22	2	0,09	0,30	0,0011
embra 2343	21	6	0,57	0,78	0,0011	22	4	0,50	0,64	0,0001

B

<i>E. globulus</i>										
Jeerlang						Flinder				
Loco	N	Na	Ho	He	p valor EHW	N	Na	Ho	He	p valor EHW
embra 3275	23	4,000	0,391	0,457	0,0429	22	4,000	0,864	0,613	0,1375
embra 3894	24	4,000	0,667	0,703	0,7239	24	2,000	0,292	0,353	0,3928
embra 2420	23	3,000	0,348	0,598	0,0201	24	2,000	0,083	0,080	0,8313
embra 3297	24	2,000	0,208	0,457	0,0076	24	2,000	0,000	0,080	0,0000
embra 2134	24	4,000	0,458	0,603	0,0000	24	2,000	0,042	0,041	0,9170
embra 2321	22	4,000	0,136	0,210	0,0011	23	4,000	0,348	0,662	0,0003
embra 2570	21	3,000	0,143	0,135	0,9888	24	5,000	0,458	0,500	0,0014
embra 2587	14	4,000	0,286	0,311	0,0001	20	4,000	0,150	0,229	0,0026
embra 3808	20	3,000	0,250	0,421	0,0001	22	4,000	0,136	0,210	0,0011
embra 2343	15	3,000	0,200	0,540	0,0080	16	3,000	0,250	0,568	0,0099

C

<i>E. urophylla</i>										
Timor						Flores				
Loco	N	Na	Ho	He	p valor EHW	N	Na	Ho	He	p valor EHW
embra 3275	23	6	0,30	0,34	0,0000	21	3	0,24	0,49	0,0185
embra 3894	21	3	0,48	0,54	0,4352	22	6	0,41	0,66	0,0000
embra 2420	14	4	0,07	0,45	0,0001	14	4	0,21	0,71	0,0006
embra 3297	23	3	0,61	0,55	0,8475	22	5	0,41	0,51	0,4960
embra 2134	24	6	0,33	0,66	0,0000	23	5	0,35	0,61	0,0019
embra 2321	19	2	0,00	0,10	0,0000	19	3	0,05	0,23	0,0003
embra 2570	18	3	0,00	0,20	0,0000	15	2	0,07	0,28	0,0032
embra 2587	18	4	0,11	0,29	0,0003	13	2	0,08	0,07	0,8853
embra 3808	0	0	0,00	0,00	monomórfico	9	2	0,11	0,10	0,8599
embra 2343	12	5	0,83	0,70	0,4645	19	6	0,37	0,59	0,0005

9. Referências bibliográficas

- Akkaya MS, Bhagwat AA e Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics*, 132: 1131-1139
- Birnboim HC, Straus NA. (1975) DNA from eukaryotic cells contains unusually long pyrimidine sequences. *Can J Biochem*. 53(5):640-643.
- Bjornstad G, Gumbo E, Roed KH (2000): Genetic structure of Norwegian horse breeds. *Journal of animal breeding genetics*, 117: 307-317.
- Brody JR, Kem SE, 2004. Sodium boric acid: a Tris free cooler conductive medium for DNA electrophoresis. *Biotechnique* 36 (2): 214-6
- Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998) Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theor Appl Genet* 97:816-827
- Brondani RPV, Brondani C, Grattapaglia D (2002) Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. *Mol Genet Genomics* 267:338-347
- Brondani RPV, Williams ER, Brondani C, Grattapaglia D (2006) A high coverage microsatellite-only consensus linkage map for *Eucalyptus* species and a novel set of 230 microsatellite markers for the genus. *BMC Genetics* (submetido).
- Butler, J.M. (2005a) *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers* (2nd Edition). Elsevier Academic Press, New York, 688 pp.

- Butler, J.M. (2005b) Constructing STR *multiplex* assays. *Methods in Molecular Biology: Forensic DNA Typing Protocols* (Carracedo, A., ed.), Humana Press: Totowa, New Jersey, 297: 53-66.
- Byrne M, Marquezgarcia MI, Uren T, Smith DS, Moran GF (1996) Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. *Aust J Bot* 44:331-341
- Cañon J, Checa ML, Carleos C, Veja-Pla JL, Vallejo M e Dunner S (2000): The genetic structure of spanish celtic horse breeds inferred from microsatellite data. *Animal genetics* 31: 39-48.
- Chagné D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio JM, Echt C, Richardson T, Plomion C. (2004) Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theor Appl Genet.* 109(6):1204-14.
- Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N e Cartinhours S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequence in rice (*Oryza sativa L.*) *Theor. Appl. Genet.* 100: 713-722.
- Collins FS, Guyer MS e Chakravarti A (1997): Variations on a theme cataloguing human DNA sequence variation. *Science* 278: 1580-1581.
- Costa e Silva C, Grattapaglia D (1997) RAPD relatedness of elite clones, applications in breeding and operational clonal forestry. In *Proceedings of the International IUFRO Conference on Eucalyptus Genetics and Silviculture, Salvador, Brazil Vol. 2*, 161-166.
- Cunningham EP, Dooley JJ, Splan RK, Bradley DG (2001): Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to Sereno, 2002).

- de Assis TF (2000) Production and use of *Eucalyptus* hybrids for industrial purposes. In: Nikles DG (ed) Hybrid Breeding and Genetics of Forest Trees Proceedings of QFRI/CRC-SPF Symposium, 9-14th April, Noosa. Department of Primary Industries, Brisbane, pp 63-74.
- Doughty RW (2000) The *Eucalyptus*. A natural and commercial history of the gum tree. The Johns Hopkins University Press, Baltimore and London.
- Eldridge K, Davidson J, Harwood C, van Wyk G (1994) Eucalypt domestication and breeding. Clarendon Press, Oxford.
- FAO (2000) Global forest resources assessment 2000 - Main report.: FAO Forestry paper <http://www.fao.org/forestry/fo/fra/main/index.jsp>.
- Ferreira M.E. e Gattapaglia D. (1998). Introducción al uso de marcadores moleculares em el análisis genético. 1 ed. Brasilia: EMBRAPA CENARGEN, p. 220 (EMBRAPA CENARGEN Documento 20).
- Fregeau, CJ, Fournay RM (1993) DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. *BioTechniques* 15: 100-119.
- Gaiotto FA, Bramucci M, Grattapaglia D (1997) Estimation of outcrossing rate in a breeding population of *Eucalyptus urophylla* s.t. Blake with dominant RAPD and AFLP markers, *Theor Appl Genet* 95:842-849.
- Glaubitz JC, Emebiri LC, Moran GF (2001) Dinucleotide microsatellite from *Eucalyptus sieberi*. Inheritance, diversity, and improved scoring of single-base differences. *Genome* 44:1014-1045.
- Grattapaglia D, Bradshaw HJ (1994) Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. *Canadian Journal of Forest Research* 24(5): 1074-1078.

- Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137:1121-1137.
- Grattapaglia D. (2000) Molecular breeding of *Eucalyptus* - State of the art, operational applications and technical challenges. In: Jain SM, Minocha SC, eds. *Molecular biology of woody plants, Forestry Sciences, Volume 64*: Kluwer Academic Publishers, The Netherlands pp. 451-474.
- Grattapaglia D, Diener PSA, Rezende GD, Pimenta D, Lins TCL, Soares CN (2003a) Aplicações operacionais da análise de vínculo genético com microssatélites em populações de melhoramento de *Eucalyptus*. Resumos do 49º Congresso Brasileiro de Genética, poster GP085.
- Grattapaglia D, Pimenta D, Campinhos EN, Rezende GDS, Assis TF (2003b) Marcadores moleculares na proteção varietal de *Eucalyptus*. Anais do 8º Congresso Florestal Brasileiro pp. 1-13 Publicado em CD.
- Grattapaglia, D (2004) Integrating genomics into *Eucalyptus* breeding. *Gen. Mol. Res.* 3: 369-379.
- Grattapaglia D, Ribeiro VJ, Rezende GD. (2004) Retrospective selection of elite parent trees using paternity testing with microsatellite markers: an alternative short term breeding tactic for *Eucalyptus*. *Theor Appl Genet* 109:192-199.
- Grattapaglia, D (2006) Marker assisted selection in *Eucalyptus*. In: Guimarães E (Ed.). *Marker-Assisted Selection (MAS) in Crops, Livestock, Forestry and Fish: Current Status and the Way Forward*. FAO Food and Agricultural Research Organization FAO, Roma (no prelo)

- Halverson J, Basten C (2005) A PCR *multiplex* and database for forensic DNA identification of dogs. J Forensic Sci. 2005 Mar;50(2):352-63.
- Hamada H, Petrino MG, Kakunaga T (1982) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc. Nat. Acad. USA 79:6465-6469.
- Henegariu O, Heerema NA, Dlouhy SR, Vance GH, Vogt PH (1997) *Multiplex* PCR: critical parameters and step-by-step protocol. BioTechniques 23, 504-511.
- Idury RM, Cardon LR, (1997) A simple method for automated allele binning in microsatellite markers. Genome Research v. 7 p1104-1109.
- ISAG International Society for Animal Genetics (Sociedade Animal de Genética animal)
<http://www.isag.org.uk>
- Keil M and Griffin AR, (1994) Use of random amplified polymorphic DNA (RAPD) markers in the discrimination and verification of genotypes in *Eucalyptus*. Theor Appl Genet 89:442-450.
- Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M. (1993) Automated DNA profiling employing *multiplex* amplification of short tandem repeat loci. PCR Methods and Applications 3:13-22.
- Kirst M, Cordeiro CM, Rezende GD, Grattapaglia D. (2005) Power of microsatellite markers for fingerprinting and parentage analysis in *Eucalyptus grandis* breeding populations. J Hered. 96:161-166.
- Legercrantz U, Ellegren H, e Anderson L. (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res.* 21: 1111-1115.

- Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 44: 397-401.
- Mamani, EM, Novaes E, Grattapaglia, D (2004) Mapeamento genético em *Eucalyptus grandis* x *E. urophylla* e extensão do mapa referência com novos microssatélites. Resumos do 50º Congresso Brasileiro de Genética, Florianópolis SC, resumo 1303.
- Marcucci Poltri SN, Zelener N, Rodriguez Traverso J, Gelid P, Hopp HE (2003). Selection of a seed orchard of *Eucalyptus dunnii* based on genetic diversity criteria calculated using molecular markers. *Tree Physiol* 23:625-32.
- Matsuoka Y, Mitchell SE, Kresovich S, Goodman M, Doebley J. (2002) Microsatellites in Zea - variability, patterns of mutations, and use for evolutionary studies. *Theor Appl Genet.* 104(2-3):436-450.
- Menotti-Raymond MA, David VA, Wachter LL, Butler JM, O'Brien SJ. (2005) An STR forensic typing system for genetic individualization of domestic cat (*Felis catus*) samples. *J Forensic Sci.* 50(5):1061-1070.
- Missiaggia AA, Grattapaglia D (2005) Genetic mapping of *Eef1*, a major effect QTL for early flowering in *Eucalyptus grandis*. *Tree Genet and Genomes* 1:79-84.
- Mora A, Garcia C (2000) A Cultura do Eucalipto no Brasil. 1.ed. São Paulo, Sociedade Brasileira de Silvicultura.
- Morgante M e Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. *Plant J.* 3: 175-182

- Morgante M, Hanafey M, Powell W. (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat Genet.* 30(2):194-200.
- Myburg A, Grattapaglia D, Kirst M, Gion J, Marques C, Grima-Pettenati J, Potts B (2006) *Eucalyptus* genome mapping and molecular breeding. In: Chitta R. Kole. (Org.). Genome Mapping & Molecular Breeding. Vol. 6: Forest Trees. Springer Publ. Heidelberg, Tokyo, New York, 2006.
- Nesbitt KA, Potts BM, Vaillancourt RE, West AK, and Reid JB, (1995) Partitioning and distribution of RAPD variation in a forest tree species, *Eucalyptus globulus* (Myrtaceae). *Heredity* 74:628-637.
- Nesbitt, KA, Potts BM, Vaillancourt RE, Reid JB (1997) Fingerprinting and pedigree analysis in *Eucalyptus globulus* using RAPDs. *Silvae Genetica* 46: 6-11.
- Nicholas, FW (1999): Introdução à genética veterinária; Trad. Rivo Fisher – Porot Alegre: Editora Artes Médicas Sul Ltda.
- Ottewell KM, Donnellan SC, Moran GF, Paton DC (2005) *Multiplexed* microsatellite markers for the genetic analysis of *Eucalyptus leucoxylon*, Myrtaceae and their utility for ecological and breeding studies in other *Eucalyptus* species. *J. Hered.* 96: 445-451.
- Peakall R, Smouse PE, (2006) GENALEX 6: Genetics analysis in Excel. Population genetics software for teaching and research. *Molecular Ecology Notes* 6: 266-295.
- Potts BM (2004) Genetic improvement of eucalypts. In: Burley J, Evans, J and Youngquist JA (ed) *Encyclopedia of Forest Science*. Elsevier Science, Oxford, pp 1480-1490
- Ramalho, M.A.P.; Santos, J.B.; Pinto, C.A.B.P. (2000). *Genética na Agropecuária*. 2ª Edição. Lavras: Editora UFLA, p. 427.

[Redd AJ](#), [Agellon AB](#), [Kearney VA](#), [Contreras VA](#), [Karafet T](#), [Park H](#), [de Knijff P](#), [Butler JM](#),
[Hammer MF](#).(2002). "Forensic value of 14 novel STRs on the human Y chromosome,"

Forensic science international 130, no. 2-3

Sarkar N, Kashyap VK, (2002) "Genetic diversity at two pentanucleotide STR and thirteen tetranucleotide STR loci by multiplex PCR in four predominant population groups of central India," *Forensic science international* 128, no. 3

Scotti I, Magni F, Paglia GP, Morgante M. (2002) Trinucleotide microsatellites in Norway spruce (*Picea abies*): their features and the development of molecular markers. *Theor Appl Genet.* 106(1):40-50.

Shuber, A.P., Grondin, V.J., Klinger, K.W. (1995) A simplified procedure for developing *multiplex* PCRs. *Genome Res.* 5, 488-493.

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB. (2004) A new integrated genetic linkage map of the soybean. *Theor Appl Genet.* 109(1):122-8.

Sreenu VB, Alevoor V, Nagaraju J, Nagarajaram HA. (2003) MICdb: database of prokaryotic microsatellites. *Nucleic Acids Res.* 31(1):106-8.

Steane DA, Vaillancourt RE, Russell J, Powell W, Marshall D, Potts BM (2001) Development and characterization microsatellite loci in *Eucalyptus globulus* (Myrtaceae). *Silvae Genet* 50(2):89

Steane DA, Conod N, Jones RC, Vaillancourt RE, Potts BM (2006) A comparative analysis of population structure of a forest tree, *Eucalyptus globulus* (Myrtaceae), using

microsatellite markers and quantitative traits *Tree Genetics and Genomes* online first
DOI: 10.1007/s11295-005-0028-7

Stone RT, Pulido JC, Duyk GM, Kappes SM, Keele JW, Beattie CW. (1995) A small-insert bovine genomic library highly enriched for microsatellite repeat sequences. *Mamm Genome*. 6(10):714-24.

Tandon M, Trivedi R, Kashyap V K. (2002) "Genomic diversity at 15 fluorescent labeled short tandem repeat loci in few important populations of State of Uttar Pradesh, India," *Forensic science international* 128, no. 3

Tang S, Kishore VK, Knapp SJ. (2003) PCR-*multiplexes* for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower. *Theor Appl Genet*. 107(1):6-19.

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res*. 12:4127-4138.

Tautz D (1989) Hypervariability of simple sequence as a general resource for polymorphic DNA markers. *Nucleic Acids Res*. 17: 6463-6471.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S e McCouch S (2001) Computational and Experimental Analysis of Microsatellites in Rice (*Oryza sativa* L.): frequency, Length Variation, Transposon Associations, and Genetics Marker Potential. *Genome Res*. 11: 1441-1452

Tereba A (1999) Tools for analysis of population statistics. *Profiles in DNA* 203:14-16.

Urquhart A, Oldroyd NJM, Kimpton CP and Gill P (1995) Highly discriminating heptaplex short tandem repeat PCR system for forensic identification. *BioTechniques* 18: 116-121.

Vallone PM, Butler JM. (2004) AutoDimer: a screening tool for *primer-dimer* and hairpin structures. *Biotechniques* 37(2):226-231.

Wang Z, Weber JL, Zhong G e Tanksley SD (1994) Survey of plant short tandem DNA repeat. *Theor. Appl. Genet.* 88: 1-6

Waugh G (2004) Growing *Eucalyptus globulus* for high-quality sawn products. In: Tomé M (ed) IUFRO Conference: *Eucalyptus* in a changing world, 11-15 October 2004. RAIZ, Instituto Investigação de Floresta e Papel, Aveiro, Portugal, pp 79-84

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the Polymerase Chain Reaction. *Am. J. Hum. Genet.* 44:388-396.

[Yi G](#), [Lee JM](#), [Lee S](#), [Choi D](#), [Kim BD](#) (2006) Exploitation of pepper EST-SSRs and an SSR-based linkage map. *Theor Appl Genet.* 114(1):113-30.

Yu J, Mangor J, Thompson L, Edwards KJ, Slabaugh MB e Knapp SJ (2002) Allelic diversity of simple sequence repeats among elite inbred lines of cultivated sunflower. *Genome* 45: 652-660.