

**JORGE LUIZ OLIVEIRA DE CASTRO**

**DESEMPENHO FORENSE DE MICROSSATÉLITES PARA A  
INVESTIGAÇÃO DA ORIGEM DE *Cannabis sativa*  
NO BRASIL E PARAGUAI**

Dissertação apresentada ao Programa de Pós-Graduação “*Stricto Sensu*” em Ciências Genômicas e Biotecnologia, da Universidade Católica de Brasília, como requisito para a obtenção do título de Mestre em Ciências Genômicas e Biotecnologia.

**Orientador: Dr. Rinaldo Wellerson Pereira**

**Co-orientador: Dr. Dario Grattapaglia**

**Brasília  
Distrito Federal  
2006**

C355d Castro, Jorge Luiz Oliveira de  
Desempenho forense de microssatélites para a investigação da origem de  
*Cannabis sativa* no Brasil e no Paraguai /Jorge Luiz Oliveira de Castro. – 2006.  
78 f. : il. ; 30 cm.

Dissertação (mestrado) – Universidade Católica de Brasília, 2006.  
Orientação: Rinaldo Wellerson Pereira

1. Maconha. 2. Biotecnologia. 3. Tráfico de drogas. 4. Microssatélites – Brasil -  
Paraguai. I. Pereira, Rinaldo Wellerson, orient. II. Título.

CDU 615.32

## DEDICATÓRIA

Ao Deus Soberano sobre todas as coisas, ao qual agradeço todos os dias por estar sempre comigo.

Ao meu filho Emanuel Erico (*in memoriam*) que apesar da pouca idade, muitas vezes suportou a minha ausência com bravura e compreensão, para que este trabalho chegasse a termo.

À minha amada esposa, grande incentivadora, pelo apoio durante toda minha formação profissional.

## AGRADECIMENTOS

Ao Deus Eterno, soberano e imortal, a Ele minha gratidão, louvor e adoração.

Ao Prof. Dr. Rinaldo W. Pereira por aceitar orientar-me, pela atenção, suporte e paciência dispensados durante todas as fases da dissertação, meu muito OBRIGADO.

Ao Prof. Dr. Dario Grattapaglia um agradecimento especial, pois com seu entusiasmo peculiar, ainda nas aulas de Genética e Evolução, onde foi concebido o projeto, ofereceu incentivo e apoio sem os quais este trabalho não teria sucesso.

A todos os membros e colegas do Laboratório de Biotecnologia da UCB, pelo apoio técnico e amizade durante o trabalho, em especial aos colegas de bancada: Suzana Neiva e Alexandre Póvoa.

Ao Dr Anísio Soares Vieira, Coordenador Geral de Prevenção e Repressão a Entorpecentes do Departamento de Polícia Federal (CGPRE/DPF), pelo interesse no projeto e apoio fundamental na obtenção das amostras.

Aos colegas do Instituto Nacional de Criminalística, em especial ao seu Diretor, o Dr. Octávio Brandão Caldas Neto, grande amigo e apoiador incondicional.

À Universidade Católica de Brasília e ao Instituto Nacional de Criminalística do Departamento de Polícia Federal, que tornaram possível o desenvolvimento deste estudo.

## RESUMO

O mercado ilícito de *Cannabis sativa* (maconha) no Brasil é basicamente suprido por plantios ilegais do Paraguai, norte e nordeste do Brasil. Sem um sistema de inteligência bem estabelecido é difícil determinar a origem correta de pequenas e grandes apreensões. Qualquer nova tecnologia que ajude nesta tarefa tem um valor inestimável no combate ao comércio ilícito de maconha. Tecnologia baseada em polimorfismos de DNA tem potencial para realizar um importante papel como ferramenta de inteligência para estabelecer o relacionamento entre apreensões e a origem da droga. Este potencial é maximizado pelo desenvolvimento recente de microssatélites em *C. sativa*, que são marcadores informativos para estudos de genética, sendo especialmente úteis na obtenção de dados de diversidade e de estrutura genética de populações naturais ou cultivadas. O trabalho apresentado aqui mostra o desenvolvimento de duas reações PCR-multiplex que amplificam 11 microssatélites, genotipados em uma única corrida eletroforética. Este sistema multiplex foi utilizado para genotipar 477 plantas de 12 áreas de cultivo no Brasil e no Paraguai. A análise dos dados mostrou que as plantas das 12 áreas cultivadas podem ser estruturadas em 5 grupos que representam áreas do estado de Pernambuco no Brasil, duas áreas no estado do Maranhão, uma área no Paraguai e um grupo representando cinco áreas no Paraguai. O excesso de homozigotos foi encontrado em diversos locos e as análises sugerem fortemente a presença de alelos nulos como o fator principal para a obtenção deste resultado. Os testes de designação aplicados às plantas agrupadas como brasileiras ou paraguaias mostraram que usando as frequências alélicas dos 11 locos 94% das 477 plantas foram designadas corretamente. Este valor cai para 90% quando as plantas são agrupadas em cinco grupos. Os resultados mostraram-se promissores no sentido de se continuar investigando novos microssatélites e se iniciar a amostragem de apreensões. Até onde sabemos este é o primeiro trabalho que investiga formalmente o potencial de microssatélites em *C. sativa* com uma perspectiva forense que visa aplicação prática.

**PALAVRAS-CHAVE:** *Cannabis sativa*; Origem geográfica, Microssatélites, Relacionamento genético, Combate ao narcotráfico.

## ABSTRACT

The black market of *Cannabis sativa* (marijuana) in Brazil is basically supplied by illegal plantations from Paraguay, North and Northeastern Brazil. Without a well established intelligence service is difficult to state the right origin of large and small apprehensions. Any new technology that helps this task will be of inestimable value for the war against the illegal marijuana dealing. Technology based on DNA variation could play an important role as an intelligence tool to establish relationship between apprehensions and drug origin. This potentiality is maximized by the recent development of *C. sativa* microsatellites, which are informative markers for genetics studies, being especially useful in obtaining diversity and genetic structure data of natural or cultivated populations. The work presented here shows the development of two PCR-multiplex reactions that amplify 11 microsatellites, genotyped in only one electrophoresis run. These multiplex system where used to genotype 477 plants from 12 cultivated areas from Brazil and Paraguay. The data analysis showed that plants from the 12 cultivated areas can be structured in 5 groups representing areas from Pernambuco state in Brazil, two areas in Maranhão state, one area in Paraguay and one group representing five areas in Paraguay. Excess of homozygous genotypes was found in several loci and analyses strongly support null alleles as the main explanation. Assignment tests applied to plants stratified as Brazilians or Paraguayans showed that using allele frequencies from the 11 loci 94% of the 477 plants were correctly assigned. This value drops to 90 % when plants are stratified in five groups. The results showed to be very promising to continue investigating new microsatellites and start moving to samples from apprehensions. As far as we are aware this is the first work that investigates formally the use of *C. sativa* microsatellites with a forensic perspective that envisages practical applications.

**KEYWORDS:** *Cannabis sativa*; Geographical origin, Microsatellites, Genetic relationship; Combat to the drug traffic.

# SUMÁRIO

<b>INTRODUÇÃO</b> .....	9
<b>1.1 A CANNABIS SATIVA</b> .....	9
<b>1.2 O CULTIVO ILEGAL DE C. SATIVA NA AMÉRICA DO SUL</b> .....	11
<b>1.3 IDENTIFICAÇÃO DE C. SATIVA NA ROTINA DE LABORATÓRIOS FORENSES</b> .....	14
<b>1.4 NOVAS PERSPECTIVAS PARA IDENTIFICAÇÃO E INDIVIDUALIZAÇÃO DE C. SATIVA.</b> ..	14
1.4.1 <i>MÉTODOS BASEADOS EM ASSINATURAS QUÍMICAS</i> .....	15
1.4.2 <i>MÉTODOS BASEADOS EM POLIMORFISMOS DE DNA</i> .....	16
<b>2 OBJETIVOS</b> .....	22
<b>2.1 OBJETIVO GERAL</b> .....	22
<b>2.2 OBJETIVOS ESPECÍFICOS</b> .....	22
<b>3 MATERIAIS E MÉTODOS</b> .....	22
<b>3.4 AMOSTRAS E MATERIAL GENÉTICO</b> .....	22
<b>3.5 EXTRAÇÃO DE DNA</b> .....	26
<b>3.6 LOCOS E INICIADORES MICROSSATÉLITES</b> .....	27
<b>3.7 OTIMIZAÇÃO DA PCR-MULTIPLEX</b> .....	28
<b>3.8 GENOTIPAGEM DOS MICROSSATÉLITES</b> .....	32
<b>3.9 ANÁLISES DOS DADOS</b> .....	32
3.9.1 <i>FREQÜÊNCIA ALÉLICA</i> .....	33
3.9.2 <i>DIVERSIDADE GÊNICA</i> .....	33
3.9.3 <i>EQUILÍBRIO DE HARDY-WEINBERG</i> .....	33
3.9.4 <i>ANÁLISE DE DESEQUILÍBRIO DE LIGAÇÃO ENTRE OS LOCOS</i> .....	34
3.9.5 <i>ANÁLISE DOS MICROSSATÉLITES QUANTO A ERROS DE GENOTIPAGEM, ALELOS NULOS E AMPLIFICAÇÃO PREFERENCIAL DE ALELOS CURTOS</i> .....	34
3.9.6 <i>ESTIMATIVA DO COEFICIENTE DE PARENTESCO</i> .....	35
3.9.7 <i>ANÁLISE DE VARIÂNCIA MOLECULAR (AMOVA)</i> .....	35
3.9.8 <i>ANÁLISE DE COORDENADAS PRINCIPAIS (PCA)</i> .....	36
3.9.9 <i>TESTE DE DESIGNAÇÃO</i> .....	37
<b>4 RESULTADOS</b> .....	38

<b>4.1 OTIMIZAÇÃO DOS MULTIPLEXES.....</b>	<b>38</b>
<b>4.2 ANÁLISE DOS DADOS.....</b>	<b>39</b>
4.2.1 <i>DIVERSIDADE ALÉLICA .....</i>	<b>39</b>
4.2.2 <i>EQUILÍBRIO DE HARDY-WEINBERG.....</i>	<b>51</b>
4.2.3 <i>INDEPENDÊNCIA ENTRE LOCOS (DESEQUILÍBRIO DE LIGAÇÃO) .....</i>	<b>55</b>
4.2.4 <i>ANÁLISE DOS MICROSSATÉLITES QUANTO A ERROS DE GENOTIPAGEM, ALELOS NULOS E AMPLIFICAÇÃO PREFERENCIAL DE ALELOS CURTOS .....</i>	<b>55</b>
4.2.5 <i>ESTIMATIVA DO COEFICIENTE DE PARENTESCO .....</i>	<b>56</b>
4.2.6 <i>PROBABILIDADE DE IDENTIDADE E PODER DE EXCLUSÃO .....</i>	<b>59</b>
4.2.7 <i>ANÁLISE DE ESTRUTURA GENÉTICA .....</i>	<b>60</b>
4.2.7.1 <b>ANÁLISE DE VARIÂNCIA MOLECULAR (AMOVA) .....</b>	<b>60</b>
4.2.7.2 <b>DISTÂNCIA GENÉTICA UTILIZANDO <math>R_{ST}</math>.....</b>	<b>62</b>
4.2.7.3 <b>ANÁLISE DE COORDENADAS PRINCIPAIS (PCA).....</b>	<b>62</b>
4.2.8 <i>TESTE DE DESIGNAÇÃO.....</i>	<b>64</b>
<b>5 DISCUSSÃO.....</b>	<b>68</b>
<b>6 CONSIDERAÇÕES FINAIS .....</b>	<b>73</b>
<b>7 REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>75</b>



# INTRODUÇÃO

## 1.1 A *Cannabis sativa*

A *Cannabis sativa* LINNEU é uma planta de origem asiática, cuja distribuição ao redor do mundo segue as migrações humanas. A partir da Ásia oriental, a *C. sativa* difundiu-se gradualmente para a Índia e Oriente Médio (Pérsia e Arábia) e mais tardiamente para a Europa, norte da África e Américas. Seu cultivo é fonte de fibra e óleo há milhares de anos. A mais antiga descrição do uso de *C. sativa* diz respeito à utilização de sua fibra para a confecção de roupas, redes de pesca, redes para caça, cordas, etc. A utilização das sementes como fonte de alimento pela cultura Yang Chao, datam de aproximadamente 6.500 anos (IVERSEN, 2000). O livro de medicina mais antigo que se conhece, o Pên-Ts'ao Ching, remonta há 4 mil anos e fala do uso mágico das inflorescências femininas da planta devido ao seu caráter alucinógeno (IVERSEN, 2000). O uso recreacional da *C. sativa* acompanhou sua distribuição ao redor do mundo.

De acordo com as classificações taxonômicas atuais a *C. sativa* existe como uma única espécie, da qual há algumas variedades ou subespécies. A planta é unissexuada e dióica, com flores características e individuais, apresentando-se como um arbusto que pode atingir de 1,5 m a 3,0 m de altura, sendo a planta masculina mais baixa e delgada que a feminina. As informações sobre o cariótipo de *C. sativa* dão conta que o número de cromossomos é  $2n=20$  (HONG, 1996). O par de cromossomos sexuais é heteromórfico, sendo a planta feminina XX e a planta masculina XY. No entanto, diferentemente dos mamíferos o cromossomo Y em *C. sativa* tem o dobro do tamanho do cromossomo X (SAKAMOTO, 2000 e PEIL, 2003).

As folhas e as inflorescências da planta contêm um grupo de substâncias denominadas canabinóides (a *C. sativa* é a única espécie que o possui), sendo o principal

componente psicoativo o  $\Delta^9$ -Tetrahydrocannabinol ( $\Delta^9$ -THC) (IVERSEN, 2000). A utilização recreacional da *C. sativa* via a inalação do  $\Delta^9$ -THC, por meio dos vapores produzidos pela queima de folhas e sumidades floridas secas, leva o usuário a experimentar sensações de bem-estar, as quais vêm acompanhadas de estados alucinatórios e exaltação imaginativa. Há também relatos de entorpecimento gradativo, perturbação psicomotora, aumento de apetite, aumento do sono e perda da noção de espaço e tempo. O usuário crônico pode experimentar sérios problemas pulmonares (asma, bronquite, câncer, etc.), cardíacos, neurológicos, reprodutivos, hepáticos, imunológicos e gastrintestinais (IVERSEN, 2000).

A reação ao consumo da *C. sativa* e outras substâncias psicotrópicas ganhou força a partir do final do século XVIII e início do século XIX. Em se tratando de *C. sativa*, contribuiu para este quadro os relatos de quadros depressivos e o desenvolvimento de características psicóticas entre os usuários. A partir de 1910, diversas substâncias foram proibidas dentro do território norte-americano, sendo esta conduta seguida por países da Europa e demais países das Américas. Ao final da década de 30, a *C. sativa* estava proibida em vários países do mundo (BONNIE, 1970). Atualmente, em função de benefícios econômicos e ambientais sobre outras culturas tradicionais fontes de fibras, tem ressurgido o interesse no cultivo de *C. sativa* com baixos teores de canabinóides, levando alguns países a permitirem o cultivo destas variedades. No entanto, a despeito das várias possibilidades de aplicações para a *C. sativa*, o cultivo e a posse da planta, independente do teor de canabinóides, são proibidos pela lei em muitos países, inclusive no Brasil. Esta proibição leva em conta as propriedades psicotrópicas capazes de produzir dependência. A despeito da proibição, existe um mercado ilegal local e internacional, o qual envolve organizações criminosas compostas por plantadores, grandes e pequenos distribuidores.

## 1.2 O cultivo ilegal de *C. sativa* na América do Sul

No caso da América do Sul, segundo dados da Coordenação Geral de Prevenção e Repressão a Entorpecentes do Departamento de Polícia Federal (CGPRE/DPF) as plantações ilegais ocorrem principalmente no Brasil e no Paraguai, existindo também plantios na Colômbia e Bolívia (UNITED NATIONS, 2003).

A propagação da *C. sativa* ao redor do mundo ou é realizada de forma vegetativa ou através de sementes, normalmente coletadas de um plantio anterior na mesma região antes da prensagem para comercialização ilegal da droga. As plantas propagadas pela forma vegetativa devem ter genótipos idênticos, como gêmeos monozigóticos tornando o rastreamento dos clones de *C. sativa* através da análise de DNA uma tarefa relativamente simples (COYLE, 2003). Entretanto, ainda segundo dados da CGPRE/DPF, a propagação vegetativa não é a forma de propagação utilizada nos plantios ilegais da América do Sul. Durante as operações de erradicação de plantios ilegais no Brasil realizadas pela CGPRE/DPF e durante as operações conjuntas realizadas em cooperação com a Direção Nacional de Narcóticos da Secretaria Nacional Antidrogas do Paraguai (DINAR/SENAD) (Figura 1), observa-se a técnica de coleta e secagem ao sol das folhas e inflorescências (que contêm as sementes) para posterior prensagem e distribuição. Antes da prensagem as sementes são colhidas e utilizadas para dar origem a outros plantios na mesma região ou exportadas para outras regiões. Esta metodologia também é a realidade em outros países da América do Sul onde há plantios ilegais relevantes como a Colômbia e a Bolívia (Figura 2).



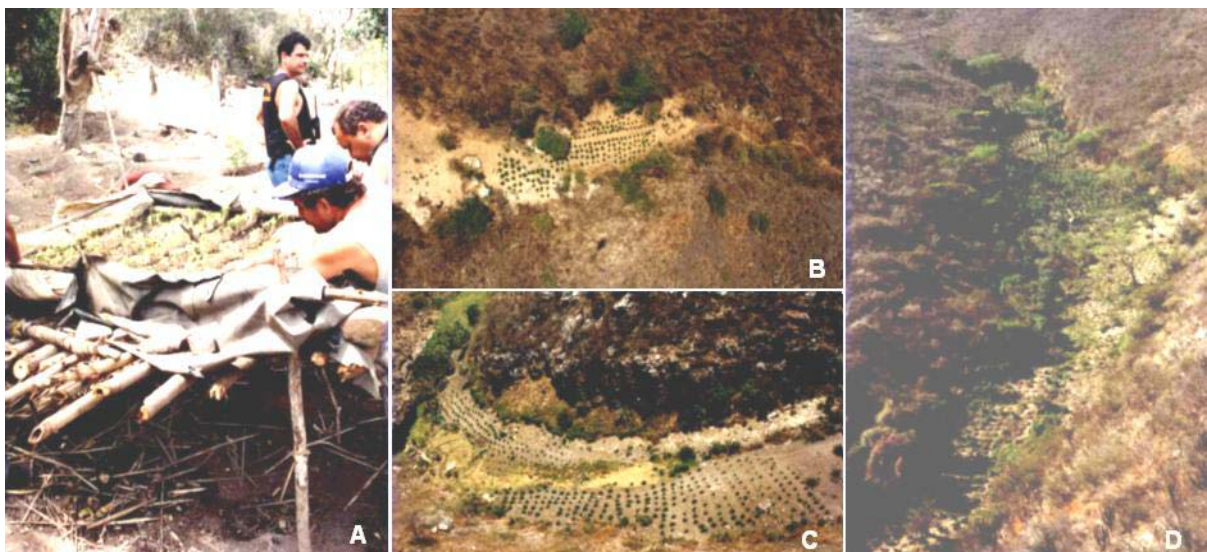
**Figura 1** – Plantios ilegais em território paraguaio, localizados durante operação conjunta de erradicação, realizada pela Secretaria Nacional Antidrogas do Paraguai em cooperação com o Departamento de Polícia Federal do Brasil.



**Figura 2** – Forma de plantio de *C. sativa* praticada no Brasil e no Paraguai. Sementeiras (A e B) e sementes (C e D) encontradas junto a plantios ilegais em operação de erradicação realizada no território paraguaio.



No Brasil, a produção de *C. sativa* é inteiramente voltada para o consumo interno e os maiores plantios ocorrem principalmente no nordeste brasileiro, mais especificamente no interior dos estados de Pernambuco e Bahia. A destruição manual dos plantios, por meio de operações policiais nestas regiões, levou nos últimos anos ao aparecimento de novas plantações no interior do estado do Maranhão, ainda na região nordeste e nos estados do Pará e Amazonas, na região norte do país (Figura 3). Também em consequência à regularidade das operações policiais focadas na região nordeste brasileira, algumas organizações criminosas no Brasil, especializadas no tráfico de *C. sativa*, passaram nos últimos anos a satisfazer a demanda interna importando o produto, principalmente do Paraguai. Após entrar no país, a droga é transportada para os mercados consumidores - principalmente para as regiões Sul, Sudeste e Centro-Oeste do país. A *C. sativa* entra no Brasil principalmente por via terrestre, mas também por via aérea utilizando as fronteiras dos estados do Mato Grosso do Sul e do Paraná. Especificamente, a droga entra no país pelos municípios de Ciudad del Este/Foz do Iguaçu, Sal del Guayra/Guaíra, Pedro Juan Caballero/Ponta-Porã ou Forte Olimpo/Porto Murtinho, sendo posteriormente levada para os mercados consumidores (UNODC/BRASIL, 2003). Assim, de uma forma geral sabe-se que a *C. sativa* presente no mercado consumidor brasileiro é oriunda de plantios ilegais em solo brasileiro e/ou é importada pelas organizações criminosas de regiões localizadas no Paraguai. Entretanto, não há como se afirmar a origem geográfica de uma apreensão, quando não foi possível ao trabalho de inteligência policial rastrear as rotas do tráfico.



**Figura 3** – Plantios ilegais de *C. sativa* localizados em operações de erradicação realizadas na região nordeste do Brasil. **A)** sementeira localizada em plantio. **B, C e D)** vista aérea de plantios.

### 1.3 Identificação de *C. sativa* na rotina de laboratórios forenses

A análise definitiva para identificação de drogas de abuso como a *C. sativa*, envolve em geral uma seqüência de testes químicos de coloração (baseada na sistemática de identificação de compostos orgânicos) e aplicação de métodos cromatográficos e/ou espectroscópicos/espectrométricos. Na grande maioria dos casos, mesmo na ausência de instrumentação analítica é perfeitamente factível determinar a natureza química de uma dada substância, desde que a sistemática indicada para sua análise seja seguida com todos os rigores exigidos. No entanto, há casos raros em que a identificação completa da substância não é possível somente pela observação de suas reatividades químicas. Nestes casos o emprego de técnicas instrumentais, como por exemplo, a Cromatografia Gasosa-Espectrometria de Massas (GC-MS) ou Espectroscopia na região do Ultravioleta-Visível (UV/VIS) é necessário (SILVERSTEIN, 2003).

No caso da *C. sativa*, embora a determinação das características botânicas da planta seja suficiente para sua identificação inequívoca, a jurisprudência atual exige a realização de exames químicos para a detecção do  $\Delta^9$ -THC e outros canabinóides. A identificação química torna-se importante, sobretudo quando a identificação botânica não é possível devido à ausência ou dificuldade de visualização dos caracteres buscados. Os ensaios clássicos mundialmente aceitos para a identificação de *C. sativa* são: a reação com “Fast Blue Salt B”, o teste de Duquenois-Levine e a cromatografia em camada delgada (UNITED NATIONS, 1987).

### 1.4 Novas perspectivas para identificação e individualização de *C. sativa*

Como apresentado no tópico anterior, os métodos para identificação de *C. sativa* dentro da rotina dos laboratórios forenses estão bem estabelecidos. No entanto, os métodos químicos apresentam limitações quando as amostras apresentam-se em quantidades diminutas

(COYLE, 2003) ou foram submetidas ao armazenamento prolongado (VARGAS, 2000). Nestes casos os testes químicos não são capazes de detectar o  $\Delta^9$ -THC. Além disso, os métodos químicos tradicionais não são capazes de individualizar e/ou correlacionar amostras de *C. sativa*. A possibilidade de individualizar e/ou correlacionar amostras de *C. sativa* constitui uma ferramenta em potencial de grande utilidade à inteligência policial, pois permitiria estabelecer relações entre diferentes apreensões, apontando para possíveis rotas de tráfico. Apesar de métodos capazes de individualizar e/ou correlacionar amostras de *C. sativa* não constarem das rotinas de laboratórios forenses, eles estão sendo desenvolvidos e avaliados. Estes métodos baseiam-se em padrões de assinaturas químicas associadas aos perfis orgânicos, inorgânicos e/ou isotópicos das amostras de *C. sativa* e em polimorfismos de DNA.

#### *1.4.1 Métodos baseados em assinaturas químicas*

Em trabalho recente (SHIBUYA, 2005), realizado em parceria entre o Instituto de Pesquisas Energéticas e Nucleares (IPEN), autarquia associada à Universidade de São Paulo e o Instituto de Criminalística da Polícia Civil do Estado de São Paulo, foi desenvolvida uma metodologia capaz de discriminar amostras de *C. sativa* por meio de suas assinaturas químicas. Os procedimentos são baseados nos constituintes inorgânicos das amostras (sendo os mais importantes o cobre, o cobalto, o bário, o lantânio, o zinco, o ferro, o ítrio e o manganês) e na distribuição isotópica do carbono e do nitrogênio. Os constituintes inorgânicos refletem características dos solos nos quais as plantas que deram origem às amostras foram desenvolvidas, enquanto o carbono e o nitrogênio são fatores ligados ao clima. No total foram analisadas 255 amostras. Do total, 157 amostras com origem em apreensões nos centros urbanos das regiões centro-oeste, nordeste e norte do país, foram consideradas como referência de plantios locais, baseando-se em informações dos Institutos

de Criminalística das Polícias Cíveis de cada Estado. Os resultados das análises dessas amostras indicaram compatibilidade entre as mesmas e as características climáticas e geológicas das regiões de origem, ratificando a informação de cultivo local. Ainda, um conjunto de 80 amostras apreendidas no Estado de São Paulo foi analisado a fim de se obter informações capazes de conduzir a inferências sobre suas prováveis origens geográficas. Os resultados indicaram que 82,5% dessas amostras apresentam origem similar àquelas de Mato Grosso do Sul (região centro-oeste), enquanto que somente 5% parecem ser originárias da região nordeste. O trabalho conclui que apesar das limitações relacionadas ao procedimento de amostragem, é possível o desenvolvimento de um banco de dados de assinaturas químicas de *C. sativa* partindo-se de amostras apreendidas nos centros urbanos. No entanto, a análise de amostras obtidas diretamente em plantios erradicados viria a complementar o banco de dados permitindo a identificação inequívoca (SHIBUYA, 2005).

#### *1.4.2 Métodos baseados em polimorfismos de DNA*

O desenvolvimento e aperfeiçoamento das técnicas de biologia molecular para análise de DNA foram aplicados inicialmente em organismos modelo, organismos de importância econômica e na promoção da saúde humana. No entanto, o nível atual de desenvolvimento e mesmo de custos para utilização destas técnicas permite sua ampla utilização. Neste âmbito, organismos de importância social como a *C. sativa* passaram também a ser alvo de aplicação dessas técnicas. Assim, os trabalhos onde se aplicam as metodologias de análise de DNA em *C. sativa* são recentes, valendo-se principalmente das técnicas que utilizam a reação da PCR (do inglês, *polymerase chain reaction*).

Os trabalhos iniciais visavam a identificação e confirmação da natureza de pequenas amostras de *C. sativa*. No final da década de 1990, foram desenvolvidos iniciadores espécie-específicos na região entre os genes *trnL* e *trnF* no DNA cloroplástico de *C. sativa*,



permitindo a distinção de outras espécies vegetais por meio da reação da PCR (LINACRE, 1998). Outros trabalhos exploraram polimorfismos na região ITS1 e ITS2 do DNA nuclear ribossomal (GIGLIANO, 1997, GIGLIANO, 1998 e GIGLIANO, 1999) e o polimorfismo de comprimento na região intergênia em DNA cloroplástico utilizando PCR-SSCP (do inglês, *single strand conformation polymorphism*) (KOHJYUUMA, 2000). Estes trabalhos permitiram a distinção entre amostras de *C. sativa* e outros vegetais, inclusive *Humulus lupulus* LINNEU, representante do gênero *Humulus*, que é o único da família *Cannabaceae* além do gênero *Cannabis*.

Além da identificação, trabalhos buscando distinção entre amostras de *C. sativa* tem surgido na literatura. Utilizando um conjunto de quatro iniciadores produzidos pela Operon Technologies (OPA-01, OPA-02, OPA-03 e OPA-04), pesquisadores da Austrália obtiveram padrões reprodutíveis de bandas RAPD (do inglês, *random amplification of polymorphic DNA*), com os quais analisaram 53 amostras vegetais, sendo 43 amostras de *C. sativa* com três origens distintas no território australiano, oito amostras também de *C. sativa* com origem em Papua-Nova Guiné e duas amostras de *H. lupulus*. Os resultados apontaram para padrões de bandas capazes de distinguir claramente as amostras de *H. lupulus* das amostras de *C. sativa*, possibilitando também a classificação destas amostras em quatro grupos distintos de acordo com a sua origem. (JAGADISH, 1996). Outro trabalho, utilizando padrões de bandas RAPD e RFLP (do inglês, *restriction fragment length polymorphism*) foi capaz de distinguir amostras cultivadas legalmente para a produção de fibras de amostras cultivadas ilegalmente para uso como psicotrópico (SHIROTA, 1998). Ainda empregando a metodologia de RAPD, pesquisadores italianos realizaram estudo para investigar a estrutura genética e a variabilidade de seis variedades de *C. sativa*, sendo quatro cultivadas na Europa para a produção de fibras e duas com alto teor de  $\Delta^9$ -THC, sendo evidenciada alta variabilidade e um baixo *Fst* entre as variedades (FORAPANI, 2001).

Diversos trabalhos também foram recentemente publicados utilizando AFLP (do inglês, *amplified fragment length polymorphism*) em *C. sativa*, mostrando suas aplicações na análise de variação genética, inclusive com resultados em investigação forense. Em 2003, foi realizado trabalho utilizando marcadores AFLP a fim de obter informações sobre a estrutura dos cromossomos sexuais de *C. sativa*, sendo encontrados cinco marcadores localizados em ambos os cromossomos, sendo dois completamente ligados ao cromossomo Y; as análises de recombinação levaram a conclusão de que existe uma região pseudoautossômica entre os cromossomos sexuais de *C. sativa*, permitindo a recombinação entre os cromossomos X e Y (PEIL, 2003). Ainda no mesmo ano, trabalho utilizando um único par de iniciadores, ao testar um método de extração de DNA de amostras de *C. sativa*, demonstrou a reprodutibilidade dos perfis AFLP, sendo possível utilizá-los na distinção e na confirmação da reprodução clonal entre indivíduos (COYLE, 2003). Em 2005, foi publicado trabalho demonstrando que a análise de marcadores AFLP é uma metodologia útil no estabelecimento de padrões genéticos de amostras de *C. sativa* e pode ser utilizada como evidência de propagação clonal (PALMBACH, 2005). Este ano, foi publicado o primeiro trabalho relatando inconsistências no perfil de bandas AFLP em *C. sativa*, uma discussão importante no contexto da admissibilidade em um contexto forense da evidência AFLP em *C. sativa*. O problema foi associado com a baixa intensidade das bandas AFLP e o trabalho conclui pela necessidade de padronizar a quantidade de DNA genômico ao utilizar AFLP em *C. sativa*, em investigações forenses. O estudo também comparou amostras de *C. sativa* cultivadas para a produção de fibras e cultivadas ilegalmente para uso como psicotrópico, demonstrando a capacidade dos marcadores AFLP em distingui-las (DATWYLER, 2006).

Apesar das metodologias de análise baseadas em RFLPs, em RAPDs, e em AFLPs, terem demonstrado potencialidades para a individualização de amostras de *C. sativa*, os marcadores ideais para esta análise são aqueles baseados em seqüências repetitivas em tandem (STRs – do inglês, *short tandem repeats*), também denominados seqüências simples repetitivas (SSRs - do inglês, *simple sequence repeats*) ou microssatélites. Estas seqüências

consistem em motivos de seqüências de 1 a 6 pares de bases que se repetem em tandem. Estas regiões ocorrem freqüente e randomicamente nos genomas de plantas e animais e tipicamente exibem uma grande variação. De maneira geral são os marcadores moleculares mais informativos para estudos de genética ao nível de espécies e têm sido aplicados em um grande número de estudos em plantas, animais e em populações humanas. São marcadores codominantes e altamente multialélicos, apresentando maior conteúdo informativo por loco gênico entre todas as classes de marcadores moleculares (GOLDSTEIN e SCHLÖTTERER, 1998).

Em 2003 e de forma independente, três grupos de pesquisa publicaram os primeiros trabalhos de desenvolvimento de microsatélites em *C. sativa* e os esforços na obtenção destes marcadores somam até agora um total de 27 (vinte e sete) locos polimórficos publicados. Um dos trabalhos descreve um loco microsatélites em *C. sativa*, apresentando uma seqüência simples repetitiva cujo motivo é um hexanucleotídeo e analisaram 108 amostras, sendo 89 sementes e 19 folhas, identificando o loco como altamente polimórfico, exibindo alelos na faixa de três até 40 unidades repetitivas (18 a 240 pb) e uma heterozigose observada de 0,87% (HSIEH, 2003). Um segundo trabalho relata a identificação de 15 locos microsatélites em *C. sativa* e a análise com estes marcadores de 48 amostras, representando cinco diferentes acessos cultivados para a produção de fibra, obtendo uma faixa de dois a 28 alelos por loco, com o conjunto exibindo uma heterozigose média de 0,68, mostrando-se hipervariáveis e informativos (GILMORE e PEAKALL, 2003). Em trabalho seqüencial, a partir dos 15 marcadores microsatélites identificados no trabalho anterior, foi selecionado um subconjunto de cinco locos, com os quais foram analisadas 93 amostras, demonstrando-se sua capacidade de discriminação entre indivíduos e variedades de *C. sativa* (GILMORE, 2003). Paralelamente, ocorreu o desenvolvimento, por pesquisadores da Universidade da Flórida, de 11 marcadores também desta classe, sendo analisadas 41 amostras, obtendo-se uma faixa de três a nove alelos por loco e uma heterozigose observada média de 0,52 (ALGHANIM e ALMIRALL, 2003).

Este trabalho, baseou-se nas publicações referentes aos dados de desenvolvimento de microssatélites em *C. sativa* descritas anteriormente, apesar de recentemente, em fevereiro de 2006, ter sido obtido junto ao *United States Patent and Trademark Office* (USPTO), agência do Departamento de Comércio dos Estados Unidos o registro de patente nº 20060035236 de um conjunto de marcadores microssatélites em *C. sativa*. No trabalho, os pesquisadores relatam o teste de oito marcadores microssatélites em três reações de PCR-multiplex, contendo de dois a quatro pares de iniciadores por reação. No registro de patente, relatam ainda a análise de 295 amostras de *C. sativa* realizada com o conjunto de oito marcadores. Foram obtidos um total de 63 alelos no conjunto de amostras testado, com o número de alelos por loco variando de dois a 16, uma heterozigose média observada de 0,41 e uma heterozigose média esperada de 0,58, com a heterozigose por loco variando de 0,21 a 0,79 (KEIM, 2006). A Tabela 1 apresenta um resumo dos dados de microssatélites em *C. sativa* publicados em 2003, nos quais este trabalho está baseado.

Atualmente, os questionamentos forenses encaminhados à Polícia Científica do Departamento de Polícia Federal têm sido sobre a natureza da substância vegetal encaminhada a exames. No caso da *C. sativa*, as análises químicas tradicionais e análises instrumentais como Cromatografia Gasosa/Espetrografia de Massa (CG/EM) e Cromatografia Líquida de Alta Pressão (HPLC) têm sido suficientes para produzir a prova material nos casos de tráfico, plantio e/ou posse ilegal da planta. Entretanto, a possibilidade de indicar a origem geográfica, individualizar ou mesmo correlacionar amostras de *C. sativa* apreendidas no Brasil abre possibilidades de geração de informações úteis à inteligência policial no combate ao narcotráfico. Neste objetivo, apesar de alguns resultados na correlação entre fontes e na diferenciação entre amostras de *C. sativa*, as análises químicas não são suficientes para a determinação da origem geográfica de uma amostra (COYLE, 2003), sendo os métodos de análise baseados em DNA muito mais informativos e eficientes. Sendo assim, este trabalho buscou testar uma bateria de microssatélites de *C. sativa* a serem amplificados

em multiplex e utilizá-la para investigar a diversidade genética em 477 amostras coletadas no Paraguai e no Brasil, avaliando sua potencialidade para diferenciar plantas oriundas destas regiões como uma nova ferramenta a ser utilizada pela inteligência policial.

**Tabela 1** - Lista dos marcadores microsatélites em *C. sativa* publicados.

Nº	Loco	Motivo	Número de amostras testadas	Número de alelos observados	Heteroz Obs. (Ho)	Heteroz Esp. (He)	Referência
1	ANUCS201	(GA) <sub>26</sub>	48	18	0,51	0,86	GILMORE, 2003
2	ANUCS202	(GA) <sub>20</sub>	48	14	0,63	0,86	GILMORE, 2003
3	ANUCS203	(CT) <sub>50</sub>	48	28	0,67	0,94	GILMORE, 2003
4	ANUCS204	(CT) <sub>26</sub>	48	14	0,14	0,87	GILMORE, 2003
5	ANUCS205	(CT) <sub>21</sub>	48	18	0,59	0,87	GILMORE, 2003
6	ANUCS206	(AT) <sub>11</sub>	48	4	0,38	0,58	GILMORE, 2003
7	ANUCS301	(TTA) <sub>15</sub>	48	13	0,60	0,85	GILMORE, 2003
8	ANUCS302	(CAA) <sub>7</sub> -(CAA) <sub>4</sub>	48	10	0,56	0,68	GILMORE, 2003
9	ANUCS303	(GTG) <sub>7</sub>	48	5	0,50	0,56	GILMORE, 2003
10	ANUCS304	(TCT) <sub>8</sub> TCA(TCT) <sub>7</sub>	48	15	0,50	0,69	GILMORE, 2003
11	ANUCS305	(TGG) <sub>7</sub>	48	7	0,67	0,71	GILMORE, 2003
12	ANUCS306	(GAT) <sub>3</sub> -(GAT) <sub>6</sub>	48	2	0,67	0,49	GILMORE, 2003
13	ANUCS307	(ACC) <sub>6</sub>	48	2	0,24	0,28	GILMORE, 2003
14	ANUCS308	(TAA) <sub>3</sub> -(AT) <sub>5</sub>	48	8	0,18	0,39	GILMORE, 2003
15	ANUCS501	(TTGTG) <sub>4</sub>	48	3	0,49	0,51	GILMORE, 2003
16	B01-CANN1	(GAA) <sub>13</sub> (A)(GAA) <sub>3</sub>	41	5	0,58	0,56	ALGHANIM, 2003.
17	B02-CANN1	(AAG) <sub>10</sub>	41	3	0,61	0,60	ALGHANIM, 2003
18	B05-CANN1	(TTG) <sub>9</sub>	41	4	0,61	0,64	ALGHANIM, 2003
19	C08-CANN2	(GA) <sub>21</sub>	41	9	0,45	0,69	ALGHANIM, 2003
20	C11-CANN1	(GAT) <sub>8</sub> (GGT) <sub>7</sub>	41	5	0,60	0,54	ALGHANIM, 2003
21	D02-CANN1	(GTT) <sub>7</sub>	41	3	0,61	0,60	ALGHANIM, 2003
22	D02-CANN2	(CTT) <sub>6</sub> (ATT)(CTT) <sub>10</sub>	41	4	0,61	0,64	ALGHANIM, 2003
23	E07-CANN1	(GA) <sub>21</sub>	41	9	0,45	0,69	ALGHANIM, 2003
24	H06-CANN2	(ACG) <sub>7</sub>	41	3	0,42	0,55	ALGHANIM, 2003
25	H09-CANN2	(GA) <sub>15</sub>	41	6	0,73	0,71	ALGHANIM, 2003
26	H11-CANN1	(CT) <sub>18</sub>	41	7	0,64	0,59	ALGHANIM, 2003
27	CS1	(CACCAT) <sub>17</sub>	108	30	0,87	-	HSIEH, 2003

## 2 OBJETIVOS

### 2.1 Objetivo Geral

- Otimizar e avaliar o desempenho de um sistema de identificação genética de *C. sativa* baseado em microssatélites.

### 2.2 Objetivos específicos

- Estabelecer a amplificação de locos microssatélites de *C. sativa* em PCR multiplex;
- Construir e caracterizar um banco de dados de frequências alélicas para microssatélites em *C. sativa*, utilizando o(s) sistema(s) multiplex(es) desenvolvido(s), em plantios desenvolvidos na América do Sul, particularmente no Brasil e no Paraguai;
- Estudar a distribuição da variabilidade genética em uma amostra de 477 plantas provenientes de plantios ilegais de *C. sativa* desenvolvidos na América do Sul, particularmente no Brasil e no Paraguai;
- Avaliar o potencial dos microssatélites estudados neste trabalho como ferramenta indicativa da origem geográfica nas amostras de *C. sativa* estudadas neste projeto.

## 3 MATERIAIS E MÉTODOS

### 3.4 Amostras e material genético

As amostras de *C. sativa* utilizadas no desenvolvimento deste trabalho foram obtidas durante a realização de operações policiais para erradicação de plantios ilegais. As amostras

provenientes de quatro plantios distintos na região que engloba os municípios de Cabrobó, Orocó, Carnaubeira da Penha, Custódia, Belém de São Francisco, Mirandiba, Serra Talhada, Ibimirim, Betânia e Floresta no Estado de Pernambuco foram coletadas diretamente das plantas durante o momento da erradicação, em operação policial denominada Operação Controle XII. Esta operação foi realizada naquela região, no período de 10 a 18 de maio de 2004, em operação conjunta entre as polícias Rodoviária Federal, Militar, Civil e Bombeiros de toda a região nordeste e o Departamento de Polícia Federal.

A coleta e o transporte destas amostras foram realizados em sacos plásticos herméticos contendo sílica gel, o que permitiu que as folhas e caules coletados secassem rapidamente, mantendo sua integridade até a chegada ao laboratório e posterior extração do DNA. Esta coleta só foi possível devido à possibilidade de participação do mestrando naquela operação. A natureza da operação policial para erradicação dos plantios ilegais não permite a dedicação dos profissionais envolvidos em atividades que não àquelas com o fim de erradicação. Assim, para a identificação e individualização das plantas, nos moldes necessários para realização de trabalhos com este, foi imperativo a presença do mestrando, o qual é também Perito Criminal do Departamento de Polícia Federal.

Devido a esta limitação, a alternativa utilizada em relação às outras regiões de plantios no Brasil e no Paraguai foi a obtenção, junto ao Departamento de Polícia Federal, de sementes apreendidas em plantios ilegais e a geração de plantios em ambiente controlado. As sementes foram germinadas em algodão umedecido e desenvolvidas em bandejas de isopor, utilizando-se como substrato terra vegetal (Figura 4). Estas amostras, constituídas por caules e folhas foram embaladas em papel alumínio e mantidas congeladas a  $-20^{\circ}\text{C}$  desde o momento da coleta, assim permanecendo até o momento da extração do DNA. A distribuição e origem das amostras utilizadas neste estudo são detalhadas na Tabela 2 e na Figura 5.



**Figura 4** - Plantios de *C. sativa* realizados a partir de sementes obtidas em operações de erradicação realizadas em regiões distintas no Estado do Maranhão e no Paraguai. **A e B)** germinação das sementes em algodão umedecido. **C e D)** plantios desenvolvidos, identificados por região. **E e F)** armazenamento das amostras.

**Tabela 2** – Origem geográfica e condição das amostras de *C. sativa* utilizadas para extração de DNA.

Região	Referência (Plantio)	Origem/Localidade	Coordenadas Geográficas	Condição	Nº de Amostras
Paraguai	PY1	Cerro Corá /PY	(S) 22° 37' 00" (W) 56° 58' 59"	Sementes germinadas	44
	PY2	Cerro Sarambi /PY	(S) 22° 44' 00" (W) 56° 17' 20"	Sementes germinadas	42
	PY3	Colônia S <sup>ta</sup> M <sup>a</sup> Auxiliadora / PY	(S) 22° 50' 09" (W) 56° 18' 02"	Sementes germinadas	47
	PY4	Col. Itapopó / PY	(S) 22° 34' 00" (W) 55° 37' 00"	Sementes germinadas	35
	PY5	Colônia S <sup>ta</sup> M <sup>a</sup> Auxiliadora / PY	(S) 22° 51' 47" (W) 56° 27' 24"	Sementes germinadas	36
	PY6	Cerro Sarambi /PY	(S) 22° 32' 15" (W) 56° 17' 10"	Sementes germinadas	34
Maranhão - Brasil -	MA1	Junco do Maranhão/ MA	(S) 1° 56' 16" (W) 46° 12' 04"	Sementes germinadas	37
	MA2	Arame / MA	(S) 4° 52' 30" (W) 46° 07' 30"	Sementes germinadas	36
Pernambuco - Brasil -	PE1	Belém de São Francisco / PE	(S) 8° 24' 35,2" (W) 38° 44' 57,6"	Folhas secas	36
	PE2	Floresta / PE	(S) 8° 35' 36,7" (W) 38° 47' 48,7"	Folhas secas	40
	PE3	Sta Maria da Boa Vista / PE	(S) 8° 36' 00" (W) 39° 51' 31,6"	Folhas secas	46
	PE4	Salgueiro / PE	(S) 8° 26' 15" (W) 39° 03' 45"	Folhas secas	44
Total de amostras analisadas					477





Figura 5 – Localização geográfica dos plantios dos quais foram obtidas as amostras de *C. sativa*.

### 3.5 Extração de DNA

Folhas ou caules de plantas individuais de *C. sativa*, coletadas diretamente dos plantios ou a partir de sementes germinadas, foram utilizados para a extração de DNA de acordo com o protocolo de extração pelo método de CTAB (DOYLE, 1990) com pequenas modificações. Aproximadamente 40 mg de tecido vegetal (caule e/ou folhas), foi adicionado em microtubo descartável de 1,5 ml juntamente com 600 µl do tampão de extração CTAB 2%, adicionado de β-Mercaptoetanol 1% imediatamente antes do procedimento e triturado em agitador FastPrep® seguindo-se incubação a 60-65°C durante 40 min. Além do tampão Tris-HCl, que mantém o pH 8,0, a mistura de extração contém NaCl em alta concentração, EDTA e o CTAB. A manutenção do pH alcalino e a presença do agente quelante EDTA inibem a ação das DNases; enquanto a alta concentração de NaCl favorece a manutenção do DNA solúvel na fase aquosa e precipitação posterior frente ao iso-propanol. Antes da precipitação, uma mistura de clorofórmio-álcool isoamílico 24:1 foi utilizada para remoção de proteínas, lipídios, demais resíduos lipossolúveis e componentes celulares outros que não ácidos nucléicos. Estes últimos são extraídos com a fase orgânica. Após a precipitação e lavagens sucessivas com Etanol, o *pellet* (complexo insolúvel precipitado e concentrado por ultracentrifugação) foi solubilizado em tampão TE (Tris-HCl / EDTA) e incubado a 37°C com a enzima Ribonuclease A (1U/µl) para degradação do RNA extraído juntamente com o DNA.

O rendimento da extração foi quantificado por eletroforese em comparação com DNA de concentração conhecida, em gel de agarose a 1%, contendo 0,5 µg/ml de brometo de etídio. Os produtos de eletroforese foram visualizados em Transiluminador-UV e o DNA diluído para uma concentração de aproximadamente 2 ng/µl.

### 3.6 Locos e iniciadores microssatélites

Até o momento existem publicados 27 marcadores microssatélites caracterizados em *C. sativa* (HSIEH, 2003, GILMORE, 2003 e ALGHANIM, 2003). Para realização deste trabalho foram selecionados 13 locos a partir dos trabalhos publicados. Os critérios de seleção utilizados basearam-se nos dados publicados e foram os seguintes: a alta heterozigose, o maior número de alelos observados e preferencialmente locos com motivo de repetição acima de dois nucleotídeos. Entre os 13 microssatélites selecionados, cinco são dinucleotídeos, cinco são trinucleotídeos perfeitos e três são trinucleotídeos imperfeitos. As informações referentes aos motivos de repetição, o número de alelos observados, a heterozigose e as respectivas referências dos 13 microssatélites selecionados são apresentadas na Tabela 3.

**Tabela 3** - Lista dos marcadores microssatélites utilizados no desenho dos multiplexes para *C. sativa*.

Nº	Loco	Motivo	Número de alelos observados	Heteroz. Obs. (Ho)	Heteroz. Esp. (He)	Referência
1	ANUCS201	(GA) <sub>26</sub>	18	0,51	0,86	GILMORE, 2003
2	ANUCS202	(GA) <sub>20</sub>	14	0,63	0,86	GILMORE, 2003
3	ANUCS301	(TTA) <sub>15</sub>	13	0,60	0,85	GILMORE, 2003
4	ANUCS302	(CAA) <sub>7</sub> -(CAA) <sub>4</sub>	10	0,56	0,68	GILMORE, 2003
5	ANUCS303	(GTG) <sub>7</sub>	5	0,50	0,56	GILMORE, 2003
6	ANUCS304	(TCT) <sub>8</sub> TCA(TCT) <sub>7</sub>	15	0,50	0,69	GILMORE, 2003
7	ANUCS305	(TGG) <sub>7</sub>	7	0,67	0,71	GILMORE, 2003
8	B01-CANN1	(GAA) <sub>13</sub> (A)(GAA) <sub>3</sub>	5	0,58	0,56	ALGHANIM, 2003.
9	B05-CANN1	(TTG) <sub>9</sub>	4	0,61	0,64	ALGHANIM, 2003
10	C08-CANN2	(GA) <sub>21</sub>	9	0,45	0,69	ALGHANIM, 2003
11	H06-CANN2	(ACG) <sub>7</sub>	3	0,42	0,55	ALGHANIM, 2003
12	H09-CANN2	(GA) <sub>15</sub>	6	0,73	0,71	ALGHANIM, 2003
13	H11-CANN1	(CT) <sub>18</sub>	7	0,64	0,59	ALGHANIM, 2003

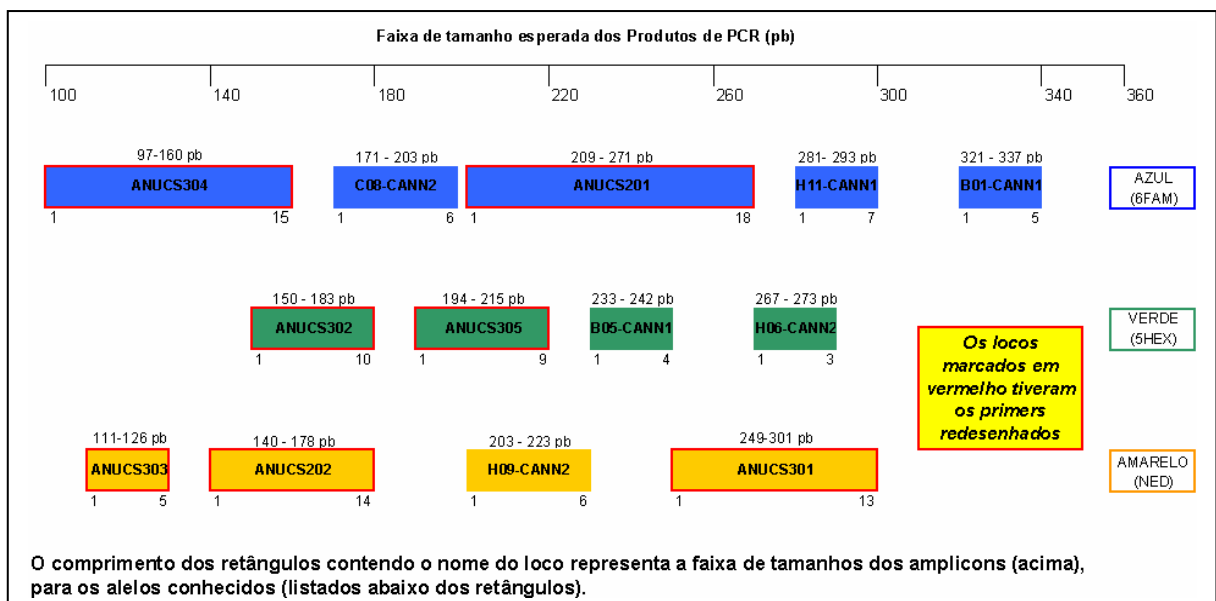
### 3.7 Otimização da PCR-multiplex

Para seis locos (locos 8 a 13 na Tabela 3) os iniciadores utilizados foram aqueles publicados na literatura. No entanto, para ajustar a faixa de tamanho dos amplicons dentro do número de alelos já observados, visando permitir a detecção dos alelos possíveis sem sobreposição entre si, novos iniciadores foram desenhados para sete locos (locos 1 a 7 na Tabela 3). As seqüências foram obtidas no Genbank e os iniciadores foram redesenhados utilizando o programa Primer3 (ROZEN, 1998), disponível no endereço <http://frodo.wi.mit.edu/cgi-bin/iniciador3/iniciador3.cgi/>. Levando-se em conta o intervalo de tamanho no qual estão distribuídos os alelos dos 13 locos, os iniciadores diretos de cada um dos pares foram marcados com os fluoróforos FAM (azul), HEX (verde) ou NED (amarelo). Um dos problemas a ser evitado quando se pretende reunir vários conjuntos de iniciadores em uma única reação PCR é a dimerização, pareamento entre os oligonucleotídeos, que pode ocorrer caso haja excesso de similaridade nas seqüências. Assim, os vinte e seis iniciadores foram comparados entre si buscando similaridade utilizando o programa AutoDimer (VALLONE, 2004), não sendo encontrada nenhuma incompatibilidade relevante.

As seqüências dos iniciadores, as respectivas temperaturas de dissociação ( $T_m$ , do inglês *melting temperature*), o intervalo esperado de distribuição dos alelos e o respectivo fluoróforo com o qual o iniciador direto foi marcado são detalhados na Tabela 4. De maneira esquemática a Figura 6 mostra a faixa de tamanho dos amplicons prevista e as cores dos marcadores fluorescentes usadas no desenho dos multiplexes.

**Tabela 4** - Lista das seqüências dos iniciadores utilizados no desenho dos multiplexes para *C. sativa*. Visando a permitir a detecção dos alelos possíveis sem interferência entre si, por meio da adequação do tamanho dos amplicons, os marcadores de 1 a 7 (seqüências em vermelho) tiveram seus iniciadores redesenhados a partir das seqüências publicadas no GenBank; os demais foram mantidos conforme a publicação original.

Nº	Loco		Seqüência dos <i>Iniciadores</i> (5' – 3')	T <sub>m</sub> (°C)	Faixa Alélica (pb)
1	ANUCS201	F	[FAM]-AGTGGGTCTTCATCCTCTGT	60	209–271
		R	CCACTAAACCAA AAGTACTCTTC	59	
2	ANUCS202	F	[NED]-TGCACATACATAGAGAGAGAGAG	54	140–178
		R	CAGAACTGAAGAGAGGGGAAG	54	
3	ANUCS301	F	[NED]-CCCAAATATGGTTGAAATCC	57	249–301
		R	CTTATGTTGCTATTTACCTATTGC	55	
4	ANUCS302	F	[HEX]-AACATAAACACCAACAACTGC	55	150–183
		R	TGTTCTGTGATGGTTGAT	55	
5	ANUCS303	F	[NED]-CGCGATTTTAAGAGACAACA	57	111–126
		R	TCGACGATAATAGGGCAAG	56	
6	ANUCS304	F	[FAM]-TCTTCACTCACCTCCTCTCT	60	97–160
		R	ATGGTGCGTTGAAGTTTATG	56	
7	ANUCS305	F	[HEX]-TTGTGGTAAAGTTGGTCTGA	56	194-215
		R	ACCAAACCACACTAAATCC	56	
8	B01-CANN1	F	[FAM]-GGAGTCAAATGAAAGGGAAC	58	321–337
		R	CATAGCATTATCCCACTCAAG	59	
9	B05-CANN1	F	[HEX]-TTGATGGTGGTGAACGG	58	233–242
		R	CCAATCTCAATCTCAACCC	58	
10	C08-CANN2	F	[FAM]-GCAAGAAGTAGAGAGACAATCC	61	171–203
		R	CCCTCAACACTCATTAACCTCAC	61	
11	H06-CANN2	F	[HEX]-TGGTTTCAGTGGTCCTCTC	60	267–273
		R	ACGTGAGTGATGACACGAG	60	
12	H09-CANN2	F	[NED]-CGTACAGTGATCGTAGTTGAG	53	203–223
		R	ACACATACAGAGAGAGCCC	53	
13	H11-CANN1	F	[FAM]-ATGTGGTTGTTTCGTACCC	58	281–293
		R	GCGAACATTCACTCTAGCTC	60	



**Figura 6** - Esquema de faixa de tamanho dos amplicons prevista e as cores dos marcadores fluorescentes utilizadas no desenho dos multiplexes.

A coamplificação de vários microssatélites simultaneamente, denominada como amplificação em multiplex, aumenta a capacidade de genotipagem, reduz os gastos com reagentes e diminui as chances de erros e a inclusão de genótipos de diferentes locos como sendo de uma mesma amostra, quando na verdade são de amostras diferentes. O desenho de um conjunto de iniciadores para uma reação PCR em multiplex e a otimização da reação, a fim de obter um conjunto equilibrado de amplicons é um desafio muito maior que a padronização de uma reação com apenas um par de iniciadores. Para a otimização de reações em multiplex é necessária uma seleção cuidadosa dos iniciadores, seguida de ajustes nas condições de reação, variando-se a concentração dos iniciadores, a concentração de  $MgCl_2$ , a concentração e pH do tampão de reação, a temperatura de pareamento ( $T_a$ , do inglês *annealing temperature*) e a concentração da Taq DNA polimerase. Em artigo recente, Schoske et al. (2003) apresentam o desenvolvimento de uma reação PCR-multiplex, cuja sistemática e estratégias de otimização, em linhas gerais, foram seguidas neste trabalho.

Iniciou-se a otimização testando os conjuntos de iniciadores em reações individuais de 15  $\mu$ l. Cada iniciador teve sua concentração final fixada em 0,5  $\mu$ M, sendo utilizado um programa de amplificação com passo inicial de incubação a 95°C por cinco minutos, seguido por dois blocos de amplificação; um primeiro, constando de um passo de desnaturação a 95°C por um minuto, seguido por um passo de pareamento dos iniciadores a 62°C por um minuto e finalizando com um passo de extensão a 72°C por um minuto, em um total de 15 ciclos e um segundo bloco de 20 ciclos, com passo de desnaturação a 95°C por um minuto, seguido por um passo de pareamento dos iniciadores a 55°C por um minuto e finalizando com um passo de extensão a 72°C por um minuto, seguindo-se ainda uma extensão final a 72°C por cinco minutos. Utilizando um gel de agarose 1% foi possível visualizar a amplificação dos segmentos alvo em onze dos treze marcadores. Foi testado em seguida reunir os conjuntos de

iniciadores em uma única reação PCR-multiplex de 20  $\mu$ l, nas mesmas condições anteriores, dessa vez, entretanto, não se obtendo sucesso na amplificação. Foram então testados ajustes na concentração de  $MgCl_2$  e pH do tampão de PCR, também sem sucesso.

Os treze marcadores microssatélites foram então analisados quanto à faixa esperada dos amplicons e distribuídos em dois blocos com faixas de tamanho esperadas próximas, buscando-se menor desequilíbrio na reação multiplex. As duas misturas de iniciadores foram preparadas e aliqüotadas para uma concentração final na reação de 0,5  $\mu$ M, procedendo-se diversas reações PCR em programas “*touchdown*”, variando-se a temperatura inicial e final de pareamento, além de se variar também a concentração de  $MgCl_2$ . Por meio de um gel de agarose 1% foi possível visualizar a amplificação discreta dos segmentos alvo em um dos blocos de iniciadores, enquanto o outro não exibiu amplificação em qualquer dos testes.

Nova análise dos conjuntos de iniciadores foi realizada, e dessa vez, focando-se na temperatura de pareamento ( $T_a$ ) dos iniciadores, estimada em 5°C abaixo da temperatura de dissociação ( $T_m$ , do inglês *melting temperature*). Após a análise, os conjuntos de iniciadores foram reunidos em dois novos blocos, reunindo iniciadores com médias de temperatura de pareamento próximas. As reações PCR seguintes, realizadas com os programas “*touchdown*”, exibiram bandas de amplificação em gel de agarose 1% em ambos os blocos de iniciadores, passando-se às corridas eletroforéticas em seqüenciador automático de DNA ABI Prism® 377, obtendo-se as primeiras genotipagens em multiplex e fazendo-se os ajustes de concentração de iniciadores e nos programas de amplificação, a fim de obter um conjunto de amplicons equilibrado e detectável.

### **3.8 Genotipagem dos microssatélites**

A genotipagem dos microssatélites amplificados pela PCR, seja durante a fase de otimização, seja durante a genotipagem do conjunto de amostras, foi realizada através de eletroforese no seqüenciador automático de DNA ABI Prism<sup>®</sup> 377. Para genotipagem 1-2 µl de produto de PCR foram adicionados a 3-4 µl de tampão de carregamento. O tampão de carregamento é uma mistura de formamida/EDTA e azul de bromofenol. Para genotipagem automatizada fragmentos de tamanhos entre 58 e 417 pares de bases marcados com o fluoróforo ROX foram aplicados em conjunto (BRONDANI, 2001).

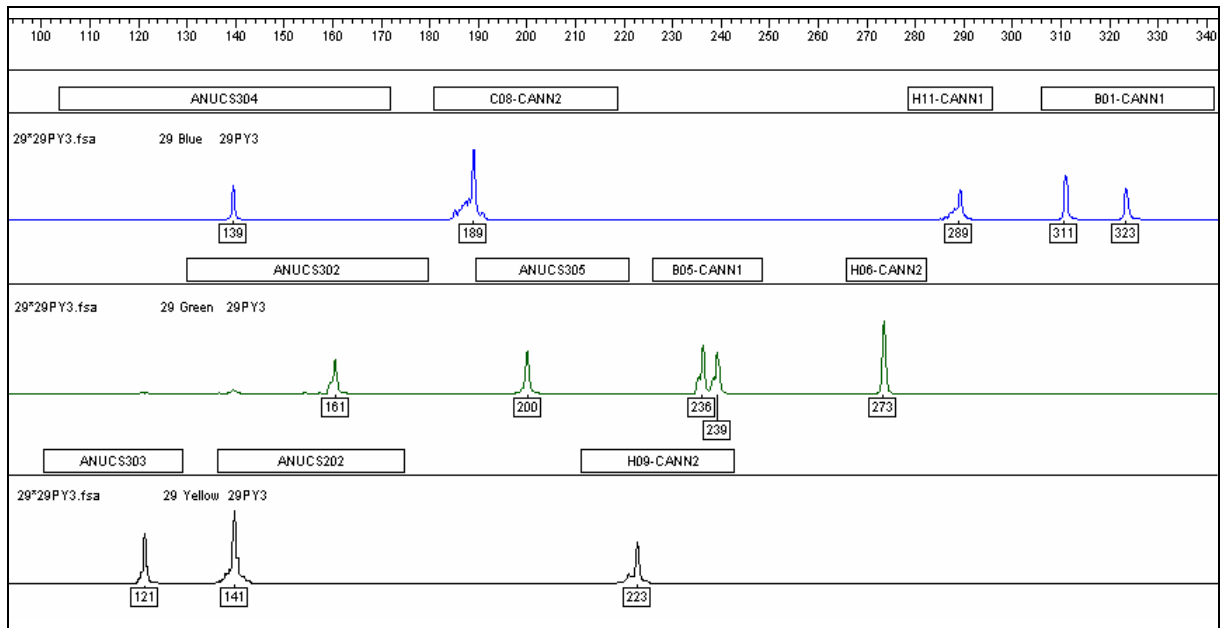
Após eletroforese os arquivos com informações de cada uma das amostras foram analisados utilizando os programas GeneScan<sup>®</sup> e Genotyper<sup>®</sup> (Applied Biosystems).

### **3.9 Análises dos dados**

Após os testes de otimização e reprodutibilidade, foram geradas genotipagens de amostras oriundas dos plantios no Brasil e no Paraguai. Os dados dos fragmentos em pares de base (pb) foram obtidos pelo programa GeneScan<sup>®</sup> (Applied Biosystems) a partir das corridas eletroforéticas em gel de poliacrilamida 5%, realizadas em seqüenciador automático de DNA ABI Prism<sup>®</sup> 377 e analisados pelo programa Genotyper<sup>®</sup> (Applied Biosystems) (Figura 7). Os genótipos homozigotos e heterozigotos foram deduzidos a partir dos padrões de bandas e as análises dos eletroferogramas permitiram a genotipagem segura das amostras com onze dos treze locos propostos.

Os genótipos obtidos para as 477 amostras foram tabulados e utilizados para gerar arquivos de entrada para os programas GENALEX V6.0 (PEAKALL, 2006) e Arlequin ver. 3.0 (EXCOFFIER, 2005).





**Figura 7** – Alelos referentes aos 11 locos da amostra 29PY3, analisados pelos programas GeneScan® e Genotyper®.

### 3.9.1 *Frequência alélica*

A análise descritiva da frequência alélica de cada um dos locos estudados foi realizada identificando o número de alelos por loco e a distribuição de frequência e foi realizada com o auxílio dos algoritmos implementados no programa GENALEX V6.0 (PEAKALL, 2006).

### 3.9.2 *Diversidade gênica*

A heterozigose ou diversidade gênica foi estimada utilizando algoritmo implementado no programa GENALEX V6.0 (PEAKALL, 2006) utilizando a fórmula:

$$He = 1 - \sum p_i^2$$

### 3.9.3 *Equilíbrio de Hardy-Weinberg*

A investigação de desvios do equilíbrio de Hardy-Weinberg foi testada utilizando o procedimento proposto por Guo e Thompson (1992) e implementado no programa Arlequin ver. 3.0 (EXCOFFIER, 2005). Esta metodologia avalia se as diferenças entre as frequências

genotípicas observadas e as frequências esperadas de acordo com a equação de Hardy-Weinberg são significativas ou não, utilizando um teste similar ao teste exato de Fisher. Com este método pode-se trabalhar com tabelas onde células com valores absolutos abaixo de cinco são encontradas, que comumente acontece com locos multialélicos.

#### *3.9.4 Análise de desequilíbrio de ligação entre os locos*

O desequilíbrio de ligação entre pares de locos foi testado utilizando uma razão de verossimilhança proposta por Slatkin e Excoffier (1996). A razão de verossimilhança é calculada dividindo a verossimilhança do par de locos estarem em equilíbrio de ligação (as frequências haplotípicas são obtidas pelo produto das frequências alélicas) pela verossimilhança do par de locos não estarem em equilíbrio de ligação. Este valor é obtido utilizando algoritmos de maximização de expectativa (EM, do inglês *expectation maximization*). O algoritmo para realização dos cálculos está implementado no programa Arlequin ver. 3.0 (EXCOFFIER, 2005).

#### *3.9.5 Análise dos microssatélites quanto a erros de genotipagem, alelos nulos e amplificação preferencial de alelos curtos*

A obtenção do genótipo correto para um dado microssatélite em uma amostra é de extrema importância para as análises populacionais utilizando estes marcadores. No entanto, em algumas circunstâncias erros de interpretação devido a presença de gaguejos (do inglês, *stutters*) a amplificação preferencial de alelos de menor tamanho e a não amplificação de alelos devido a presença de polimorfismos no sítio de alinhamento dos iniciadores levam a obtenção de resultados que não refletem a variabilidade real de um determinado loco, levando de maneira geral a um excesso de homozigotos (revisado em DEWOODY, 2006).

Oosterhout et al.(2004) desenvolveram um programa onde algoritmos para detecção dos problemas apontados acima foram implementados. Este programa chamado Micro-Checker, foi utilizado para análise dos dados obtidos neste trabalho e pode ser obtido gratuitamente no endereço eletrônico: <http://www.microchecker.hull.ac.uk/>.

### 3.9.6 *Estimativa do coeficiente de parentesco*

O histórico de que as plantações de *C. sativa* no Brasil e no Paraguai são mantidas com amostras de sementes coletadas em plantios anteriores, sugere a possibilidade de que a cada novo plantio uma grande proporção de plantas aparentadas cruzem entre si. Com o intuito de verificar se esta possibilidade realmente procede, o método desenvolvido por Wang (2002) foi aplicado para estimar o coeficiente de parentesco par a par entre indivíduos nos plantios. O algoritmo é implementado no programa Mark (RITLAND, 2004), gratuitamente disponibilizado no endereço eletrônico <http://genetics.forestry.ubc.ca/ritland/programs.html#relatedness>. A partir de simulações utilizando cadeias de Markov, o algoritmo fornece valores médios do coeficiente de parentesco para cada população analisada. Uma população de indivíduos não aparentados deve retornar valores para o coeficiente de parentesco próximo a zero. Um valor de coeficiente de parentesco de 0,5 vale para pares pais-filhos ou irmãos germanos.

### 3.9.7 *Análise de Variância molecular (AMOVA)*

A distribuição da variância em medidas de diversidade genética foi avaliada utilizando a metodologia de AMOVA proposta originalmente por Excoffier et al. (1992) e implementada no programa GENALEX V6.0 (PEAKALL, 2006). A análise de AMOVA foi realizada utilizando três níveis hierárquicos: de um grupo total, representado pelo total de indivíduos analisados, distribuídos em duas regiões distintas de plantio (Brasil e Paraguai) e

os doze plantios considerados individualmente. As distâncias genéticas foram estimadas utilizando-se os índices estatísticos  $\Phi_{ST}$  e  $R_{ST}$ , seguindo as fórmulas da Tabela 5, implementadas no programa GENALEX V6.0 (PEAKALL, 2006), com 10.000 permutações para o teste de significância. Ambos os índices são estimativas de diferenciação genética, análogas às estatísticas  $F_{ST}$  de WRIGHT. Os índices  $\Phi_{ST}$  podem ser usados para todos os marcadores moleculares e permitem a comparação de padrões de variância molecular entre marcadores dominantes e codominantes. Os índices  $R_{ST}$ , propostos por Slatkin (1995) são mais específicos para dados microssatélites e levam em conta o processo mutacional característico destes marcadores.

**Tabela 5** – Fórmulas dos índices estatísticos utilizados na análise de AMOVA e implementadas no programa GENALEX V6.0 (PEAKALL, 2006).

Índice estatístico	Fórmulas estatísticas		
$\Phi_{ST}$	$\Phi_{PT} = \frac{(V_{AP} + V_{AR})}{(V_{WP} + V_{AP} + V_{AR})}$	$\Phi_{RT} = \frac{V_{AR}}{(V_{WP} + V_{AP} + V_{AR})}$	$\Phi_{PR} = \frac{V_{AP}}{(V_{WP} + V_{AP})}$
$R_{ST}$	$R_{ST} = \frac{(V_{AP} + V_{AR})}{(V_{WP} + V_{AP} + V_{AR})}$	$R_{RT} = \frac{V_{AR}}{(V_{WP} + V_{AP} + V_{AR})}$	$R_{SR} = \frac{V_{AP}}{(V_{WP} + V_{AP})}$

Onde:

$V_{AP}$  – variância entre os plantios, dentro das regiões (Brasil e Paraguai);

$V_{AR}$  – variância entre as regiões (Brasil e Paraguai);

$V_{WP}$  – variância entre indivíduos, dentro dos plantios;

$\Phi_{PT}$ ,  $R_{ST}$  – proporção da variância entre plantios (nas regiões) em relação a variância total;

$\Phi_{RT}$ ,  $R_{RT}$  – proporção da variância entre regiões de plantios em relação a variância total;

$\Phi_{PR}$ ,  $R_{SR}$  – proporção da variância entre plantios (nas regiões) em relação as variâncias dentro e entre os plantios.

### 3.9.8 Análise de coordenadas principais (PCA)

A análise PCA (do inglês, *Principal Coordinate Analysis*) é uma técnica que permite encontrar e delinear os padrões principais em um conjunto de dados multivariados (por exemplo, múltiplos locos e múltiplas amostras) e foi utilizada neste trabalho com a finalidade de mostrar graficamente a relação genética entre os diferentes plantios. A matemática é

complexa, mas essencialmente a análise PCA é um método onde os principais eixos de variação são posicionados dentro de um conjunto multidimensional de dados. Para conjuntos de dados multidimensionais, cada eixo sucessivo explica proporcionalmente menos da variação total, tal que quando houver grupos distintos, os primeiros dois ou três eixos tipicamente revelarão a separação entre eles. A análise foi realizada no programa GENALEX V6.0 (PEAKALL, 2006), utilizando os dois primeiros eixos, por meio de um algoritmo implementado no programa e publicado por Orloci (1978). Como base de análise foi utilizada a matriz de distâncias genéticas baseadas no índice  $R_{ST}$ .

### 3.9.9 *Teste de designação*

O teste de designação, traduzido neste trabalho do termo em inglês *Assignment test*, utilizado neste estudo foi aquele desenvolvido por Paetkau (1995). O método baseia-se nas frequências alélicas a partir das quais é estimada a frequência de genótipos multilocos. O método designa um indivíduo para uma população na qual seu genótipo multiloco tenha a maior probabilidade de ocorrer. Muito embora o cálculo do teste de designação baseado em frequências tenha como premissas o Equilíbrio de Hardy-Weinberg e o equilíbrio de ligação, o método apresentou o melhor desempenho quando comparado com outros métodos que não levavam em conta as referidas premissas (Cornuet et al., 1999). O teste de designação baseado em frequências foi utilizado por meio de algoritmo implementado no programa GENALEX V6.0 (PEAKALL, 2006). Foram utilizadas configurações onde os doze plantios foram analisados como diferentes grupos e também configurações onde os plantios foram agrupados de acordo com a região de origem. Mais especificamente, os plantios da região de Pernambuco e Maranhão foram agrupados em dois grupos. Para a região do Paraguai cinco plantios foram agrupados como sendo um único grupo e o plantio PY6 foi tratado individualmente. Esta sistemática de análise foi baseada em resultados encontrados na análise de AMOVA. O teste de designação também foi aplicado distribuindo os plantios em dois grupos, um brasileiro e outro paraguaio.

## 4 RESULTADOS

### 4.1 Otimização dos multiplexes

Para otimização das reações em multiplex obtidas neste trabalho, diversos arranjos de misturas dos conjuntos de iniciadores, baseados no tamanho dos fragmentos amplificados esperados e na temperatura de pareamento ( $T_a$ ) de cada par de iniciadores foram realizados. Ainda, variações nos programas de amplificação em PCR e ajustes na concentração de  $MgCl_2$  e pH do tampão de PCR foram testados, até obter-se um conjunto de amplicons equilibrado e detectável. O processo de otimização levou ao estabelecimento de duas reações.

Durante as etapas de otimização os marcadores microssatélites ANUCS201 e ANUCS301 não exibiram amplificação nas amostras analisadas, sendo o conjunto total reduzido para 11 (onze) marcadores, com a melhor otimização obtida mostrada na Tabela 6.

**Tabela 6** - Conjuntos multiplexes organizados pela média das temperaturas de pareamento dos conjuntos de iniciadores, com suas concentrações na mistura de reação otimizadas para um equilíbrio dos amplicons.

Mix_Iniciadores I				Mix_Iniciadores II			
STRs	Iniciadores	Tm (°C)	Conc. na PCR (µM)	STRs	Iniciadores	Tm (°C)	Conc. na PCR (µM)
ANUCS202	F	54,0	0,5	ANUCS304	F	60,0	0,8
	R	54,0	0,5		R	56,0	0,8
ANUCS302	F	57,0	0,8	B01-CANN1	F	58,0	0,5
	R	56,0	0,8		R	59,0	0,5
ANUCS303	F	57,0	0,5	B05-CANN1	F	58,0	0,8
	R	56,0	0,5		R	58,0	0,8
ANUCS305	F	56,0	0,5	C08-CANN2	F	61,0	0,5
	R	56,0	0,5		R	61,0	0,5
H09-CANN2	F	53,0	0,5	H06-CANN2	F	60,0	0,5
	R	53,0	0,5		R	60,0	0,5
				H11-CANN1	F	58,0	0,8
					R	60,0	0,8
<i>Tm média (°C)</i>			<i>55,0</i>	<i>Tm média (°C)</i>			<i>59,0</i>
<i>Ta média aproximada (°C)</i>			<i>50,0</i>	<i>Ta média aproximada (°C)</i>			<i>54,0</i>

A reação de amplificação otimizada consta de um volume total de 20  $\mu$ l, contendo aproximadamente 10 ng de DNA genômico, tampão PCR 1X, 250  $\mu$ M de cada dNTP, 1 unidade de Taq polimerase (Pht, Phoneutria) e concentração de iniciadores conforme a Tabela 6. As amplificações foram conduzidas em termociclador Perkin Elmer modelo 9700 (Applied Biosystems, Foster City, CA, EUA) empregando um programa “*touchdown*”, com passo inicial de incubação a 94°C por cinco minutos, seguido de um passo de desnaturação a 94°C por um minuto, um passo de pareamento dos iniciadores a 5°C acima da temperatura média de pareamento ( $T_a$  média) de cada conjunto de iniciadores (Tabela 6) por um minuto, reduzindo 0,50°C a cada ciclo e finalizando com um passo de extensão a 72°C por um minuto, em um total de 15 ciclos; seguem-se 20 ciclos de um passo de desnaturação a 94°C por um minuto, um passo de pareamento dos iniciadores na temperatura média de pareamento ( $T_a$  média) de cada conjunto de iniciadores (Tabela 6) e finalizando com um passo de extensão a 72°C por um minuto, seguindo-se ainda uma extensão final a 72°C por 40 minutos.

## 4.2 Análise dos dados

### 4.2.1 *Diversidade alélica*

As estimativas de frequências alélicas, número de alelos e heterozigose esperada ( $H_e$ ) para cada plantio individualmente e para as duas regiões de plantio (Brasil e Paraguai), foram calculadas diretamente a partir dos genótipos obtidos usando o programa GENALEX V6.0 (PEAKALL, 2006) e são resumidas na Tabela 7. A Figura 8 mostra graficamente a distribuição de frequências alélicas para cada loco em de cada plantio e a Figura 9 a distribuição das frequências nas regiões de plantio (Brasil e Paraguai).

No total das 477 amostras de *C. sativa* genotipadas, representando plantios no Brasil e no Paraguai observa-se o padrão microssatélite típico de marcadores codominantes, com pelo menos dois alelos diferentes por loco. Um total de 75 alelos foi detectado no conjunto total de dados, com o número de alelos por loco variando de três alelos nos locos ANUCS303, ANUCS305 e H06-CANN2 a quinze alelos no loco C08-CANN2 (Tabela 7).

Para todos os onze locos, o tamanho do alelo microssatélite observado varia em múltiplos esperados de suas unidades de repetição. Por exemplo, o loco mais diverso, C08-CANN2 é um dinucleotídeo (GA)<sub>n</sub> com 15 alelos variando em múltiplos de duas unidades de pares de bases de 169 a 203 pb. A distribuição das freqüências alélicas desse loco para cada plantio individualmente e para as duas regiões de plantio (Brasil e Paraguai) é mostrada na Tabela 7 e na Figura 8. Nota-se que dentro de uma amplitude de 34 pares de bases, todos os alelos potenciais foram detectados no conjunto de amostras analisado e o loco tem uma distribuição bimodal, evidenciada nas amostras com origem em plantios do Brasil.

Com o número de amostras analisado, dentre os onze locos microssatélites, os locos ANUCS303, ANUCS305 e H06-CANN2 exibiram número reduzido de alelos, apresentando no conjunto total de dados apenas três alelos cada um, sendo que os locos ANUCS305 e H06-CANN2 mostram-se ainda monomórficos nas amostras oriundas de plantios no Brasil, indicando serem pouco informativos. Já os oito marcadores restantes, exibindo uma média de 8,25 alelos por loco, mostram-se mais informativos, sendo o loco C08-CANN2 o que apresentou o maior número de alelos, exibindo 15 alelos no conjunto total de amostras analisadas.



Comparando-se os resultados obtidos por Gilmore et al. (2003) para os locos comuns, foi observado um menor número de alelos no material avaliado neste trabalho. Nos dois locos comuns aos trabalhos: ANUCS202 e ANUCS303, o número de alelos encontrado nas 93 amostras analisadas por aqueles pesquisadores foram 18 e 6 alelos, respectivamente, enquanto em nossas 477 amostras são de oito e três alelos somente. Já se compararmos os nossos resultados com os dados de Alghanim e Almirall (2003), nos locos comuns, estes apresentam número de alelos semelhantes, um pouco maior nos locos C08-CANN2, B01-CANN1 e H09-CANN2, um pouco menor no loco H11-CANN1 e idêntica nos locos B05-CANN1 e H06- CANN2.

**Tabela 7** – Distribuição das frequências alélicas e heterozigose esperada (He), obtidas no programa GENALEX V6.0 (PEAKALL, 2006) dos locos microsatélites, para os plantios no Brasil e no Paraguai.

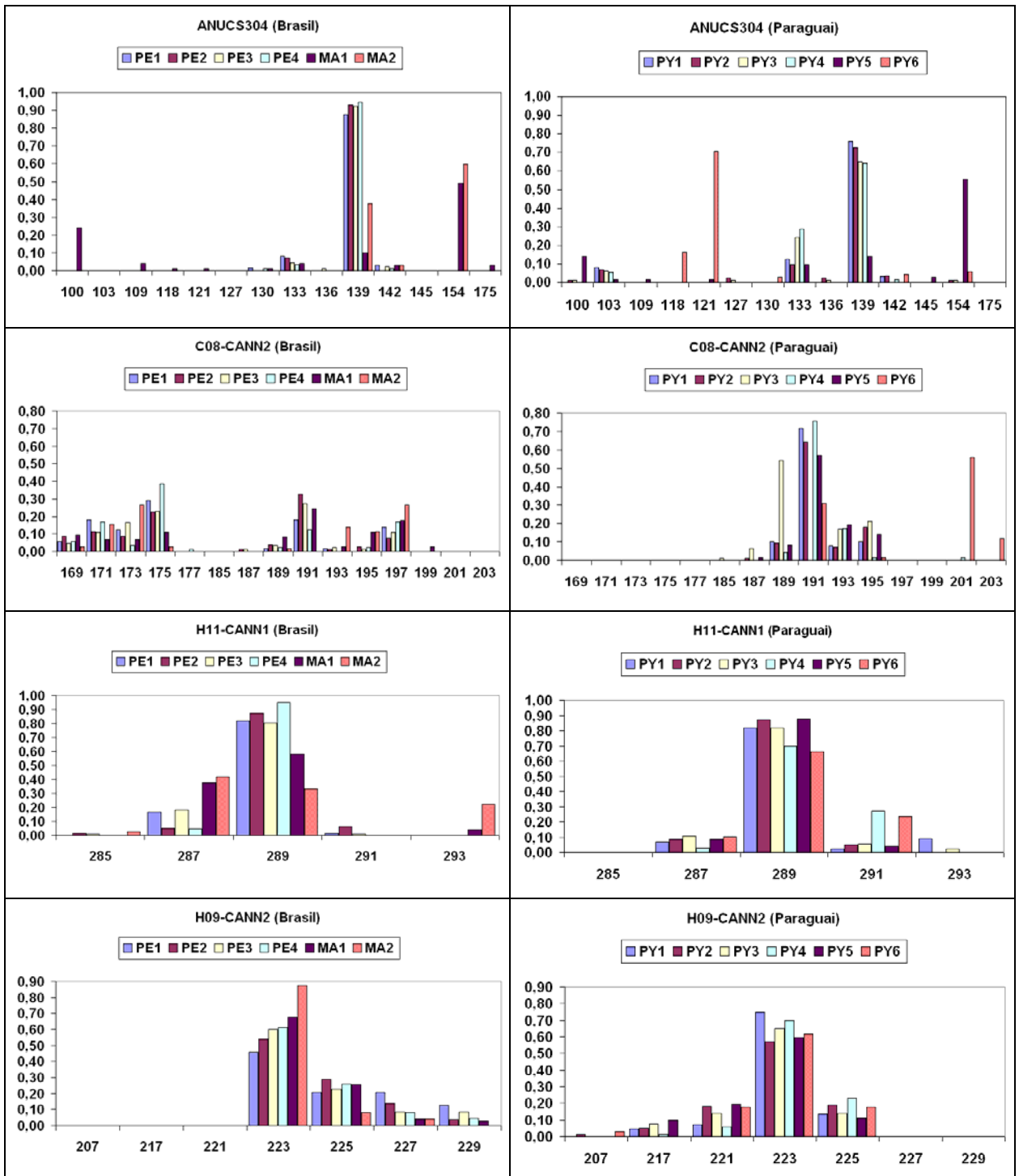
	Plantios no Brasil						Plantios no Paraguai						Total BR	Total PY
	PE1	PE2	PE3	PE4	MA1	MA2	PY1	PY2	PY3	PY4	PY5	PY6		
<i>Alelos</i> <b>ANUCS304</b>														
<b>100</b>	-	-	-	-	<b>0,24</b>	-	-	<b>0,01</b>	<b>0,01</b>	-	<b>0,14</b>	-	0,038	0,025
<b>103</b>	-	-	-	-	-	-	<b>0,08</b>	<b>0,07</b>	<b>0,06</b>	<b>0,06</b>	<b>0,01</b>	-	-	0,050
<b>109</b>	-	-	-	-	<b>0,04</b>	-	-	-	-	-	<b>0,01</b>	-	0,006	0,002
<b>118</b>	-	-	-	-	<b>0,01</b>	-	-	-	-	-	-	<b>0,16</b>	0,002	0,023
<b>121</b>	-	-	-	-	<b>0,01</b>	-	-	-	-	-	<b>0,01</b>	<b>0,71</b>	0,002	0,103
<b>127</b>	-	-	-	-	-	-	-	<b>0,02</b>	<b>0,01</b>	-	-	-	-	0,006
<b>130</b>	<b>0,01</b>	-	-	<b>0,01</b>	<b>0,01</b>	-	-	-	-	-	-	<b>0,03</b>	0,006	0,004
<b>133</b>	<b>0,08</b>	<b>0,07</b>	<b>0,04</b>	<b>0,04</b>	<b>0,04</b>	-	<b>0,13</b>	<b>0,10</b>	<b>0,25</b>	<b>0,29</b>	<b>0,10</b>	-	0,046	0,145
<b>136</b>	-	-	<b>0,01</b>	-	-	-	-	<b>0,02</b>	<b>0,01</b>	-	-	-	0,002	0,006
<b>139</b>	<b>0,88</b>	<b>0,93</b>	<b>0,92</b>	<b>0,94</b>	<b>0,10</b>	<b>0,38</b>	<b>0,76</b>	<b>0,73</b>	<b>0,65</b>	<b>0,64</b>	<b>0,14</b>	-	0,709	0,513
<b>142</b>	<b>0,03</b>	-	<b>0,02</b>	<b>0,01</b>	<b>0,03</b>	<b>0,03</b>	<b>0,03</b>	<b>0,04</b>	-	<b>0,01</b>	-	<b>0,04</b>	0,019	0,021
<b>145</b>	-	-	-	-	-	-	-	-	-	-	<b>0,03</b>	-	-	0,004
<b>154</b>	-	-	-	-	<b>0,49</b>	<b>0,60</b>	-	<b>0,01</b>	<b>0,01</b>	-	<b>0,56</b>	<b>0,06</b>	0,165	0,097
<b>175</b>	-	-	-	-	<b>0,03</b>	-	-	-	-	-	-	-	0,004	-
<b>He</b>	<i>0,226</i>	<i>0,139</i>	<i>0,144</i>	<i>0,109</i>	<i>0,690</i>	<i>0,502</i>	<i>0,397</i>	<i>0,456</i>	<i>0,514</i>	<i>0,502</i>	<i>0,642</i>	<i>0,469</i>	<i>0,466</i>	<i>0,692</i>
<i>Alelos</i> <b>C08-CANN2</b>														
<b>169</b>	<b>0,06</b>	<b>0,09</b>	<b>0,04</b>	<b>0,06</b>	<b>0,10</b>	<b>0,03</b>	-	-	-	-	-	-	0,061	-
<b>171</b>	<b>0,18</b>	<b>0,11</b>	<b>0,11</b>	<b>0,17</b>	<b>0,07</b>	<b>0,15</b>	-	-	-	-	-	-	0,132	-
<b>173</b>	<b>0,13</b>	<b>0,09</b>	<b>0,16</b>	<b>0,03</b>	<b>0,07</b>	<b>0,26</b>	-	-	-	-	-	-	0,121	-
<b>175</b>	<b>0,29</b>	<b>0,23</b>	<b>0,23</b>	<b>0,39</b>	<b>0,11</b>	<b>0,03</b>	-	-	-	-	-	-	0,218	-
<b>177</b>	-	-	-	<b>0,01</b>	-	-	-	-	-	-	-	-	0,002	-
<b>185</b>	-	-	-	-	-	-	-	-	<b>0,01</b>	-	-	-	-	0,002
<b>187</b>	-	<b>0,01</b>	<b>0,01</b>	-	-	-	-	<b>0,01</b>	<b>0,06</b>	-	<b>0,01</b>	-	0,004	0,017
<b>189</b>	<b>0,01</b>	<b>0,04</b>	<b>0,03</b>	<b>0,02</b>	<b>0,08</b>	<b>0,01</b>	<b>0,10</b>	<b>0,10</b>	<b>0,54</b>	<b>0,04</b>	<b>0,08</b>	-	0,033	0,162
<b>191</b>	<b>0,18</b>	<b>0,33</b>	<b>0,27</b>	<b>0,13</b>	<b>0,24</b>	-	<b>0,72</b>	<b>0,64</b>	-	<b>0,76</b>	<b>0,57</b>	<b>0,31</b>	0,195	0,487
<b>193</b>	<b>0,01</b>	<b>0,01</b>	<b>0,02</b>	-	<b>0,03</b>	<b>0,14</b>	<b>0,08</b>	<b>0,07</b>	<b>0,17</b>	<b>0,17</b>	<b>0,19</b>	-	0,033	0,116
<b>195</b>	-	<b>0,03</b>	<b>0,01</b>	<b>0,02</b>	<b>0,11</b>	<b>0,11</b>	<b>0,10</b>	<b>0,18</b>	<b>0,21</b>	<b>0,01</b>	<b>0,14</b>	<b>0,02</b>	0,044	0,118
<b>197</b>	<b>0,14</b>	<b>0,08</b>	<b>0,11</b>	<b>0,17</b>	<b>0,18</b>	<b>0,26</b>	-	-	-	-	-	-	0,153	-
<b>199</b>	-	-	-	-	<b>0,03</b>	-	-	-	-	-	-	-	0,004	-
<b>201</b>	-	-	-	-	-	-	-	-	-	<b>0,01</b>	-	<b>0,56</b>	-	0,082
<b>203</b>	-	-	-	-	-	-	-	-	-	-	-	<b>0,12</b>	-	0,017
<b>He</b>	<i>0,811</i>	<i>0,808</i>	<i>0,820</i>	<i>0,771</i>	<i>0,860</i>	<i>0,804</i>	<i>0,460</i>	<i>0,541</i>	<i>0,627</i>	<i>0,395</i>	<i>0,611</i>	<i>0,578</i>	<i>0,851</i>	<i>0,702</i>

Continuação da Tabela 7 – Distribuição das frequências alélicas e heterozigose esperada (He), obtidas no programa GENALEX V6.0 (PEAKALL, 2006) dos locos microssatélites, para os plantios no Brasil e no Paraguai.

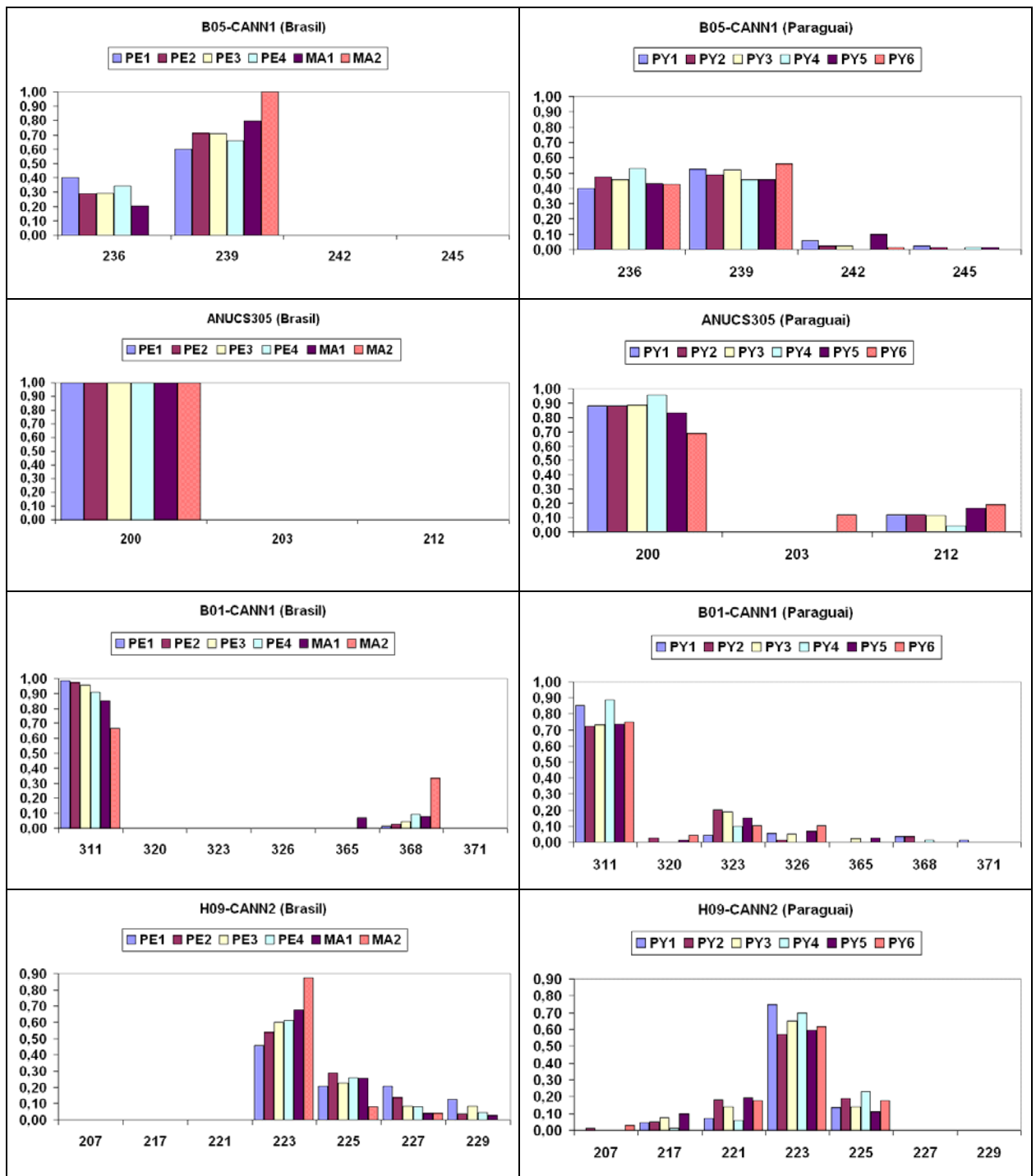
	Plantios no Brasil						Plantios no Paraguai						Total BR	Total PY	
	PE1	PE2	PE3	PE4	MA1	MA2	PY1	PY2	PY3	PY4	PY5	PY6			
<i>Alelos</i>														<i>H09-CANN2</i>	
207	-	-	-	-	-	-	-	0,01	-	-	-	0,03	-	0,006	
217	-	-	-	-	-	-	0,05	0,05	0,07	0,01	0,10	-	-	0,048	
221	-	-	-	-	-	-	0,07	0,18	0,14	0,06	0,19	0,18	-	0,134	
223	0,46	0,54	0,60	0,61	0,68	0,88	0,75	0,57	0,65	0,70	0,60	0,62	0,623	0,649	
225	0,21	0,29	0,23	0,26	0,26	0,08	0,14	0,19	0,14	0,23	0,11	0,18	0,224	0,162	
227	0,21	0,14	0,09	0,08	0,04	0,04	-	-	-	-	-	-	0,098	-	
229	0,13	0,04	0,09	0,05	0,03	-	-	-	-	-	-	-	0,054	-	
<i>He</i>	0,688	0,608	0,575	0,547	0,475	0,226	0,412	0,603	0,535	0,454	0,584	0,555	0,549	0,532	
<i>Alelos</i>														<i>B01-CANN1</i>	
311	0,99	0,98	0,96	0,91	0,85	0,67	0,85	0,73	0,73	0,89	0,74	0,75	0,895	0,779	
320	-	-	-	-	-	-	-	0,02	-	-	0,01	0,04	-	0,013	
323	-	-	-	-	-	-	0,05	0,20	0,19	0,10	0,15	0,10	-	0,134	
326	-	-	-	-	-	-	0,06	0,01	0,05	-	0,07	0,10	-	0,048	
365	-	-	-	-	0,07	-	-	-	0,02	-	0,03	-	0,010	0,008	
368	0,01	0,03	0,04	0,09	0,08	0,33	0,03	0,04	-	0,01	-	-	0,094	0,015	
371	-	-	-	-	-	-	0,01	-	-	-	-	-	-	0,002	
<i>He</i>	0,027	0,049	0,083	0,165	0,264	0,444	0,267	0,430	0,421	0,205	0,429	0,414	0,189	0,372	
<i>Alelos</i>														<i>ANUCS202</i>	
137	-	-	-	-	-	-	-	-	0,01	-	-	-	-	0,002	
139	0,14	0,15	0,32	0,17	0,18	0,19	0,22	0,06	0,10	0,49	0,14	0,07	0,195	0,172	
141	0,54	0,32	0,37	0,52	0,55	0,50	0,50	0,57	0,64	0,49	0,65	0,59	0,467	0,574	
143	0,32	0,48	0,30	0,30	0,23	0,25	0,28	0,33	0,25	0,02	0,21	0,34	0,312	0,246	
145	-	0,05	-	0,01	0,03	0,06	-	0,03	-	-	-	-	0,023	0,004	
147	-	-	-	-	0,01	-	-	-	-	-	-	-	0,002	-	
149	-	-	-	-	-	-	-	0,01	-	-	-	-	-	0,002	
173	-	-	0,01	-	-	-	-	-	-	-	-	-	0,002	-	
<i>He</i>	0,585	0,644	0,670	0,610	0,608	0,647	0,623	0,558	0,518	0,527	0,511	0,534	0,647	0,581	
<i>Alelos</i>														<i>ANUCS303</i>	
118	-	0,05	0,05	0,02	0,26	0,32	0,02	0,13	0,12	0,04	0,04	0,10	0,111	0,078	
121	1,00	0,95	0,95	0,98	0,74	0,68	0,98	0,87	0,85	0,94	0,92	0,81	0,889	0,895	
124	-	-	-	-	-	-	-	-	0,03	0,01	0,04	0,09	-	0,027	
<i>He</i>	0,000	0,095	0,103	0,044	0,382	0,435	0,044	0,228	0,261	0,109	0,156	0,327	0,197	0,192	

**Continuação da Tabela 7** – Distribuição das frequências alélicas e heterozigose esperada ( $H_e$ ), obtidas no programa GENALEX V6.0 (PEAKALL, 2006) dos locos microsatélites, para os plantios no Brasil e no Paraguai. A heterozigose média para o conjunto de onze locos também foi calculada.

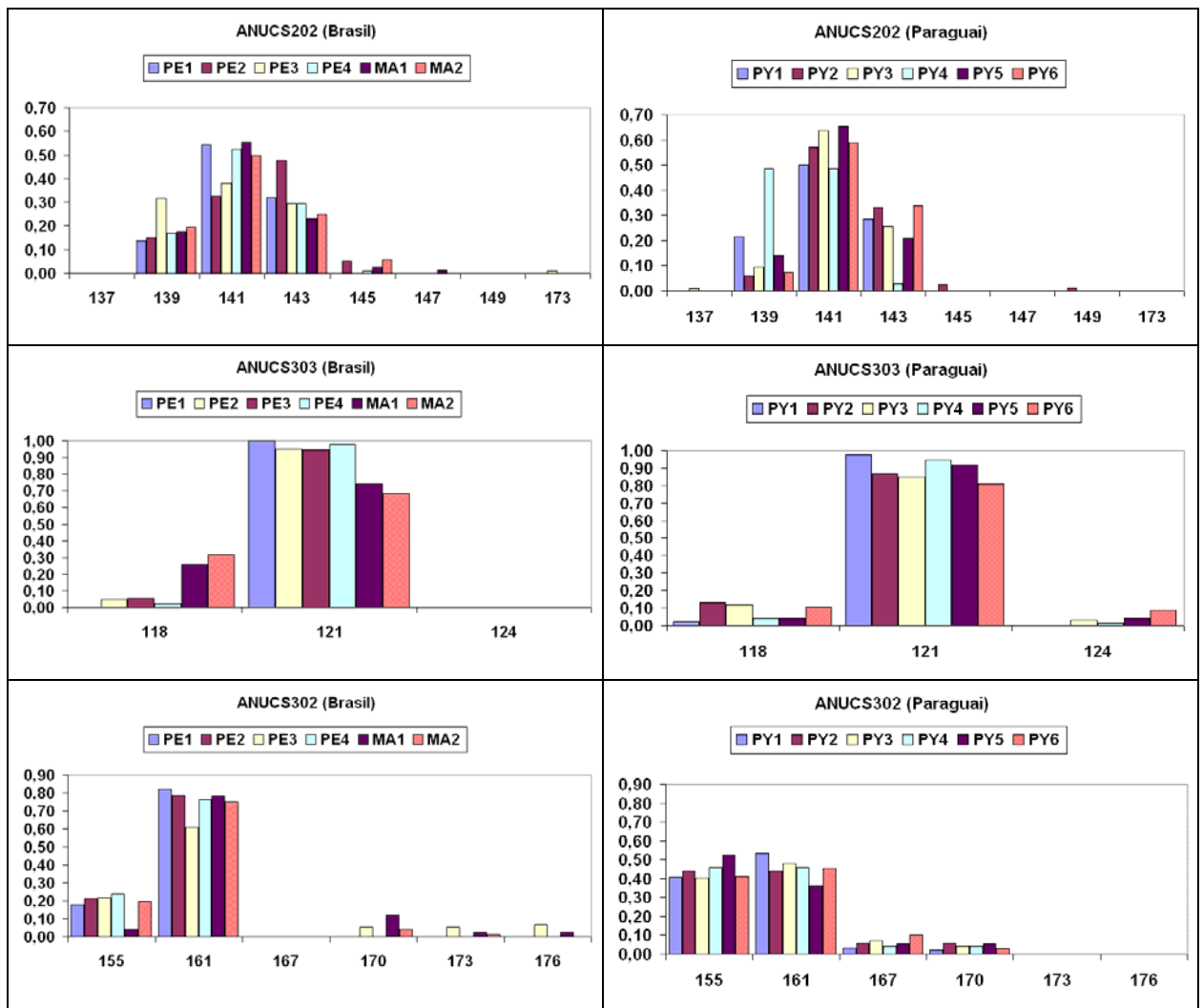
	Plantios no Brasil						Plantios no Paraguai						Total BR	Total PY	
	PE1	PE2	PE3	PE4	MA1	MA2	PY1	PY2	PY3	PY4	PY5	PY6			
<i>Alelos</i>														<b>H06-CANN2</b>	
<b>270</b>	-	-	-	-	-	-	<b>0,01</b>	<b>0,02</b>	<b>0,09</b>	-	<b>0,10</b>	<b>0,03</b>	-	0,042	
<b>273</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>0,93</b>	<b>0,94</b>	<b>0,86</b>	<b>0,97</b>	<b>0,89</b>	<b>0,94</b>	1,000	0,920	
<b>276</b>	-	-	-	-	-	-	<b>0,06</b>	<b>0,04</b>	<b>0,05</b>	<b>0,03</b>	<b>0,01</b>	<b>0,03</b>	-	0,038	
<b>He</b>	<i>0,000</i>	<i>0,000</i>	<i>0,000</i>	<i>0,000</i>	<i>0,000</i>	<i>0,000</i>	<i>0,128</i>	<i>0,114</i>	<i>0,247</i>	<i>0,056</i>	<i>0,200</i>	<i>0,112</i>	<i>0,000</i>	<i>0,150</i>	
<i>Alelos</i>														<b>ANUCS302</b>	
<b>155</b>	<b>0,18</b>	<b>0,21</b>	<b>0,22</b>	<b>0,24</b>	<b>0,04</b>	<b>0,19</b>	<b>0,41</b>	<b>0,44</b>	<b>0,40</b>	<b>0,46</b>	<b>0,53</b>	<b>0,41</b>	0,184	0,439	
<b>161</b>	<b>0,82</b>	<b>0,79</b>	<b>0,61</b>	<b>0,76</b>	<b>0,78</b>	<b>0,75</b>	<b>0,53</b>	<b>0,44</b>	<b>0,48</b>	<b>0,46</b>	<b>0,36</b>	<b>0,46</b>	0,747	0,458	
<b>167</b>	-	-	-	-	-	-	<b>0,03</b>	<b>0,06</b>	<b>0,07</b>	<b>0,04</b>	<b>0,06</b>	<b>0,10</b>	-	0,061	
<b>170</b>	-	-	<b>0,05</b>	-	<b>0,12</b>	<b>0,04</b>	<b>0,02</b>	<b>0,06</b>	<b>0,04</b>	<b>0,04</b>	<b>0,06</b>	<b>0,03</b>	0,036	0,042	
<b>173</b>	-	-	<b>0,05</b>	-	<b>0,03</b>	<b>0,01</b>	-	-	-	-	-	-	0,017	-	
<b>176</b>	-	-	<b>0,07</b>	-	<b>0,03</b>	-	-	-	-	-	-	-	0,017	-	
<b>He</b>	<i>0,296</i>	<i>0,335</i>	<i>0,572</i>	<i>0,363</i>	<i>0,368</i>	<i>0,398</i>	<i>0,546</i>	<i>0,605</i>	<i>0,600</i>	<i>0,578</i>	<i>0,585</i>	<i>0,611</i>	<i>0,406</i>	<i>0,592</i>	
<i>Alelos</i>														<b>H11-CANN1</b>	
<b>285</b>	-	<b>0,01</b>	<b>0,01</b>	-	-	<b>0,03</b>	-	-	-	-	-	-	0,008	-	
<b>287</b>	<b>0,17</b>	<b>0,05</b>	<b>0,18</b>	<b>0,05</b>	<b>0,38</b>	<b>0,42</b>	<b>0,07</b>	<b>0,08</b>	<b>0,11</b>	<b>0,03</b>	<b>0,08</b>	<b>0,10</b>	0,197	0,080	
<b>289</b>	<b>0,82</b>	<b>0,88</b>	<b>0,80</b>	<b>0,95</b>	<b>0,58</b>	<b>0,33</b>	<b>0,82</b>	<b>0,87</b>	<b>0,82</b>	<b>0,70</b>	<b>0,88</b>	<b>0,66</b>	0,741	0,796	
<b>291</b>	<b>0,01</b>	<b>0,06</b>	<b>0,01</b>	-	-	-	<b>0,02</b>	<b>0,05</b>	<b>0,05</b>	<b>0,27</b>	<b>0,04</b>	<b>0,24</b>	0,015	0,103	
<b>293</b>	-	-	-	-	<b>0,04</b>	<b>0,22</b>	<b>0,09</b>	-	<b>0,02</b>	-	-	-	0,040	0,021	
<b>He</b>	<i>0,301</i>	<i>0,228</i>	<i>0,323</i>	<i>0,087</i>	<i>0,581</i>	<i>0,665</i>	<i>0,317</i>	<i>0,236</i>	<i>0,314</i>	<i>0,436</i>	<i>0,226</i>	<i>0,496</i>	<i>0,411</i>	<i>0,349</i>	
<i>Alelos</i>														<b>B05-CANN1</b>	
<b>236</b>	<b>0,40</b>	<b>0,29</b>	<b>0,29</b>	<b>0,34</b>	<b>0,20</b>	-	<b>0,40</b>	<b>0,48</b>	<b>0,46</b>	<b>0,53</b>	<b>0,43</b>	<b>0,43</b>	0,259	0,452	
<b>239</b>	<b>0,60</b>	<b>0,71</b>	<b>0,71</b>	<b>0,66</b>	<b>0,80</b>	<b>1,00</b>	<b>0,52</b>	<b>0,49</b>	<b>0,52</b>	<b>0,46</b>	<b>0,46</b>	<b>0,56</b>	0,741	0,502	
<b>242</b>	-	-	-	-	-	-	<b>0,06</b>	<b>0,02</b>	<b>0,02</b>	-	<b>0,10</b>	<b>0,01</b>	-	0,036	
<b>245</b>	-	-	-	-	-	-	<b>0,02</b>	<b>0,01</b>	-	<b>0,01</b>	<b>0,01</b>	-	-	0,011	
<b>He</b>	<i>0,481</i>	<i>0,410</i>	<i>0,415</i>	<i>0,449</i>	<i>0,323</i>	<i>0,000</i>	<i>0,565</i>	<i>0,534</i>	<i>0,519</i>	<i>0,511</i>	<i>0,595</i>	<i>0,506</i>	<i>0,384</i>	<i>0,542</i>	
<i>Alelos</i>														<b>ANUCS305</b>	
<b>200</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>1,00</b>	<b>0,88</b>	<b>0,88</b>	<b>0,88</b>	<b>0,96</b>	<b>0,83</b>	<b>0,69</b>	1,000	0,857	
<b>203</b>	-	-	-	-	-	-	-	-	-	-	-	<b>0,12</b>	-	0,017	
<b>212</b>	-	-	-	-	-	-	<b>0,12</b>	<b>0,12</b>	<b>0,12</b>	<b>0,04</b>	<b>0,17</b>	<b>0,19</b>	-	0,126	
<b>He</b>	<i>0,000</i>	<i>0,000</i>	<i>0,000</i>	<i>0,000</i>	<i>0,000</i>	<i>0,000</i>	<i>0,219</i>	<i>0,210</i>	<i>0,207</i>	<i>0,082</i>	<i>0,278</i>	<i>0,472</i>	<i>0,000</i>	<i>0,249</i>	
<b>Heterozigose esperada média dos plantios no Brasil</b>												<b>0,406</b>			
<b>Heterozigose esperada média dos plantios no Paraguai</b>												<b>0,532</b>			



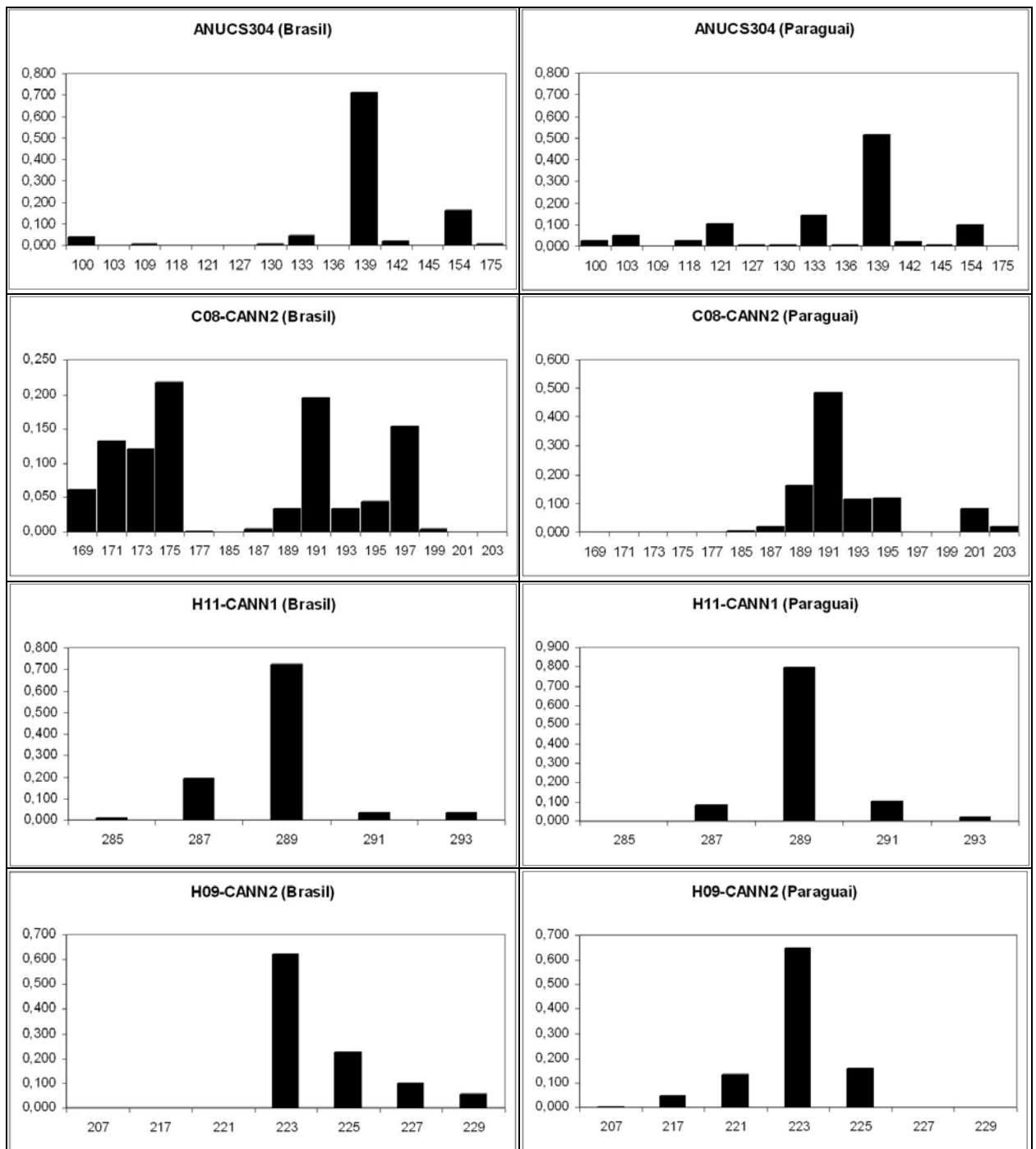
**Figura 8** – Gráficos mostrando a distribuição das freqüências alélicas dos onze locos microsatélites para cada um dos plantios no Brasil e no Paraguai.



Continuação da Figura 8 – Gráficos mostrando a distribuição das frequências alélicas dos onze locos microssatélites para cada um dos plantios no Brasil e no Paraguai.

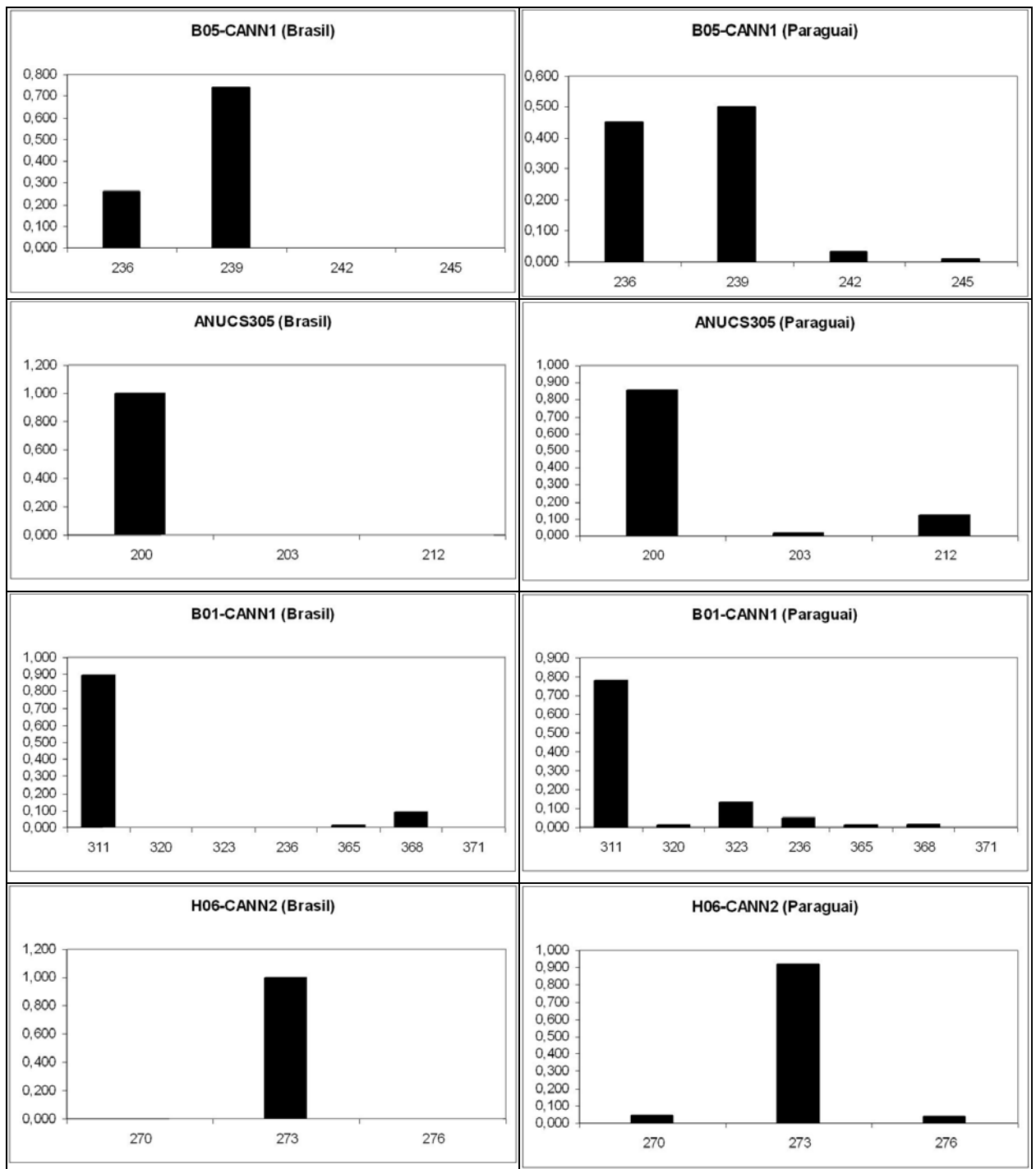


Continuação da Figura 8 – Gráficos mostrando a distribuição das freqüências alélicas dos onze locos microssatélites para cada um dos plantios no Brasil e no Paraguai.

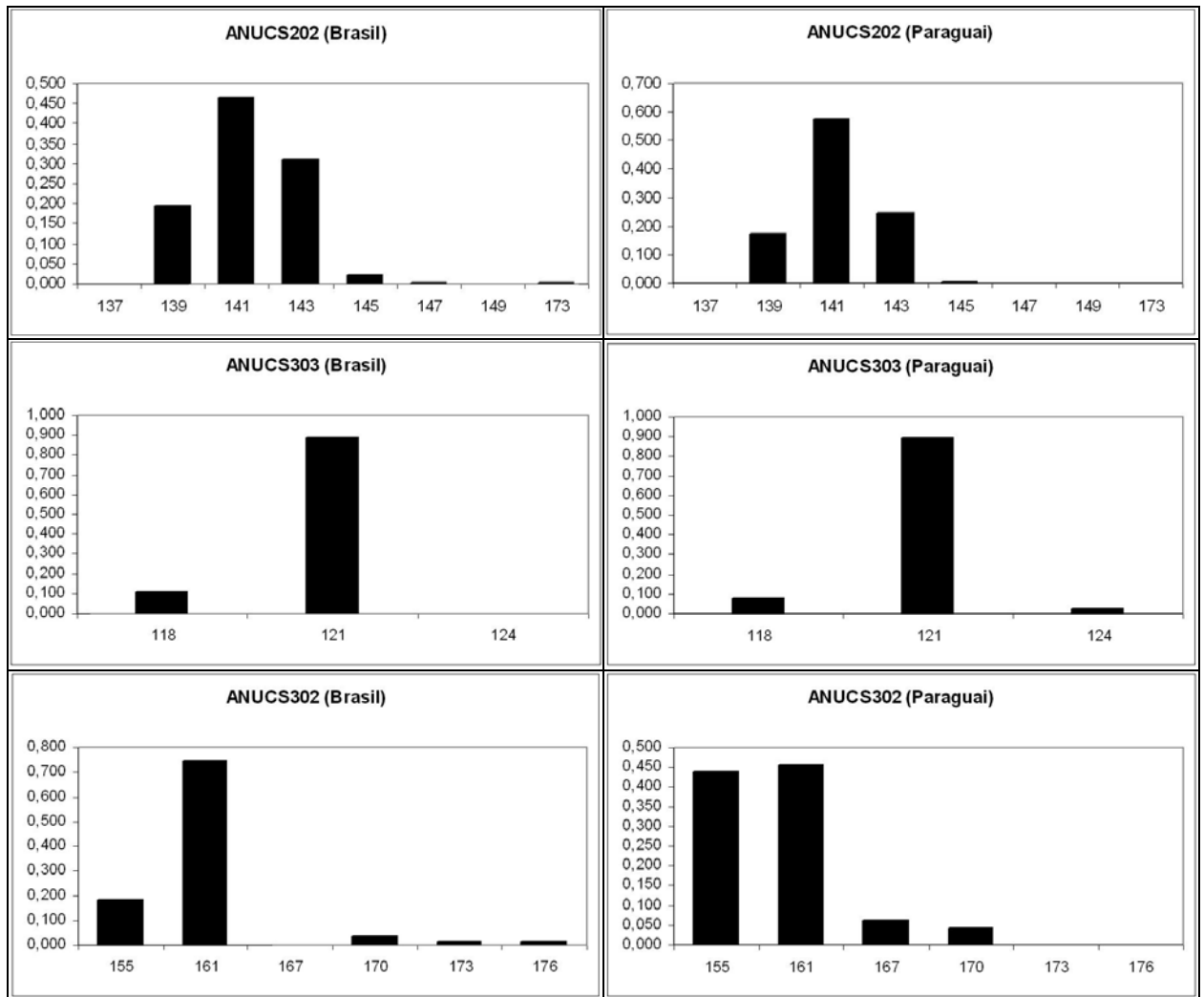


**Figura 9** – Gráficos mostrando a distribuição das freqüências alélicas dos onze locos microssatélites para plantios no Brasil e no Paraguai.





Continuação da Figura 9 – Gráficos mostrando a distribuição das frequências alélicas dos onze locos microssatélites para plantios no Brasil e no Paraguai.



**Continuação da Figura 9** – Gráficos mostrando a distribuição das frequências alélicas dos onze locos microssatélites para plantios no Brasil e no Paraguai.

#### 4.2.2 *Equilíbrio de Hardy-Weinberg*

O teste de hipótese de concordância com o equilíbrio de Hardy-Weinberg para as frequências genóticas observadas, realizado de acordo com o procedimento implementado por Guo e Thompson (1992), foi efetuado para cada um dos 12 plantios e para uma estrutura de análise onde os plantios foram agrupados de acordo com a estrutura apontada pelas distâncias de  $R_{ST}$ , ou seja, cinco agrupamentos representados pelos plantios de Pernambuco (**PE1**, **PE2**, **PE3**, **PE4** e **PE5**), os plantios do Paraguai (**PY1**, **PY2**, **PY3**, **PY4** e **PY5**), o plantio **MA1** (Maranhão), o plantio **MA2** (Maranhão) e o plantio **PY6** (Paraguai), sendo adotado o nível de significância de  $P < 0,05$ . Os resultados apontam para uma grande heterogeneidade entre os plantios no que diz respeito ao número de locos em concordância com o equilíbrio de Hardy-Weinberg. Quando os plantios são analisados separadamente os únicos locos que mostram homogeneidade em todos os plantios são os locos ANUCS304 e B01-CANN1, que respectivamente, não está de acordo e está de acordo com o equilíbrio de Hardy-Weinberg. Quando são analisados os cinco agrupamentos a heterogeneidade reduz-se um pouco, com cinco locos mostrando homogeneidade. Os locos ANUCS303 e ANUCS304 passam a discordar e os loco B01-CANN1, ANUCS302 e B05-CANN1 (exceto no plantio MA2, onde o loco é monomórfico) passam a concordar com o equilíbrio de Hardy-Weinberg. Os resultados são apresentados de maneira detalhada nas Tabelas 10 e 11, que também mostram os valores da heterozigose esperada e da heterozigose observada e os respectivos valores de  $F$ . As prováveis causas do grande número de locos que não estão de acordo com o equilíbrio de Hady-Weinberg serão apresentadas na discussão deste trabalho.

**Tabela 10** – Concordância com o equilíbrio de Hardy-Weinberg – Valores de heterozigose observada ( $H_o$ ), heterozigose esperada ( $H_e$ ), valores de  $p$  do teste estatístico utilizado para testar a concordância com o equilíbrio de Hardy-Weinberg e o valor do coeficiente de endogamia ( $F_{is}$ ). Em amarelo são evidenciados os locos onde a hipótese de concordância com o equilíbrio de Hardy-Weinberg é descartada. Em verde são evidenciados os locos monomórficos.

Plantio: PE1					Plantio: PE2				
Loco	$H_o$	$H_e$	$p$	$F_{is}$	Loco	$H_o$	$H_e$	$p$	$F_{is}$
ANUCS304	0,056	0,256	<0,001	0,783	ANUCS304	<0,001	0,165	<0,001	1,000
C08-CANN2	0,861	0,823	0,174	-0,047	C08-CANN2	0,750	0,820	0,237	0,085
H11-CANN1	0,250	0,332	0,355	0,246	H11-CANN1	0,200	0,253	0,282	0,208
B01-CANN1	0,028	0,055	1,000	0,496	B01-CANN1	0,050	0,074	1,000	0,322
ANUCS302	0,306	0,327	1,000	0,067	ANUCS302	0,375	0,339	0,659	-0,106
ANUCS305					ANUCS305				
B05-CANN1	0,472	0,515	1,000	0,084	B05-CANN1	0,275	0,433	0,049	0,364
H06-CANN2					H06-CANN2				
ANUCS303					ANUCS303	0,050	0,120	0,077	0,583
ANUCS202	0,139	0,602	<0,001	0,769	ANUCS202	0,275	0,665	<0,001	0,586
H09-CANN2	0,750	0,697	0,647	-0,076	H09-CANN2	0,625	0,616	1,000	-0,015
Multiloco	0,278	0,424	0,501	0,165	Multiloco	0,275	0,339	0,237	0,322
Plantio: PE3					Plantio: PE4				
Loco	$H_o$	$H_e$	$p$	$F_{is}$	Loco	$H_o$	$H_e$	$p$	$F_{is}$
ANUCS304	0,109	0,166	0,021	0,344	ANUCS304	0,045	0,132	<0,001	0,655
C08-CANN2	0,870	0,829	0,578	-0,049	C08-CANN2	0,795	0,784	0,270	-0,015
H11-CANN1	0,326	0,326	0,343	<0,001	H11-CANN1	0,091	0,109	1,000	0,169
B01-CANN1	0,087	0,105	1,000	0,171	B01-CANN1	0,136	0,188	0,292	0,274
ANUCS302	0,609	0,578	0,710	-0,052	ANUCS302	0,341	0,368	0,683	0,072
ANUCS305					ANUCS305				
B05-CANN1	0,500	0,419	0,286	-0,193	B05-CANN1	0,409	0,477	0,516	0,142
H06-CANN2					H06-CANN2				
ANUCS303	0,022	0,125	0,002	0,827	ANUCS303	0,045	0,067	1,000	0,323
ANUCS202	0,196	0,698	<0,001	0,720	ANUCS202	0,205	0,627	<0,001	0,674
H09-CANN2	0,500	0,582	0,415	0,140	H09-CANN2	0,523	0,567	0,662	0,078
Multiloco	0,326	0,419	0,343	0,140	Multiloco	0,205	0,368	0,516	0,169
Plantio: MA1					Plantio: MA2				
Loco	$H_o$	$H_e$	$p$	$F_{is}$	Loco	$H_o$	$H_e$	$p$	$F_{is}$
ANUCS304	0,459	0,700	<0,001	0,343	ANUCS304	0,639	0,520	0,001	-0,230
C08-CANN2	0,973	0,872	0,181	-0,115	C08-CANN2	0,917	0,815	0,012	-0,124
H11-CANN1	0,162	0,550	<0,001	0,705	H11-CANN1	0,278	0,692	<0,001	0,598
B01-CANN1	0,243	0,291	0,343	0,163	B01-CANN1	0,500	0,451	0,709	-0,109
ANUCS302	0,432	0,373	1,000	-0,160	ANUCS302	0,389	0,403	0,731	0,036
ANUCS305					ANUCS305				
B05-CANN1	0,405	0,328	0,305	-0,237	B05-CANN1				
H06-CANN2					H06-CANN2				
ANUCS303	0,243	0,387	0,034	0,371	ANUCS303	0,139	0,468	<0,001	0,703
ANUCS202	0,216	0,632	<0,001	0,658	ANUCS202	0,500	0,662	0,074	0,245
H09-CANN2	0,486	0,500	0,869	0,027	H09-CANN2	0,139	0,254	0,024	0,453
Multiloco	0,405	0,500	0,181	0,163	Multiloco	0,444	0,494	0,018	0,140

Continuação da Tabela 10

Plantio: PY1					Plantio: PY2				
Loco	Ho.	He.	$p$	$F_{is}$	Loco	Ho.	He.	$p$	$F_{is}$
ANUCS304	0,159	0,402	<0,001	0,604	ANUCS304	0,238	0,483	<0,001	0,507
C08-CANN2	0,409	0,468	0,111	0,125	C08-CANN2	0,548	0,547	0,667	-0,001
H11-CANN1	0,045	0,343	<0,001	0,867	H11-CANN1	0,119	0,238	0,003	0,501
B01-CANN1	0,227	0,270	0,120	0,159	B01-CANN1	0,476	0,435	0,716	-0,095
ANUCS302	0,591	0,552	0,115	-0,071	ANUCS302	0,452	0,622	0,008	0,273
ANUCS305	0,114	0,241	0,011	0,529	ANUCS305	0,190	0,212	0,457	0,103
B05-CANN1	0,523	0,592	0,293	0,117	B05-CANN1	0,571	0,541	0,882	-0,057
H06-CANN2	0,114	0,151	0,063	0,247	H06-CANN2	0,071	0,137	0,047	0,480
ANUCS303	<0,001	0,067	0,011	1,000	ANUCS303	0,119	0,251	0,013	0,526
ANUCS202	0,500	0,641	0,013	0,220	ANUCS202	0,476	0,586	0,216	0,187
H09-CANN2	0,341	0,418	0,001	0,185	H09-CANN2	0,571	0,610	0,613	0,063
Multiloco	0,227	0,402	0,013	0,220	Multiloco	0,452	0,483	0,216	0,187
Plantio: PY3					Plantio: PY4				
Loco	Ho.	He.	$p$	$F_{is}$	Loco	Ho.	He.	$p$	$F_{is}$
ANUCS304	0,298	0,540	<0,001	0,448	ANUCS304	0,143	0,528	<0,001	0,730
C08-CANN2	0,617	0,637	0,504	0,032	C08-CANN2	0,314	0,401	0,262	0,216
H11-CANN1	0,106	0,319	<0,001	0,666	H11-CANN1	0,057	0,469	<0,001	0,878
B01-CANN1	0,468	0,441	0,934	-0,061	B01-CANN1	0,229	0,234	1,000	0,021
ANUCS302	0,638	0,607	0,790	-0,052	ANUCS302	0,657	0,587	0,399	-0,120
ANUCS305	0,149	0,228	0,097	0,346	ANUCS305	0,086	0,111	1,000	0,225
B05-CANN1	0,574	0,534	0,494	-0,076	B05-CANN1	0,714	0,519	0,029	-0,377
H06-CANN2	0,234	0,268	0,125	0,128	H06-CANN2	0,057	0,084	1,000	0,320
ANUCS303	0,213	0,282	0,038	0,245	ANUCS303	0,114	0,137	1,000	0,169
ANUCS202	0,511	0,524	0,199	0,025	ANUCS202	0,143	0,549	<0,001	0,740
H09-CANN2	0,447	0,556	0,122	0,196	H09-CANN2	0,486	0,487	0,682	0,003
Multiloco	0,447	0,524	0,125	0,128	Multiloco	0,143	0,469	0,399	0,216
Plantio: PY5					Plantio: PY6				
Loco	Ho.	He.	$p$	$F_{is}$	Loco	Ho.	He.	$p$	$F_{is}$
ANUCS304	0,083	0,676	<0,001	0,877	ANUCS304	0,235	0,482	<0,001	0,512
C08-CANN2	0,611	0,644	0,470	0,052	C08-CANN2	0,147	0,599	<0,001	0,754
H11-CANN1	0,139	0,253	0,024	0,451	H11-CANN1	0,118	0,529	<0,001	0,778
B01-CANN1	0,500	0,435	0,920	-0,149	B01-CANN1	0,382	0,442	0,258	0,136
ANUCS302	0,667	0,593	0,601	-0,124	ANUCS302	0,618	0,620	0,628	0,004
ANUCS305	0,222	0,309	0,232	0,281	ANUCS305	0,324	0,482	<0,001	0,329
B05-CANN1	0,528	0,603	0,448	0,125	B05-CANN1	0,529	0,529	1,000	<0,001
H06-CANN2	0,167	0,203	0,355	0,179	H06-CANN2	0,059	0,142	0,029	0,586
ANUCS303	0,167	0,158	1,000	-0,052	ANUCS303	0,088	0,361	<0,001	0,755
ANUCS202	0,361	0,536	0,002	0,327	ANUCS202	0,471	0,569	0,640	0,173
H09-CANN2	0,444	0,608	0,027	0,269	H09-CANN2	0,588	0,564	0,325	-0,044
Multiloco	0,361	0,536	0,355	0,179	Multiloco	0,324	0,529	0,029	0,329

**Tabela 11** – Concordância com o equilíbrio de Hardy-Weinberg – Valores de heterozigose observada ( $H_o$ ), heterozigose esperada ( $H_e$ ), valores de  $p$  do teste estatístico utilizado para testar a concordância com o equilíbrio de Hardy-Weinberg e o valor do coeficiente de endogamia ( $F_{IS}$ ). Em amarelo são evidenciados os locos onde a hipótese de concordância com o equilíbrio de Hardy-Weinberg é descartada. Em verde são evidenciados os locos monomórficos. A análise aqui envolve cinco grupos. PE representa todos os plantios de Pernambuco e PY(1-5) representam os plantios Paraguaiois de 1 a 5.

Plantio: PE					Plantio: MA1				
Loco	$H_o$	$H_e$	$p$	$F_{IS}$	Loco	$H_o$	$H_e$	$p$	$F_{IS}$
ANUCS304	0,054	0,159	<0,001	0,659	ANUCS304	0,459	0,700	0,010	0,343
C08-CANN2	0,819	0,821	0,188	0,002	C08-CANN2	0,973	0,872	0,168	-0,115
H11-CANN1	0,217	0,247	0,082	0,122	H11-CANN1	0,162	0,550	<0,001	0,705
B01-CANN1	0,078	0,092	0,283	0,151	B01-CANN1	0,243	0,291	0,360	0,163
ANUCS302	0,416	0,416	0,380	0,000	ANUCS302	0,432	0,373	1,000	-0,160
ANUCS305					ANUCS305				
B05-CANN1	0,416	0,448	0,489	0,073	B05-CANN1	0,405	0,328	0,304	-0,237
H06-CANN2					H06-CANN2				
ANUCS303	0,030	0,070	<0,001	0,570	ANUCS303	0,243	0,387	0,033	0,371
ANUCS202	0,205	0,653	<0,001	0,686	ANUCS202	0,216	0,632	<0,001	0,658
H09-CANN2	0,590	0,610	0,822	0,032	H09-CANN2	0,486	0,500	0,876	0,027
Multiloco	0,217	0,416	0,188	0,122	Multiloco	0,405	0,500	0,168	0,163
Plantio: MA2					Plantio: PY(1-5)				
Loco	$H_o$	$H_e$	$p$	$F_{IS}$	Loco	$H_o$	$H_e$	$p$	$F_{IS}$
ANUCS304	0,639	0,520	0,001	-0,230	ANUCS304	0,191	0,600	<0,001	0,681
C08-CANN2	0,917	0,815	0,007	-0,124	C08-CANN2	0,505	0,662	<0,001	0,238
H11-CANN1	0,278	0,692	<0,001	0,598	H11-CANN1	0,093	0,323	<0,001	0,711
B01-CANN1	0,500	0,451	0,711	-0,109	B01-CANN1	0,382	0,364	0,913	-0,050
ANUCS302	0,389	0,403	0,751	0,036	ANUCS302	0,598	0,590	0,492	-0,014
ANUCS305					ANUCS305	0,152	0,209	0,002	0,272
B05-CANN1					B05-CANN1	0,578	0,554	0,283	-0,045
H06-CANN2					H06-CANN2	0,132	0,161	0,010	0,179
ANUCS303	0,139	0,468	<0,001	0,703	ANUCS303	0,123	0,173	0,001	0,290
ANUCS202	0,500	0,662	0,072	0,245	ANUCS202	0,412	0,589	<0,001	0,301
H09-CANN2	0,139	0,254	0,021	0,453	H09-CANN2	0,456	0,529	0,003	0,138
Multiloco	0,444	0,494	0,014	0,140	Multiloco	0,382	0,529	0,002	0,238
Plantio: PY6									
Loco	$H_o$	$H_e$	$p$	$F_{IS}$					
ANUCS304	0,235	0,482	<0,001	0,512					
C08-CANN2	0,147	0,599	<0,001	0,754					
H11-CANN1	0,118	0,529	<0,001	0,778					
B01-CANN1	0,382	0,442	0,277	0,136					
ANUCS302	0,618	0,620	0,637	0,004					
ANUCS305	0,324	0,482	<0,001	0,329					
B05-CANN1	0,529	0,529	1,000	0,000					
H06-CANN2	0,059	0,142	0,031	0,586					
ANUCS303	0,088	0,361	<0,001	0,755					
ANUCS202	0,471	0,569	0,622	0,173					
H09-CANN2	0,588	0,564	0,325	-0,044					
Multiloco	0,324	0,529	0,031	0,329					

#### 4.2.3 *Independência entre locos (desequilíbrio de ligação)*

A análise de desequilíbrio de ligação para os 11 locos, resultou em 55 comparações par a par quando realizada individualmente nos 12 plantios e apresentou resultados heterogêneos. Com nível de significância estabelecido para valores  $\leq 0,05$ , o número de pares de locos em desequilíbrio de ligação variou de um par de locos no plantio **PE2**, localizado no Estado de Pernambuco/Brasil, até 37 pares de locos em desequilíbrio no plantio **PY1**, localizado no território paraguaio. As Tabelas 12 e 13 mostram os dados da análise de desequilíbrio de ligação para cada plantio, exibindo a significância e o valor de  $p$  para cada par de locos.

#### 4.2.4 *Análise dos microssatélites quanto a erros de genotipagem, alelos nulos e amplificação preferencial de alelos curtos*

Os resultados obtidos com os algoritmos implementados no programa Micro-Checker, apontam alelos nulos ou erros de interpretação de heterozigotos como causa do excesso de homozigotos em alguns locos quando analisados dentro dos doze plantios ou quando analisados dentro da estrutura de cinco grupos. Os cinco grupos seriam representados por todos os plantios de Pernambuco, MA1, MA2, plantios PY1-P5 e plantio PY6. A possibilidade de erro de interpretação de heterozigotos devido a presença de “*stutters*”, vem do fato de alguns locos apresentarem uma reduzida proporção de heterozigotos onde os alelos diferem por uma unidade repetitiva. A possibilidade de alelos nulos explicarem o excesso de homozigotos vem do fato dos homozigotos estarem distribuídos ao longo de todas as classes de alelos. A análise com os algoritmos implementados no Micro-Checker não apontam a não amplificação de alelos longos como causa do excesso de homozigotos. A Tabela 14 mostra quais locos apresentam evidências de erro de interpretação de heterozigotos e alelos nulos.

**Tabela 14** – Locos com evidências de alelos nulos e/ou com suspeita de erro de interpretação de heterozigotos – A letra **N** indica aqueles locos com evidências de alelos nulos e a letra **S** indica aqueles locos com suspeita de erro de interpretação de heterozigoto devido a presença de “*stutters*”.

Locos/Plantios	12 plantios												Plantios em 5 grupos				
	PE1	PE2	PE3	P34	MA1	MA2	PY1	PY2	PY3	PY4	PY5	PY6	PE	MA1	MA2	PY1-5	PY6
ANUCS304	N	N		N	N		N	N, S	N	N	N	N	N, S	N		N, S	N
C08-CANN2												N, S					N, S
H11-CANN1					N, S	N, S	N, S	N, S	N, S	N, S	N	N, S		N, S	N, S		N, S
B01-CANN1																	
ANUCS302																	
ANUCS305							N					N, S					N, S
B05-CANN1																	
H06-CANN2												N					N
ANUCS303			N, S			N, S	N, S	N, S				N, S	N, S		N, S	N, S	N, S
ANUCS202	N, S	N, S	N, S	N, S	N, S				N, S	N, S	N, S		N, S	N, S		N, S	
H09-CANN2											N, S				N		

#### 4.2.5 Estimativa do coeficiente de parentesco

Os valores médios do coeficiente de parentesco calculados de acordo com o método proposto por Wang (2002), utilizando algoritmo implementado no programa Mark são apresentados na Tabela 15.

**Tabela 15** – Coeficiente de Parentesco – Valores médios dos coeficientes de parentesco ( $r$ ) par a par entre indivíduos dentro dos plantios e a variância ( $V_r$ ) de sua estimativa são apresentados para os 12 plantios e para os cinco grupos.

	12 plantios												Plantios em 5 grupos				
	PE1	PE2	PE3	P34	MA1	MA2	PY1	PY2	PY3	PY4	PY5	PY6	PE	MA1	MA2	PY1-5	PY6
$r$	0,13	0,15	0,14	0,15	0,12	0,11	0,14	0,13	0,13	0,1	0,12	0,12	0,15	0,12	0,11	0,13	0,12
$V_r$	0,02	0,02	0,02	0,02	0,01	0,01	0,02	0,02	0,02	0,01	0,02	0,01	0,02	0,01	0,01	0,02	0,01





**Tabela 13**– Dados de desequilíbrio de ligação para os 06 plantios em território paraguaio. Acima da diagonal são apresentados os valores de  $p$  para cada par de locos e abaixo a significância. Legenda dos locos: 1=ANUCS304; 2=C08-CANN2; 3=H11-CANN1; 4=B01-CANN1; 5=ANUCS302; 6=ANUCS305; 7=B05-CANN1; 8=H06-CANN2; 9=ANUCS303; 10=ANUCS202; 11=H09-CANN2. Significância: (+) = par de locos onde  $p \leq 0,05$  (em vermelho).

PY1												PY2											
	1	2	3	4	5	6	7	8	9	10	11		1	2	3	4	5	6	7	8	9	10	11
1		<0,001	0,001	<0,001	0,001	<0,001	0,001	<0,001	1,000	0,344	<0,001	1		0,015	0,001	0,025	0,066	0,016	0,078	0,054	0,006	0,308	0,011
2	+		0,006	<0,001	0,001	<0,001	0,002	<0,001	1,000	0,706	<0,001	2	+		<0,001	0,004	0,082	0,010	0,099	0,221	0,009	0,624	0,011
3	+	+		0,011	<0,001	0,001	0,001	0,001	1,000	0,198	<0,001	3	+	+		0,002	0,002	0,007	0,029	1,000	0,040	0,104	0,001
4	+	+	+		<0,001	<0,001	<0,001	<0,001	1,000	0,267	<0,001	4	+	+	+		0,005	0,034	0,004	0,059	0,001	0,042	0,024
5	+	+	+	+		<0,001	<0,001	<0,001	1,000	0,076	<0,001	5	-	-	+	+		0,006	0,555	0,175	0,445	0,327	0,283
6	+	+	+	+	+		<0,001	<0,001	1,000	0,049	<0,001	6	+	+	+	+	+		0,695	0,284	0,024	0,327	0,616
7	+	+	+	+	+	+		0,003	1,000	0,654	<0,001	7	-	-	+	+	-	-		0,801	0,319	0,457	0,013
8	+	+	+	+	+	+	+		1,000	0,561	<0,001	8	-	-	-	-	-	-	-		0,040	0,145	0,423
9	-	-	-	-	-	-	-	-		1,000	1,000	9	+	+	+	+	-	+	-	+		0,622	0,141
10	-	-	-	-	-	+	-	-	-		0,171	10	-	-	-	+	-	-	-	-	-		0,254
11	+	+	+	+	+	+	+	+	-	-		11	+	+	+	+	-	-	+	-	-	-	

PY												PY4											
	1	2	3	4	5	6	7	8	9	10	11		1	2	3	4	5	6	7	8	9	10	11
1		0,617	0,083	0,167	0,552	0,444	0,247	0,017	0,006	0,437	0,017	1		0,367	0,837	0,207	0,951	0,375	0,943	0,620	0,017	0,131	0,452
2	-		0,424	0,029	0,034	0,143	0,122	0,012	0,094	0,004	0,105	2	-		0,959	0,339	0,767	0,251	0,451	0,699	0,243	0,882	0,123
3	-	-		0,059	0,056	0,019	0,125	<0,001	<0,001	0,138	0,194	3	-	-		0,146	0,004	0,006	0,700	0,619	0,178	0,265	0,424
4	-	+	-		0,082	0,524	0,291	<0,001	0,023	0,659	0,005	4	-	-	-		0,054	0,001	0,440	0,401	<0,001	0,128	0,116
5	-	+	-	-		0,017	0,828	0,028	0,066	0,114	0,059	5	-	-	+	-		0,021	0,153	0,738	0,060	0,380	0,837
6	-	-	+	-	+		0,014	0,164	0,141	0,076	0,003	6	-	-	+	+	+		0,528	1,000	0,049	0,220	0,025
7	-	-	-	-	-	+		0,277	0,569	0,374	0,329	7	-	-	-	-	-	-		1,000	0,812	0,795	0,587
8	+	+	+	+	+	-	-		<0,001	0,345	0,103	8	-	-	-	-	-	-	-		1,000	0,469	1,000
9	+	-	+	+	-	-	-	+		0,239	0,078	9	+	-	-	+	-	+	-	-		0,657	0,104
10	-	+	-	-	-	-	-	-	-		0,357	10	-	-	-	-	-	-	-	-	-		0,313
11	+	-	-	+	-	+	-	-	-	-		11	-	-	-	-	-	+	-	-	-	-	

PY5												PY6											
	1	2	3	4	5	6	7	8	9	10	11		1	2	3	4	5	6	7	8	9	10	11
1		0,356	0,031	0,294	0,666	0,465	0,317	0,104	0,015	0,023	0,268	1		0,296	0,228	0,115	0,358	0,361	0,690	0,658	<0,001	0,573	0,399
2	-		0,822	0,350	0,365	0,759	0,044	0,091	0,292	0,658	0,779	2	-		0,621	0,241	0,505	<0,001	0,382	0,620	0,351	0,258	0,051
3	+	-		0,412	0,013	0,379	0,109	0,325	0,176	0,141	0,559	3	-	-		0,594	0,805	0,090	0,786	0,619	0,040	0,424	0,559
4	-	-	-		0,396	0,331	0,266	0,268	0,917	0,356	0,753	4	-	-	-		0,004	0,001	0,505	0,985	<0,001	0,953	0,006
5	-	-	+	-		0,215	0,222	0,451	0,146	0,789	0,562	5	-	-	-	+		0,365	0,323	0,881	0,056	0,403	0,002
6	-	-	-	-	-		0,888	0,119	0,130	0,778	0,290	6	-	+	-	+	-		0,837	0,931	0,014	0,585	0,005
7	-	+	-	-	-	-		0,508	0,528	0,897	0,504	7	-	-	-	-	-	-		0,288	1,000	0,733	0,932
8	-	-	-	-	-	-	-		0,034	0,018	0,466	8	-	-	-	-	-	-	-		0,416	0,120	0,640
9	+	-	-	-	-	-	-	+		0,238	0,005	9	+	-	+	+	-	+	-	-		0,706	0,045
10	+	-	-	-	-	-	-	+	-		0,085	10	-	-	-	-	-	-	-	-	-		0,201
11	-	-	-	-	-	-	-	-	+	-		11	-	-	-	+	+	+	-	-	+	-	

#### 4.2.6 Probabilidade de identidade e Poder de exclusão

A probabilidade de identidade (PI) expressa a probabilidade de encontrar dois indivíduos com o mesmo genótipo para determinado loco na população. Os cálculos mostraram que com o conjunto de 477 amostras analisado a probabilidade de identidade varia de um mínimo de 0,04 (loco C08-CANN2) a um máximo de 0,85 (loco H06-CANN2). Admitindo-se por hipótese que cada loco segregasse independentemente, assumindo, portanto que estariam em equilíbrio de ligação – o que não está de acordo com o encontrado neste trabalho - a chance de encontrar genótipos idênticos, quando os 11 locos microssatélites são incluídos na análise corresponde ao produto das probabilidades de identidade de cada loco. A estimativa combinada é de  $3,65 \times 10^{-6}$ , significando que a probabilidade de encontrar dois indivíduos tomados ao acaso com o mesmo genótipo multilocos para os 11 locos é de aproximadamente 1 em cada 274.000 indivíduos. Ao analisar os dados deste trabalho encontramos 94% dos indivíduos com genótipos únicos, o equivalente a 449 das 477 plantas analisadas. Entre as 28 amostras que não apresentaram genótipos multilocos únicos, encontramos 14 pares de genótipos idênticos. Quatro pares do plantio PE4, três pares de PY1, dois pares de PY6, três pares envolvendo plantas dos plantios PY1 e PY2, um par PE2/PE4 e um par envolvendo plantas de PE4 e PY2. O poder de exclusão (PE) indica a probabilidade de excluir um indivíduo erroneamente apontado como genitor. Com o conjunto de microssatélites utilizado o poder de exclusão varia de 0,039 (loco H06-CANN2) a 0,692 (loco C08-CANN2), com uma média de 0,220. O poder de exclusão combinado, o qual é a probabilidade de exclusão considerando todos os 11 locos é de 98%.

#### 4.2.7 *Análise de estrutura genética*

##### **4.2.7.1 Análise de Variância molecular (AMOVA)**

Como esperado, devido à clara variação no número de alelos e na distribuição das frequências alélicas entre os plantios, a análise de variância molecular (AMOVA), procedimento estatístico que permite a partição hierárquica das variações genéticas entre populações e regiões e a estimativa de sua extensão, revela uma diferenciação genética significativa entre as regiões de plantio de *C. sativa* amostradas. A Tabela 16 mostra os dados de  $\phi_{ST}$  e  $R_{ST}$ , enquanto a Figura 12 exhibe graficamente a distribuição da variância entre os três níveis hierárquicos. A maior variabilidade encontra-se entre indivíduos dentro dos plantios sendo de 79% ou 77% da variabilidade total, quando utiliza-se respectivamente os índices  $R_{ST}$  e  $\phi_{ST}$ .

Tabela 16 – Resultados da AMOVA.

Fonte de Variação	$R_{ST}$			$\Phi_{ST}$		
	Variância	Índices estatísticos	Proporção da Variância	Variância	Índices estatísticos	Proporção da Variância
Entre regiões (Brasil e Paraguai)	59,825	$R_{RT} = 0,120$	12%	0,531	$\Phi_{RT} = 0,083$	8%
Entre plantios dentro das regiões	46,837	$R_{SR} = 0,107$	9%	0,937	$\Phi_{PR} = 0,159$	15%
Entre indivíduos dentro dos plantios	391,649	$R_{ST} = 0,214$	79%	4,962	$\Phi_{PT} = 0,228$	77%

Onde:

$\Phi_{PT}$ ,  $R_{ST}$  – proporção da variância entre plantios (nas regiões) em relação a variância total;

$\Phi_{RT}$ ,  $R_{RT}$  – proporção da variância entre regiões de plantios em relação a variância total;

$\Phi_{PR}$ ,  $R_{SR}$  – proporção da variância entre plantios (nas regiões) em relação as variâncias dentro e entre os plantios.

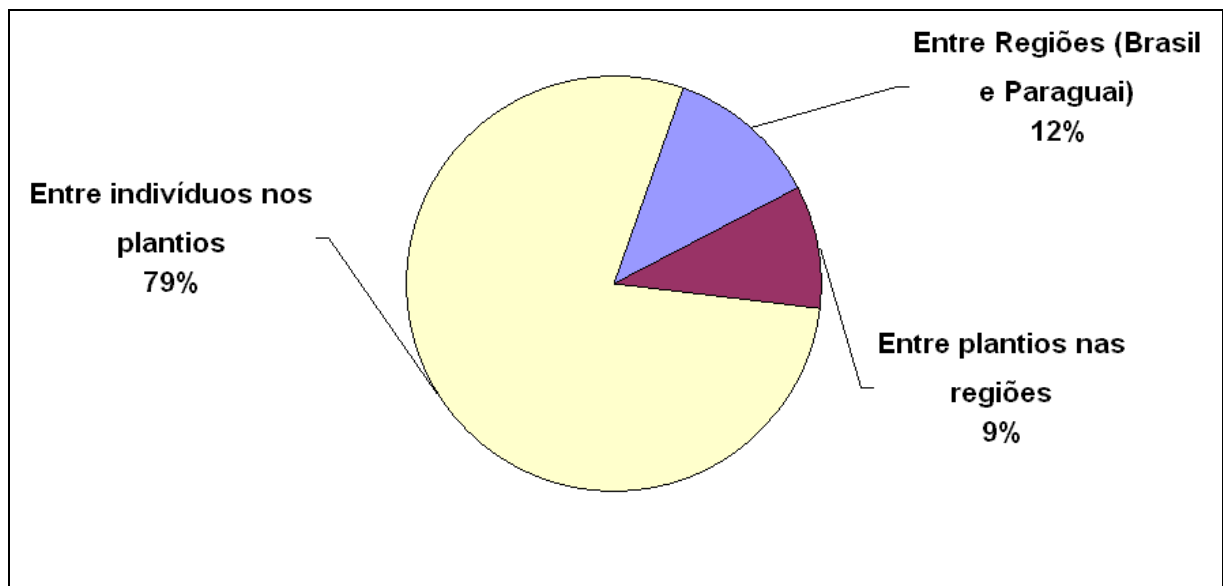


Figura 12 – Resultado da AMOVA, utilizando o índice  $R_{ST}$ , que mostra a porcentagem da variância molecular total particionada entre as diferenças entre as regiões de plantio (Brasil e Paraguai), entre plantios dentro das regiões e entre indivíduos dentro dos plantios.

#### 4.2.7.2 Distância genética utilizando $R_{ST}$

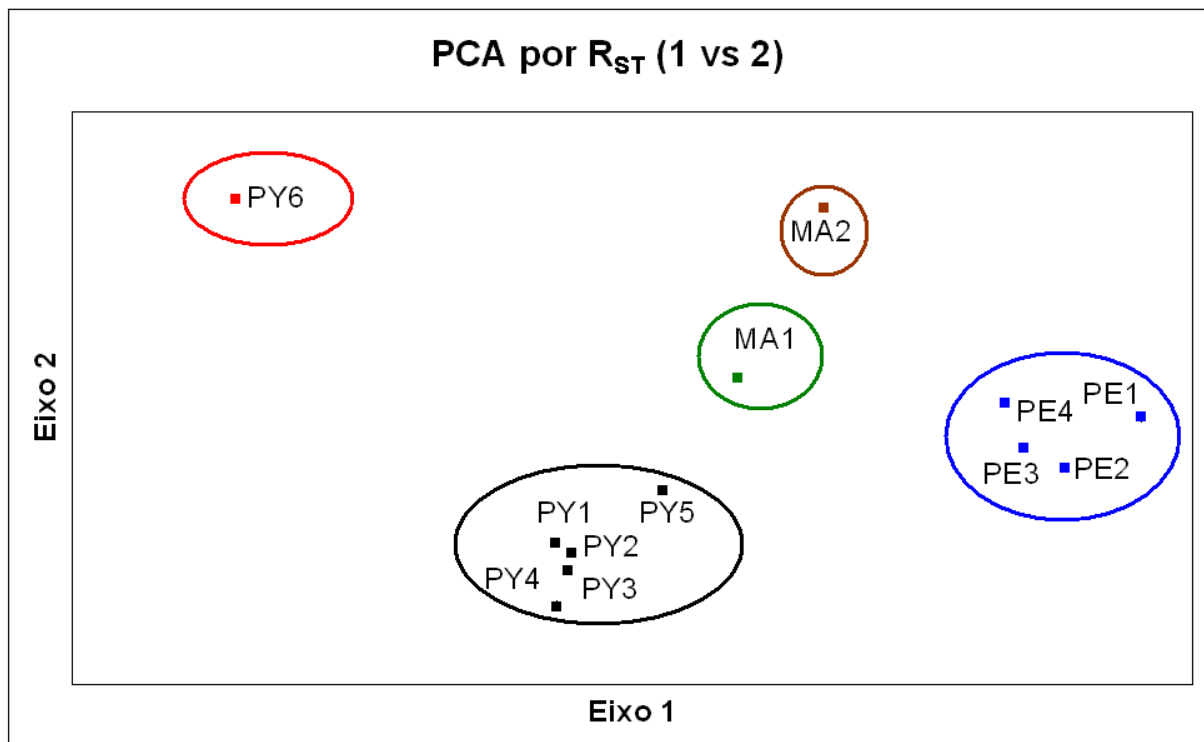
A análise par a par entre os 12 plantios foi realizada com 10.000 permutações e a Tabela 17 mostra abaixo da diagonal os valores das distâncias genéticas e acima da diagonal os valores de  $p$ . Os dados marcados referem-se aqueles com valor de  $p \leq 0,05$  e apontam para a presença de cinco grupos de acordo com as distâncias significativas de  $R_{ST}$ . De acordo com o nível de significância adotado, os dados indicam que os plantios de Pernambuco, identificados como **PE1**, **PE2**, **PE3** e **PE4** formam um grupo e os plantios do Paraguai, identificados como **PY1**, **PY2**, **PY3**, **PY4** e **PY5** formam o segundo grupo. Os plantios **MA2**, do Maranhão e **PY6**, do Paraguai não apresentam  $R_{ST}$  significativo com nenhum dos demais plantios, constituindo o terceiro e quarto grupos. O plantio **MA1** exibe similaridade com alguns plantios do Paraguai, porém próxima ao nível de significância, sendo também agrupado separadamente, constituindo-se no quinto grupo.

#### 4.2.7.3 Análise de coordenadas principais (PCA)

A análise de PCA (do inglês, *Principal Coordinates Analysis*) utilizando a matriz de distâncias genéticas entre os plantios mensuradas como  $R_{ST}$  e realizada no programa GENALEX V6.0 (PEAKALL, 2006), utilizando os dois primeiros eixos, evidencia a presença de pelos menos cinco agrupamentos representados por todos os plantios de Pernambuco (**PE1**, **PE2**, **PE3** e **PE4**), os plantios **MA1** e **MA2** do Maranhão, o plantio **PY6** do Paraguai de forma isolada e os cinco plantios do Paraguai (**PY1**, **PY2**, **PY3**, **PY4** e **PY5**). A análise é mostrada graficamente na Figura 13.

**Tabela 17** – Dados de análise de distância genética par a par entre os 12 plantios. Abaixo da diagonal temos os valores das distâncias genéticas e acima da diagonal os valores de  $p$ . Os valores em vermelho representam comparações par a par onde as distâncias são significativas.

	PE1	PE2	PE3	PE4	MA1	MA2	PY1	PY2	PY3	PY4	PY5	PY6		
Distâncias genéticas	PE1		0,363	0,171	0,075	0,003	0,000	0,000	0,000	0,000	0,000	0,000	0,000	PE1
	PE2	0,002		0,386	0,105	0,010	0,000	0,000	0,000	0,000	0,000	0,000	0,000	PE2
	PE3	0,010	0,000		0,179	0,024	0,000	0,000	0,000	0,000	0,000	0,000	0,000	PE3
	PE4	0,022	0,017	0,010		0,045	0,000	0,000	0,000	0,000	0,000	0,000	0,000	PE4
	MA1	0,068	0,049	0,035	0,030		0,000	0,045	0,057	0,051	0,023	0,062	0,000	MA1
	MA2	0,280	0,263	0,235	0,179	0,108		0,000	0,000	0,000	0,000	0,000	0,000	MA2
	PY1	0,238	0,170	0,156	0,156	0,033	0,264		0,898	0,907	0,453	0,188	0,000	PY1
	PY2	0,259	0,188	0,168	0,163	0,029	0,255	-0,009		0,904	0,138	0,157	0,000	PY2
	PY3	0,260	0,186	0,165	0,163	0,030	0,282	-0,008	-0,008		0,230	0,092	0,000	PY3
	PY4	0,304	0,217	0,189	0,193	0,049	0,305	0,001	0,014	0,010		0,107	0,000	PY4
	PY5	0,182	0,129	0,122	0,122	0,039	0,174	0,018	0,020	0,032	0,037		0,000	PY5
	PY6	0,608	0,554	0,500	0,470	0,191	0,469	0,265	0,263	0,288	0,339	0,296		PY6



**Figura 13** – Distribuição das duas primeiras coordenadas de análise PCA entre os plantios de *Cannabis sativa* estudados.

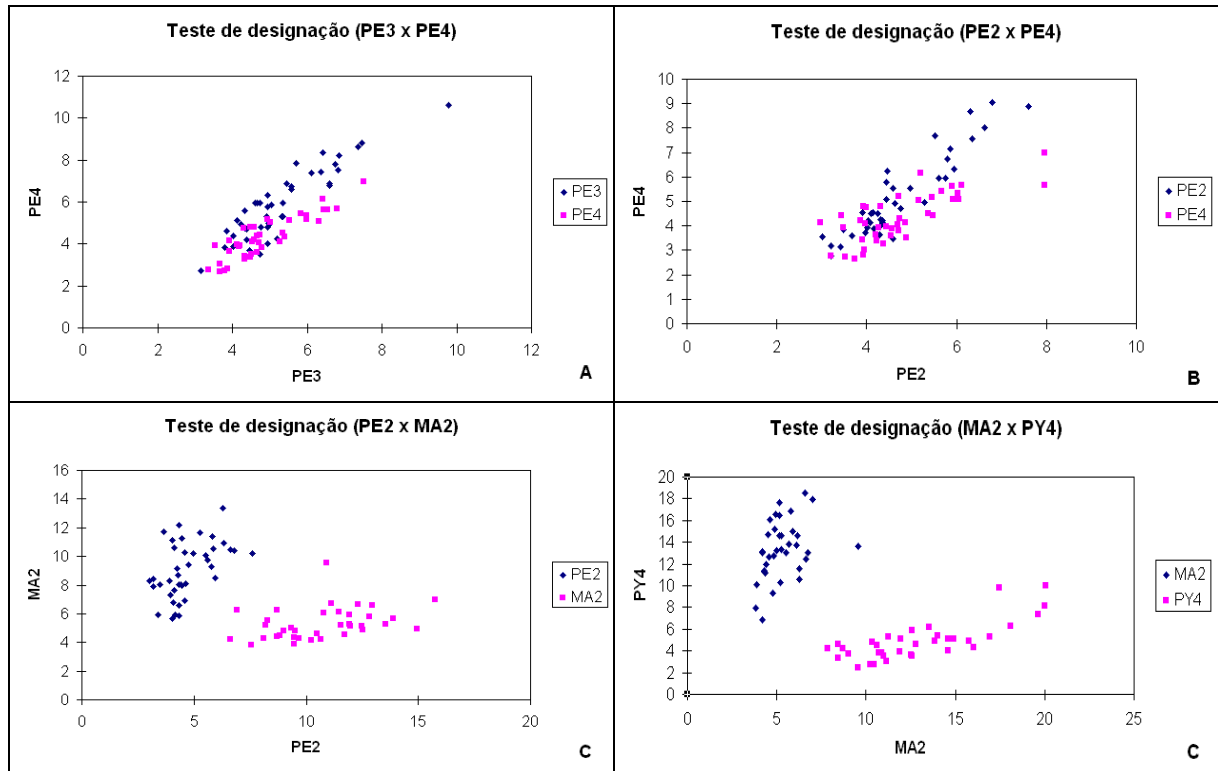
#### 4.2.8 *Teste de designação*

A análise dos dados por meio do teste de designação (traduzido do inglês, *assignment test*) baseado nas frequências foi realizada em diferentes níveis de agrupamento dos 12 plantios. A aplicação do teste utilizando os doze plantios, resultou em 59% das plantas designadas corretamente para o plantio de origem e conseqüentemente 41% designadas com pertencentes a um outro plantio. Entre os doze plantios, aqueles que apresentaram maior proporção de designação em outra população que não a de origem foram os plantios de Pernambuco. Neste caso 63% das plantas foram designadas em outros plantios. No entanto, a grande maioria foi designada em outro plantio da região de Pernambuco. O teste de designação para uma estrutura envolvendo cinco grupos formados por todos os plantios de Pernambuco (**PE1**, **PE2**, **PE3** e **PE4**), os plantios **MA1** e **MA2** do Maranhão e **PY6** do Paraguai de forma isolada e cinco plantios do Paraguai (**PY1**, **PY2**, **PY3**, **PY4** e **PY5**) mostrou que 90% das plantas foram designadas corretamente no grupo de origem e conseqüentemente 10% foram designadas em outro grupo que não o seu grupo de origem. A análise com grupos formados por plantios do Paraguai e plantios do Brasil apontou para uma designação correta de 96% das plantas em seu grupo de origem e 4 % foram designadas incorretamente. Em números absolutos os 4% representam 21 plantas designadas em outro grupo que não o de origem. Entre as 21 plantas, 14 plantas de plantios paraguaios foram designadas como sendo do grupo de plantas de plantios brasileiros e sete plantas de plantios brasileiros designadas como plantas de plantios paraguaios. Estes dados são mostrados na Tabela 18. O algoritmo do teste de designação implementado no programa GENALEX V6.0 (PEAKALL, 2006) permite a geração de gráficos que permitem visualizar o teste na forma de figuras. As Figuras 14, 15 e 16 mostram alguns desses gráficos.

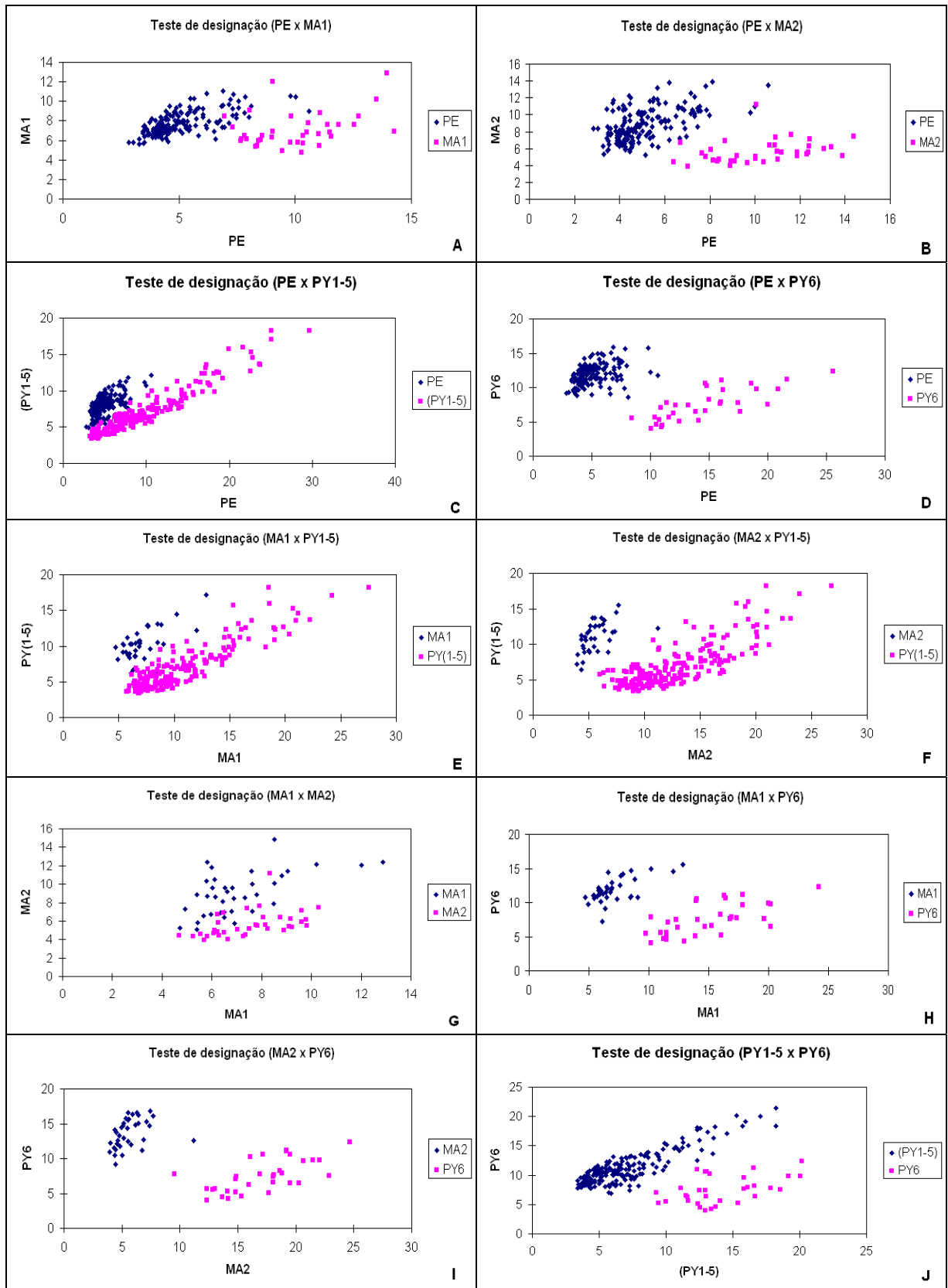


Tabela 18 – Dados do teste de designação (*Assignment test*).

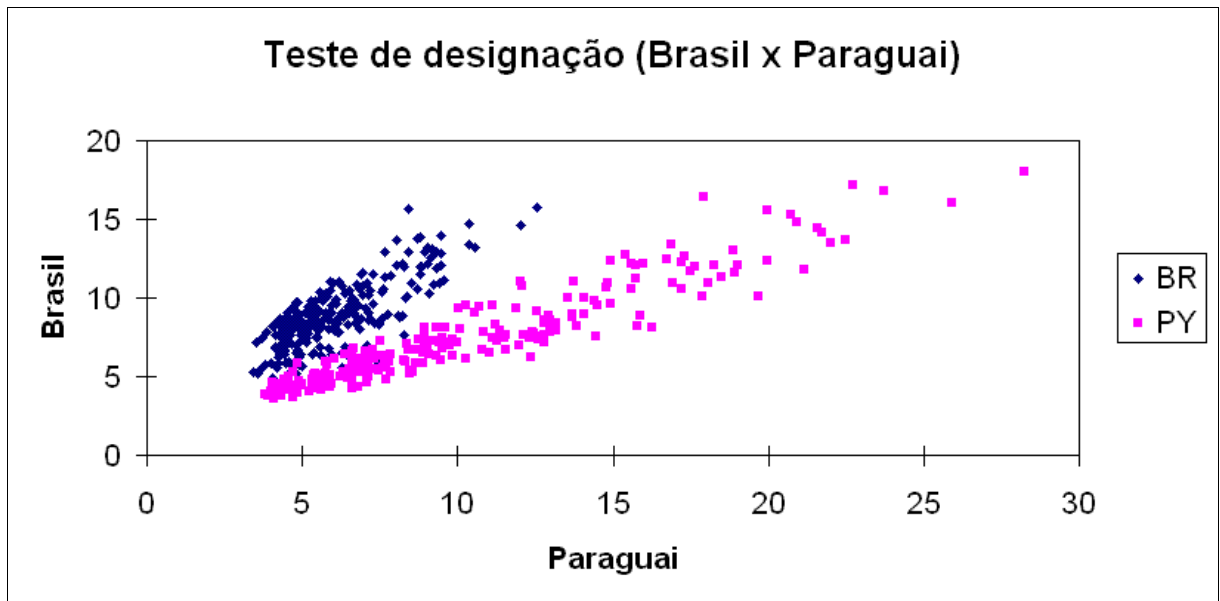
	Indivíduos corretamente designados		Indivíduos designados em outros plantios	
	Nº plantas	%	Nº plantas	%
<b>Designação entre os 12 plantios</b>				
PE1	13	36%	23	64%
PE2	10	25%	30	75%
PE3	14	30%	32	70%
PE4	24	55%	20	45%
MA1	27	73%	10	27%
MA2	33	92%	3	8%
PY1	19	43%	25	57%
PY2	13	31%	29	69%
PY3	42	89%	5	11%
PY4	27	77%	8	23%
PY5	26	72%	10	28%
PY6	34	100%	0	0%
<b>Total</b>	<b>282</b>	<b>59%</b>	<b>195</b>	<b>41%</b>
<b>Designação entre os 5 grupos de plantios</b>				
PE	160	96%	6	4%
MA1	28	76%	9	24%
MA2	33	92%	3	8%
PY(1-5)	175	86%	29	14%
PY6	34	100%	0	0%
<b>Total</b>	<b>430</b>	<b>90%</b>	<b>47</b>	<b>10%</b>
<b>Designação entre as 2 Regiões de plantios</b>				
Brasil	232	97%	7	3%
Paraguai	224	94%	14	6%
<b>Total</b>	<b>456</b>	<b>96%</b>	<b>21</b>	<b>4%</b>



**Figura 14** – Gráficos do teste de designação realizados no nível de grupamento constituído pelos 12 plantios. **A e B** – Mostram a análise par a par entre plantios de Pernambuco, onde diversos indivíduos são designados em um plantio que não de origem (mas na mesma região). **C e D** – mostram análises par a par onde há designação correta para todos os indivíduos.



**Figura 15** – Gráficos do teste de designação par a par realizados para a estrutura de cinco grupos. 90% das plantas são designadas corretamente. Os 10% de designação incorreta são mostrados nos gráficos A, B, C, E e G.



**Figura 16** – Gráfico do teste de designação realizados para a estrutura de duas regiões de plantio: Brasil e Paraguai; apenas 4% dos indivíduos são designados incorretamente.

## 5 DISCUSSÃO

A utilização de marcadores do tipo microssatélites em *C. sativa* é um campo ainda pouco explorado, onde os trabalhos publicados até o momento tratam da identificação e caracterização destes marcadores (HSIEH, 2003, GILMORE, 2003 e ALGHANIM, 2003). O trabalho apresentado nesta dissertação é certamente o primeiro que utiliza estes marcadores em um grande número de amostras coletadas em plantios que potencialmente abastecem um grande mercado consumidor, testando o potencial para uma aplicação prática como ferramenta de auxílio à inteligência policial no rastreamento de rotas de tráfico.

A reduzida quantidade de informação em termos de seqüências de DNA para o genoma de *C. sativa* reduz em muito as possibilidades de mineração de novos microssatélites a serem testados. Em termos de seqüências depositadas em bancos de dados, somente os 15 microssatélites desenvolvidos pelo grupo australiano liderado por Simon Gilmore, estão disponíveis. As seqüências dos microssatélites desenvolvidos por Keim et al. também estão disponíveis no texto de descrição da Patente nº 20060035236. Para desenvolvimento desta dissertação, treze microssatélites foram escolhidos inicialmente para constituírem uma única bateria a ser amplificada em multiplex. Novos iniciadores para a reação em cadeia da polimerase foram redesenhados para microssatélites onde as seqüências flanqueadoras estavam disponíveis em bancos de dados. A impossibilidade de redesenhar iniciadores para todos os treze microssatélites escolhidos fez com que as temperaturas de pareamento não fossem homogêneas, o que dificultou a otimização da co-amplificação dos treze microssatélites em uma única reação. Os treze microssatélites foram amplificados em dois conjuntos diferentes organizados de acordo com a temperatura de pareamento dos iniciadores. Após a otimização das concentrações de reagentes e programa de amplificação não foi possível obter amplificação para dois microssatélites. Para ambos, foram utilizados

iniciadores redesenhados para o desenvolvimento desta dissertação. A explicação mais provável é a de que estes iniciadores, desenhados com base em uma seqüência disponível em bancos de dados, não apresente homologia total com o sítio de pareamento nas amostras utilizadas durante a otimização.

A análise dos onze locos de microssatélites aponta para uma menor diversidade nas amostras estudadas neste trabalho para os locos descritos por Gilmore et al. (2003) e para níveis de diversidade semelhantes para os locos descritos por Alghanim e Almirall (2003), quando comparados com os dados de diversidade publicados pelos referidos autores. A menor diversidade dos dados obtidos em relação ao trabalho de Gilmore et al. (2003) não surpreende, já que aqueles pesquisadores analisaram amostras oriundas de 15 regiões de plantio distribuídas pelo mundo, sendo seis variedades cultivadas para a produção de fibra e nove de plantios ilegais para o uso como droga, enquanto este trabalho se restringiu aos plantios realizados para o comércio ilícito de drogas e com origem em apenas duas regiões de plantio (Brasil e Paraguai). As semelhanças de diversidade alélica observadas em relação ao trabalho de Alghanim e Almirall (2003) pode ser explicada de forma semelhante, porém diversa. Estes pesquisadores utilizaram um número reduzido de amostras (41 amostras) e com apenas duas origens distintas.

A análise da concordância com o equilíbrio de Hardy-Weinberg utilizando os plantios como unidade de análise mostra que vários locos não apresentam a distribuição de genótipos de acordo como o esperado assumindo o equilíbrio de Hardy-Weinberg. As principais possibilidades para explicação destes resultados passam pelos erros de genotipagem, número de amostras reduzido, presença de alelos nulos, gargalo genético ou elevados níveis de endogamia. Com o conjunto de resultados obtidos neste trabalho, determinar exatamente qual dos três mecanismos desempenharia um papel mais importante

não é tarefa simples. A presença de alelos nulos poderia ser identificada com a genotipagem de famílias, as quais não estão disponíveis. A análise dos dados utilizando algoritmos implementados no programa Micro-Checker aponta para a presença de alelos nulos e potencialmente erros de interpretação de heterozigoto como causas dos excessos de homozigotos para alguns locos. Para aqueles locos onde a possibilidade de erro de interpretação de heterozigotos foi apontada, baseado em uma redução de heterozigotos onde os alelos diferem por uma unidade repetitiva, uma nova observação dos eletroferogramas não apresenta indícios que levaria a concordar com esta observação. Esta diminuição de heterozigotos com alelos diferindo em uma unidade repetitiva ocorre provavelmente como consequência dos alelos nulos.

O conhecimento disseminado de que os plantios são gerados por amostragens de plantios anteriores pode exercer um efeito de gargalo genético e também elevar os níveis de endogamia. Este efeito poderia ser classificado como demográfico e a expectativa é a de que seus efeitos fossem verificados em todos os locos e a expectativa é a de que o excesso de homozigotos se concentrasse em determinada classe de alelos. No entanto, os resultados encontrados mostram locos que estão de acordo com as expectativas previstas pelo equilíbrio de Hardy-Weinberg, o que pode ser utilizado como uma evidência favorável a uma forte presença de alelos nulos. Muito embora os valores médios de coeficiente de parentesco apontem para um desvio significativo da panmixia, a análise com algoritmos implementados no programa Micro-Checker, não identificam a endogamia como fator principal na determinação do excesso de homozigotos.

A análise de independência entre os locos tem como premissa a conformidade com as expectativas do equilíbrio de Hardy-Weinberg. Assim, os resultados de

independência encontrados neste trabalho sofrem pelo fato de vários locos entre os onze investigados não estarem de acordo com o equilíbrio de Hardy-Weinberg.

A porcentagem da variabilidade explicada por diferenças entre plantios nas regiões (Brasil e Paraguai) é da ordem de 21%. A análise dos valores de  $R_{ST}$  par a par entre os plantios aponta para a presença de cinco grupos, levando em conta valores de  $R_{ST}$  não significativos (Tabela 17). A representação gráfica dos valores de  $R_{ST}$  através da análise de componentes principais (Figura 12) mostra que os cinco grupos de certa maneira representam a proximidade geográfica entre os plantios. Este padrão de estruturação pode ser explicado pela forma de geração dos plantios a partir de sementes obtidas em plantios anteriores no mesmo local ou de plantios que estão mais próximos geograficamente. O fato do plantio PY6 aparecer separado dos demais plantios do Paraguai poderia ser explicado pela introdução de novos conjuntos de sementes na região. Apesar desta hipótese carecer de qualquer confirmação formal, a observação direta dos plantios, feita por policiais paraguaios e brasileiros durante operações de erradicação recentes, tem chamado a atenção para o surgimento de plantas mais robustas e de ciclo de vida mais rápido. Estas observações tem levado inclusive a atribuição da denominação de “transgênica” a estas possíveis variedades, o que sugere a introdução de novas sementes naquela região.

Muito embora o cálculo do teste de designação baseado em frequências tenha como premissas o Equilíbrio de Hardy-Weinberg e o equilíbrio de ligação, o método apresentou o melhor desempenho quando comparado com outros métodos que não levavam em conta as referidas premissas (Cornuet et al., 1999). Os resultados apresentados neste trabalho apontam para uma alta porcentagem de amostras designadas corretamente, quando trabalhando com plantios de origem no Brasil e Paraguai. No total 96% das amostras foram designadas corretamente. O teste de designação aplicado a uma estrutura de análise onde os

plantios foram agrupados de acordo com a estrutura apontada pelas distâncias de  $R_{ST}$ , ou seja, cinco grupos representados pelos formados por todos os plantios de Pernambuco (**PE1**, **PE2**, **PE3** e **PE4**), os plantios **MA1** e **MA2** do Maranhão e **PY6** do Paraguai de forma isolada e cinco plantios do Paraguai (**PY1**, **PY2**, **PY3**, **PY4** e **PY5**), aponta também para uma alta porcentagem de designações corretas. Neste caso 90% das plantas foram designadas corretamente.



## 6 CONSIDERAÇÕES FINAIS

*“A droga e a arma têm o mesmo procedimento, seguem a mesma estrutura. Entendendo a rota das drogas vamos entender também a das armas, cercando-as.”*

Luiz Fernando Corrêa  
Secretário Nacional de Segurança Pública.

A descoberta das rotas do tráfico da *C. sativa* (maconha) no Brasil requer um pesado e oneroso trabalho realizado pela polícia. Com a pouca utilização de fontes humanas, a determinação dessas rotas é feita, sobretudo, com o trabalho dos Serviços de Inteligência Policial. Tais operações demandam tempo, têm custo elevado, uma vez que os agentes de inteligência são mantidos longe da base por um longo período e, muitas vezes, não obtêm o sucesso esperado.

A avaliação dos 13 microssatélites de *C. sativa* amplificados em PCR multiplex quanto a seu potencial em distinguir amostras de plantios ilegais do Brasil e Paraguai foi o objetivo geral deste trabalho. Ao concluir este trabalho, 11 microssatélites foram amplificados e mostram-se eficientes quanto a este objetivo geral.

Os marcadores microssatélite mostram-se uma poderosa ferramenta forense em casos de individualização de *C. sativa* e os dados deste trabalho demonstraram claramente o potencial discriminatório destes entre indivíduos de *C. sativa* com origem em plantios ilegais no Brasil (Pernambuco e Maranhão) e no Paraguai, principais fontes da droga presente no mercado consumidor brasileiro. Utilizando os sistemas multiplexes desenvolvidos, foi construído e caracterizado um banco de dados de frequências alélicas microssatélites em *C. sativa*, com possibilidades de identificação individual ou associação de plantas e/ou resíduos

de plantas e inferências sobre origem geográfica de apreensões no Brasil, informações úteis para a inteligência policial no combate ao narcotráfico.

A transposição do presente trabalho para a prática policial, pretende produzir informações úteis à investigação. Conhecer a origem da *C. sativa* presente no mercado consumidor e nas varias apreensões tem o potencial de facilitar o planejamento de operações, diminuir custos, e proporcionar redução no tempo despendido em investigações, ajudando a estabelecer as rotas do tráfico de uma maneira mais confiável. Além disso, poderá facilitar o intercâmbio das informações de repressão ao tráfico de *C. sativa* entre o Brasil e países da América do Sul, sobretudo com o Paraguai, país que produz a maior quantidade da maconha atualmente comercializada em nosso País.

As palavras do Secretário Nacional de Segurança Pública, Dr. Luiz Fernando Corrêa, ilustram a importância do conhecimento das rotas das drogas. Os protocolos desenvolvidos e os dados obtidos neste trabalho podem ser úteis como ferramenta forense neste objetivo, em relação à *Cannabis sativa* presente no mercado consumidor brasileiro.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

ALGHANIM, H. J; ALMIRALL J. R. **Development of microsatellite markers in *Cannabis sativa* for DNA typing and genetic relatedness analyses.** Anal Bioanal Chem, 2003. 376:1225-1233.

BONNIE, Richard J; WHITEBREAD II, Charles H. **The forbidden fruit and the tree of knowledge: An inquiry into the legal history of american marijuana prohibition.** Virginia law review, 1970. 56:6.

BRONDANI, R.P; GRATTAPAGLIA, D. **Cost-effective method to synthesize a fluorescent internal DNA standard for automated fragment sizing.** Biotechniques. 2001 Oct; 31(4):793-5, 798, 800.

COYLE, Heather Miller et al. **An overview of DNA methods for the identification and individualization of marijuana.** Croatian Medical Journal, 2003. 44:315-321.

COYLE, Heather Miller et al. **A simple DNA extraction method for marijuana samples used in amplified fragment length polymorphism (AFLP) analysis.** J Forensic Sci, 2003. 48:343-7.

DATWYLER, S.L; WEIBLEN G.D. **Genetic variation in hemp and marijuana (*Cannabis sativa* L.) according to amplified fragment length polymorphisms.** J Forensic Sci, 2006. 51(2):371-5.

DEWOODY, J. et al. **Mitigating scoring errors in microsatellite data from wild populations.** Molecular Ecology Notes (2006) doi: 10.1111/j.1471-8286.2006.01449.x.

DOYLE, J.J.; DOYLE, J.L. **Isolation of plant DNA from fresh tissue.** Focus, 1990. 12:13-15.

EXCOFFIER, L.; SMOUSE, P.E.; QUATTRO, J.M. **Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction sites.** Genetics, 1992. 131, 479-491.

EXCOFFIER, L.; LAVAL, G.; SCHNEIDER, S. **Arlequin ver. 3.0: An integrated software package for population genetics data analysis.** Evolutionary Bioinformatics Online, 2005. 1:47-50.

FORAPANI, Silvia et al. **Comparison of hemp varieties using random amplified polymorphic DNA markers.** Crop Science, 2001; 41:1682-1689.

GIGLIANO, G. Siniscalco; DI FINIZIO, A. **The *Cannabis sativa* L. fingerprint as a tool in forensic investigations.** Bulletin-on-Narcotics, 1997; 49-50 (1-2):129–137.

GIGLIANO, G. Siniscalco; CAPUTO, P.; COZZOLINO, S. **Ribosomal DNA analysis as a tool to identify specimens of *Cannabis sativa* L. of forensic interest.** Sci. Justice, 1997; 37:171–174.

GIGLIANO, G. Siniscalco. **Identification of *Cannabis sativa* L. (*Cannabaceae*) using restriction profiles of the Internal Transcribed Spacer II (ITS2).** Sci. Justice, 1998; 38:225–230.

GIGLIANO, G. Siniscalco. **Preliminary data on the usefulness of Internal Transcribed Spacer I (ITS1) sequence in *Cannabis sativa* L. identification.** J. Forensic Sci., 1999; 44(3):475–477.

GILMORE, S.; PEAKALL, R. **Isolation of microsatellite markers in *Cannabis sativa* L. (marijuana).** Molecular Ecology Notes, 2003; 3:105–107.

GILMORE, S.; PEAKALL, R.; ROBERTSON, J. **Short tandem repeat (STR) DNA markers are hypervariable and informative in *Cannabis sativa*: implications for forensic investigations.** Forensic Sci. Int., 2003; 131:65–74.

GOLDSTEIN, David B; SCHLÖTTERER, Christian. **Microsatellites evolution and applications.** Oxford University Press, New York, 1998.

GUO, S.; THOMPSON, E. **Performing the exact test of Hardy-Weinberg proportion for multiple alleles.** Biometrics, 1992; 48:361-372.

HONG, Shao; CLARKE, Robert C. **Taxonomic studies of *Cannabis* in China.** Journal of the International Hemp Association, 1996; 3(2): 55-60.

HSIEH, Hsing-Mei et al. **A highly polymorphic STR locus in *Cannabis sativa*.** Forensic Sci. Int., 2003; 131:53-58.

IVERSEN, Leslie L. **The science of marijuana.** New York, Oxford University Press, 2000.

JAGADISH, V.; ROBERTSON, J.; GIBBS, A. **RAPD analysis distinguishes *Cannabis sativa* samples from different sources.** Forensic Sci. Int., 1996; 79:113–121.

KEIM, Paul S. et al. **DNA fingerprinting for *Cannabis sativa* (marijuana) using short tandem repeat (STR) markers.** Patente nº 20060035236, registrada em 16 de fevereiro de 2006, no United States Patent and Trade Mark Office. Disponível em: <<http://www.uspto.gov/>>. Acesso em: 16 ago 2006.

KOHJYUMA, M. et al. **Intraspecific variation in *Cannabis sativa* L. based on intergenic spacer region of chloroplast DNA.** Biological & Pharmaceutical Bulletin, 2000; 23(6):727-730.

LINACRE, Adrian; THORPE, James. **Detection and identification of *Cannabis* by DNA.** Forensic Sci. Int., 1998; 91:71–76.

OOSTERHOUT, C. V. et al. **Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data.** Molecular Ecology Notes, 2004; 4, 535-538.

ORLOCI, L. **Multivariate analysis in vegetation research.** The Hague: Dr W. Junk B. V., 1978.

PAETKAU, D. et al. **Microsatellite analysis of population structure in Canadian polar bears.** Mol Ecol, 1995; 4:347-54.

PALMBACH, T H M; LADD, C; LEE, HC. **Tracking clonally propagated marijuana using amplified fragment length polymorphism (AFLP) analysis.** In: Coyle HM, editor. Forensic botany: principles and applications to criminal casework. Boca Raton, FL: CRC Press; 2005.

PEAKALL, R.; SMOUSE, P.E. **Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research.** Molecular Ecology Notes, 2006; 6, 288-295.

PEIL, A et al. **Sex-linked AFLP markers indicate a pseudoautosomal region in hemp (*Cannabis sativa* L.).** Theor Appl Genet., 2003; 107(1):102-9.

ROZEN, Steve; SKALETSKY, Helen J. **Primer3 on the WWW for general users and for biologist programmers.** In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, 2000; pp 365-386

SAKAMOTO, K. et al. **Site-specific accumulation of a LINE-like retrotransposon in a sex chromosome of the dioecious plant *Cannabis sativa*.** Plant Mol Biol., 2000; 44(6):723-32.

SCHOSKE, Richard et al. **Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci.** Anal Bioanal Chem, 2003; 375 : 333–343.

SHIBUYA, Elisa Kayo. **Rastreamento da origem geográfica de amostras de maconha apreendidas nas ruas de São Paulo por meio de assinaturas químicas.** São Paulo, 2005. 168 f. Tese (Doutorado em Ciências na área de tecnologia nuclear – materiais) – Universidade de São Paulo, São Paulo, 2005.

SHIROTA, O. et al. **Random amplified polymorphic DNA and restriction fragment length polymorphism analyses of *Cannabis sativa***. *Natural Medicines*, 1998; 52(2):160-166.

SILVERSTEIN, Robert M. et al. **Spectrometric Identification of Organic Compounds (7th edition)**. John Wiley & Sons Inc., October 2003; ISBN: 0471393622.

SLATKIN, M. **A measure of population subdivision based on microsatellite allele frequencies**. *Genetics*, 1995; 139, 1463.

SLATKIN, M; EXCOFFIER, L. **Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm**. *Heredity*, 1996 Apr;76 ( Pt 4):377-83.

UNITED NATIONS, Office on drugs and crime. **Global Illicit Drug Trends 2003**. Disponível em: [http://www.unodc.org/pdf/trends2003\\_www\\_E.pdf](http://www.unodc.org/pdf/trends2003_www_E.pdf)

UNITED NATIONS, Division of Narcotic Drugs. **Recommended methods for testing Cannabis: manual for use by national narcotics laboratories**. New York, 1987.

UNODC /BRASIL (ESCRITÓRIO DAS NAÇÕES UNIDAS CONTRA DROGAS E CRIME NO BRASIL). **Perfil de país: Brasil**. C.P. 07-0285 Brasília, DF Brasil 2003.

VALLONE, P.M.; BUTLER, J.M. (2004), **AutoDimer**: a screening tool for iniciador-dimer and hairpin structures. *Biotechniques*, 2004; 37(2): 226-231.

VARGAS, R. M.; CAMARGO, M. A.; TALHAVINI, M. **Modificações nos teores de tetra-hidro-canabinol e canabinol em amostras de maconha**. Trabalho apresentado na 23ª Reunião Anual da Sociedade Brasileira de Química, Poços de Caldas-MG, 2000.

WANG, J. **An estimator for pairwise relatedness using molecular markers**. *Genetics*, 2002;160(3):1203-15.