

Universidade Estadual de Campinas
Instituto de Biologia

Prianda Rios Laborda

**Diversidade Genética entre Linhagens de Milho Tropical:
Estudo com Base em Marcadores Moleculares**

Tese apresentada ao Instituto de
Biologia para obtenção do Título de
Mestre em Genética e Biologia
Molecular na área de Genética Vegetal
e Melhoramento.

Orientadora: Dra. Anete Pereira de Souza
Co-orientador: Dr. Antônio Augusto Franco Garcia

Campinas - SP

2003

Data da defesa: 05.novembro.2003

Banca Examinadora:

Profa. Dra. Anete Pereira de Souza (orientadora)

Prof. Dr. Louis Bernard Klaczko

Prof. Dr. Isaias Olivio Geraldi

Profa. Dra. Vera Nisaka Solferini

Aos meus pais, **Jorge e Paraguassu**,
ao meu irmão **Igor** e
ao meu namorado **Rogério**,

dedico esse trabalho, fruto do **apoio**.

À minha orientadora **Anete** e
ao meu co-orientador **Augusto**,

ofereço esse trabalho, fruto do **aprendizado**.

Agradecimentos

Aos professores Dr. Isaias Olivio Geraldi, Dr. Louis Bernard Klaczko, Dra. Maria Elisa A.G.Z. Paterniani, Dr. Michel A.G. Vincentz e Dra. Vera Nisaka Solferini, membros de pré-banca e/ou banca, pela disponibilidade em contribuir na revisão desse trabalho e pelos ensinamentos dentro e fora de sala de aula.

Aos amigos de Brasília e Campinas, em quem sempre encontrei alegria e descontração para olhar para o cotidiano de forma mais tranqüila e divertida.

Aos colegas do Laboratório de Análises Genética e Molecular, pelas proveitosas conversas, pelos conselhos, pelas experiências trocadas, pela preciosa ajuda na condução dos experimentos. É um prazer trabalhar com vocês.

Aos meus familiares, em Bauru, Cuiabá, Porto Velho, Santos e São Luis, que, mesmo à distância, sempre se preocuparam com o andamento das minhas atividades, além de também se orgulharem delas.

Às instituições, CAPES, CNPq e FAPESP, que possibilitaram a realização desse trabalho, financiando, de uma forma ou de outra, nossos projetos de pesquisa.

À Universidade Estadual de Campinas - UNICAMP - e mais especificamente ao Instituto de Biologia - IB, onde encontrei a infra-estrutura necessária para estudar e o ambiente apropriado para sentir o conforto de uma casa.

Agradecimentos especiais

No trabalho...

À Karine Miranda Oliveira, responsável pelos dados de AFLP e amiga para momentos de tristeza e felicidade.

Ao meu co-orientador Dr. Antonio Augusto Franco Garcia, com quem venho aprendendo que a Genética também requer cuidadosa abordagem estatística e que, além disso, me ensinou a ver com mais tranquilidade alguns aspectos da “vida de pesquisador”.

À minha orientadora Dra. Anete Pereira de Souza, com quem aprendo diariamente a viver no mundo da pesquisa. São admiráveis o seu vigor como administradora e a competência e alegria com que nos ensina a lidar com a ciência.

e na vida pessoal...

Ao meu namorado Rogério, pelo companheirismo, pelo apoio incondicional, pela torcida, por me ouvir em horas de tristeza e por me ajudar a buscar as melhores saídas.

Ao meu irmão Igor, que sempre se orgulhou de ter uma irmã “pós-graduanda”. Imagino que você não saiba como isso é importante para mim.

Aos meus pais, Jorge e Paraguassu. O mestrado não teria nenhum valor para mim se eu não pudesse ver a satisfação de vocês com essa conquista. Se eu não tivesse recebido os importantes valores morais que vocês me deram, certamente não teria concluído esse trabalho. Obrigada por acreditarem nos meus sonhos.



“Columbus did not realize that the gift of maize was far more valuable than the spices or gold he hoped to find. He had no way of knowing that the history of maize traced back some 8,000 years or that it represented the most remarkable plant breeding accomplishment of all time. He might have been embarrassed if he had understood that then, as now, this plant developed by peoples he judged poor and uncivilized far outstripped in productivity any of the cereals bred by Old World farmers --wheat, rice, sorghum, barley, and rye. Were he alive today, he would certainly be astonished to see the extent to which the advent of maize has affected land use, food production, cuisine, and population growth around the world.”

Walton Galinat, 1992

Índice

Resumo	10
Abstract	11
Lista de tabelas e figuras	12
1) Introdução	14
2) Revisão bibliográfica	16
2.1) Importância, origem e evolução do milho	16
2.2) Diversidade do milho: dos estudos fenotípicos às análises moleculares	29
2.2.1) Da morfologia às isozimas	31
2.2.2) Do RFLP aos microsatélites	35
2.2.2.1) Milho de clima temperado	38
2.2.2.2) Milho de clima tropical	43
2.3) Métodos estatísticos de análise da diversidade	45
3) Objetivos	48
4) Artigo	49
5) Discussão	82
6) Conclusões	87
7) Referência citadas	88
8) Anexo	96

Resumo

Conhecer a variabilidade genética de plantas de interesse econômico permite manipular recursos naturais de forma eficiente, visando o melhoramento vegetal. A milho é uma das mais importantes plantas cultivadas e a diversidade que o germoplasma tropical apresenta ainda não está bem estabelecida. Atualmente, os marcadores moleculares são a principal maneira de analisar essa diversidade, e muitas possibilidades estatísticas foram desenvolvidas para organizar informações de dados moleculares. Este trabalho usou os marcadores AFLP e SSR para avaliar a diversidade genética em linhagens de milho tropical pertencentes ao Banco de Germoplasma do Instituto Agronômico de Campinas (IAC), e revelou grande nível de variabilidade. Os três tipos de coeficiente usados (Jaccard, MRD e coeficiente de parentesco molecular) exibiram baixos valores de similaridade entre as linhagens e, dependendo do marcador usado, diferentes agrupamentos foram obtidos. Entretanto, altas correlações entre as matrizes geradas por Jaccard e pelo coeficiente de parentesco molecular com AFLP, SSR e AFLP/SSR foram observadas. As correlações entre AFLP e SSR foram baixas, independentemente do coeficiente de similaridade aplicado. Devido à habilidade dos SSRs em revelar vários aspectos da variação genômica, essa técnica foi considerada mais apropriada para análises de diversidade. Notou-se que a grande diversidade existente no material tropical levou a resultados distintos se marcadores diferentes foram usados, e que o uso simultâneo de AFLPs e SSRs pode não ser a maneira mais eficiente de avaliar a variabilidade em materiais altamente divergentes.

Abstract

Knowledge about genetic variability of agronomically important crops allows efficient manipulation of resources in order to achieve plant improvement. Maize is one of the most important crops, but the extent of its diversity in the tropical germplasm has still not been well established. Nowadays, molecular markers are the principal way of analyzing genetic diversity and many statistical possibilities have been developed to compile data from molecular assays. This work used AFLP and SSR markers to access genetic diversity in tropical maize inbred lines from the Maize Genebank of the Agronomic Institute of Campinas (IAC), and revealed a great level of variability, with high rates of polymorphism. All three types of coefficients applied (Jaccard, MRD and molecular coefficient of coancestry) exhibited very low genetic similarities among the inbreds, and depending on the marker used, different clusters were obtained. Nevertheless, high correlations between matrices generated by Jaccard and by the molecular coefficient of coancestry with AFLP, SSR and AFLP/SSR were seen. AFLP and SSR were poorly correlated with both similarity coefficients. Because of SSR's ability of uncovering many aspects of genomic variation, it was considered more appropriate for diversity analysis. It was observed that the great amount of diversity found in tropical maize led to distinct results with different markers, and that using both AFLP and SSR together may not be the most efficient manner of accessing variability in highly diverse materials.

Lista de tabelas e figuras

Tabela 1. Produção e área cultivada de acordo com as diferentes regiões mundiais.	17
Tabela 2. Os dez maiores produtores mundiais de milho e as respectivas áreas de plantio usadas.	17
Table 1 Maize inbred lines maintained in the Agronomic Institute of Campinas, the material they were selected from and their origins	76
Table 2 Microsatellites markers used in the diversity analyses and their genomic location, class of repeat, number of alleles, PIC values and number of heterozygotes	77
Table 3 Correlations among the seven ways used to access diversity by means of Mantel's test and comparison of mean similarity/distance values according to the type of coefficient used to generate the similarity/distance matrix	78
Figura 1. Distribuição mundial da área plantada (cada ponto vermelho representa 100 mil toneladas produzidas).	18
Figura 2. Diferenças morfológicas entre as partes vegetativas de teosinte e milho e entre as suas espigas (detalhe).	21
Figura 3. Hipótese “tripartite” proposta por Paul Mangelsdorf para explicar a origem do milho. (1) O milho moderno teria surgido do “pod corn”, (2) o teosinte seria resultado da hibridação entre milho e <i>Tripsacum</i> e (3) os híbridos entre essas duas plantas justificariam a diversidade atual do milho.	23
Figura 4. Distribuição dos 50 locos de microssatélites entre dos 100 <i>bins</i> dos cromossomos do milho. A boa dispersão dos marcadores proporciona maior cobertura do genoma e maior confiabilidade dos dados, pois mais partes genômicas podem ser amostradas.	97
Figura 5. Agrupamento (UPGMA) das 85 linhagens com base nos dados de AFLP-Jaccard (valor cofenético = 0,78).	98
Figura 6. Agrupamento (UPGMA) das 85 linhagens com base nos dados de SSR-Jaccard (valor cofenético = 0,66).	99

Figura 7. Agrupamento (UPGMA) das 85 linhagens com base nos dados de AFLP/SSR-Jaccard (valor cofenético = 0,79).	100
Figura 8. Agrupamento (UPGMA) das 85 linhagens com base nos dados de SSR-MRD (valor cofenético = 0,70).	101
Figura 9. Agrupamento (UPGMA) das 85 linhagens com base nos dados de AFLP/SSR- f_{AB}^M (valor cofenético = 0,83).	102
Figura 10. Bootstrap com dados de marcadores microssatélites feito calculado a partir da distância modificada de Rogers (MRD) e aplicando medianas para avaliar o comportamento da variação em função do aumento do número de locos. Percebe-se que o coeficiente de variação se aproxima de 6% com os 50 locos usados no estudo, o que garante a boa qualidade dos dados gerados por esse marcador.	103
Figure 1 UPGMA showing the clusters obtained by the molecular coefficient of coancestry applied to the AFLP data (cophenetic value = 0.81)	79
Figure 2 UPGMA showing the clusters obtained by the molecular coefficient of coancestry applied to the SSR data (cophenetic value = 0.72)	80
Figure 3 Distribution of frequencies of the 3570 pairwise values in the similarity matrices. (A) Markers analyzed by Jaccard's coefficient (and also a comparison with the distances obtained by modified Rogers' distance for SSR data) (B) Markers analyzed by the molecular coefficient of coancestry	81

1) Introdução

Os recursos genéticos vegetais constituem um dos principais patrimônios que a humanidade deve preservar. A variabilidade genética, que é inerente às populações vegetais, é a base de programas de melhoramento, os quais buscam de forma contínua novas variedades de plantas que possam servir direta ou indiretamente para a alimentação humana. Conhecer a diversidade genética é fator limitante para o progresso da agricultura, principalmente no que concerne a plantas de denotado interesse econômico.

O milho é uma das espécies vegetais de maior importância cultural e econômica, e a sua produção ocupa o segundo lugar no *ranking* mundial dos cereais. A diversidade do milho é reconhecidamente a maior entre as plantas, e preocupações com a sua manutenção vêm desde épocas remotas. Diversos são os trabalhos que objetivaram organizar a variabilidade do milho, mas o material tropical, do qual faz parte o germoplasma brasileiro, ainda não foi estudado com profundidade.

O Instituto Agrônomo de Campinas (IAC) possui um importante banco de germoplasma de milho, o qual dá sustentação à produção de novos híbridos dessa espécie. Esse conjunto de genótipos é constituído por linhagens brasileiras de grande relevância histórica assim como por linhagens introduzidas do México, centro de origem do milho. Há algumas décadas, esse banco vem sendo mantido, mas estudos detalhados sobre a diversidade existente entre as linhagens não haviam sido feitos. Em adição, análises moleculares que pudessem certificar a existência de grupos divergentes de linhagens, a partir dos quais poderiam ser escolhidos pares de genótipos para produção de híbridos, ainda não haviam sido conduzidas. Na escolha de pares de linhagens que pudessem atuar

como genitores, apenas análises fenotípicas baseadas em caracteres morfológicos eram utilizadas e nem sempre bons resultados eram obtidos.

Para otimizar a escolha de genitores, reconheceu-se a importância de avaliar a disponibilidade de variação genética entre as linhagens de seu banco de germoplasma. Além disso, estudar a variabilidade entre essas linhagens representa avançar no conhecimento mais organizado da diversidade do milho tropical, ainda pouco estudado e explorado. Portanto, esse trabalho visou contribuir para o conhecimento mais sistematizado do milho tropical, que baseia programas de melhoramento dessa espécie em nosso país e faz dele o terceiro maior produtor mundial desse cereal. Por meio de dois tipos de marcadores moleculares, 85 linhagens do Bando de Germoplasma do IAC foram avaliadas e pode-se conhecer uma parcela da diversidade genética que o material tropical apresenta.

2) Revisão bibliográfica

2.1) Importância, origem e evolução do milho

O milho (*Zea mays* L.) é um cereal pertencente à família *Poaceae*, sendo uma das mais importantes plantas cultivadas atualmente e a espécie mais produzida nos países em desenvolvimento. Devido à sua característica de alta adaptabilidade a diversos ambientes, o milho é o cereal mais cultivado em termos de número de países (cerca de 70), o que totaliza mais de 100 milhões de hectares em todo o mundo. O cultivo do milho pode ser encontrado na amplitude latitudinal de 50°N a 50°S - o que compreende climas tropicais, subtropicais e temperados - e do nível do mar a altitudes superiores a 3000 metros.

Apenas no ano de 2002, foram produzidas aproximadamente 600 milhões de toneladas de milho em cerca de 140 milhões de hectares (Tabela 1). Os Estados Unidos respondem por quase 40% dessa produção, e os dez maiores produtores totalizam aproximadamente 80% (Tabela 2). O Brasil ocupa a terceira posição no *ranking* mundial, com quase 6% da produção e 8% da área plantada. A distribuição mundial da área plantada está apresentada na Figura 1.

O interesse na produção do milho vem crescendo a cada ano, o que provavelmente reflete a importância desse cereal nos variados âmbitos da alimentação. A possibilidade do uso de praticamente todo o vegetal para fins alimentícios diretos ou indiretos faz do milho um dos cereais mais versáteis em termos de aproveitamento da planta e de diversidade de produtos. O seu uso passa (1) pela dieta humana com o consumo direto do grão, ou por meio de vários alimentos industrializados a partir do processamento somente do grão ou de

Tabela 1. Produção e área cultivada de acordo com as diferentes regiões mundiais.

Região	Produção (milhões de toneladas)	Área (milhões de hectares)
África	42,56	26,94
América Latina e Caribe	79,07	26,24
América do Norte	237,87	29,25
Ásia	165,92	42,72
Europa	76,48	13,50
Oceania	0,69	0,11
Total	602,59	138,76

Fonte: FAOSTAT Database, junho de 2003.

Tabela 2. Os dez maiores produtores mundiais de milho e as respectivas áreas de plantio usadas.

País	Produção (milhões de toneladas)	Área (milhões de hectares)
EUA	228,81	28,05
China	123,18	24,53
<i>Brasil</i>	<i>35,48</i>	<i>11,87</i>
México	17,50	7,18
França	16,01	1,82
Argentina	14,71	2,42
Índia	10,57	6,20
Iugoslávia	9,41	1,91
Romênia	8,50	2,90
Rússia	8,49	2,57
Total	472,66	89,45

Fonte: FAOSTAT Database, junho de 2003.

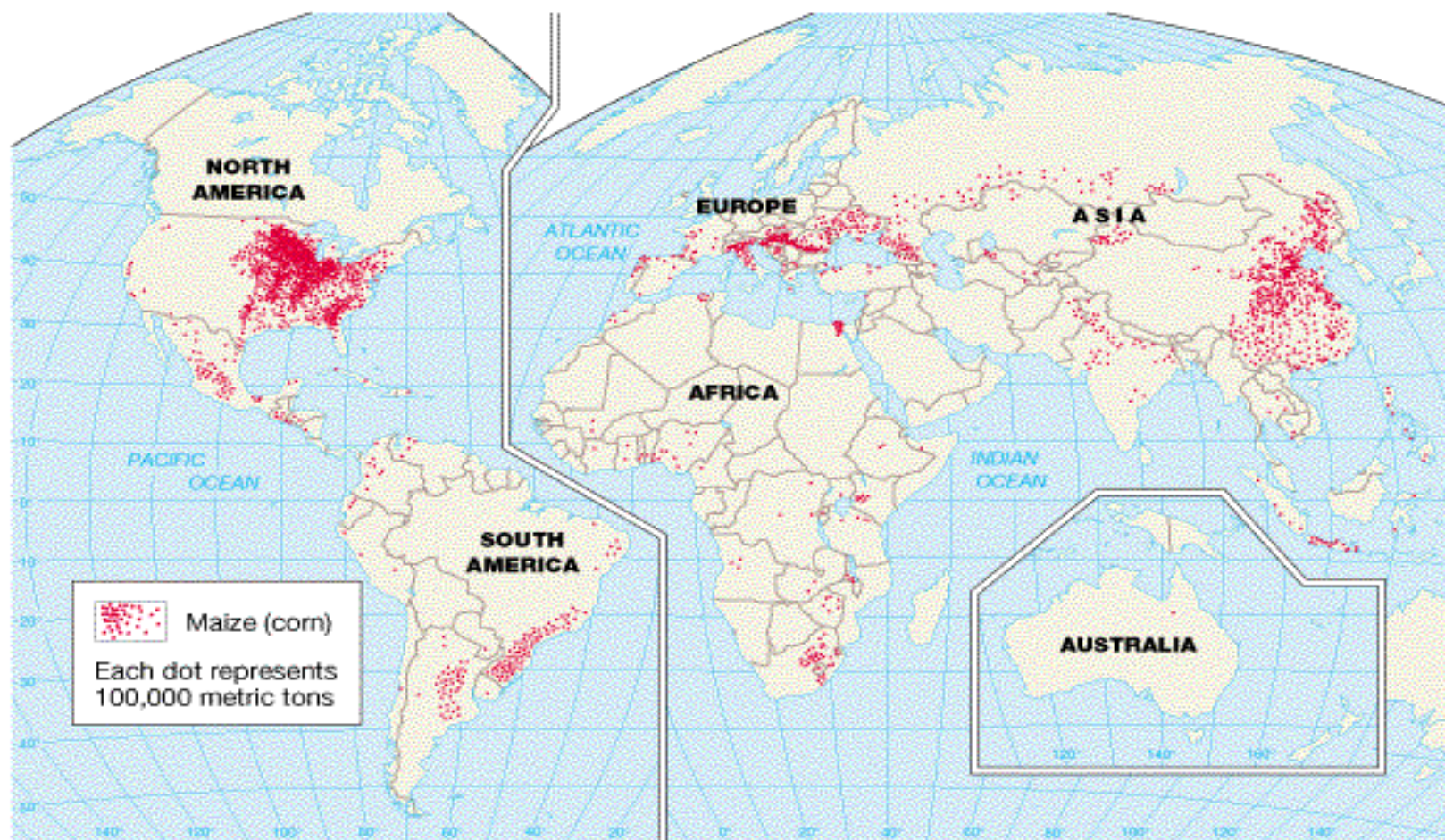


Figura 1. Distribuição mundial da área plantada (cada ponto vermelho representa 100 mil toneladas produzidas).

toda a espiga, (2) pela alimentação de animais com os grãos (suínos e aves) e com o caule e as folhas (ruminantes), (3) pela produção de bebidas por meio da fermentação e da destilação e (4) pela obtenção de produtos secundários, como adoçantes, óleos de cozinha e farinha. Além dessa vantagem de utilização de várias partes da planta, o seu valor altamente nutritivo, devido à presença de grandes quantidades de ácidos graxos e carboidratos, e a sua já mencionada adaptabilidade de cultivo em diversos ambientes dão a essa espécie um grande potencial econômico.

Por ser uma das culturas cereais de maior importância econômica no mundo, o milho é uma das espécies vegetais mais estudadas. Programas de melhoramento estão constantemente buscando novos procedimentos para a obtenção de variedades mais produtivas e economicamente rentáveis e, para tanto, muito esforço é feito em prol do entendimento mais amplo sobre as suas origens e evolução. Informações providas de diversas áreas, como a arqueologia, a citologia, a história e a biosistemática corroboraram e ainda corroboram para o esclarecimento da longa trajetória evolutiva que levou à constituição do fenótipo da planta de milho como atualmente a conhecemos. Algumas controvérsias que acompanharam o início dos estudos sobre o milho persistem ainda hoje, o que se percebe facilmente se nos reportarmos à questão do parentesco entre milho (*Zea mays* spp. *mays*), teosinte (*Zea mays* spp. *mexicana* ou *Zea mays* ssp. *parviglumis*) e *Tripsacum*.

Os registros mais antigos de amostras de milho foram encontrados na década de 1950, na Cidade do México, e datam de aproximadamente 7000 anos. O material encontrado constituía-se de pólen e grãos de milho, teosinte e do possível ancestral comum, e gerou muita discussão, pois foram cogitadas as possibilidades de contaminação e de erros de identificação dos grãos. Documentos arqueológicos sobre o milho já

domesticado datam também desse período (5000 a.C.) e foram encontrados também no México, porém, na localidade de Tehuacán (Mangelsdorf 1974). Acredita-se que o melhoramento do milho fazia-se desde essa época (por meio de seleção massal das espigas maiores e com características mais desejáveis), o que coincide parcialmente com a época do surgimento da agricultura. O sedimento analisado revelou dois tipos de milho: nos níveis inferiores, mais antigos, um possível milho selvagem e, nos níveis superiores, um milho resultante da introgressão com o teosinte. Na América do Sul, fósseis de partes da planta de milho são muito escassos. O material mais antigo foi localizado em território peruano e data de aproximadamente 1000 a.C.; espigas completas datam de 500 a.C..

Em adição às pesquisas feitas com o objetivo de identificar o período do surgimento do milho, vários foram os estudos com o intento de desvendar a sua jornada evolutiva. As tentativas de explicar a origem dessa espécie e a sua relação com as principais plantas da família *Poaceae* culminaram na elaboração de duas principais hipóteses, as quais dividiram grupos de especialistas no assunto por muito tempo. Registros botânicos e antigas descrições de diferentes variedades, feitas desde a época da chegada de Cristóvão Colombo à América, foram usados para dar suporte a uma ou outra hipótese, mas, em muitas ocasiões, não foram conclusivos. Isso se explica também com base no fato de que mesmo as poucas evidências arqueológicas, muitas vezes, se mostraram controversas.

Nos parágrafos seguintes, estão apresentadas as principais hipóteses formuladas, seus autores e o embasamento científico que estes utilizaram para justificá-las. Em adição, serão brevemente apresentados os estudos mais recentes sobre a evolução do milho e as conclusões para que apontam essas análises.

A origem do milho está intimamente ligada ao teosinte, planta considerada a sua mais próxima parente (Figura 2) e com a qual cruza livremente, produzindo progênie fértil.

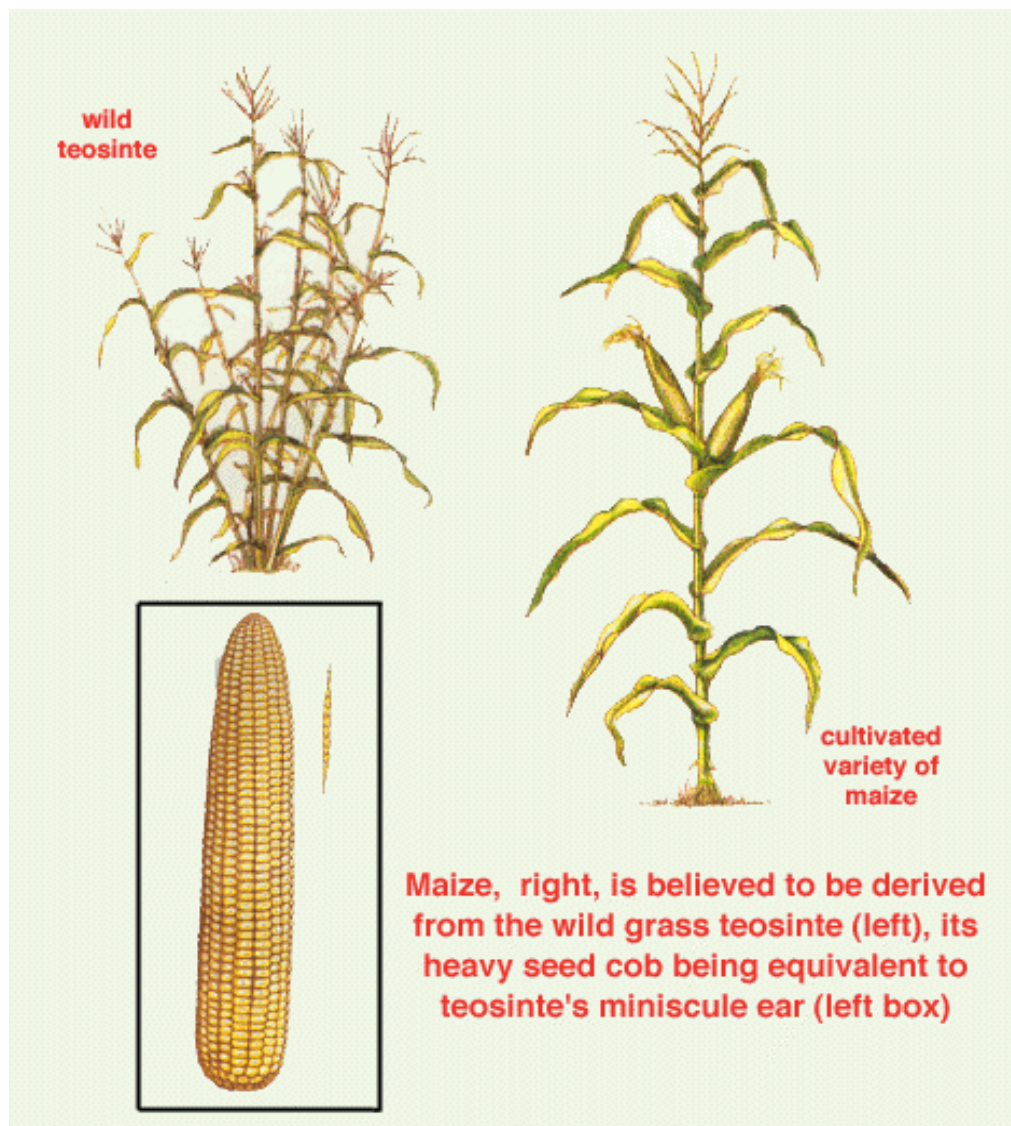


Figura 2. Diferenças morfológicas entre as partes vegetativas de teosinte e milho e entre as suas espigas (detalhe).

Desse modo, as hipóteses que surgiram tentaram sempre correlacionar essas duas plantas. A primeira, denominada “tripartite”, foi defendida por Paul C. Mangelsdorf (Mangelsdorf e Reeves 1939, Mangelsdorf *et al.* 1964, Galinat 1984) e, como o próprio nome sugere, baseava-se em três suposições. De modo bem simplificado, esse autor propôs que (1) o milho moderno teria surgido de um tipo de milho ancestral denominado “pod corn”, (2) o teosinte seria uma espécie recente, que teria surgido da hibridação entre o milho e uma gramínea do gênero *Tripsacum* e (3) a grande diversidade atual do milho seria fruto dos híbridos entre essas duas plantas. A segunda hipótese, conhecida como “descendência do teosinte”, foi proposta por um conjunto de autores dos quais se destacou George W. Beadle (Beadle 1939). De acordo com Beadle, o milho teria surgido diretamente do teosinte, como resultado do processo de seleção feito pelo homem ao longo dos anos de prática da agricultura na era pré-Colombiana.

O principal argumento de Mangelsdorf para justificar a hipótese da origem híbrida do teosinte foram alguns caracteres morfológicos de teosinte que são intermediários entre milho e *Tripsacum*. Entretanto, estudos que foram reportados após o anúncio de Mangelsdorf e Reeves, e até mesmo trabalhos que antecederam essa data, terminaram por derrubar essa hipótese, a qual foi revisada em 1973. Ainda em 1937, Albert E. Longley (Longley 1937) já havia apresentado resultados de citogenética que comprovavam a grande semelhança entre os dez cromossomos do teosinte e do milho e a grande diferença entre os genomas dessas duas plantas e os dezoito cromossomos de *Tripsacum*. Além disso, não eram conhecidas progênies resultantes do cruzamento entre milho e *Tripsacum*, e posteriores tentativas de hibridar essas plantas não foram bem sucedidas. Mais de trinta anos depois, Banerjee e Barghoorn (1972) reuniram às evidências já existentes resultados de palinologia obtidos com microscopia eletrônica e deram suporte à inverossimilhança da

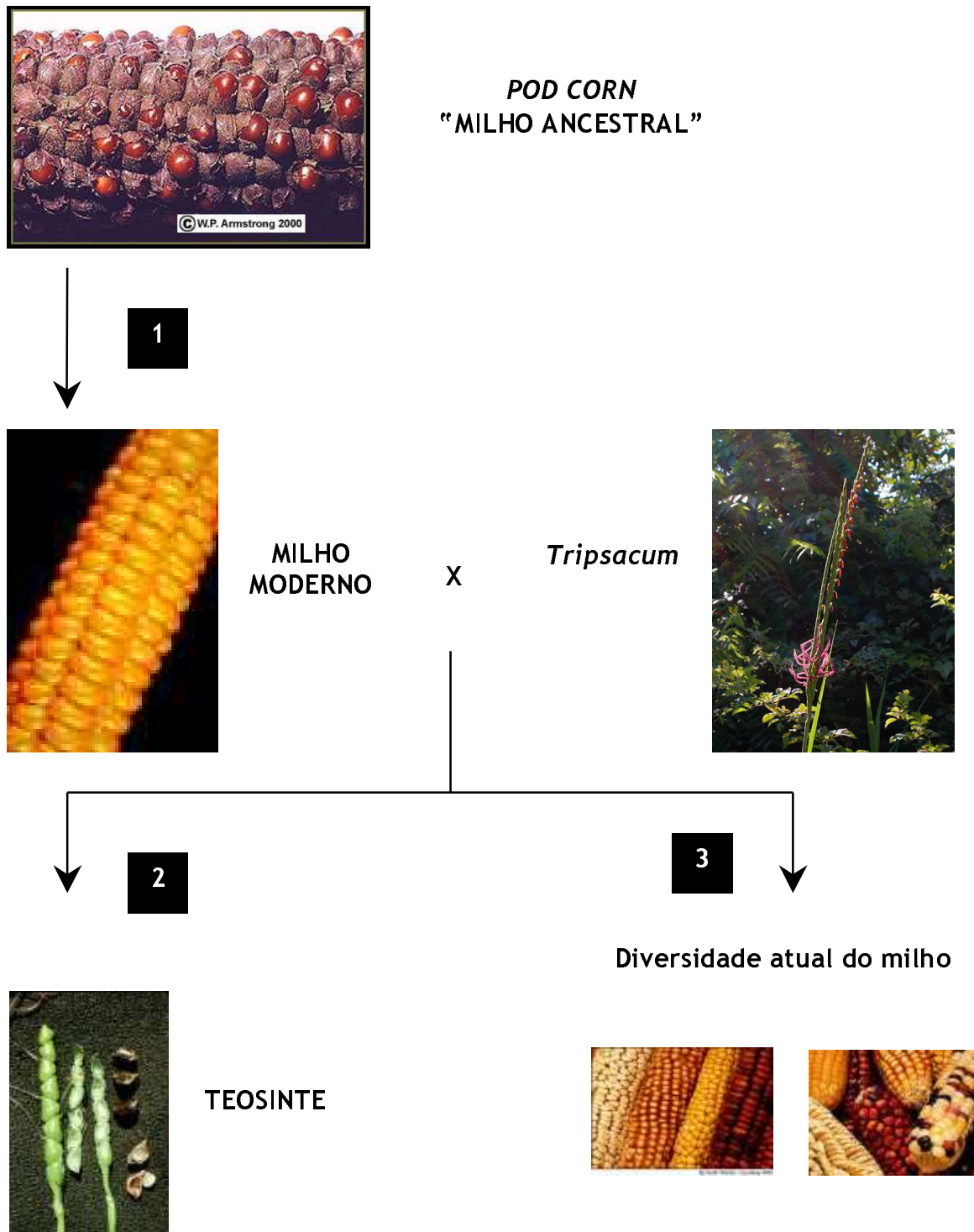


Figura 3. Hipótese “tripartite” proposta por Paul Mangelsdorf para explicar a origem do milho. (1) O milho moderno teria surgido do “pod corn”, (2) o teosinte seria resultado da hibridação entre milho e *Tripsacum* e (3) os híbridos entre essas duas plantas justificariam a diversidade atual do milho.

hipótese de Mangelsdorf. As micrografias revelaram pólen de milho e teosinte com espínulas regularmente espaçadas e pólen de *Tripsacum* com as mesmas estruturas em padrão agrupado. Como híbridos entre milho e *Tripsacum* são intermediários para esse caráter, teosinte não poderia ser resultante de um cruzamento entre as duas espécies.

Com base nessas fortes evidências, Mangelsdorf abandonou a defesa da origem híbrida do teosinte e reestruturou sua hipótese e, então, propôs a possibilidade de o teosinte e de o milho atual descenderem diretamente de uma planta de milho selvagem ancestral. A falta de evidência arqueológica que justificasse a hipótese de Beadle levou Mangelsdorf a defender essa idéia durante algumas décadas. Com efeito, os mais antigos registros de milho e teosinte são aproximadamente da mesma época, o que, portanto, dificultaria o surgimento de milho a partir de teosinte. Durante trinta anos, Mangelsdorf e outros autores reuniram evidências que pudessem comprovar a hipótese “tripartite” reformulada. Além da falha no registro arqueológico, argumentos como (1) maior especialização de alguns caracteres – comuns às duas plantas - no teosinte (adaptação a menor número de ambientes, redução da espiga de polística para dística, endurecimento das glumas, etc), (2) falta de evidência de que o uso do teosinte tenha antecedido o do milho (ou que até mesmo essa planta, com sementes tão fortemente revestidas, tenha sido usado na agricultura como fonte de alimento) e (3) pequena probabilidade de que as populações primitivas que habitavam Tehuacán tenham conseguido, por meio de seleção artificial, desenvolver uma planta de milho a partir do teosinte foram usados por muito tempo para sustentar suas idéias.

Entretanto, com o passar do tempo, Beadle também reuniu evidências que fortaleceram a corrente contrária, de descendência do milho a partir do teosinte. Como já mencionado, as morfologias do milho e do teosinte, sob alguns aspectos, em muito se equivalem. Estruturas como o caule, as folhas, a raiz e a inflorescência masculina, além de

aspectos genéticos, não deixam incertezas quanto à estreita relação entre essas duas plantas. Primeiramente, o argumento que usa a maior especialização do teosinte também foi usado por Beadle para defender a sua hipótese: (1) a espiga primitiva do teosinte é uma estrutura simultaneamente rígida e frágil que tem grande capacidade de dispersão de suas sementes, ao contrário do milho moderno, (2) o duro revestimento das sementes proporciona maior proteção a elas se compararmos aos grãos desprotegidos característicos do milho e, além disso, seus variados padrões de cores permitem camuflagem também protegendo contra o ataque animal, (3) as sementes possuem um mecanismo de dormência que, em condições favoráveis, garante a germinação da planta e (4) o teosinte responde adaptativamente a mudanças ambientais de forma eficaz, o que não acontece com o milho que depende da manipulação humana para sobreviver e que, provavelmente, se possuísse um ancestral com as mesmas caracteres da planta moderna, teria várias desvantagens com relação ao teosinte se tivesse dividido os mesmos *habitats* no passado (revisão em Beadle 1980). Em adição a essas análises que concernem as vantagens adaptativas do teosinte, um dos argumentos que Mangelsdorf usava para defender a hipótese do teosinte como descendente do milho – a possibilidade de o teosinte ser uma espécie recente – foi derrubado com a descoberta de grãos de aproximadamente 7000 anos, semelhantes ao teosinte moderno na parte sudeste da Cidade do México, e proporcionou maior credibilidade a hipótese de Beadle. Diante disso e das provas que se acumularam para contrapor a hipótese do milho como antepassado do teosinte, novas especulações sobre a ascendência do milho surgiram.

Na década de 80, Hugh H. Iltis (Iltis 1983), considerando o conceito darwiniano de mudança gradual no surgimento de novas espécies e diante da escassa evidência arqueológica de intermediários entre o teosinte e o milho, sugeriu que teria havido uma “transmutação sexual” no teosinte. Esse evento único teria transformado a inflorescência

masculina do teosinte na inflorescência feminina do milho por ocasião de uma mudança súbita no ambiente. Esta alteração, então, teria sido fixada e a planta, domesticada. Apesar de não haver dúvida sobre a homologia entre tais estruturas, posteriores evidências arqueológicas e genéticas (Galinat 1985) não indicam como possível a hipótese de macromutações em épocas evolutivas mais recentes.

Os trabalhos mais atuais ainda concentram esforços na procura de mutações, do número mínimo de locos que diferencia milho e teosinte e das possíveis ligações entre esses locos. John Doebley é um dos autores com maior expressão nessa área, assinando diversos estudos de alta informatividade no que concerne à identificação de possíveis locos cujos alelos mutantes deram origem ao milho. Em seu trabalho com Adrian Stec (Doebley e Stec 1991), baseando-se em estudos anteriores que propunham que a maioria das mudanças morfológicas em vegetais era iniciada por mutações com grandes efeitos no fenótipo, Doebley analisou várias características em plantas F₂ obtidas a partir de uma F₁ híbrida (milho-teosinte) e as correlacionou com a segregação de marcadores moleculares. Os resultados indicaram que as características-chave que distinguem as inflorescências de milho e teosinte, por exemplo, são reguladas por um grupo de genes que varia em número de quatro a oito. Entretanto, os autores reconhecem que maior importância deve ser dada à questão dos efeitos que cada região do genoma controlando um caráter é capaz de exercer. Nesse caso, apenas um ou dois genes seriam responsáveis por grande parte da variação fenotípica encontrada no material que estudaram. Todavia, evidências definitivas sobre a existência de muitos ou poucos genes interferindo nas alterações morfológicas não foram apresentadas.

Dois anos depois, Jane Dorweiler e colaboradores (Dorweiler *et al.* 1993) descobriram um gene (*tga1- teosinte glume architecture 1*), responsável pela transformação

da espiga do teosinte. Alelos mutantes para este loco promovem modificações na arquitetura da gluma, acabando por expor o grão na espiga. Os dados então obtidos confirmaram as especulações feitas no trabalho anterior: o controle das transformações que deram origem ao milho é feito por um pequeno número de genes e não por meio de uma herança do tipo poligênica, como havia sido cogitado antes dos estudos de Doebley e Stec. Em novas investigações no ano seguinte, Doebley e colaboradores (Doebley *et al.* 1994) analisaram a característica peso do grão em populações híbridas de teosinte-milho e, mais uma vez, concluíram que são vários os locos envolvidos na determinação da diferença desse caráter entre as duas plantas, mas somente poucos genes têm efeito expressivo.

Em outra publicação, Doebley e Stec, agora com a colaboração de Lauren Hubbard (Doebley *et al.* 1997), mais uma vez, reforçaram suas idéias. Após terem identificado o gene *tb1*- *teosinte branched 1* (Doebley *et al.* 1995), neste trabalho eles apresentaram resultados sobre a sua possível função: repressão do crescimento de órgãos axilares e promoção da formação das inflorescências femininas, o que reduzia o aspecto ramificado do teosinte a uma planta com grande dominância apical, sem ramificações laterais longas e terminadas por espigas. Além disso, relataram a expressão em dobro do alelo *tb1* do milho se comparado ao alelo do teosinte. Essas análises derrubaram definitivamente as especulações de Iltis, e provaram a existência de regulação genética para esse caráter que diferencia teosinte e milho. Ainda detalhando as questões que a descoberta do gene *tb1* trouxe, Wang *et al.* (1999) avaliaram a possibilidade de ter havido seleção para esse alelo durante a domesticação do milho. Genes envolvidos na evolução da planta deveriam exibir menor taxa de polimorfismo e agruparem-se de forma consistente em um único clado. Esse grupo mostrou que no milho, a região 5'-flanqueadora, responsável pela regulação de *tb1*, possui apenas 3% da diversidade encontrada no teosinte e que a filogenia feita com regiões

regulatórias de várias amostras exibia um único clado com pouca variação entre elas. Portanto, comprovou-se que a região regulatória, a qual determina a dosagem de expressão do gene *tb1*, foi alvo de seleção, tendo esse gene participado ativamente da evolução do milho a partir do teosinte. Com essas informações, mais uma vez reforçaram-se as hipóteses de descendência do teosinte e de que as alterações morfológicas que foram responsáveis por tal evolução são governadas pela regulação diferencial em poucos locos.

Mais recentemente, um novo grupo, também coordenado por Doebley (Matsuoka *et al.* 2002), estudou um outro aspecto da evolução dessas plantas e testou a possibilidade de a grande diversidade do milho atual ser decorrente de domesticações múltiplas a partir de plantas de teosinte. Seus resultados comprovaram uma origem monofilética para as plantas atuais de milho, sugerindo apenas um único evento de domesticação, ocorrido com base na subespécie *parviglumis*.

O volume de dúvidas que cercou as origem e evolução do milho aparenta hoje ser bem menor do que na época dos primeiros trabalhos de Mangelsdorf e Beadle. O avanço da biotecnologia, disponibilizando diversas técnicas e métodos de análise, corroborou de forma decisiva para o salto que foi dado na investigação da história do milho. Atualmente, os estudos publicados não cogitam outra possibilidade para o México como centro de origem do milho e para o surgimento dessa espécie senão a partir da derivação do teosinte, e isso se deve em muito à adoção de práticas genéticas mais aprimoradas. A técnica do DNA recombinante, o uso de marcadores moleculares e de análises de QTLs (*Quantitative Trait Loci*) proporcionaram uma outra perspectiva na busca de informações sobre a evolução do milho. Cada vez mais a metodologia molecular está sendo empregada na descoberta de novos locos responsáveis pelo diferencial entre milho e teosinte, pois, dessa maneira, poder-se-á esclarecer os rumos evolutivos que levaram à sua diferenciação. Essas

informações serão de grande relevância para o conhecimento mais aprofundado da variabilidade dessa espécie e para o delineamento de programas de melhoramento. No processo de obtenção de híbridos, onde é necessário um grande número de plantas produzidas por meio de sucessivos eventos de autofertilização, é comum o aparecimento de linhagens susceptíveis a estresses bióticos e abióticos e a correta identificação da origem genética do milho pode suprir importantes informações sobre as mais valiosas fontes da diversidade que podem contornar esse problema.

2.2) Diversidade do milho: dos estudos fenotípicos às análises moleculares

Coletar, caracterizar, nomear, catalogar e analisar as inúmeras variedades de plantas conhecidas sempre foi uma preocupação do homem. Sabe-se que manter viáveis diferentes expressões fenotípicas de uma mesma espécie é o mesmo que dispor de diversidade genética em larga escala e poder manipular características de interesse rumo a plantas com maior vigor em vários aspectos. O milho é a espécie que exibe maior diversidade entre os cereais, sendo atualmente conhecidas cerca de 300 raças que se apresentam na forma de milhares de cultivares adaptados a diferentes regiões, climas, qualidades de solo e altitudes, e resistentes a diversos tipos de doenças. Essa imensa disponibilidade de diversidade genética atual é resultado, em primeira análise, do contínuo processo de seleção feito pelos pequenos produtores mexicanos em épocas remotas, e da paralela seleção natural imposta pela natureza. A incessante busca por plantas mais produtivas ou com características que fossem de maior interesse para uma ou outra pequena propriedade proporcionou a divergência de uma única espécie rumo a plantas com exuberante variação genotípica e fenotípica. A palavra “landrace”, modo como são denominadas as raças de milho que há

muito vêm sendo desenvolvidas e mantidas por pequenos produtores e que são artificialmente selecionadas para suprirem preferências locais, define com rigor o “pool” genético de onde divergiu grande quantidade de genes e, conseqüentemente, de caracteres morfológicos. Também a chegada do milho ao Velho Mundo, após a descoberta da América, submeteu essa planta a novos ambientes e mais diversidade resultou dessa interação. Em adição, nos dias de hoje, com o desenvolvimento de novos métodos de melhoramento vegetal, não somente em termos de raças é avaliado o potencial de reserva genética do milho, mas também em termos de populações de polinização aberta e de linhagens, plantas sucessivamente autofecundadas durante várias gerações e que tendem a revelar quase todos os seus locos em estado de homozigose.

O início do uso de linhagens em programas de melhoramento está associado à busca do vigor de híbrido, ou heterose, fenômeno que proporciona grande produtividade em plantas F_1 provenientes do cruzamento de genitores que exibem alta divergência entre si. Na tentativa de escolha dos melhores genitores, ensaios vêm sendo realizados desde o século XIX e mostram que cruzamentos entre linhagens com relativa divergência entre si resultam em híbridos com alto vigor. As causas para essa correlação entre heterose e diversidade ainda não estão bem estabelecidas, e algumas teorias que tentam explicar as causas genéticas da heterose ainda hoje são discutidas. As duas principais correntes são (1) dominância, segundo a qual alelos recessivos potencialmente deletérios ficariam ocultos nos heterozigotos obtidos em F_1 e os prejuízos decorrentes da homozigose para esses alelos seriam evitados e (2) sobredominância, que credita o vigor apresentado pelas plantas F_1 à elevada heterozigosidade, a união das duas formas alélicas do heterozigoto sendo superior à ação separada de qualquer um dos alelos em homozigose; a variabilidade das formas protéicas em plantas heterozigotas as tornaria mais eficientes frente à variação de condições

ambientais (Crow 1948). Independentemente das causas genéticas que baseiam a ocorrência da heterose, tornou-se importante avaliar os limites de divergência que geram maiores níveis de vigor na F_1 e que podem resultar em bons híbridos em programas de melhoramento. Estudos em milho (Prasad e Singh 1986) mostraram que, apesar de não haver uma correlação linear entre diversidade genética e heterose, é importante selecionar linhagens que apresentem moderada divergência genética. Portanto, mais uma vez o conhecimento sistematizado dos recursos genéticos do milho comprovou ser necessário e maiores avaliações sobre a diversidade do milho foram propostas.

A seguir, são apresentados os principais estudos que contribuíram e ainda contribuem para a análise da extensa diversidade do milho. Estes estudos buscam prover programas de melhoramento dessa espécie com valiosas informações sobre as plantas mais importantes para formação de híbridos, bem como, sobre plantas que possuem genes de reconhecida relevância na adaptação a estresses bióticos e abióticos. Dos estudos mais antigos, que se basearam apenas na observação de caracteres morfológicos até os atuais, que são realizados por meio de sequenciamento de DNA, todos compõem um esforço mundial de organização e entendimento da diversidade do milho e conseqüente melhor aproveitamento das reservas genéticas disponíveis.

2.2.1) Da morfologia às isozimas

As primeiras tentativas de classificar e organizar a diversidade do milho envolveram estudos com base na localidade original das plantas e também na descrição de características fenotípicas como o tipo de endosperma ou a cor do grão. As características morfológicas foram por muito tempo usadas por Carl Linnaeus em seu sistema de

classificação. Todavia, posteriormente, comprovou-se que a morfologia representa um meio muito limitado no que concerne à análise da diversidade. Isso se deve principalmente à grande influência ambiental que recebem e, dessa forma, podem não revelar obrigatoriamente relações genéticas (Smith e Smith 1989). Coe *et al.* (1988) já haviam mostrado a inadequação do uso da morfologia do endosperma com objetivos de estabelecer grupos heteróticos entre linhagens de milho. Foi demonstrado que as diferenças entre alguns tipos de endosperma podem estar associadas a apenas um único gene. Entretanto, a reduzida cobertura do genoma, devido à existência de poucos locos disponíveis e polimórficos, constituiu um fato limitante à aplicação generalizada dessa técnica. Em adição, comprovou-se que a utilização de alguns tipos de proteínas, como as zeínas, produzidas apenas no endosperma, pode fornecer informações viesadas, pois mostram em maior proporção a herança ligada a apenas um dos progenitores (Smith 1988).

Apesar disso, os primeiros estudos, os quais não dispunham dessas informações, usaram a morfologia como método para agrupar as várias raças de milho conhecidas, e contribuíram para o início da organização da variabilidade dessa espécie. Em 1977, Major Goodman e Robert Bird (Goodman e Bird 1977) realizaram um importante trabalho separando mais de duzentas raças latino-americanas de milho em quatorze grupos, partindo de vinte caracteres morfológicos da espiga e das localidades originais das plantas. Nesse trabalho, os autores apresentaram uma vasta discussão sobre a importância do rigor que deve acompanhar as diversas fases de um estudo de diversidade. Goodman e Bird analisaram os problemas que a classificação por meio da morfologia traz e mostraram como isso pode inserir erros nos agrupamentos. Questões como a possibilidade de uma mesma planta ser classificada com nomes diferentes, a necessidade de seleção de características que não recebam grande influência do ambiente, e a posição intermediária de algumas raças

com relação a dois grupos próximos foram abordadas como alguns dos principais problemas em estudos morfológicos. Além de discutirem pertinentes assuntos metodológicos, os autores elaboraram conclusões muito relevantes sobre a distribuição e o uso do germoplasma nas Américas: (1) a principal fonte de variabilidade do milho concentra-se nos “dents” e nos “flints” caribenhos; (2) o grande número de introduções, vindas na América do Norte, Europa e México, torna difícil a caracterização do germoplasma brasileiro, raças indígenas e exóticas não sendo distinguidas, e (3) a produção de híbridos na América do Norte até meados de 1970 foi feita com base em pequeno número de linhagens.

Alguns anos mais tarde, Brown (1985) apresentou estudo semelhante, com a descrição de oito raças de milho temperado usadas nos Estados Unidos. O autor usou raças contrastantes, objetivando abranger a maior parte da variabilidade possível, e sugeriu que uma maior amplitude da diversidade deve ser utilizada nos programas de melhoramento.

Além dos germoplasmas latino-americano e norte-americano, recentemente, um estudo de diversidade fenotípica foi apresentado por Li *et al.* (2002) com raças e linhagens chinesas de milho. A China, segunda maior produtora mundial de milho, possui grandes áreas destinadas à essa cultura onde são mantidas “landraces” chinesas e linhagens usadas em programas de melhoramento. O estudo mostrou que a maior parte da diversidade genética do milho chinês está presente nas “landraces”, devido à forte seleção humana, e que a conservação desse tipo de germoplasma é fundamental para a manutenção da variabilidade do milho na China. Em adição, o trabalho sugere que o centro de origem do milho chinês é a região sudoeste do país e que a possibilidade de adaptação a várias condições gerou maiores níveis de diversidade.

Com o desenvolvimento de técnicas bioquímicas, um outro método de avaliar a diversidade genética - as isozimas - passou a ser amplamente empregado. Essa metodologia permite avaliar a diversidade comparando-se diferentes variantes protéicas entre as plantas e, com este tipo de informação, pôde-se superar significativamente o problema da interferência do ambiente, pois as proteínas retratam com maior fidelidade as bases genéticas.

Com relação ao milho tropical, Doebley *et al.* (1985) avaliaram 34 raças de milho mexicanas, distribuídas em 94 coleções, com 23 locos de isozimas. Foi encontrada alta diversidade no material, sendo que a maior parte da variabilidade estava presente dentro das coleções. Entretanto, considerável diversidade entre coleções foi verificada, o que pode fazer de uma coleção uma amostra não representativa de sua raça. Os níveis de diversidade no milho mexicano, segundo esse estudo, equiparam-se aquele encontrado nos teosintes, e estão correlacionados a determinadas regiões e altitudes.

Em contrapartida à pequena quantidade de estudos com milho tropical, o germoplasma norte-americano, na década de 1980, foi extensivamente estudado por meio de isozimas. Smith *et al.* (1985a; 1985b) apresentaram duas importantes análises sobre a variabilidade genética de linhagens dos EUA, a primeira abordando linhagens historicamente relevantes e a segunda abordando as principais linhagens usadas na década de 1970. O estudo que usou linhagens históricas revelou que a diversidade genética nos EUA é maior do que se esperava se for considerado apenas o número de raças envolvidas na formação desse germoplasma. Os autores argumentam que a capacidade das isozimas de exibir heterogeneidade intra-racial pode justificar a obtenção de altos níveis de polimorfismo e conseqüente variabilidade. Além disso, é argumentado que os três maiores grupos de linhagens que compuseram a fonte de germoplasma elite nesse país (“Reid

Yellow Dent”, “Lancaster Sure Crop” e “Iowa Stiff Stalk Synthetic”) possuem, cada um, alta diversidade, e os cruzamentos entre esses grupos disponibilizaram ainda mais variação. Os autores, entretanto, alertam para o fato de que a maioria dos híbridos produzidos deriva quase exclusivamente de cruzamentos entre “Lancaster Sure Crop” e “Iowa Stiff Stalk Synthetic”. No segundo trabalho, o objetivo foi investigar se o uso contínuo do mesmo germoplasma havia promovido erosão genética. Os resultados obtidos pelos autores mostraram que a redução no número de locos monomórficos e o aumento no número de alelos, ao contrário do esperado, proporcionaram aumento da diversidade nesse período. As linhagens “Reid Yellow Dent” tiveram um pequeno decréscimo em variabilidade, mas no grupo “Iowa Stiff Stalk Synthetic” foi detectado acréscimo, com novos genótipos sendo identificados. Esse evento, segundo os autores, provavelmente decorreu da introdução gradual de novos alelos por meio das linhagens que não eram usadas como o parental recorrente em programas de retrocruzamento. Um estudo com variedades de polinização aberta (Smith 1986) também concluiu que a diversidade disponível no germoplasma americano é superior àquela esperada, mas menciona que os estudos com locos de isozimas detectaram queda de 30% na variabilidade genética, ocasionado por perda de alguns genótipos na década de 1970. O trabalho ainda mostra que os níveis de diversidade das variedades de polinização aberta e das linhagens elite, obtido em termos de número médio de alelos de isozimas, não foram discrepantes, e traduzem a manutenção da variabilidade em materiais de grande importância no melhoramento do milho de clima temperado.

2.2.2) Do RFLP aos microssatélites

A partir da segunda metade da década de 1980, uma grande revolução nas

metodologias moleculares ocorreu e novas técnicas começaram a ser aplicadas com objetivos de retratar com maior precisão a diversidade entre genótipos vegetais. RLFPs (*Restriction Fragment Length Polymorphisms*), RAPDs (*Random Amplified Polymorphic DNAs*) (Welsh e McClelland 1990; Newbury e Ford-Lloyd 1993), AFLPs (*Amplified Fragment Length Polymorphisms*) (Vos *et al.* 1995), e SSRs (*Simple Sequence Repeats*) (Litt e Luty 1989) revolucionaram a análise da variabilidade genética vegetal pois revelaram métodos mais confiáveis e mais práticos na obtenção de informações sobre diversidade genotípica. Apesar de serem vários os métodos de estudo da diversidade, nenhum deles mostra-se tão eficaz como os marcadores moleculares. Dados obtidos por meio de técnicas moleculares superam a maioria das limitações existentes nas demais formas de análise. Características como (1) número de marcadores praticamente ilimitado, (2) falta de influência de fatores ambientais, (3) grande quantidade de locos polimórficos, (4) acesso indistinto à contribuição de ambos genitores e, principalmente, (5) o fato de permitirem comparações entre genótipos, considerando-se o DNA propriamente dito, tornam esse tipo de marcador diferenciado no que diz respeito à estimativa de diversidade genética.

O RFLP foi o primeiro marcador de DNA amplamente usado para estimativas de diversidade em milho. O padrão de herança codominante e a grande quantidade de locos disponíveis tornaram esse marcador o principal método de avaliação genotípica no começo da década de 1990, quando diversos estudos em milho e outras espécies foram apresentados. Ainda hoje o RFLP é usado para avaliar diferentes germoplasmas, porém, por ser uma técnica muito trabalhosa e dispendiosa, o RFLP cedeu espaço a metodologias mais rápidas e acessíveis principalmente após o advento do PCR.

Entre os marcadores de DNA, o que primeiro se destacou foi a técnica do RAPD. A

possibilidade de análise genômica, sem a necessidade do prévio conhecimento das seqüências do organismo sob estudo disseminou rapidamente o RAPD. Apesar de não exibir muitos polimorfismos, de ser um marcador dominante – não permite que, em espécies diplóides, por exemplo, os alelos dos dois homólogos sejam distinguidos - e de ser uma técnica muito sensível a alterações nas condições de amplificação, por um longo período o RAPD foi, com sucesso, usado na análise da diversidade.

O AFLP, uma outra técnica recentemente desenvolvida, superou problemas apresentados pelo RFLP e pelo RAPD, e rapidamente passou a ser usada em trabalhos de diversidade. Unindo características desses dois últimos marcadores, o AFLP é considerada a metodologia mais eficiente em termos de obtenção de polimorfismo por experimento. Seu padrão de herança é dominante e a sua aplicabilidade se estende a todas as espécies vegetais.

Conhecidos desde o final da década de 1980, os microssatélites têm mostrado diversas vantagens sobre os demais marcadores moleculares. Também conhecidos como SSRs, os microssatélites são seqüências de 2 a 6 pares de bases, repetidas *em tandem* no genoma e flanqueadas por regiões altamente conservadas. Apresentam como características principais (1) a distribuição uniforme e randômica, (2) a herança mendeliana do tipo codominante, (3) a alta reprodutibilidade via PCR, (3) a abundância e (4) o multialelismo (Chin *et al.* 1996). Apesar da necessidade de conhecimento prévio de seqüências genômicas, após o desenvolvimento de *primers*, o uso desse tipo de marcador genético é o mais simples entre todos já apresentados. Baseando-se apenas em reações de PCR e eletroforese dos produtos de amplificação, é possível realizar genotipagem em larga escala de forma rápida e fácil.

A função dos microssatélites ainda não foi elucidada, mas algumas pesquisas revelam que podem estar associados, por exemplo, à adaptação de bactérias a ambientes diversos e também a doenças neurológicas no homem (Moxon e Wills 1999), além de servirem como sítios de ligação para proteínas responsáveis pela regulação gênica durante a intérfase em *Drosophila melanogaster* (Csink e Henikoff 1998). Mesmo com função desconhecida, o uso dos microssatélites é muito disseminado no que concerne à identificação genotípica em plantas.

Para análises de plantas como o milho, os microssatélites vêm se consolidando como a técnica mais informativa. A disponibilidade de um grande banco de dados, que contém mais de 1800 seqüências de pares de *primers* já desenvolvidas para amplificação de microssatélites nessa espécie (*Maize Genetics and Genomics Data Base* – <http://www.agron.missouri.edu/ssr.html>), torna fácil a exploração do seu genoma e a obtenção de dados sobre a sua variabilidade genotípica. Atualmente, praticamente todos os estudos sobre diversidade do milho se baseiam prioritariamente nos microssatélites e muitos fornecem informações sobre correlações entre os quatro principais tipos de marcadores moleculares. Nos parágrafos seguintes, foi elaborada uma revisão com os mais importantes estudos sobre a diversidade do milho no âmbito mundial. O milho de clima temperado, representado em sua maioria por material usado no Cinturão do Milho nos EUA e pelas linhagens européias, foi por várias vezes analisado e diversas foram as abordagens dadas a esse tipo de germoplasma. Entretanto, percebe-se com facilidade a carência de trabalhos que abordem o germoplasma tropical, o qual é tratado em muito poucas análises.

2.2.2.1) Milho de clima temperado

A década de 1990 foi, sem dúvida, o período quando mais trabalhos sobre a diversidade do milho foram apresentados. Como mencionado anteriormente, o RFLP foi a primeira técnica baseada em DNA a auxiliar os pesquisadores na obtenção de dados. Melchinger *et al.* (1990; 1991) iniciaram o uso dos fragmentos de restrição na análise da diversidade usando as linhagens do Cinturão de Milho americano, as quais, mais tarde, seriam muitas outras vezes estudadas. Esses pesquisadores pretendiam avaliar a possibilidade de uso dos RFLPs na predição de heterose na F_1 e a diversidade entre algumas linhagens dos três principais grupos de milho norte-americano. Foi encontrada variabilidade moderada entre os genótipos e as correlações com a produtividade foram consideradas pequenas. No segundo trabalho, os dados moleculares concordaram com as informações de *pedigree* já existentes, separando as linhagens em grupos heteróticos como se previa com base nas suas origens. Alguns anos mais tarde, Mumm e Dudley (1994) apresentaram um trabalho mais amplo, com 148 linhagens, que corroborou as informações anteriormente obtidas por Melchinger e colaboradores. Mais uma vez, a separação das linhagens por meio do RFLP foi semelhante à esperada partindo-se de dados de *pedigree*. Os autores ainda acrescentaram que o fator genético que proporcionou a diferenciação entre os grupos heteróticos foi a frequência diferencial dos alelos e não a presença/ausência de alelos distintos.

Com a evolução dos marcadores moleculares, novas metodologias foram incorporadas à análise do germoplasma norte-americano e surgiu a possibilidade de comparação entre o potencial das técnicas. Nos anos de 1996 e 1997, Taramino e Tingey (1996) e Smith *et al.* (1997) estudaram pela primeira vez a utilidade dos microssatélites como marcadores moleculares em milho e revelaram que os SSRs possuem capacidade discriminatória superior a dos RFLPs. O primeiro trabalho calculou as médias de

heterozigosidade para microssatélites e RFLP (respectivamente 0,76 e 0,58) para vinte linhagens. Como a capacidade de um marcador molecular em promover diferenciação entre indivíduos é tanto maior quanto maior for o número de alelos que ele gerar, verificou-se que os microssatélites possuem significativa vantagem sobre a outra técnica. No estudo de Smith e colaboradores, que utilizaram um número maior de linhagens e cálculos semelhantes aos de Taramino e Tingey para a heterozigosidade, observaram-se conclusões semelhantes. Em adição, a correlação entre os dados de microssatélites com o *pedigree* das linhagens foi alta para esse germoplasma ($r=0,81$), o que os autores justificaram com base na boa cobertura genômica que os SSRs provêm.

Com o objetivo de correlacionar a diversidade genética com a performance de híbridos, Ajmone-Marsan *et al.* (1998) usaram, além dos dados de RFLP, informações de AFLP. Assim como no trabalho de Melchinger *et al.* (1990), agrupamentos como os feitos por meio do *pedigree* foram obtidos e a correlação entre as distâncias genéticas e a produtividade na F_1 não foram consideradas satisfatórias para que a predição de cruzamentos pudesse ser feita somente com base nos dados moleculares. Dados de Drinic *et al.* (2002), com base apenas em microssatélites, também mostraram resultados semelhantes para correlação com heterose.

Ainda no ano de 1998, Pejic *et al.* (1998) apresentaram um dos principais estudos comparativos entre os marcadores moleculares em milho. Usando as quatro técnicas anteriormente descritas, os autores concluíram que, de fato, há diferença entre o nível de polimorfismo revelado em cada uma delas. Por meio do cálculo de heterozigosidade e do número médio de alelos, mostraram que, em média, a quantidade de informação obtida por microssatélites é duas vezes maior que por AFLPs e RAPDs, além de gerar 40% mais dados quando comparado aos RFLPs. Os SSRs mostraram a maior taxa de polimorfismo e

os AFLPs, a menor, apesar dessa técnica ter sido considerada o sistema mais eficiente. Na comparação com informações de *pedigree*, apenas os resultados obtidos com os RAPDs foram discrepantes, os demais estando bem correlacionados e ratificando a possibilidade de recuperação do *pedigree* a partir de dados moleculares. As correlações entre cada tipo de marcador, com exceção do RAPD, foram de moderada a alta ($r=0,59$ para RFLP-SSR e $r=0,70$ para AFLP-RFLP).

O primeiro grande estudo de diversidade envolvendo apenas microssatélites e as reconhecidas linhagens dos EUA foi feito por Senior *et al.* (1998), e seguido por Lu e Bernardo (2001) e Gethi *et al.* (2002). Senior e colaboradores usaram 94 linhagens e 70 locos de microssatélites. Os resultados, mais uma vez, alocaram os genótipos em grupos já esperados, e moderada diversidade foi encontrada, com média de alelos de 5,0 e heterozigosidade de 0,59. Com objetivos que iam um pouco além da simples avaliação da diversidade, Lu e Bernardo (2001) compararam a variabilidade entre linhagens históricas e aquelas atualmente usadas no programas de melhoramento americanos. Com 32 genótipos históricos e oito genótipos elite, os pesquisadores mostraram que houve redução da diversidade no nível gênico, com diminuição do número médio de alelos de 4,9 no material histórico para 3,2, nas linhagens atualmente usadas. Todavia, no nível populacional, medido em termos de distâncias genéticas, esse decréscimo não foi observado. Os autores argumentaram que a redução no número de alelos não implica necessariamente a redução das distâncias genéticas entre linhagens de grupos heteróticos distintos, principalmente se as frequências de alguns dos alelos na população base for baixa. Finalizando os estudos mais recentes que usam os três principais grupos de linhagens norte-americanas (“Reid Yellow Dent”, “Lancaster Sure Crop” e “Iowa Stiff Stalk Synthetic”), Gethi *et al.* (2002) analisaram a variabilidade existente entre os estoques de sementes de algumas linhagens de

grande interesse. Foi identificada pequena, porém significativa variação, entre os estoques de uma mesma linhagem e, até mesmo, dentro dos estoques de um mesmo genótipo. Os autores discutiram a importância de uniformidade entre os estoques para efeitos de conservação genética, mapeamento e estudo de frequência gênica e alertaram para o cuidado que melhoristas devem ter ao usar as linhagens que exibiram esse tipo de variação.

Ainda sobre o milho de clima temperado, estudos com germoplasma europeu e japonês também foram realizados. Assim como nos trabalhos com material norte-americano, a análise da diversidade com linhagens européias também iniciou com o RFLP e incorporou gradativamente outros tipos de marcadores. Boppenmaier *et al.* (1993) apresentaram um trabalho semelhante aos de Melchinger *et al.* (1990) e Smith *et al.* (1997), buscando correlações entre informações moleculares e a performance de híbridos. Os resultados não diferiram daqueles obtidos para o milho norte-americano: os dados obtidos com RFLPs separaram os grupos heteróticos já conhecidos, apresentaram moderada diversidade, mas não foram úteis na predição de cruzamentos entre linhagens de grupos diferentes. A identidade entre o comportamento dos germoplasmas dos EUA e da Europa deve ser vista com cautela, uma vez que várias foram as introduções de material norte-americano em território europeu. Um estudo comparativo entre RAPD, RFLP e informações de *pedigree* foi elaborado por Hahn *et al.* (1995) e apresentou baixa diversidade entre antigas linhagens européias - similaridade genética (GS) variando de 0,57 a 0,80. Os autores ainda concluíram que os RFLPs superam os RAPDs quando se trata de correlações entre cada tipo de marcador e as informações de *pedigree*, o que provavelmente acontece devido a maior amostragem genômica que os primeiros proporcionam. Complementando esse trabalho, Lübberstedt *et al.* (2000) incluíram análises de AFLP às linhagens previamente avaliadas e também esse marcador comprovou ser mais adequado

que o RAPD para alocar genótipos em diferentes grupos. Com relação ao germoplasma japonês anteriormente mencionado, o qual também recebeu influência de material norte-americano, Enoki *et al.* (2002) trabalhou com linhagens adaptadas a regiões de baixa temperatura e revelou alta diversidade com média de alelos superior a 7,0 e heterozigosidade média próxima de 0,7 e os dados de SSRs usados por esses pesquisadores reconstituíram grupos heteróticos já conhecidos, o que gerou uma correlação entre *pedigree* e microssatélites alta ($r=0,70$).

2.2.2.2) Milho de clima tropical

Importantes trabalhos sobre o milho tropical têm sido desenvolvidos em nosso laboratório (Laboratório de Análises Genética e Molecular – CBMEG/UNICAMP) e tentam avaliar a variabilidade genética existente no material brasileiro. O primeiro estudo com marcadores moleculares foi apresentado em 1997 e usou RAPDs para avaliar a diversidade do germoplasma brasileiro e para prever cruzamentos entre duas populações (Lanza *et al.* 1997). Alguns anos depois estudo semelhante com as mesmas duas populações envolveu dados de RFLP (Benchimol *et al.* 2000) e, em 2003, Barbosa *et al.* (2003) apresentaram resultados com AFLP e SSR. Lanza *et al.* (1997) encontraram diversidade moderada entre as linhagens das duas populações, com distâncias genéticas (GD) variando de 0,47 a 0,75, observaram que o RAPD aloca de forma distinta os genótipos, se comparado ao agrupamento feito somente com base no *pedigree*, e mostraram que, ao contrário das informações providas por este, existe correlação significativa entre os dados moleculares e estimativas de produtividade. Acredita-se que a grande base genética que deu origem ao germoplasma brasileiro tenha proporcionado a falta de correlação entre

os dados de *pedigree* e os moleculares. O posterior trabalho com RFLPs, assim como observado nos estudos com material temperado, mostrou que esse marcador molecular reconstituiu os agrupamentos da mesma forma como esperado a partir do *pedigree* dos genótipos (duas populações). Além disso, foi observada alta correlação entre as GDs e os valores de produtividade na F₁ quando linhagens de uma mesma população estavam sendo analisadas. As recentes análises de Barbosa *et al.* (2003) também revelaram o AFLP como metodologia a partir de que se pode agrupar genótipos de diferentes grupos heteróticos e prever o potencial de cruzamentos intrapopulacionais. Com relação aos microssatélites, os quais exibiram maior nível de diversidade que os AFLPs, o mesmo não se verificou, pois a sua grande capacidade de revelar polimorfismo exacerba as diferenças entre os genótipos e pode levar a agrupamentos diferentes do previsto somente com base no *pedigree*.

Ainda sobre o germoplasma brasileiro, um trabalho ainda não publicado (Oliveira *et al.* 2004) usou AFLPs para avaliar a diversidade entre 96 importantes linhagens de milho pertencentes ao Banco de Germoplasma do Instituto Agrônomo de Campinas (IAC). A técnica usada nesse estudo provou ser muito eficaz na detecção de polimorfismo e exibiu grande diversidade genética. A alta divergência entre as linhagens não permitiu a formação de grupos heteróticos bem definidos e trouxe como perspectivas a possibilidade de o germoplasma tropical não se comportar como o temperado e a necessidade de aprofundamento dos conhecimentos a respeito do germoplasma brasileiro.

O recente estudo de Warburton *et al.* (2002) constitui a primeira grande tentativa de classificar o germoplasma por meio de SSRs. Com base em 57 importantes linhagens mexicanas e sete populações mantidas pelo Centro Internacional de Milho e Trigo (CIMMYT), os autores mostraram a alta variabilidade de que o material mexicano possui. As populações formaram grupos de acordo com o previsto pelo *pedigree*, mas apenas as

linhagens de origem muito similar agruparam-se nos dendrogramas. Os autores discutem que a característica intrínseca que o milho mexicano tem de ser originário de uma ampla base genética pode dificultar a sistematização da diversidade nesse tipo de germoplasma, porém acreditam que o uso de marcadores moleculares é fundamental para o aprimoramento da organização dessa diversidade.

2.3) Métodos estatísticos de análise da diversidade

Muitos métodos de análise de dados de marcadores moleculares já foram desenvolvidos e facilitam a interpretação das informações obtidas com cada tipo de técnica. Por meio desses métodos, é possível determinar a capacidade informativa dos locos marcadores e estimar o grau de proximidade entre as amostras em estudo.

No que se refere aos marcadores codominantes, a avaliação do poder informativo de cada loco é facilmente estimada pela diversidade gênica (anteriormente referida como “heterozigosidade média” e também conhecida como PIC - *Polymorphism Information Content*) e calculada com base na soma dos quadrados das frequências alélicas. Os termos heterozigosidade média e PIC são comumente usados como sinônimos na literatura. Entretanto, se maior rigor for necessário, é possível fazer a diferenciação entre ambas: (1) a heterozigosidade média representa a frequência de heterozigotos na população em estudo e, portanto, se aplica a indivíduos que sofreram fecundação cruzada e (2) o PIC pode ser calculado para indivíduos resultantes de um grande número de autofecundações, em que há poucos heterozigotos e em que a variação se revela na existência de homozigotos para diferentes alelos.

A utilização dos polimorfismos revelados por marcadores moleculares com o objetivo de estimar divergências entre indivíduos requer a escolha de coeficientes, geralmente baseados em diferentes pressupostos. Coeficientes de similaridade, como o de Jaccard (Jaccard 1908) desconsideram as combinações do tipo 0-0, quando nenhum dos dois genótipos comparados apresenta bandas. Segundo Dias (1998), esse coeficiente é muito útil para avaliação de indivíduos de uma mesma espécie, onde a probabilidade de serem encontradas concordâncias é maior. Estudos que envolvem marcadores dominantes têm aplicado com sucesso esse coeficiente de similaridade em várias espécies (Doldi *et al.* 1997; Lanza *et al.* 1997; Lima *et al.* 2002).

Medidas de distância genética combinam conceitos geométricos e genéticos e devem possuir como características a metricidade, o uso de frequências alélicas e a capacidade de refletir predomínio dos fatores genéticos sobre os ambientais. Tais medidas se distinguem das de dissimilaridade pelo fato dessas últimas não resguardarem as propriedades anteriormente mencionadas (Dias 1998) e de serem calculadas apenas como o complemento das medidas de similaridade. A distância de Rogers modificada (Wright 1978) é apropriada para o estudo de informações obtidas com marcadores codominantes como os microssatélites, uma vez que usa como base de cálculo frequências alélicas e não somente a informação de presença e ausência como no coeficiente de similaridade de Jaccard. Esse coeficiente de distância é definido como a média aritmética entre os valores obtidos para cada loco individualmente, possibilitando comparações diretas entre duas ou mais medidas desse tipo.

Uma nova metodologia foi recentemente proposta por Bernardo (1993) para compilar dados de marcadores moleculares. Denominada coeficiente de parentesco molecular, é uma medida de similaridade e a sua principal característica é diminuir o viés

que as medidas de similaridade possuem devido à semelhança em estado que resulta em uma combinação 1-1 entre dois indivíduos sendo comparados. A inserção de termos que prevêm essa condição no cálculo do valor de similaridade (baseado em genótipos indicados como não relacionados ao grupo) ocasiona a diminuição do viés e gera informações mais confiáveis sobre a relação entre os indivíduos sob estudo. Bernardo (1993) e Bernardo *et al.* (1996) usaram esse coeficiente para avaliar o já bem conhecido germoplasma norte-americano, mas nenhum estudo com o milho tropical foi apresentado até hoje.

Como facilmente se pode notar, a literatura traz enorme quantidade de análises sobre o milho de clima temperado, as quais praticamente saturam questões sobre a diversidade dessa classe de germoplasma. Todavia, o milho tropical, dispersor originário da variabilidade em todo o mundo, ainda não tem consolidada a classificação de suas reservas genéticas. Estudos que disponham da tecnologia de marcadores moleculares para avaliar de forma mais eficaz a verdadeira diversidade que o milho tropical são fundamentais para o manejo mais proveitoso dos recursos vegetais. O trabalho de Warburton *et al.* (2002) é o primeiro que visa a acessar a diversidade das importantes linhagens tropicais, porém ainda há diversos germoplasmas a serem explorados e organizados. O germoplasma brasileiro que faz do Brasil o terceiro maior produtor mundial de milho certamente detém significativos genótipos que ainda não são aproveitados em nossos programas de melhoramento. As melhores formas de análise genética para um germoplasma ainda não muito bem conhecido surgirão de estudos que objetivem conhecer a sua diversidade e podem, talvez, revelar importantes aspectos que o diferenciem do material temperado.

3) Objetivos

O trabalho que essa dissertação apresentará a seguir objetivou:

- (1) conhecer a diversidade genética de linhagens de milho tropical do Banco de germoplasma do IAC;
- (2) avaliar o uso de marcadores AFLP e SSR na amostragem da variabilidade genotípica do milho;
- (3) investigar a capacidade de alguns coeficientes estatísticos em expor e organizar a diversidade do milho tropical e
- (4) estudar a possibilidade de usar simultaneamente os dados desses dois tipos de marcadores.

4) Artigo

TROPICAL MAIZE GERMPLASM: WHAT CAN WE SAY ABOUT AFLP AND SSR IN ANALYZING GENETIC DIVERSITY?

Laborda P.R., Oliveira K.M., Garcia A.A.F., Zagatto-Paterniani M.E.A.G. and Souza A.P.

(Artigo enviado para a revista *Theoretical and Applied Genetics*)

TROPICAL MAIZE GERMPLASM: WHAT CAN WE SAY ABOUT AFLP AND SSR WHEN ANALYZING GENETIC DIVERSITY?

P. R. LABORDA · K. M. OLIVEIRA · A. A. F. GARCIA · M. E. A. G. ZAGATTO-PATERNIANI ·
A. P. DE SOUZA

P. R. Laborda · K. M. Oliveira · A. P. de Souza(✉)

Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), Cidade Universitária "Zeferino Vaz", CP 6010, CEP 13083-970, Campinas, SP, Brasil.

A. A. F. Garcia

Departamento de Genética, Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ), Universidade de São Paulo (USP), CP 83, CEP 13400-970, Piracicaba, SP, Brasil.

M. E. A. G. Zagatto-Paterniani

Instituto Agronômico de Campinas (IAC), Av. Theodureto de Almeida Camargo, 1500, CP 28, CEP 13020-902, Campinas, SP, Brasil.

E-mail: anete@unicamp.br

Phone: (55-19) 3788-1132

Fax: (55-19) 3788-1089

Abstract

Knowledge about genetic variability of agronomically important crops allows efficient manipulation of resources in order to achieve plant improvement. Temperate maize has been largely studied, but diversity in tropical maize has still not been well established. Brazilian maize germplasm represents a very important pool of genetic diversity due to many past introductions of exotic material. Nowadays, molecular markers are the principal way of analyzing genetic diversity and many statistical possibilities have arisen to compile data from molecular assays. This work used 569 bands of AFLP and 50 loci of microsatellites to access genetic diversity in tropical maize inbred lines, and revealed a great level of variability, with high rates of polymorphism. All three types of coefficients (Jaccard, MRD and molecular coefficient of coancestry) applied exhibited very low genetic similarities among the inbreds, and depending on the marker used, different clusters were obtained. Nevertheless, high correlations between matrices generated by Jaccard and by the molecular coefficient of coancestry with AFLP ($r = 0.88$), SSR ($r = 0.87$) and AFLP/SSR ($r = 0.89$) were seen. AFLP and SSR were poorly correlated with both similarity coefficients ($r = 0.43$ and $r = 0.48$). Because of SSR's ability of uncovering many aspects of genomic variation, it was considered more appropriate for diversity analysis. It was observed that the great amount of diversity found in tropical maize led to distinct results with different markers, and that using both AFLP and SSR together may not be the most efficient manner of accessing variability in highly diverse materials.

Key words *Tropical maize · Genetic diversity · SSR · AFLP · Molecular coefficient of coancestry*

Introduction

Maize genetic diversity has been the topic of various kinds of studies due to the extreme economic importance showed by this species. The improvement of maize is the principal aim of most breeders and different ways of enhancing the features of this crop have already been developed. A fundamental aspect that lies on the basis of improvement processes is the availability of genetic diversity. Having a huge genic pool makes the manipulation of different genotypes that may lead to high performance hybrids in relation to yield, resistance to diseases and many other characteristics possible.

During the beginning of the 1990s, some researches used RFLP (restriction fragment length polymorphism) to disclose diversity in temperate European maize (Boppenmaier et al. 1993) and US maize (Melchinger et al. 1991), and revealed the ability of this marker to assign lines to different heterotic groups. RFLPs, nevertheless, proved to be a rather time consuming methodology, and molecular technology improvements made available other kinds of markers, which can be analyzed by means of PCR (polymerase chain reaction). RAPDs (random amplified polymorphic DNA), AFLPs (amplified fragment length polymorphism) and SSRs (simple sequence repeat) were then adopted for use in genetic analyses providing more data in less time. Even though being a simple technique, it was observed that RAPDs may present some repeatability problems, and not always correlate with other markers in maize (Hahn et al. 1995; Pejic et al. 1998; Garcia et al. 2004) and other species (Doldi et al. 1997; Russell et al. 1997).

AFLPs and SSRs are very reproducible techniques and show different features that are very suitable for genetic analyses. AFLPs, which may be applied to any plant species without previous knowledge of DNA sequences, have already been proven to be the most

efficient marker, disclosing the major number of bands per single assay (Russell et al. 1997; Pejic et al. 1998; Garcia et al. 2004). SSRs are very abundant and dispersed throughout the genome, and they can uncover great amounts of polymorphism, as multiallelic loci are very common and codominant inheritance is displayed (Chin et al. 1996). As a consequence of these special characteristics, many studies were carried out using these two markers to access diversity in various species: sugarcane (Lima et al. 2002), wheat (Plaschke et al. 1995; Manifesto et al. 2001), cotton (Liu et al. 2000; Abdalla et al. 2001), soybean (Priolli et al. 2002) and sorghum (Smith et al. 2000; Ghebru et al. 2002).

Maize was also extensively studied and many works have already revealed how important temperate maize germplasm is. Lines from the U.S. Corn Belt were analyzed in a large number of investigations (Smith et al. 1997; Senior et al. 1998; Lu and Bernardo 2001; Gethi et al. 2002). These studies showed the moderate diversity of temperate maize and were able to recover the relationships among the inbred lines. Moreover, Enoki et al. (2002) worked with 51 Japanese lines and, by means of SSR analysis, compared them to genotypes introduced from Europe and the U.S..

Despite the well-analyzed temperate material, tropical inbreds lack information about diversity and very few studies involving a great number of lines are available (Warburton et al. 2002; Oliveira et al. 2004). Lanza et al. (1997) used RAPDs and Benchimol et al. (2000) used RFLPs to predict hybrid performance among 18 inbred lines and Barbosa et al. (2003), with the same lines, made a comparative study of genetic distances and single-cross performance using AFLPs and SSRs. The Agronomic Institute of Campinas (IAC) owns an important tropical maize genebank, having a great deal of lines obtained from germplasm introduced from the CIMMYT. Nevertheless, no previous work

has accessed its real variability. Knowing the genetic diversity of tropical maize means to make available more genic variation for breeding with the already known lines.

To study the data provided by molecular markers, several similarity/dissimilarity coefficients have been developed both for dominant and codominant markers. Jaccard's similarity (Jaccard 1908) is one of the most applied coefficients for dominant data (Doldi et al. 1997; Lanza et al. 1997; Lima et al. 2002; Barbosa et al. 2003) and it can also analyze information from codominant markers (Li et al. 2001). It is considered very useful in studies that aim to analyze genotypes from a single species, when more concordant bands are seen. In addition to Jaccard's similarity, a modification suggested by Wright (1978) for the Rogers' distance (Rogers 1972) is largely used for SSRs, considering allelic frequencies and not only the presence/absence of bands. Recently, Bernardo (1993) proposed the molecular coefficient of coancestry, whose principal characteristic is removing the bias in similarity values that are due to likeness in state, privileging similarities that reflect identity by descent. This coefficient may reveal accurate relationships, but it still has not been used for accessing variability in very diverse materials, as the tropical germplasm. The molecular coefficient of coancestry brings up a possibility of compiling data from different systems when some coefficients such as Jaccard or modified Rogers' distance do not show high correlations between different markers. Until the present work, no large study comparing the abilities of AFLPs and SSRs in disclosing variability had been reported, and no work evaluating the statistical methods available for diversity analyses had been presented.

The objectives of this study were to (1) access genetic diversity in tropical maize inbred lines, (2) compare the capacity of AFLPs and SSRs in displaying variability and (3) evaluate the power of Jaccard's similarity, modified Rogers' distance and the molecular coefficient of coancestry to expose diversity in tropical material.

Material and Methods

Plant Material

Eighty-five inbred lines were chosen from the Agronomic Institute of Campinas (IAC) Genebank (Table 1). These inbreds are currently used in the Maize Hybrid Program of the IAC, and among them, there are recent lines, which were obtained from populations introduced from the CIMMYT (named “L-“) as well as historical Brazilian lines that have been used for a long time in Brazil.

DNA isolation and AFLP/SSR procedures

Young leaves from 15 plants of each genotype were collected, lyophilized and grounded to powder. DNA extraction followed the CTAB method described by Hoisington et al. (1994).

AFLP

AFLP profiles analyses were performed as described by Vos et al. (1995), using the “AFLP Analysis Kit” (Life Technologies - GIBCO BRL). Genomic DNA (400 ng) was double-digested with *EcoRI* and *MseI*. Restriction fragments were then linked to adaptors and the ligation product was pre-amplified for 20 cycles (94°C for 30 s, 56°C for 1 min, 72°C for 1 min), using primers carrying one selective nucleotide. *EcoRI* primers carrying three selective nucleotides were end-labeled with $\gamma[^{33}\text{P}]\text{-ATP}$ (4,000 Ci/mmol), and mixed with unlabelled *MseI* primer, for hot selective amplification following the cycles: 94°C for 30 s, 65°C (-0.7°C/cycle) for 30 s and 72°C for 1 min during 12 cycles, until the optimal

annealing temperature of 56°C was reached. All amplifications were carried out in a PTC™-100 (MJ Research, Inc.). Nine primer combinations were used and are similar of those described by Vos et al. (1995). Six percent denaturing polyacrylamide gels were loaded with 3.5 µl of a mixture of reaction product and formamide dye (1:1) and submitted to electrophoresis (Sequi-Gen® GT- Nucleic Acid- Electrophoresis Cell/BIO RAD Apparatus of Electrophoresis) for 4 h at 75 W. Samples were visualized by autoradiography, and manually scored.

SSR

To select the best SSRs, three sets of tests were performed: the first among 3 lines, the second among 8 lines and the third among 28 lines. A total of 215 microsatellite were tested, and classifications according to amplification quality and genotyping difficulty were made. From the 215 microsatellites tested, 50 were chosen for the diversity analysis. The primer sequences were obtained in the Maize Data Bank site (<http://www.agron.missouri.edu>). The amplification protocol included 50 ng of DNA, 1X reaction buffer (20 mM Tris-HCl, 50 mM KCl pH 8.4), 2 mM MgCl₂, 100 µM of each dNTP, 0.5 U of *Taq* DNA polymerase (GIBCO BRL) and 0.2 µM of each primer. Amplifications were carried out using a PTC™-100 (MJ Research, Inc.) and the “touchdown” program described by Senior et al. (1998). Samples were electrophoresed in either 0.5X TBE (Sambrook et al. 1989) (1:1) 4% agarose/Metaphor gels (Ultrapure agarose, GIBCO BRL / Metaphor agarose, FCM BioProducts) or 1X TBE 6% polyacrylamide gels, depending on the genotyping difficulty of each microsatellite. Horizontal electrophoreses were proceeded at 170 V for 1.5 h in HORIZON 20:25 gel

system (GIBCO BRL) and products were visualized by ethidium bromide (0.5 µg/ml). Vertical electrophoreses were proceeded at 90 W for 2 h using a Model S2001 Sequencing Gel Electrophoresis Apparatus (Life Technologies - GIBCO BRL) and samples were detected by silver staining according to Creste et al. (2001).

Data Analysis

AFLP and SSR gels were scored for the presence/absence of bands, generating two binary matrices. The SSR matrix was also converted to a table of allelic frequencies. Polymorphism information content (PIC) (Smith et al. 1997), also named genic diversity or expected heterozigosity, was calculated for SSRs according to the following formula:

$$[1] \quad \text{PIC} = 1 - \sum_i^n f_i^2 \quad i = 1 \dots n,$$

where f_i is the frequency of the i^{th} allele. PIC values show how powerful a locus can be to discriminate samples, considering not only the number of alleles but also their frequencies.

Preliminary diversity analyses among all the lines were made based on Jaccard's similarity (Jaccard 1908) and on modified Rogers' distance (MRD) (Wright 1978). Jaccard's similarity was used to analyze AFLP, SSR and AFLP and SSR together (all based on binary matrices), owing to study the possibility of using the data of different marker types simultaneously. MRD was applied only to SSR, since this coefficient requires allelic frequencies. In order to compare the ability of different ways in "quantifying" genetic diversity in tropical maize, the molecular coefficient of coancestry (Bernardo 1993) was also used for AFLP, SSR, and both together. This coefficient, named f_{AB}^M , represents the probability of inbreds A and B being identical by descent (*ibd*) and it is a measure of

similarity. Inbreds that are *ibd* have the same allele (at a given locus) inherited from a common ancestral allele. f_{AB}^M was obtained by the formula:

$$[2] \quad f_{AB}^M = \frac{S_{AB} - \frac{1}{2}(\delta_A + \delta_B)}{1 - \frac{1}{2}(\delta_A + \delta_B)},$$

where S_{AB} is the similarity value obtained by Jaccard's coefficient, δ_A is the average similarity between inbred A and unrelated inbreds, and δ_B is the average similarity between inbred B and unrelated inbreds. For this work, three unrelated inbreds (PM-308, PM-624 and PM-2837) were chosen to calculate δ_A and δ_B , which show the proportion of alleles that are alike in state (*ais*), not *ibd*. As recommended by Bernardo (1993), negative values were set to zero.

For obtaining the similarity matrices using Jaccard's coefficient, NTSYSpc v. 2.1 (Rohlf 1997) was applied. For computing the distance matrix, TFGPA v. 1.3 (Miller 1997) was used. The UPGMA method was applied to all of the clustering procedures and the respective cophenetic values were calculated to each of the dendrograms.

Correlations between AFLP and SSR matrices were obtained by means of Mantel's test (Mantel 1967). Bootstrap procedure (Tivang et al. 1994) with 1000 units of re-sampling was carried out for SSR data using medians instead of means, as recommended by Garcia et al. (2004). Modified Rogers' distance was used and each locus was considered a unit of re-sampling. Bootstrap procedures and results for AFLP are described in Oliveira et al. (2004).

Results

Polymorphism and bootstrap validation

Twenty AFLP primer-enzyme combinations were previously tested and 9 combinations were selected. AFLP assays generated 638 bands, from which 569 were polymorphic. Although using only 9 combinations out of the 64 possible, the AFLP profiles could distinguish undoubtedly each inbred line. The mean number of bands detected per combination was 63, ranging from 42 to 98, and the bootstrap procedures guaranteed the good quality of the data (mean CV = 5.43%) (Oliveira et al. 2004).

Two hundred fifteen microsatellite primers were pre-screened in order to select the best amplification and genotyping conditions. The set of tests that included 8 inbreds was the most efficient, being able to reveal the polymorphism among the inbreds and not wasting much effort and time. Out of the 215 SSR primers pairs, 109 were tested using 8 lines, and 35 primer pairs were chosen. The other 15 were selected based on the test set using 28 lines (58 pairs tested). None of the primers tested in 8 or 28 lines was monomorphic. The 50 loci studied were well distributed in the 10 maize chromosomes, ensuring a good sampling of the genome. A total of 262 bands were obtained and a mean of 5.2 alleles per locus was reached, ranging from 2 to 14. Di- and tri-nucleotides motifs were the most abundant. They represented 36% and 26%, respectively. Tetra-, penta- and hexa-nucleotides represented the remaining 38%. AG repeats corresponded to 72% of dinucleotides, which agrees with the results obtained by Chin et al. (1996) and Taramino and Tingey (1996). PIC values for SSR data ranged from 0.24 to 0.90 with a mean of 0.61. Although this study used inbred lines that have been maintained for many years in auto-fertilizing conditions, a high frequency of heterozygotes was found. Five loci showed more

than 20 heterozygotes out of the 85 genotypes and a mean of 8.8 was observed for the 50 loci. The results concerning number of alleles, class of nucleotide motif, PIC values and frequency of heterozygotes are displayed in Table 2.

Bootstrap analysis showed that the number of loci used ($n = 50$) was appropriate to access diversity reliably among the 85 lines. If a mean coefficient of variation of 10% was used, as recommended in some works (Halldén et al. 1994; Tivang et al. 1994), 20 microsatellite loci would be necessary to analyze the 85 genotypes with accuracy. The 50 loci in this work brought the mean coefficient of variation to approximately 6%, which made the set of SSR data very consistent.

Clustering, correlations and diversity levels

Clusters analyses were all carried out using the UPGMA method and showed some differences according to the marker/coefficient applied, which resulted in some low correlation values. In addition, distinct levels of divergence were reached also considering the marker used. Even though these disagreements were detected, all dendrograms were together in not showing clear clustering of lines. Well-defined divergent groups could not be seen in any of the ways used to visualize clusters, corroborating what Warburton et al. (2002) had already reported for other tropical inbreds. CIMMYT-derived inbreds tended to form groups, but AFLP data grouped more lines together than did SSR data. As CIMMYT-derived lines, “AL-“ lines also constituted a cluster when AFLP/Jaccard was applied and were dispersed in SSR/MRD tree (data not shown). Not even inbreds that were considered identical with one marker behaved the same way with the other. AFLP analyses showed that L-158 and L-162 had a high degree of similarity, which was not observed within the

SSR analyses. On the other hand, microsatellites showed that AL-491 and AL-516 were genetically similar, what could not be seen in the AFLP dendrograms. Taking into account the lines that were the most divergent, AFLP and SSR brought out one of the few congruencies found among these markers. Whatever the marker type or the sort of coefficient, a group of specific lines was always indicated to be the most divergent (data not shown). From this group, PM-308, PM-624 and PM-2837 were nominated the unrelated lines used in the calculation of molecular coefficients of coancestry.

In order to compare all the similarity/distance matrices with each other, correlations were established to visualize whether the different markers and coefficients showed the same results (Table 3). As expected for homozygotes lines, the comparison of results obtained by Jaccard and MRD for SSR data only showed a high correlation ($r = -0.95$, a negative value as a consequence of a correlation made between a similarity and a distance matrix), indicating that these coefficients grouped lines in a very similar way. Nevertheless, as mentioned above, AFLP and SSR disagreed in some groupings and a very low correlation ($r = 0.43$) was obtained for both markers using Jaccard, what is probably due to the different marker characteristics. The dendrograms constructed on the basis of molecular coefficient of coancestry for AFLP and microsatellites are shown in Figures 1 and 2, respectively, and it can be noticed that this coefficient also showed discrepancies when different marker systems were used. The correlation for both markers considering this coefficient was slightly higher ($r = 0.48$), but still demonstrating lack of agreement. AFLP-Jaccard / AFLP- f_{AB}^M and SSR-Jaccard / SSR- f_{AB}^M correlations ($r = 0.88$ and $r = 0.87$, respectively) revealed that the coefficient is, probably, not what leads the clusters to be different, but the information provided by each sort of marker. Furthermore, an interesting

characteristic was noticed when matrices containing both data were compared to AFLP and SSR separately. Correlations among AFLP/SSR-Jaccard and AFLP-Jaccard, and AFLP/SSR- f_{AB}^M and AFLP- f_{AB}^M were extremely high ($r = 0.96$ and $r = 0.88$, respectively), but correlations for AFLP/SSR-Jaccard and SSR-Jaccard, and AFLP/SSR- f_{AB}^M and SSR- f_{AB}^M were much lower ($r = 0.64$ in both cases). This indicates that clusters made with AFLP and SSR data together were very similar only to those made with AFLP data. As a matter of fact, we could observe this among the seven dendrograms constructed: groupings with AFLP were almost the same as those obtained by means of both markers (data not shown).

Table 3 also lists mean similarity/distance values obtained in the seven ways used to compare the clustering of the 85 inbred lines. The analyses made on the basis of Jaccard's coefficient presented different means, revealing different levels of diversity. It was observed that microsatellites (mean = 0.26) exposed more diversity than did AFLP (mean = 0.51). The range of variation was also greater for SSR (0.08-1.00) when contrasted with AFLP (0.35-0.87). Even when being analyzed as a dominant marker (using a binary matrix and not allelic frequencies), the microsatellites were more effective in showing divergence as a result of higher degree of polymorphism and multiallelism. The molecular coefficient of coancestry proposed by Bernardo (1993) revealed very low mean similarity values, owing to the great number of negative similarities that this coefficient provides. One reason for this is that means calculation for this coefficient was made considering negative values as zero. The matrices obtained by this coefficient showed very proximate similarity means for different data, but the clusters displayed were not the same for AFLP and SSR, just as mentioned above. In Figure 3, it is possible to graphically visualize the frequencies of each similarity interval. The great majority of AFLP similarities are restricted to values that go

from 0.4 to 0.6 while, for SSR, the similarities go from 0.1 to 0.4 (Jaccard). For similarities obtained with the molecular coefficient of coancestry, very similar results were reached for AFLP and SSR, all values ranging from 0 to 0.3. The graph exhibits the high genetic divergence that the tropical maize accessed here has, since very low similarity values were seen.

Discussion

The results concerning polymorphism revealed interesting aspects of genomic divergence used to analyze variability. AFLP is undoubtedly a very powerful technique to display polymorphic bands per single assay, despite having the lowest level of polymorphism (Pejic et al. 1998; Russell et al. 1997). In our study, up to 98 polymorphic bands could be seen in a single gel and an overall rate of 89% of polymorphism was reached. If we consider each AFLP band a locus, it is definitely an efficient manner of exposing genomic variation. In addition, contrasting with the work of Russell et al. (1997), who found a very low level of polymorphism for AFLP in barley (48%), it can be concluded that the tropical inbreds here analyzed are, as a matter of fact, very divergent.

In contrast to the very efficient AFLP assays, SSR assays are able to exhibit only one locus at a time. Multiplexed gels are an option of optimizing the SSR technique (Mitchell et al. 1997), but not always can they be applied, due to high costs and machinery requirements. Although it is possible to analyze one locus per gel, the codominant behavior of this marker makes it extremely capable of showing polymorphism among few genotypes. The capacity of uncovering alleles of a single locus, and not only showing if a specific genic variation is present or not, makes microsatellites more appropriate for

diversity analyses. SSR multiallelism made successful tests undertaken with only 8 inbreds possible. Thirty-five loci were chosen out of the 109 tests made with 8 lines, but more loci would have been genotyped if the amplification conditions had been better; polymorphism was not the striking problem. The great number of available maize SSR loci made the screening much easier. Nonetheless, the fact that these loci have been developed for temperate maize must have caused some amplifications to fail when tropical maize DNA was used.

Another problem faced during the SSR genotyping procedures was the little difference between fragment lengths. After having selected the best microsatellites in an agarose/metaphor system, many of them could not be solved in this sort of gel. The polymorphism was evident, but genotyping all the 85 samples was very difficult. This justifies the change of technique from agarose to polyacrylamide gels that occurred during SSR analyses. Moreover, the microsatellites with best amplification and genotyping conditions had few alleles, which probably biased the mean number of alleles (mean = 5.2) and the PIC values (mean = 0.61). During the microsatellite tests, many loci with high number of alleles were found, but many could not be applied to the analyses due to amplification problems. Some works with temperate maize, which does not have a genetic base as diverse as the tropical maize, revealed mean values for these parameters that are similar (Smith et al. 1997; Senior et al. 1998) or even higher (Pejic et al. 1998) than those obtained in this study. Not having found a larger number of alleles in tropical maize, when compared to temperate maize, may indicate likeness in state for some alleles (and not necessarily identity by descent), which may have evolved from different genetic sources. This condition would make the parameter “number of alleles” not to be the most appropriate way to compare diversity among germplasm pools since it would be hiding the

original variability from which the actual alleles have arisen. Comparing our results with the study of Warburton et al. (2002) with tropical lines (mean number of alleles = 4.9), a slightly higher number of alleles was reached. PIC values, which reveal not only the presence of many alleles but also their frequencies, may also have been lowered by the presence of some specific alleles. The presence of alleles with very low frequency may be due to high rates of mutation of the microsatellite (Henderson and Petes 1992) or to the introduction of exotic germplasm (Senior et al. 1998). Matsuoka et al. (2002) found that the percentage of specific alleles in tropical inbreds was higher than that found in temperate maize, and proposed that these alleles are interesting guides to identify inbreds. These authors also revealed that the size variation in microsatellites is not always due to addition or deletion of motifs. Most researches assume that length ranging is only caused by reduction or expansion of the SSR alleles, but, for diversity studies, all kinds of variation are important to identify genotypes since they expose a sample of genomic divergence.

This great mutation capacity of the SSRs may have led to the differences observed between markers data. According to Li et al. (2001), especially for inbred lines, the inconsistency among different molecular markers data results from the fact that they evaluate different components of DNA variation, which may evolve in different ways. AFLPs, which do not mutate as fast as SSRs, showed higher similarities values and this may have resulted in more consistent groups, as observed for the CIMMYT-derived lines. SSRs maximize the differences among inbreds because they can disclose the diversity in a very strong way, which allows them to be really trustworthy in analyzing genetic variability. This was also seen in very closely related wheat lines, when Plaschke et al. (1995) observed very low similarities values and could distinguish 40 genotypes using only 23 microsatellite loci. On the other hand, this great amount of SSR polymorphism may not

be appropriate for establishing pedigree among cultivars, as suggested by Russell et al. (1997). Markers that privilege the similarity among genotypes and that do not vary much genetically are more suitable for pedigree analyses, as shown by Lima et al. (2002) in a work involving sugar cane pedigree studies with AFLP.

As a consequence of these different characteristics displayed by AFLP and SSR, low correlation values were obtained. Manifesto et al. (2001) also described a low correlation between AFLP and SSR for wheat ($r = 0.27$), but they suggested that this value was influenced by the few loci used in their analyses ($n = 10$). This does not justify the low correlations obtained in our study, once a great number of markers was applied. Previous works (Pejic et al. 1998; Barbosa et al. 2003) demonstrated moderate high correlation among AFLP and SSR, but it should be noticed that the materials surveyed did not have the great amount of variability shown by this work, which probably influenced the correlations. In Barbosa et al. (2003), tropical maize was analyzed; however, it must be considered that the maize inbreds that were studied by this group, even being tropical, did not show as much diversity when compared with the data here revealed. Probably, disagreements in analyzing inbreds by one marker type or another reflect the great amount of diversity within the material and the different levels of variation uncovered by each sort of marker.

An optimum number of SSR loci that would be appropriate for accessing genetic diversity with accuracy has been discussed in many works. The soybean studies of Doldi et al. (1997) suggested that the use of few markers from both RAPD and SSR would generate reliable results, even considering that the breeding material used was shown to have a narrow genetic base and, consequently, a low level of polymorphism. Similarly, Plaschke et al. (1995) and Manifesto et al. (2001), as mentioned above, were also able to distinguish

lines using only few microsatellites. This last group of authors, however, used loci with previous known high PIC values. SSR loci chosen at random would probably not bring as much information as the ones intentionally selected. Considering temperate maize, Pejic et al. (1998) suggested that a set of 20-30 would be enough to analyze the variability in maize with confidence. In this study, 50 loci were employed in tropical maize and the bootstrap analysis proved that this number was above what was necessary (20 loci would be enough with a 10% variation coefficient). In very diverse materials, it is expected that fewer markers would be able to distinguish genotypes, however, so as to provide more credible data, many parts of the genome must be surveyed. Moreover, clusters based on a small number of markers have limited value and may exhibit disagreement with pedigree data when this is available. Warburton et al. (2002) suggest that a set of approximately 50 loci would be appropriate for analysis in high diverse materials.

Despite the importance of the number of markers that must be applied to get reliable results, the choice of statistical coefficients is a rather fundamental step in studying genetic diversity. Our work presented three types of coefficients, which displayed different information. Modified Rogers' distance, one of the most appropriate to analyze codominant markers, such as SSR, was not used for AFLP data because it requires allelic frequencies. Nonetheless, this parameter of genetic distances showed very similar results to those obtained by Jaccard and f_{AB}^M for SSR, which could be compared to AFLP by means of these last two coefficients. Using Jaccard for a codominant marker did not influence the results when this marker was analyzed separately. The correlations for Jaccard and the molecular coefficient of coancestry considering AFLP, SSR and AFLP/SSR were high, indicating that they revealed proximate results. However, f_{AB}^M is a coefficient that reveals

relationships among genotypes and does not focus on the diversity among them. According to Bernardo (1993), similarity values obtained by molecular markers are overestimates of true relationships among genotypes, the amount of bias depending on how divergent the material being studied is. The greater the distance among inbreds, the greater the bias created. “Molecular similarities” may provide wrong relationships among inbreds because of their inability in discriminating lines that are only *ais* from the ones that are *ibd*. Therefore, it can be concluded that similarities, such as the ones obtained by Jaccard, simply display diversity, no matter if they come from real relations or from likeness in state. The molecular coefficient of coancestry circumvents this problem and gives more reliable information about relationships, but may hide a part of the diversity revealed by the molecular markers. Considering mean similarity values, those obtained by f_{AB}^M were much lower than were those shown by Jaccard. This was a consequence of the great number of negative values. Bernardo (1993) foreshadowed that a weakness in his model would cause their existence. Furthermore, the great variability of the lines may have contributed to an increase in the negative estimates, which reveal no shared relation by the pair of lines. In relation to the discrepancies among matrices from AFLP/SSR and AFLP, and AFLP/SSR and SSR, some observations must also be cited. The use of Jaccard’s coefficient, which supposes that each band in the similarity matrix is a locus, may justify the high correlations seen for AFLP/SSR and AFLP, and the low correlations for AFLP/SSR and SSR. To consider a band of SSR data as a locus, and not as an allele of a locus, decreases the value of microsatellites as a codominant marker. Analyzing it separately does not seem to imply loss of information, but when together with AFLP, the same weight of an AFLP band is attributed to a SSR allele. Owing to this, the great number of AFLP bands hide the

relationships among SSR alleles and push the results towards AFLP as it was being analyzed alone.

In conclusion, it was observed that tropical maize is a fantastic source of variability, resulting in an outstanding potential for breeding programs. Moreover, we suggest that, when very diverse material is under analysis, appropriate choice of molecular marker should be taken. All types of markers are able to reveal diversity, but as noted in this work, some do not access high levels of genomic variation, which may hide a part of divergence exhibited by the genotypes. The use of Jaccard, for SSR data only, has proven to be as efficient as MRD for our lines, but its use in matrices joining AFLP and SSR data prevented this last to show all the genetic variability that it can reveal. The molecular coefficient of coancestry decreased the intrinsic bias that Jaccard may have, and behave as the latter when both markers were together. However, its use does not seem to have been able to manage all the genetic diversity in our tropical lines. Known coefficients, apparently, do not accurately quantify highly diverse materials when data from more than one sort of marker is available. The development of new coefficients that would be able to analyze AFLP and SSR together, preserving the dominant/codominant features of each one, and giving them appropriate weights, would be ideal, and would also enhance the possibilities of studying species with a wide range of variability.

Acknowledgments

The authors wish to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the financial

support during the development of this research. A.P.S. received a research fellowship (00/11517-4) from CNPq.

References

Abdalla AM, Reddy OUK, El-Zik KM, Pepper AE (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. *Theor Appl Genet* 102:222-229

Barbosa AMM, Geraldi IO, Benchimol LL, Garcia AAF, Souza Jr CL, Souza AP (2003) Relationship of intra- and interpopulational tropical maize single cross hybrid performance and genetic distances computed from AFLP and SSR markers. *Euphytica* 130:87-99

Benchimol LL, Souza Jr CL, Garcia AAF, Kono PMS, Mangolin CA, Barbosa AAM, Coelho ASG, Souza AP (2000) Genetic diversity in tropical maize inbred lines: heterotic group assignment and hybrid performance determined by RFLP markers. *Plant Breeding* 119:491-496

Bernardo R (1993) Estimation of coefficient of coancestry using molecular markers in maize. *Theor Appl Genet* 85:1055-1062

Boppenmaier J, Melchinger AE, Seitz G, Geiger HH, Herrmann RG (1993) Genetic diversity for RFLPs in European maize inbreds. III. Performance of crosses within versus between heterotic groups for grain traits. *Plant Breeding* 111:217-226

Chin ECL, Senior ML, Shu H, Smith JSC (1996) Maize simple repetitive DNA sequences: abundance and allele variation. *Genome* 39:866-873

Creste S, Tulmann Neto A, Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol Biol Rep* 19:299-306

Doldi M-L, Vollmann J, Lelley T (1997) Genetic diversity in soybean as determined by RAPD and microsatellite analysis. *Plant Breeding* 116:331-335

Enoki H, Sato H, Koinuma K (2002) SSR analysis of genetic diversity among maize inbred lines adapted to cold regions of Japan. *Theor Appl Genet* 104:1270-1277

Garcia AAF, Benchimol LL, Barbosa AAM, Geraldi IO, Souza Jr CL, de Souza AP (2004) Comparing of RAPDs, RFLPs, AFLPs and SSRs for computation of genetic divergence in tropical maize inbred lines. *Genet Mol Biol (in press)*

Gethi JG, Labate JA, Lamkey KR, Smith ME, Kresovich S (2002) SSR variation in important US maize inbred lines. *Crop Sci* 42:951-957

Ghebru B, Schmidt RJ, Bennetzen JL (2002) Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. *Theor Appl Genet* 105:229-236

Hahn V, Blankenhorn K, Schwall M, Melchinger AE (1995) Relationships among early European maize inbreds. III. Genetic diversity revealed with RAPD markers and comparison with RFLP and pedigree data. *Maydica* 40:299-310

Halldén C, Nilsson N-O, Rading IM, Säll T (1994) Evaluation of RFLP and RAPD markers in a comparison of *Brassica napus* breeding lines. *Theor Appl Genet* 88:123-128

Henderson ST, Petes TD (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. *Mol Cell Biol* 12:2749-2757

Hoisington D, Khairallah M, Gonzalez-de-Leon D (1994) Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. 2nd edn. CIMMYT, Mexico DF, pp 2-3

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44:223-270

Lanza LLB, de Souza Jr CL, Ottoboni LMM, Vieira MLC, de Souza AP (1997) Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. Theor Appl Genet 94:1023-1030

Li C, Fatokun CA, Ubi B, Singh BB, Scoles GJ (2001) Determining genetic similarities and relationships among cowpea breeding lines and cultivars by microsatellite markers. Crop Sci 41:189-197

Lima MLA, Garcia AAF, Oliveira KM, Matsuoka S, Arizono H, de Souza Jr CL, de Souza AP (2002) Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). Theor Appl Genet 104:30-38

Liu S, Cantrell RG, McCarty Jr JC, Stewart JMcD (2000) Simple sequence repeat-based assessment of genetic diversity in cotton race stock accessions. Crop Sci 40:1459-1469

Lu H, Bernardo R (2001) Molecular marker diversity among current and historical maize inbreds. Theor Appl Genet 103:613-617

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209-220

Manifesto MM, Schlatter AR, Hopp HE, Suárez EY, Dubcovsky J (2001) Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. Crop Sci 41:682-690

- Matsuoka Y, Mitchell SE, Kresovich S, Goodman M, Doebley J (2002) Microsatellites in *Zea* – variability, patterns of mutations, and use for evolutionary studies. *Theor Appl Genet* 104:436-450
- Melchinger AE, Messmer MM, Lee M, Woodman WL, Lamkey KR (1991) Diversity and relationships among US maize inbreds revealed by Restriction Fragment Length Polymorphisms. *Crop Sci* 31:669-678
- Miller MP (1997) TFPGA - Tools For Population Genetic Analyses, v. 1.3 Northern Arizona University
- Mitchell SE, Kresovich S, Jester CA, Hernandez CJ, Szewc-McFadden AK (1997) Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. *Crop Sci* 37:617-624
- Oliveira KM, Laborda PR, Garcia AAF, Zagatto-Paterniani MEAG, Souza AP (2004) Evaluating genetic relationships between tropical maize inbred lines by means of AFLP profiling. *Hereditas (in press)*
- Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs and AFLPs. *Theor Appl Genet* 97:1248-1255
- Plaschke J, Ganai MW, Röder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. *Theor Appl Genet* 91:1001-1007
- Priolli RHG, Mendes-Junior CT, Arantes NE, Contel EPB (2002) Characterization of Brazilian soybean cultivars using microsatellite markers. *Genetics and Molecular Biology* 25:185-193
- Rogers JS (1972) IV. Measures of genetic similarity and genetic distance. *Studies in genetics VII. Univ. of Texas Publ.* 7213:145-153

Rohlf FJ (1997) NTSYSpc - Numerical Taxonomy and Multivariate Analysis System, v. 2.1 Exeter Publications, New York

Russell JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. *Theor Appl Genet* 95:714-722

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory Press, New York, pp. B.23

Senior ML, Murphy JP, Goodman MM, Stuber CW (1998) Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. *Crop Sci* 38:1088-1098

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegler J (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays L.*): comparisons with data from RFLPs and pedigree. *Theor Appl Genet* 95:163-173

Smith JSC, Kresovich S, Hopkins MS, Mitchell SE, Dean RE, Woodman WL, Lee M, Porter K (2000) Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. *Crop Sci* 40:226-232

Taramino G, Tingey S (1996) Simple sequence repeats for germplasm analysis and mapping in maize. *Genome* 39:277-287

Tivang JG, Nienhuis J, Smith OS (1994) Estimation of sampling variance of molecular marker data using the bootstrap procedure. *Theor Appl Genet* 89:259-264

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Fritjers A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23:4407-4414

Warburton ML, Xianchun X, Crossa J, Franco J, Melchinger AE, Frisch M, Bohn M, Hoisington D (2002) Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. *Crop Sci* 42:1832-1840

Wright S (1978) *Evolution and the Genetics of Populations*, vol. 4. Variability Within and Among Natural Populations. University of Chicago Press, Chicago

Table 1 Maize inbred lines maintained in the Agronomic Institute of Campinas, the material they were selected from and their origins

Inbred line	Selected from:	Origin	Inbred line	Selected from:	Origin
AL - 124	CATETO	IAC - Brazil	L - 111	Pop. 26	CIMMYT - Mexico
AL - 218	CATETO	IAC - Brazil	L - 112	Pop. 26	CIMMYT - Mexico
AL - 491	CATETO	IAC - Brazil	L - 114	Pop. 26	CIMMYT - Mexico
AL - 516	CATETO	IAC - Brazil	L - 116	Pop. 27	CIMMYT - Mexico
AL - 535	CATETO	IAC - Brazil	L - 117	Pop. 24	CIMMYT - Mexico
AL - 604	CATETO	IAC - Brazil	L - 118	Pop. 27	CIMMYT - Mexico
AL - 614	CATETO	IAC - Brazil	L - 120	Pop. 28	CIMMYT - Mexico
AL - 673	CATETO	IAC - Brazil	L - 121	Pop. 27	CIMMYT - Mexico
AL - 745	CATETO	IAC - Brazil	L - 123	Pop. 27	CIMMYT - Mexico
AL - 758	CATETO	IAC - Brazil	L - 126	Pop. 27	CIMMYT - Mexico
IA - 278	CATETO	IAC - Brazil	L - 128	Pop. 24	CIMMYT - Mexico
IA - 606	CATETO	IAC - Brazil	L - 130	ACROSS 7543	CIMMYT - Mexico
IA - 2938	CATETO	IAC - Brazil	L - 131	ACROSS 7543	CIMMYT - Mexico
IA - 3040	CATETO	IAC - Brazil	L - 132	Pool 23	CIMMYT - Mexico
IAC - B	Pop. TX 303	IAC - Brazil	L - 134	Pop. 24	CIMMYT - Mexico
IP - 48	CATETO	IAC - Brazil	L - 137	Pop. 36	CIMMYT - Mexico
IP - 301	CATETO	IAC - Brazil	L - 155	Pop. 25	CIMMYT - Mexico
IP - 330	CATETO	IAC - Brazil	L - 156	Pop. 36	CIMMYT - Mexico
IP - 365	CATETO	IAC - Brazil	L - 157	Pop. 27	CIMMYT - Mexico
IP - 398	CATETO	IAC - Brazil	L - 158	Pop. 27	CIMMYT - Mexico
IP - 661	CATETO	IAC - Brazil	L - 160	Pop. 28	CIMMYT - Mexico
IP - 701	TUXPEÑO	IAC - Brazil	L - 161	Pop. 26	CIMMYT - Mexico
IP - 3644	CATETO	IAC - Brazil	L - 162	Pop. 26	CIMMYT - Mexico
IP - 3668	CATETO	IAC - Brazil	L - 163	Pop. 26	CIMMYT - Mexico
IP - 3854	CATETO	IAC - Brazil	L - 164	Pop. 27	CIMMYT - Mexico
IP - 3855	CATETO	IAC - Brazil	L - 165	Pop. 27	CIMMYT - Mexico
IP - 3999	CATETO	IAC - Brazil	L - 166	Pop. 28	CIMMYT - Mexico
IP - 4022	CATETO	IAC - Brazil	L - 167	Pop. 36	CIMMYT - Mexico
L - 1	MJ 268	CIMMYT - Mexico	L - 168	Pop. 24	CIMMYT - Mexico
L - 2	MJ 274	CIMMYT - Mexico	L - 169	Pop. 26	CIMMYT - Mexico
L - 3	Pop. 24	CIMMYT - Mexico	L - 170	Pop. 27	CIMMYT - Mexico
L - 4	Pop. 24	CIMMYT - Mexico	L - 171	Pop. 28	CIMMYT - Mexico
L - 5	Pop. 26	CIMMYT - Mexico	L - 172	Pop. 28	CIMMYT - Mexico
L - 6	Pop. 26	CIMMYT - Mexico	PM - 219	TUXPEÑO	IAC - Brazil
L - 8	Pop. 28	CIMMYT - Mexico	PM - 308	TUXPEÑO	IAC - Brazil
L - 9	Pop. 36	CIMMYT - Mexico	PM - 518	TUXPEÑO	IAC - Brazil
L - 10	Pop. 36	CIMMYT - Mexico	PM - 624	TUXPEÑO	IAC - Brazil
L - 11	Pop. 27	CIMMYT - Mexico	PM - 888	TUXPEÑO	IAC - Brazil
L - 13	Pop. 26	CIMMYT - Mexico	PM - 2837	TUXPEÑO	IAC - Brazil
L - 100	Pool 27	CIMMYT - Mexico	SLP - 103	TUXPEÑO	IAC - Brazil
L - 101	Pool 27	CIMMYT - Mexico	SLP - 365	TUXPEÑO	IAC - Brazil
L - 105	Pop. 26	CIMMYT - Mexico	VER - 266	TUXPEÑO	IAC - Brazil
L - 110	Pop. 24	CIMMYT - Mexico			

Table 2 Microsatellites markers used in the diversity analyses and their genomic location, class of repeat, number of alleles, PIC values and number of heterozygotes

Locus	Genomic location	Repeat	N° alleles	PIC	N° heterozygotes
bnlg 149	1.00	?	7	0.75	1
bnlg 1484	1.03	di	6	0.52	2
umc 1397	1.03	penta	4	0.55	12
umc 1297	1.05	di	6	0.63	7
umc 1395	1.05	di	3	0.53	10
umc 1122	1.06	tri	4	0.51	9
umc 1774	1.10	di	3	0.60	3
umc 1630	1.11	penta	5	0.45	11
umc 1422	2.02	tri	4	0.64	1
bnlg 1621b	2.03	di	10	0.87	16
phi 083	2.04	tetra	4	0.66	12
umc 2019	2.07	tri	3	0.49	11
bnlg 2077	2.07-2.08	di	13	0.90	7
umc 1230	2.09	tri	6	0.79	8
umc 1252	2.09	tri	2	0.31	32
nc 030	3.04	di	3	0.56	2
bnlg 1951	3.06	di	6	0.75	6
bnlg 2241	3.06	di	6	0.68	4
phi 046	3.08	tetra	2	0.49	0
umc 1639	3.09	penta	5	0.76	23
phi 072	4.00-4.01	tetra	7	0.61	14
nc 135	4.01	?	4	0.56	1
umc 1943	4.02	?	4	0.32	6
phi 021	4.03	di	5	0.56	2
umc 1650	4.09	tri	2	0.42	12
umc 1325	5.00	di	6	0.72	4
umc 1416	5.00	tri	2	0.50	10
umc 1221	5.04	di	9	0.80	21
umc 1524	5.06	hexa	5	0.67	9
umc 1646	5.07	penta	5	0.36	8
umc 1792	5.08	tri	6	0.79	9
bnlg 1600	6.00	di	5	0.77	0
umc 1857	6.04	tri	6	0.80	16
umc 1653	6.07	tetra	10	0.84	5
umc 1426	7.00	penta	3	0.54	14
phi 057	7.01	tri	3	0.49	5
phi 112	7.01	di	3	0.24	0
bnlg 1666	7.04	di	12	0.88	4
umc 1161	8.06	hexa	5	0.72	6
umc 1069	8.08	penta	8	0.72	10
umc 1638	8.09	hexa	6	0.65	4
bnlg 1724	9.01	di	6	0.48	22
phi 028	9.01	tri	4	0.58	1
phi 022	9.03	tetra	3	0.52	21
umc 1357	9.05	tri	4	0.65	19
umc 1733	9.06	tetra	4	0.52	3
umc 1804	9.07	di	14	0.87	13
phi 059	10.02	tri	2	0.50	16
bnlg 2336	10.04	di	4	0.43	2
umc 1640	10.07	penta	3	0.55	5

Table 3 Correlations among the seven ways used to access diversity by means of Mantel's test and comparison of mean similarity/distance values according to the type of coefficient used to generate the similarity/distance matrix

	AFLP- Jaccard	SSR- Jaccard	AFLP/ SSR- Jaccard	SSR- MRD*	AFLP- f_{AB}^M	SSR- f_{AB}^M	AFLP/ SSR- f_{AB}^M	Mean	Range
AFLP- Jaccard	1.00							0.506	0.345- 0.869
SSR- Jaccard	0.43	1.00						0.264	0.085- 1
AFLP/SSR -Jaccard	0.96	0.64	1.00					0.459	0.302- 0.876
SSR- MRD*	-0.51	-0.95	-0.69	1.00				0.742	0- 0.894
AFLP- f_{AB}^M	0.88	0.46	0.86	-0.51	1.00			0.071	0- 0.748
SSR- f_{AB}^M	0.39	0.87	0.56	-0.83	0.48	1.00		0.058	0- 0.663
AFLP/SSR - f_{AB}^M	0.84	0.64	0.89	-0.67	0.96	0.68	1.00	0.072	0- 0.789

* Distance matrix; all the others represent similarity matrices.

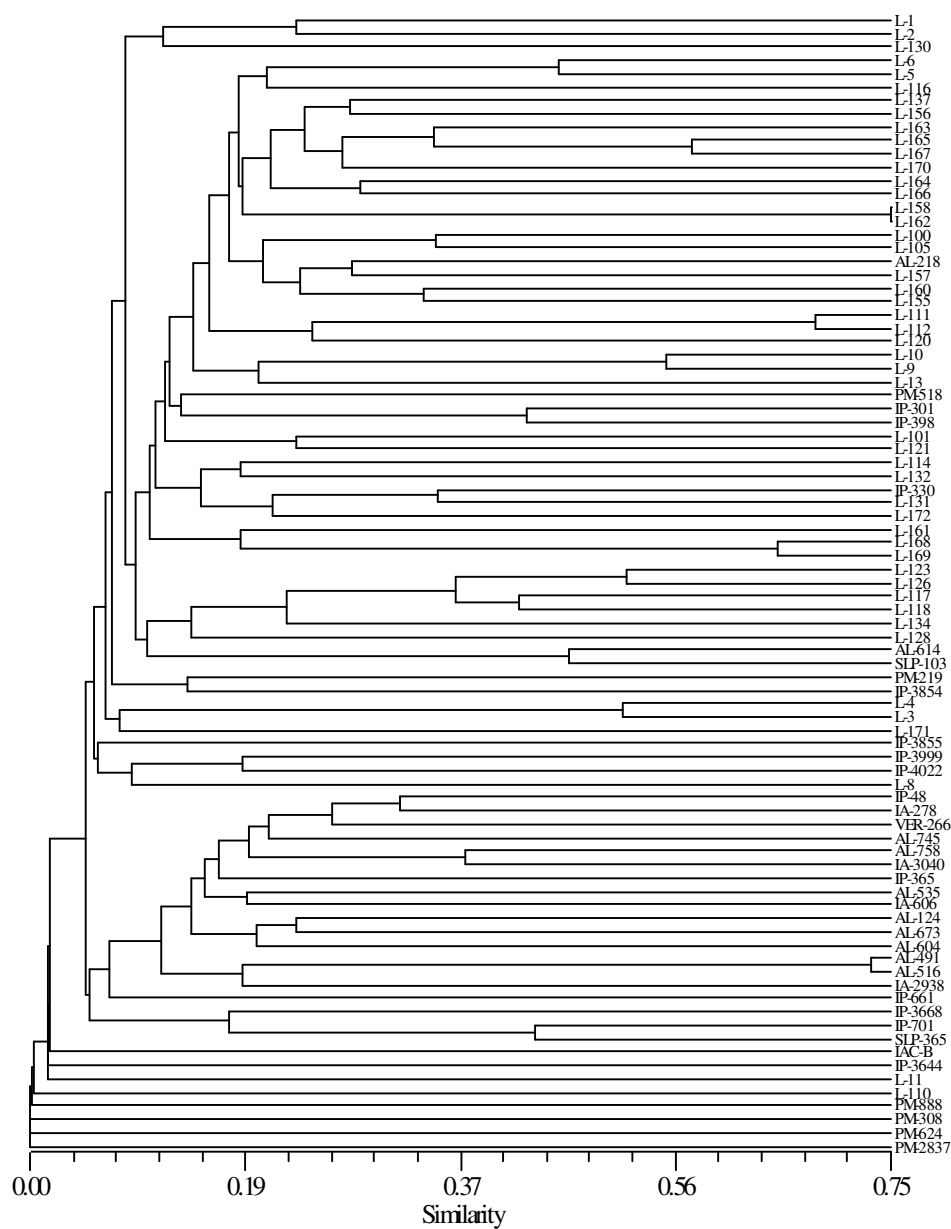


Figure 1 UPGMA showing the clusters obtained by the molecular coefficient of coancestry applied to the AFLP data (cophenetic value = 0.81)

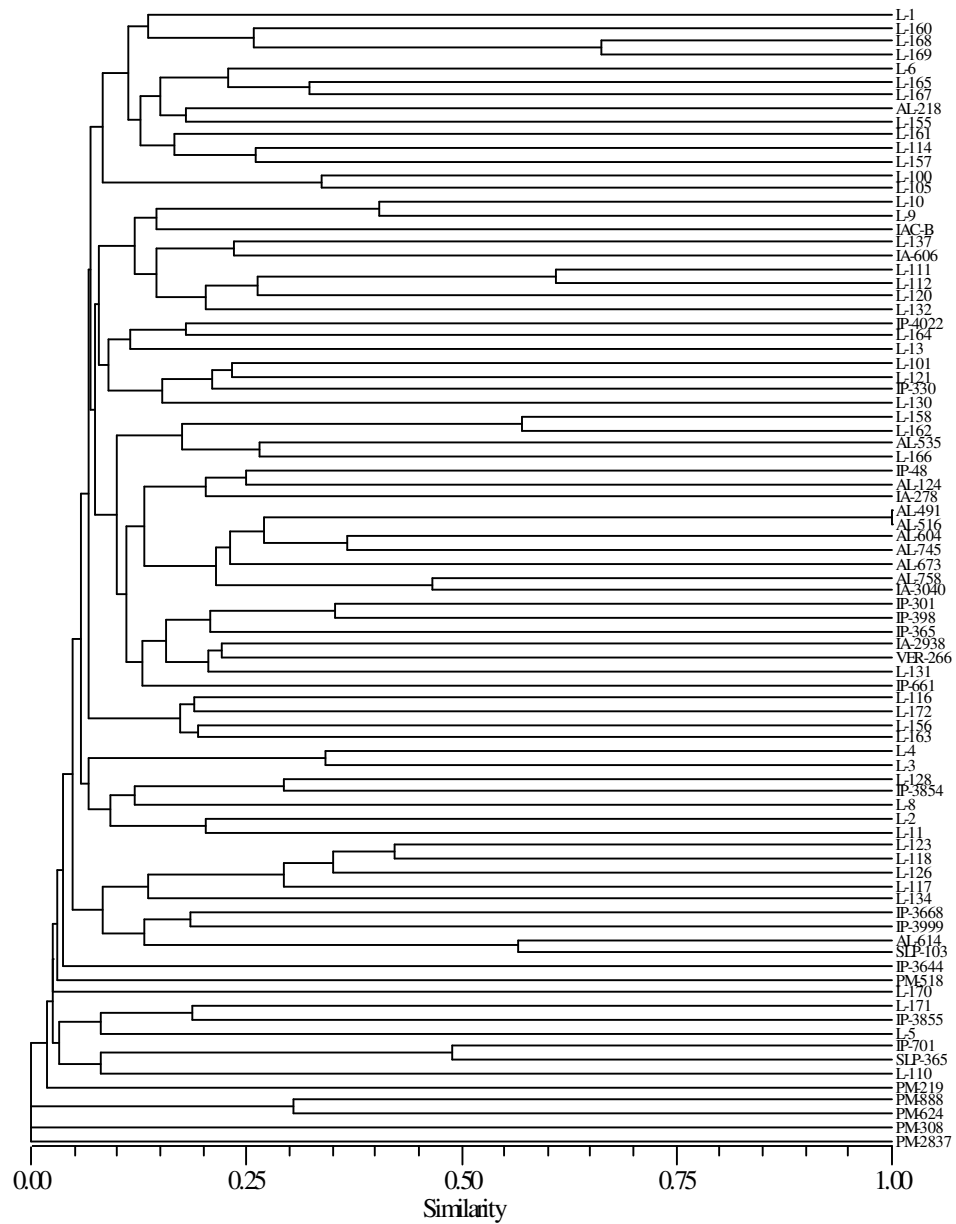
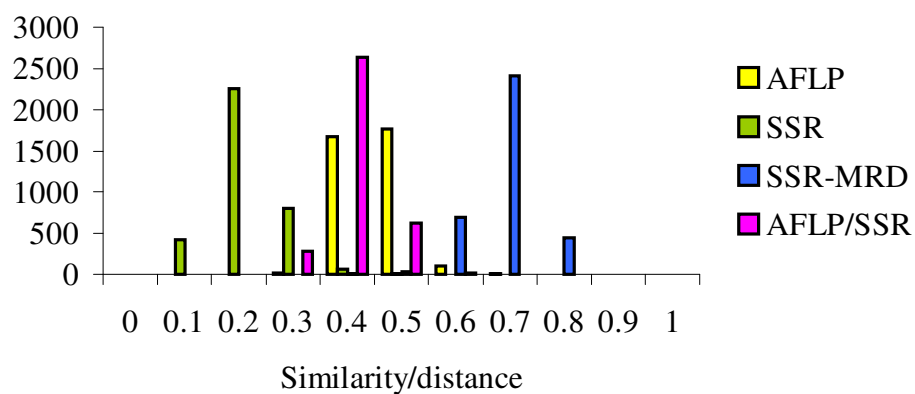


Figure 2 UPGMA showing the clusters obtained by the molecular coefficient of coancestry applied to the SSR data (cophenetic value = 0.72)

A



B

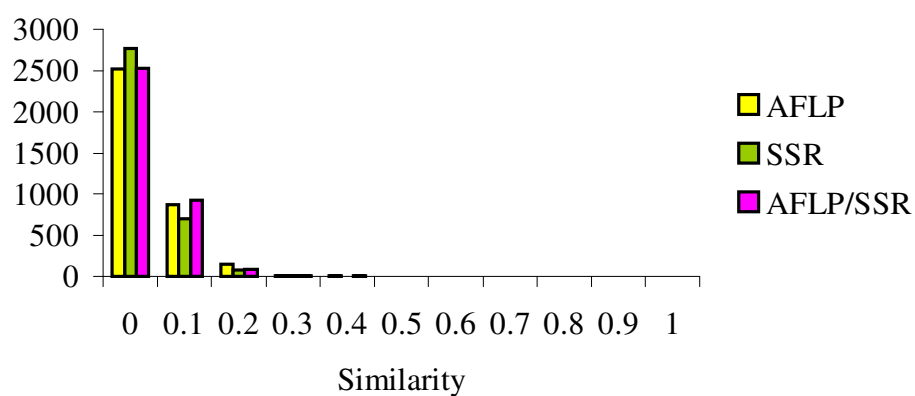


Figure 3 Distribution of frequencies of the 3570 pairwise values in the similarity matrices. (A) Markers analyzed by Jaccard's coefficient (and also a comparison with the distances obtained by modified Rogers' distance for SSR data) (B) Markers analyzed by the molecular coefficient of coancestry

5) Discussão

“Germoplasma do milho tropical: o que os marcadores AFLP e SSR podem nos dizer quando estamos analisando a diversidade genética?” Os resultados que este trabalho apresentou geraram interessantes informações sobre a análise da variabilidade do milho tropical, mas suscitaram também algumas dúvidas quanto à comparação desse tipo de germoplasma com o milho de clima temperado além de questionamentos quanto à classe de marcador molecular a ser usado nesse tipo de estudo. Nessa discussão pretende-se apontar alguns aspectos que vieram de encontro às expectativas do início desse estudo e possíveis explicações para resultados que não corresponderam a essas expectativas. Além disso, são acrescentadas informações que não foram incluídas no corpo do artigo mas que, talvez, sejam interessantes para a discussão mais ampla do assunto.

Apesar de poucos trabalhos terem sido conduzidos com o milho tropical, acreditava-se que o nível de variabilidade nele existente fosse alto, principalmente se comparado à variabilidade observada para o milho de clima temperado. No que concerne ao material brasileiro – e mais especificamente às linhagens do IAC –, as diversas introduções de genótipos e a mistura dessas às raças e variedades típicas do Brasil sugeriam que a diversidade entre as linhagens fosse grande, pois a base genética a partir de que foram desenvolvidas é muito ampla. O milho de clima temperado, principalmente o material largamente usado no Cinturão do Milho nos EUA, por sua vez, reconhecidamente, foi obtido a partir de apenas três grupos heteróticos bem definidos. Portanto, pressupunha-se que, na comparação desses dois germoplasmas, o material tropical exibiria muito maior diversidade. Com base nos resultados obtidos com os SSRs, todavia, se o parâmetro “número médio de alelos” for considerado, vê-se que o milho tropical talvez exiba mesmo

nível de variabilidade que o milho de clima temperado ou talvez nível inferior (em comparação com os trabalhos de Smith *et al.* 1997).

Entretanto, o número médio de alelos pode ser um parâmetro muito relativo, pois está em grande extensão vinculado à escolha dos SSRs usados para a análise. Nesse trabalho, a seleção dos marcadores SSRs foi feita exclusivamente com base na sua qualidade de amplificação, e não a partir de informações de polimorfismo. Locos que apresentaram grande polimorfismo, exibindo vários alelos, porém que não mostraram boa repetibilidade e resolução, foram descartados da análise, e locos que apresentaram apenas dois alelos, porém que mostraram boa qualidade de amplificação e eletroforese, foram integrados à avaliação. Essa conduta na escolha dos SSRs, a qual privilegiou características para facilidade de genotipagem e não para alta taxa de polimorfismo, pode ter embutido um viés nos cálculos do número médio de alelos, assim como nos valores de PIC, os quais também não revelaram diferenças com relação ao milho de clima temperado. Além disso, deve-se considerar que todos os SSRs que puderam ser genotipados foram usados nas análises de diversidade e não somente aqueles que revelaram maior quantidade de alelos. Como exemplo, pode-se citar o seguinte cálculo: se tivéssemos utilizado apenas os vinte locos (número de locos suficiente para um CV = 10%, valor adotado na literatura) com maiores valores de PIC, a média de alelos que seria encontrada subiria de 5,2 para 7,4 e o PIC médio subiria de 0,61 para 0,77, valores que superam amplamente os resultados encontrados para o milho temperado.

É importante considerar também que os SSRs usados nesse estudo foram desenvolvidos para o milho de clima temperado. O fato de esses marcadores não terem sido desenvolvidos especificamente para os genótipos que foram utilizados provavelmente diminuiu a porcentagem de marcadores que puderam ser avaliados. Observou-se que a

classe “bnlg”, que representou apenas 20% dos SSRs usados, apresentou média de alelos de 7,5, contrapondo-se aos quase 60% da classe “umc”, com média de alelos de 5,1. Como esses marcadores (“bnlg”) foram desenvolvidos para um outro tipo de germoplasma, a qualidade de amplificação que foi observada para essa classe muitas vezes não foi satisfatória e SSRs potencialmente muito informativos foram descartados da análise. Caso tivessem sido utilizados mais SSRs “bnlg”, mais polimórficos para o material tropical, em vez de SSRs “umc”, menos polimórficos, o número médio de alelos encontrado possivelmente teria sido superior.

Uma avaliação mais precisa das diferenças entre milho tropical e o temperado, poderia ser obtida por meio da comparação entre o número de alelos por loco, considerando-se os mesmos SSRs. Todavia, não se pode realizar essa comparação, pois foram poucos os locos em comum entre esse trabalho e os marcadores usados nos grandes estudos com genótipos de clima temperado. De qualquer forma, é fundamental compreender que, na comparação da diversidade de dois tipos de material, não apenas alguns parâmetros, como o “número de alelos” que aqui foi discutido, podem ser estudados. Deve-se avaliar diversos aspectos da variabilidade genética e promover correlações com a metodologia escolhida para exibi-la.

Além dos dados que mostraram certa identidade entre a diversidade do milho tropical e do temperado, os resultados que revelaram diferentes agrupamentos de linhagens dependendo do tipo de marcador utilizado (AFLP ou SSR) e a conseqüente baixa correlação ($r = 0,43$) entre essas técnicas também levaram a conclusões não esperadas.

As linhagens que foram utilizadas nesse trabalho podem ser divididas de acordo com a raça/população a partir de que foram desenvolvidas: Cateto e Tuxpeño (raças brasileiras-linhagens antigas) e Antigua Vera Cruz, Mezcla Cristalina, Cogollero, Amarillo

Cristalino e Amarillo Dentado (populações mantidas pelo CIMMYT-linhagens recentes). Nos dendrogramas obtidos com dados de AFLP, pode-se observar que a formação, ainda que parcial, de alguns *clusters*, foi suficiente para unir linhagens pertencentes a cada um desses grupos, com exceção de Tuxpeño, Amarillo Cristalino e Amarillo Dentado. Em adição, notou-se uma separação mais ampla entre linhagens recentes e antigas. Entretanto, nos três dendrogramas construídos com marcadores SSRs, não se observou qualquer tipo de agrupamento ou qualquer organização das linhagens recentes e antigas. Por consequência, a correlação entre as informações das duas técnicas foi muito baixa e dúvidas quanto à capacidade investigativa de cada marcador surgiram.

Uma importante característica dos marcadores moleculares que pode justificar a discrepância dos resultados quando técnicas diferentes foram usadas é a taxa de evolução que cada tipo de marcador apresenta. A grande velocidade com que os SSRs mutam (aproximadamente 10^{-3}) (Vigouroux *et al.* 2002), variando o número de repetições em suas seqüências, em comparação com a taxa de mutação muito inferior dos AFLPs (aproximadamente 10^{-6}) (Mariette *et al.* 2002) ocasiona avaliação do genoma de modo diferencial por cada uma das duas técnicas.

Provavelmente devido a isso, nas análises preliminares, quando valores de similaridade/distância foram verificados, observou-se que a variabilidade exibida pelos AFLPs era inferior àquela exibida pelos SSRs, uma vez que as médias dos valores de similaridade nas matrizes com SSRs eram inferiores àsquelas observadas para as matrizes com AFLPs (p. ex., média = 0,26 para SSR e média = 0,51 para AFLP , ambos com Jaccard). Em adição, a amplitude de variação dos valores foi maior nas matrizes com SSRs em comparação com as matrizes com AFLPs. Esses resultados revelam que os SSRs

possuem maior capacidade de explorar a diversidade, mostrando maior polimorfismo e conseqüentemente maior variabilidade.

Uma outra característica típica dos SSRs, o multialelismo, confere a esse marcador a possibilidade de revelar a diversidade genética de forma muito ampla, pois mostra não apenas a presença ou ausência de uma determinada variável alélica (como vê-se nos AFLPs), mas as diferentes formas alélicas que um dado loco pode conter. Por exibir de maneira tão intensa a diversidade genética, os SSRs privilegiam diferenças gênicas entre dois genótipos em estudo, e não as suas semelhanças. Essa característica pode ter causado o grande embaralhamento das linhagens como observado nos dendrogramas construídos por SSRs e a dificuldade de organização dos genótipos em possíveis grupos. Os AFLPs, em contrapartida, exibindo nível moderado de polimorfismo, facilitam a identificação de similaridades entre pares de genótipos sendo analisados e, em última análise, proporcionam a formação de grupos onde linhagens com alguma porcentagem de semelhança se unem.

Portanto, a escolha de um ou outro marcador em estudos de diversidade genética deve ser norteadada pelo objetivo do trabalho: para agrupar genótipos, o uso de AFLPs parece ser mais apropriado por mostrar as similaridades entre indivíduos, porém para realizar identificações de genótipos, os SSRs talvez sejam mais propícios, pois exibem de maneira mais extensa as “peculiaridades” genéticas de cada indivíduo. Saber explorar as qualidades de cada marcador molecular talvez seja a forma mais adequada de usá-los em estudos como este. As particularidades metodológicas e de análise estatísticas que cada técnica tem, aparentemente, não sugerem o uso simultâneo de AFLPs e SSRs em trabalhos de variabilidade; contudo, a correta escolha e utilização de um ou outro marcador molecular certamente acarretará interessantes resultados quando se pretende investigar a diversidade genética de um conjunto de genótipos.

6) Conclusões

Esse trabalho apresentou as seguintes conclusões:

- (1) a variabilidade existente entre as linhagens de milho tropical analisadas é vasta e grupos divergentes bem definidos não são formados;
- (2) os marcadores AFLP e SSR são ferramentas eficientes para mostrar diversidade genética em linhagens de milho tropical, mas possuem capacidade diferencial de revelar polimorfismo e agrupar genótipos:
 - a. AFLPs exibiram nível moderado de divergência e
 - b. SSRs exibiram alto nível de divergência;
- (3)
 - a. o coeficiente de parentesco molecular aparentemente não consegue expor toda a diversidade que era esperada com base nos dados de polimorfismo dos SSRs;
 - b. Jaccard e MRD agrupam de forma muito semelhante dados de SSR separadamente;
 - c. Jaccard e coeficiente de parentesco molecular não são úteis para unir informações de mais de um marcador, privilegiando o volume de dados gerados pelos AFLPs em detrimento da qualidade de informações fornecidas pelos SSRs;
- (4) o uso simultâneo de marcadores AFLPs e SSRs – o que implicaria, em primeira instância, maior quantidade de informações e maior amostragem do genoma – não é a melhor opção para exibir a diversidade, mostrando agrupamentos semelhantes aos de AFLP separadamente.

7) Referências citadas

- Ajmone-Marsan, P., Castiglioni, P., Fusari, F., Kuiper, M. and Motto, M. (1998) Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. **Theor. Appl. Genet.** 96: 219-227.
- Barbosa, A.M.M., Geraldi, I.O., Benchimol, L.L., Garcia, A.A.F., Souza Jr., C.L. and Souza, A.P. (2003) Relationship of intra- and interpopulational tropical maize single cross hybrid performance and genetic distances computed from AFLP and SSR markers. **Euphytica** 130: 87-99.
- Banerjee, U.C. & Barghoorn, E.S.(1972) *Fine structure f pollen grain ektexine of maize, teosinte and tripsacum*. Em 30th Ann. Proc. Electron Microscopy Soc. Amer., C.J. Arceneaux (ed.), Los Angeles, USA. Pg. 226-227.
- Beadle, G.W. (1939) Teosinte and the origin of maize. **J. Hered.** 30: 245-247.
- Beadle, G.W. (1980) The ancestry of corn. **Sci. Am.** 242: 96-103.
- Bernardo, R. (1993) Estimation of coefficient of coancestry using molecular markers in maize. **Theor. Appl. Genet.** 85: 1055-1062.
- Bernardo, R., Murigneux, A. and Karaman, Z. (1996) Marker-based estimates of identity by descent and likeness among maize inbreds. **Theor. Appl. Genet.** 93: 262-267.
- Benchimol, L.L., de Souza Jr, C.L., Garcia, A.A.F., Kono, P.M.S., Mangolin, C.A., Barbosa, A.A.M., Coelho, A.S.G. and de Souza, A.P. (2000) Genetic diversity in tropical maize inbred lines: heterotic group assignment and hybrid performance determined by RFLP markers. **Plant Breeding** 119: 491-496.
- Boppenmaier, J., Melchinger, A.E., Seitz, G., Geiger, H.H. and Herrmann, R.G. (1993)

- Genetic diversity for RFLPs in European maize inbreds. III. Performance of crosses within versus between heterotic groups for grain traits. **Plant Breeding** 111: 217-226.
- Brown, W.L. (1985) Maize variability of potential interest to plant molecular geneticists. **Maydica** 30: 225-233.
- Chin, E.C.L., Senior, M.L., Shu, H. and Smith, J.S.C. (1996) Maize simple repetitive DNA sequences: abundance and allele variation. **Genome** 39: 866-873.
- Coe, E.H., Neuffer, M.G. and Hoisington, D.A. (1988) *The genetics of corn*. Em *Corn and Corn Improvement*, G.F. Sprague e J.W. Dudley (ed.) 3º ed. Madison, USA. Pg 81-258.
- Crow, J.F. (1948) Alternative hipótesis of hybrid vigor. **Genetics** 33: 477-487.
- Csink, A.K. & Henikoff, S. (1998) Something from nothing: the evolution and utility of satellite repeats. **Trends Genet.** 14: 200-204.
- Dias, L.A.S. (1998) *Análises Multidimensionais*. Em *Eletroforese de Isoenzimas e Proteínas Afins*, A.C. Alfenas (ed) Editora UFV. Viçosa, Brasil. Pg 405-475.
- Doebley, J. & Stec, A. (1991) Genetic analysis of the morphological differences between maize and teosinte. **Genetics** 129: 285-295.
- Doebley, J.F., Goodman, M.M. and Stuber, C.W. (1985) Isozyme variation in the races of maize from Mexico. **Amer. J. Bot.** 72: 629-639.
- Doebley, J., Bacigalupo, A. and Stec, A. (1994) Inheritance of kernel weight in two maize-teosinte hybrid populations: implications for crop evolution. **J. Hered.** 85: 191-195.
- Doebley, J., Stec, A. and Gustus, C. (1995) *Teosinte branched 1* and the origin of maize: evidence for epistasis and the evolution of dominance. **Genetics** 141: 333-346.

- Doebley, J., Stec, A. and Hubbard, L. (1997) The evolution of apical dominance in maize. **Nature** 386: 485-488.
- Doldi, M.-L., Vollmann, J. and Lele, T. (1997) Genetic diversity in soybean as determined by RAPD and microsatellite analysis. **Plant Breeding** 116: 331-335.
- Dorweiler, J., Stec, A., Kermicle, J. and Doebley, J. (1993) *Teosinte glume architecture 1*: a genetic locus controlling a key step in maize evolution. **Science** 262: 233-235.
- Drinic, S.M., Trifunovic, S., Drinic, G. and Konstantinov, K. (2002) Genetic divergence and its correlation to heterosis in maize as revealed by SSR-based markers. **Maydica** 47: 1-8.
- Enoki, H., Sato, H. and Koinuma, K. (2002) SSR analysis of genetic diversity among maize inbred lines adapted to cold regions of Japan. **Theor. Appl. Genet.** 104: 1270-1277.
- Galinat, W.C. (1984) The origin of maize. **Science** 225: 1093-1094.
- Galinat, W.C. (1985) The missing links between teosinte and maize: a review. **Maydica** 30: 137-160.
- Gethi, J.G., Labate, J.A., Lamkey, K.R., Smith, M.E. and Kresovich, S. (2002) SSR variation in important US maize inbred lines. **Crop Sci.** 42: 951-957.
- Goodman, M.M. & Bird, R.McK. (1977) The races of maize IV: tentative grouping of 219 Latin American races. **Econ. Bot.** 31: 204-221.
- Hahn, V., Blankenhorn, K., Schwall, M. and Melchinger, A.E. (1995) Relationships among early European maize inbreds. III. Genetic diversity revealed with RAPD markers and comparison with RFLP and pedigree data. **Maydica** 40: 299-310.

- Iltis, H.H. (1983) From teosinte to maize: the catastrophic sexual transmutation. **Science** 222: 886-8894.
- Jaccard, P. (1908) Nouvelles recherches sur la distribution florale. **Bull. Soc. Vaud. Sci. Nat.** 44: 223-270.
- Lanza, L.L.B., de Souza Jr., C.L., Ottoboni, L.M.M., Vieira, M.L.C. and de Souza, A.P. (1997) Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. **Theor. Appl. Genet.** 94: 1023-1030.
- Li, Y., Shi, Y.S., Cao, Y.S. and Wang, T.Y. (2002) A phenotypic diversity analysis of maize germplasm preserved in China. **Maydica** 47: 107-114.
- Lima, M.L.A., Garcia, A.A.F., Oliveira, K.M., Matsuoka, S., Arizono, H., de Souza Jr., C.L. and de Souza, A.P. (2002) Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). **Theor. Appl. Genet.** 104: 30-38.
- Litt, M. & Luty, J.A. (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. **Am. J. Hum. Genet.** 44: 397-401.
- Longley, A.E. (1937) Morphological characters of teosinte chromosomes. **J. Agric. Res.** 54: 835-862.
- Lu, H. & Bernardo, R. (2001) Molecular marker diversity among current and historical maize inbreds. **Theor. Appl. Genet.** 103: 613-617.
- Lübberstedt, T., Melchinger, A.E., Duble, C., Vuylsteke, M. and Kuiper, M. (2000) Relationships among early European maize inbreds: IV. Genetic diversity revealed

- with AFLP markers and comparison with RFLP, RAPD, and pedigree data. **Crop Sci.** 40: 783-791.
- Mangelsdorf, P.C. (1974) *Corn, Its origin, evolution and improvement*. Belknap press, Harvard University Press, Cambridge MA.
- Mangelsdorf, P.C. & Reeves, R.G. (1939) The origin of Indian corn and its relatives. **Texas Agric. Expt. Sta. Bull.** 574:1-315.
- Mangelsdorf, P.C., MacNeish, R.S. and Galinat, W.C. (1964) Domestication of corn. **Science** 143: 538-545.
- Mariette, S., Le Corre, V., Austerlitz, F. and Kremer, A. (2002) Sampling within the genome for measuring within population diversity: trade-offs between markers. **Mol. Ecol.** 11: 1145-1156.
- Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez G., J., Buckler, E. and Doebley, J. (2002) A single domestication for maize shown by multilocus microsatellite genotyping. **P. Natl. Acad. Sci. USA** 99: 6080-6084.
- Melchinger, A.E., Lee, M., Lamkey, K.R. and Woodman, W.L. (1990) Genetic diversity for Restriction Fragment Length Polymorphisms: relation to estimated genetic effects in maize inbreds. **Crop Sci.** 30: 1033-1040.
- Melchinger, A.E., Messmer, M.M., Lee, M., Woodman, W.L. and Lamkey, K.R. (1991) Diversity and relationships among US maize inbreds revealed by Restriction Fragment Length Polymorphisms. **Crop Sci.** 31: 669-678.
- Moxon, E.R. & Wills, C. (1999) DNA Microsatellites: Agents of Evolution? **Sci. Am.** 280: 94-99.

- Mumm, R.H. & Dudley, J.W. (1994) A classification of 148 U.S. maize inbreds: I. Cluster analysis based on RFLPs. **Crop Sci.** 34: 842-851.
- Newbury, H.J. & Ford-Lloyd, B.V. (1993) The use of RAPD for assessing variation in plants. **Plant Growth Regulation** 12: 43-51.
- Oliveira, K.M., Laborda, P.R., Garcia, A.A.F., Zagatto-Paterniani, M.E.A.G. and Souza, A.P. (2004) Evaluating genetic relationships between tropical maize inbred lines by means of AFLP profiling. *Hereditas (in press)*
- Pejic, I., Ajmone-Marsan, P., Morgante, M., Kozumplick, V., Castiglioni, P., Taramino, G. and Motto, M. (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs and AFLPs. **Theor. Appl. Genet.** 97: 1248-1255.
- Prasad, S.K. & Singh, T.P. (1986) Heterosis in relation to genetic divergence in maize (*Zea mays* L.). **Euphytica** 35: 919-924.
- Senior, M.L., Murphy, J.P., Goodman, M.M. and Stuber, C.W. (1998) Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. **Crop Sci.** 38: 1088-1098.
- Smith, J.S.C. (1986) Genetic diversity within the Corn Belt Dent racial complex of maize (*Zea mays* L.). **Maydica** 31: 349-367.
- Smith, J.S.C. (1988) Identification of pedigrees of hybrid maize (*Zea mays* L.) cultivars by isozyme electrophoresis and reversed-phase high-performance liquid chromatography. **Euphytica** 39: 199-205.
- Smith, J.S.C. & Smith, O.S. (1989) The description and assessment of distances between inbred lines of maize: I. The use of morphological traits as descriptors. **Maydica** 34:

141-150.

Smith, J.S.C., Goodman, M.M. and Stuber, C.W. (1985a) Genetic variability within U.S. maize germplasm. I. Historically important lines. **Crop Sci.** 25: 550-555.

Smith, J.S.C., Goodman, M.M. and Stuber, C.W. (1985b) Genetic variability within U.S. maize germplasm. II. Widely-used inbred lines 1970 to 1979. **Crop Sci.** 25: 681-685.

Smith, J.S.C., Chin, E.C.L., Shu, H., Smith, O.S., Wall, S.J., Senior, M.L., Mitchell, S.E., Kresovich, S. and Ziegler, J. (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. **Theor. Appl. Genet.** 95: 163-173.

Taramino, G. & Tingey, S. (1996) Simple sequence repeats for germplasm analysis and mapping in maize. **Genome** 39: 277-287.

Vigouroux, Y, Jaqueth, J.S., Matsuoka, Y., Smith, O.S., Beavis, W.D., Smith, J.S.C. and Doebley, J. (2002) Rate and pattern of mutation at microsatellite loci in maize. **Mol. Biol. Evol.** 19: 1251-1260.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Fritjers, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting. **Nucleic Acids Res.** 23: 4407-4414.

Wang, R., Stec, A., Hey, J., Lukens, L. and Doebley, J. (1999) The limits of selection during maize domestication. **Nature** 398: 236-239.

Warburton, M.L., Xianchun, X., Crossa, J., Franco, J., Melchinger, A.E., Frisch, M., Bohn, M. and Hoisington D (2002) Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. **Crop Sci.** 42: 1832-1840.

Welsh, J. & McClelland, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. **Nucleic Acids Res.** 18: 7213-7218.

Wright, S. (1978) Evolution and the Genetics of Populations, vol. 4. Variability Within and Among Natural Populations. University of Chicago Press, Chicago

8) Anexo

Importantes figuras que foram produzidas durante a realização do trabalho não puderam ser exibidas no corpo do artigo, apesar de terem constituído dados fundamentais para a elaboração da discussão e das conclusões. Esses resultados, complementares aqueles já exibidos no artigo, são exibidos a seguir.

1.00	<i>bnlg 149</i>	2.00		3.00		4.00		5.00	<i>umc 1325</i>
1.01		2.01		3.01		4.01	<i>nc 135</i>		<i>umc 1416</i>
1.02		2.02	<i>umc 1422</i>	3.02		4.02	<i>umc 1943</i>	5.01	
1.03	<i>bnlg 1484</i>	2.03	<i>bnlg 1621b</i>	3.03		4.03	<i>phi 021</i>	5.02	
	<i>umc 1397</i>	2.04	<i>phi 083</i>	3.04	<i>nc 030</i>	4.04		5.03	
1.04		2.05		3.05		4.05		5.04	<i>umc 1221</i>
1.05	<i>umc 1297</i>	2.06		3.06	<i>bnlg 1951</i>	4.06		5.05	
	<i>umc 1395</i>	2.07	<i>umc 2019</i>		<i>bnlg 2241</i>	4.07		5.06	<i>umc 1524</i>
1.06	<i>umc 1122</i>	2.08		3.07		4.08		5.07	<i>umc 1646</i>
1.07		2.09	<i>umc 1230</i>	3.08	<i>phi 046</i>	4.09	<i>umc 1650</i>	5.08	<i>umc 1792</i>
1.08			<i>umc 1252</i>	3.09	<i>umc 1639</i>	4.10		5.09	
1.09		2.10		3.10		4.11			
1.10	<i>umc 1774</i>								
1.11	<i>umc 1630</i>		<i>bnlg 2077</i>				<i>phi 072</i>		
1.12			<i>2.07-2.08</i>				<i>4.00-4.01</i>		

6.00	<i>bnlg 1600</i>	7.00	<i>umc 1426</i>	8.00		9.00		10.00	
6.01		7.01	<i>phi 057</i>	8.01		9.01	<i>bnlg 1724</i>	10.01	
6.02			<i>phi 112</i>	8.02			<i>phi 028</i>	10.02	<i>phi 059</i>
6.03		7.02		8.03		9.02		10.03	
6.04	<i>umc 1857</i>	7.03		8.04		9.03	<i>phi 022</i>	10.04	<i>bnlg 2336</i>
6.05		7.04	<i>bnlg 1666</i>	8.05		9.04		10.05	
6.06		7.05		8.06	<i>umc 1161</i>	9.05	<i>umc 1357</i>	10.06	
6.07	<i>umc 1653</i>	7.06		8.07		9.06	<i>umc 1733</i>	10.07	<i>umc 1640</i>
6.08				8.08	<i>umc 1069</i>	9.07	<i>umc 1804</i>		
				8.09	<i>umc 1638</i>	9.08			

Figura 4. Distribuição dos 50 locos de microssatélites entre dos 100 *bins* dos cromossomos do milho. A boa dispersão dos marcadores proporciona maior cobertura do genoma e maior confiabilidade dos dados, pois mais partes genômicas podem ser amostradas.

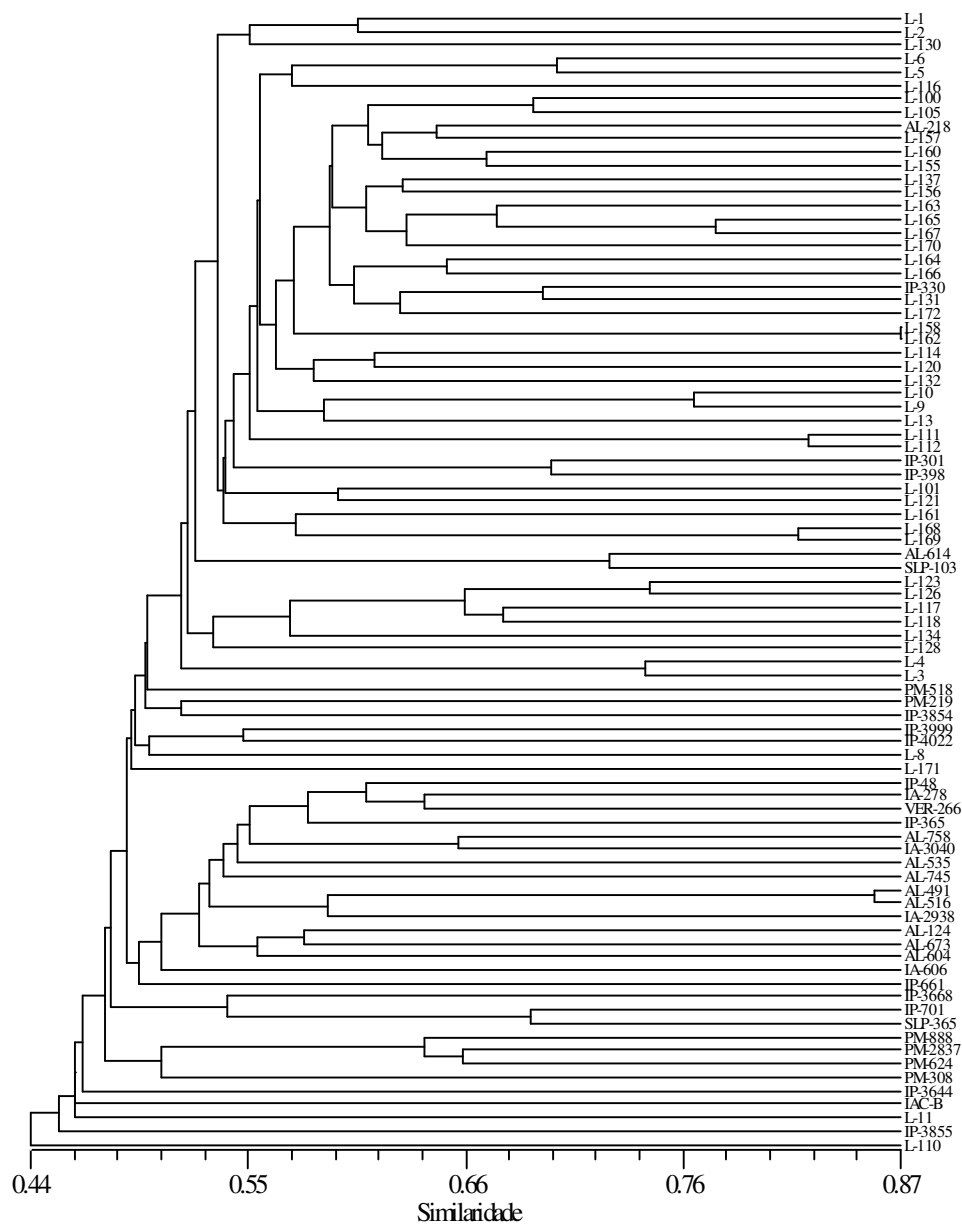


Figura 5. Agrupamento (UPGMA) das 85 linhagens com base nos dados de AFLP-Jaccard (valor cofenético = 0,78).

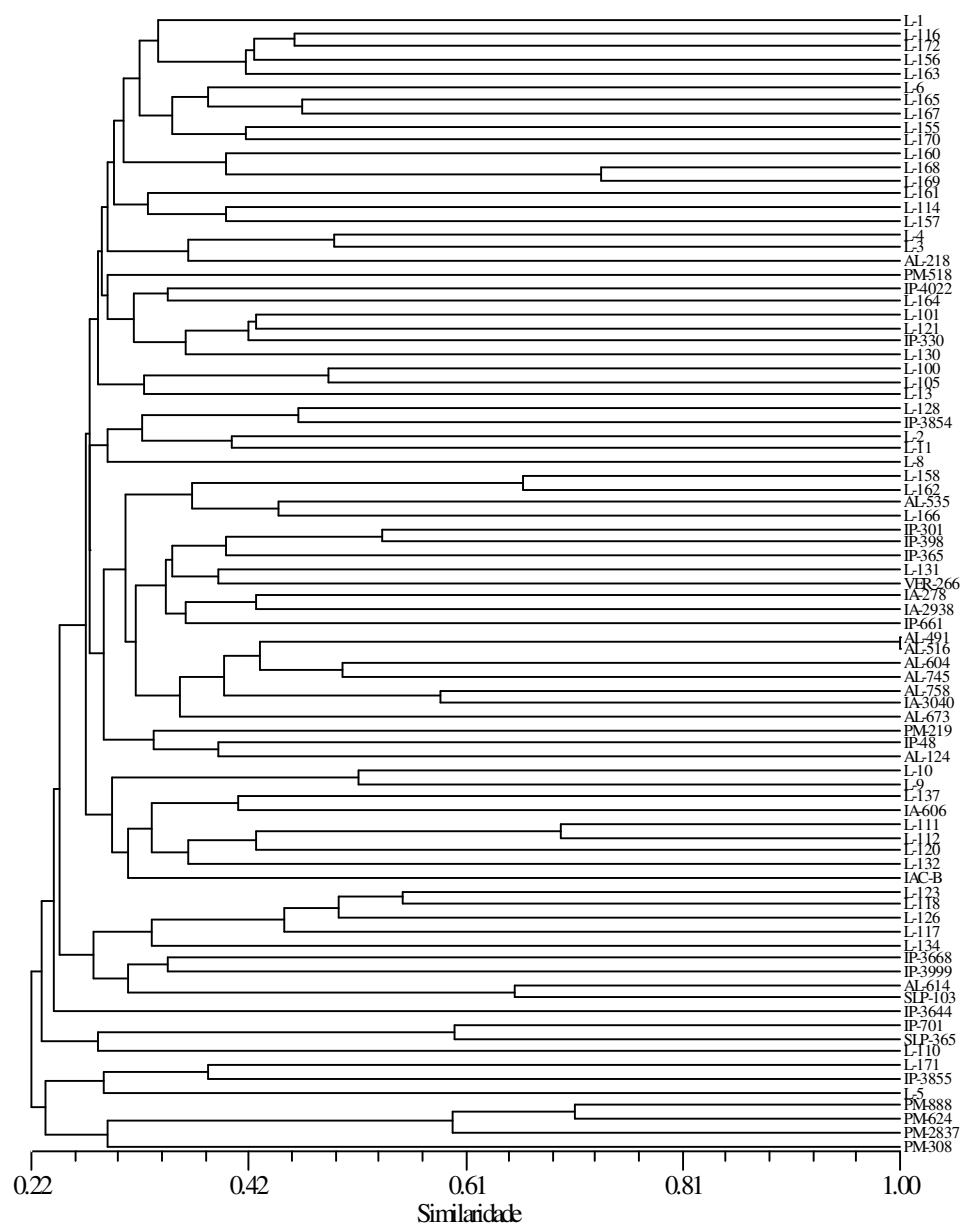


Figura 6. Agrupamento (UPGMA) das 85 linhagens com base nos dados de SSR-Jaccard (valor cofenético = 0,66).

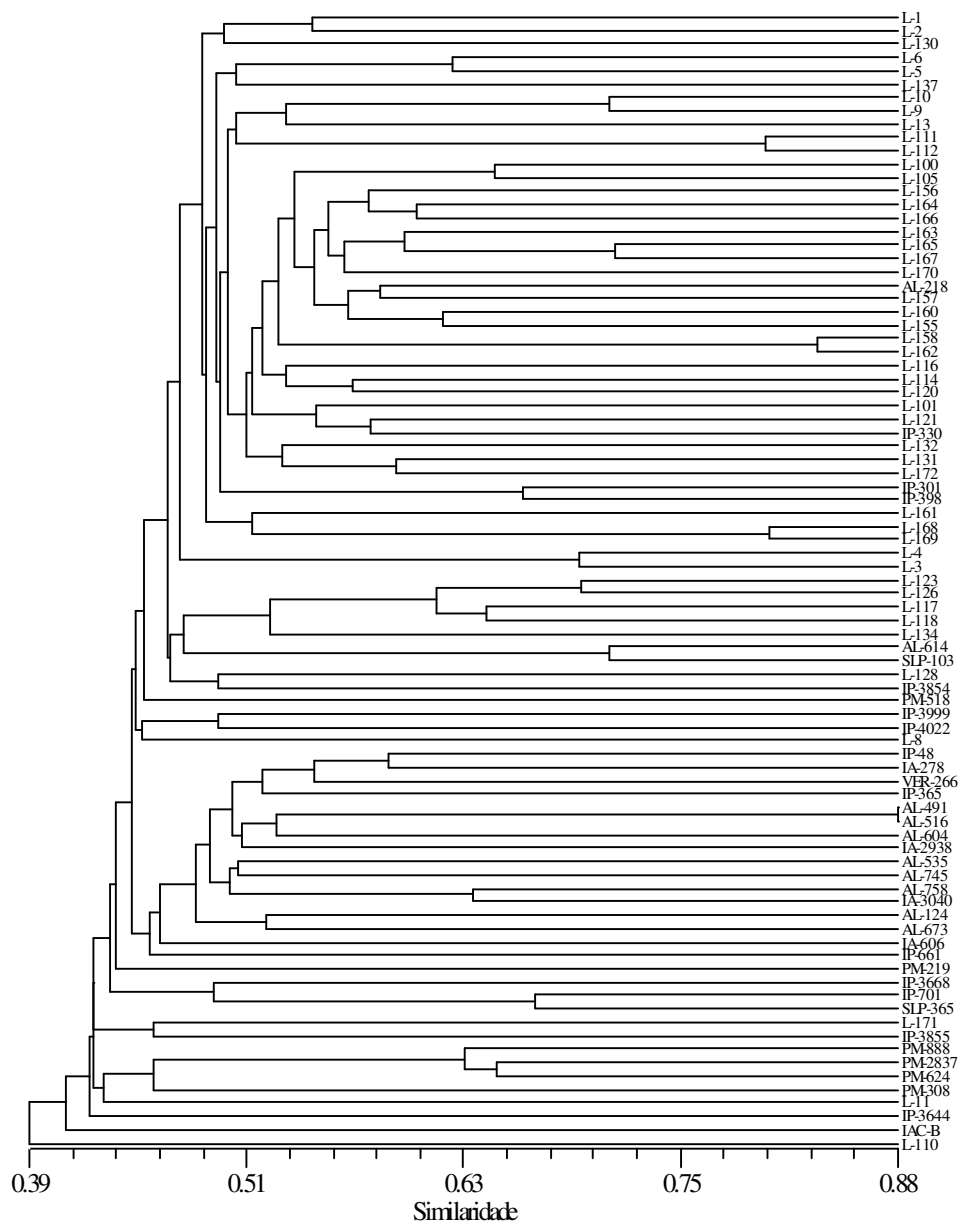


Figura 7. Agrupamento (UPGMA) das 85 linhagens com base nos dados de AFLP/SSR-Jaccard (valor cofenético = 0,79).

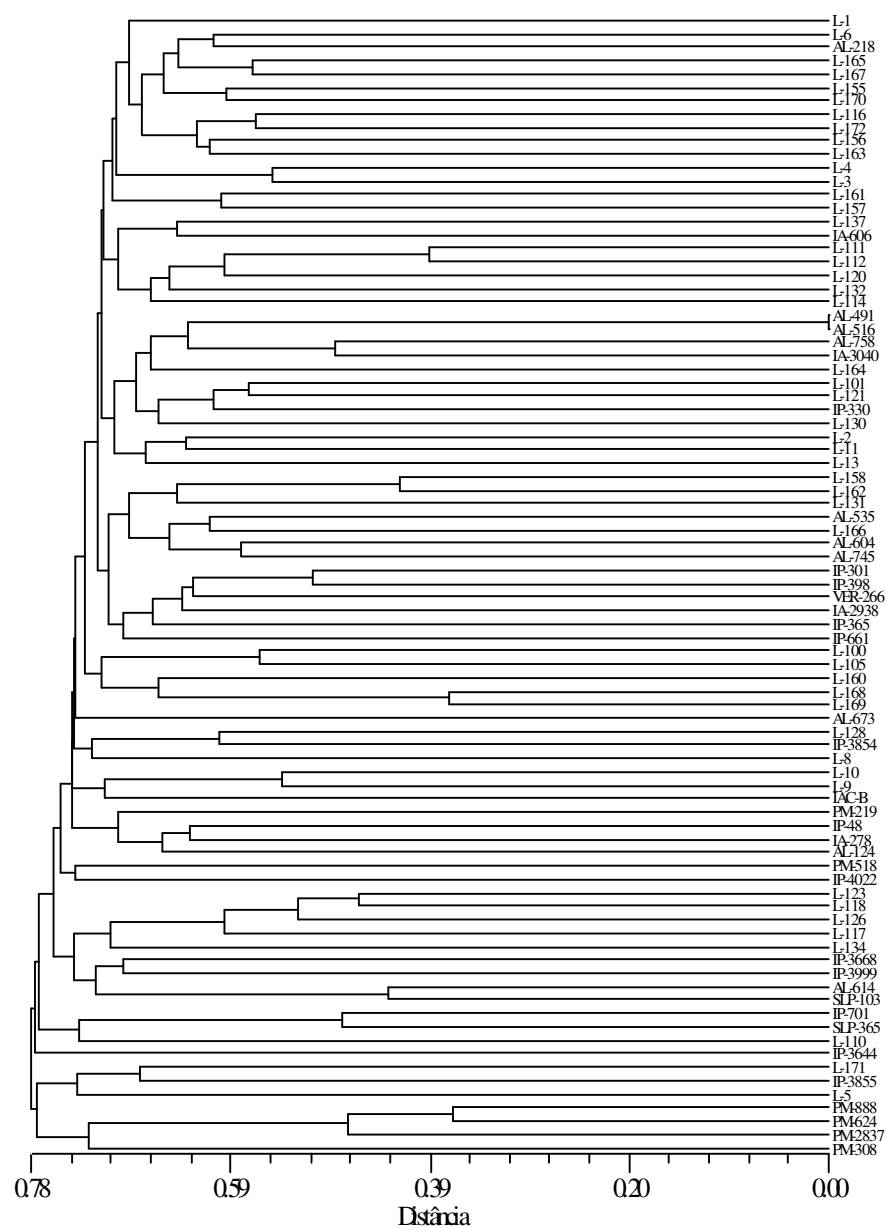


Figura 8. Agrupamento (UPGMA) das 85 linhagens com base nos dados de SSR-MRD (valor cofenético = 0,70).

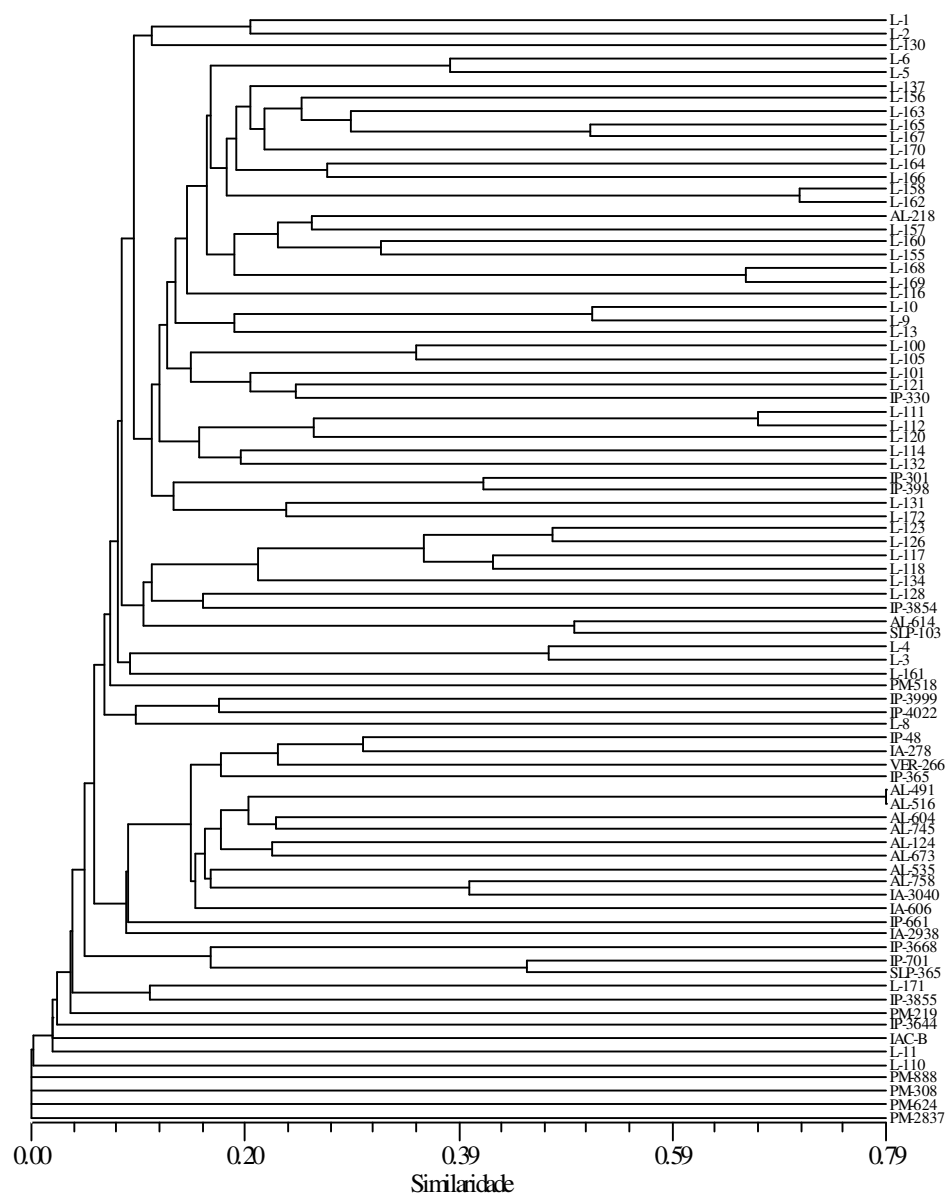


Figura 9. Agrupamento (UPGMA) das 85 linhagens com base nos dados de AFLP/SSR- f_{AB}^M (valor cofenético = 0,83).

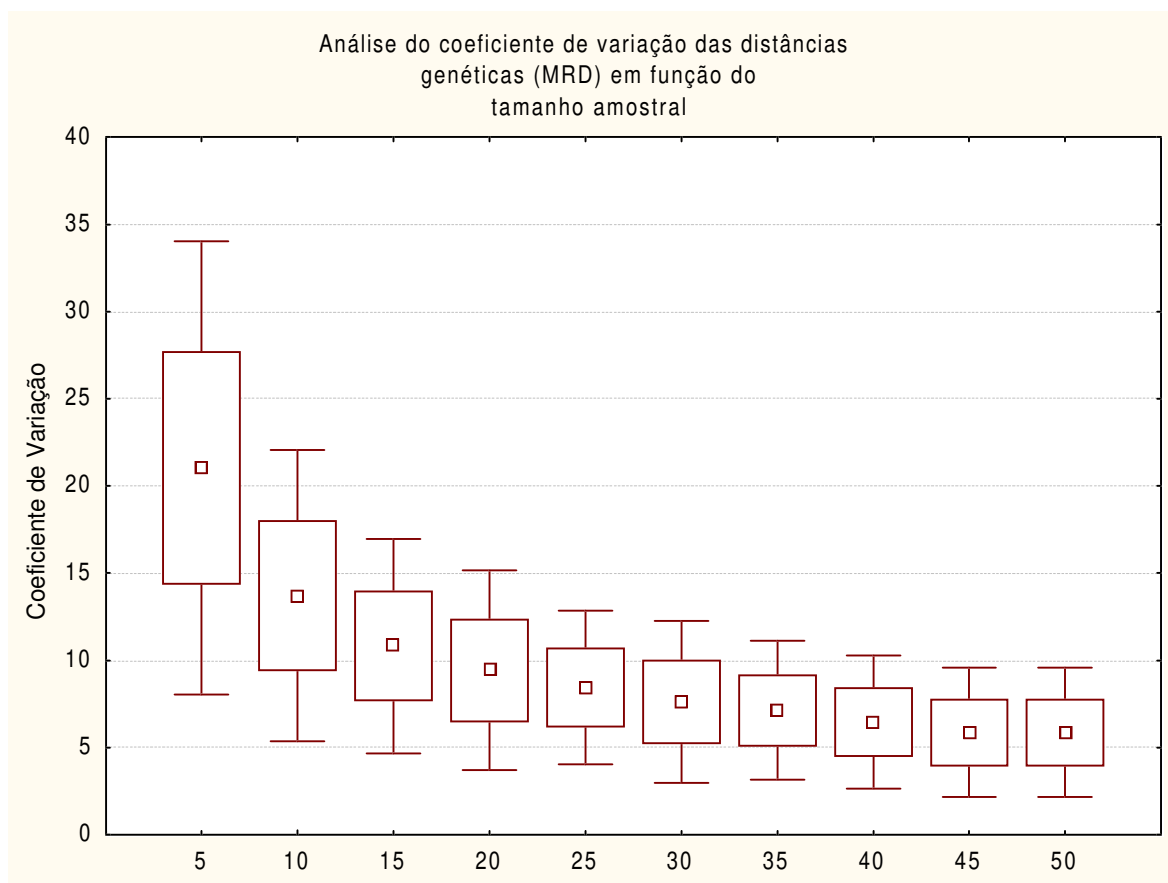


Figura 10. Bootstrap com dados de marcadores microssatélites feito calculado a partir da distância modificada de Rogers (MRD) e aplicando medianas para avaliar o comportamento da variação em função do aumento do número de locos. Percebe-se que o coeficiente de variação se aproxima de 6% com os 50 locos usados no estudo, o que garante a boa qualidade dos dados gerados por esse marcador.