

**VALIDAÇÃO E VALORAÇÃO DA COLEÇÃO
NÚCLEO DE MILHO, SUBGRUPO
ENDOSPERMA DURO**

DÉA ALECIA MARTINS NETTO

2003

DÉA ALECIA MARTINS NETTO

**VALIDAÇÃO E VALORAÇÃO DA COLEÇÃO NÚCLEO DE
MILHO, SUBGRUPO ENDOSPERMA DURO**

Tese apresentada à Universidade Federal de Lavras,
como parte das exigências do programa de
Agronomia, área de concentração em Genética e
Melhoramento de Plantas, para obtenção do título de
"Doutor".

Orientador

Prof. Ph.D. César Augusto Brasil Pereira Pinto

LAVRAS
MINAS GERAIS – BRASIL
2003

Ficha Catalográfica Preparada Pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA

Netto, Déa Alecia Martins

Validação e valoração da coleção núcleo de milho, subgrupo endosperma
duro / Déa Alecia Martins Netto. -- Lavras : UFLA, 2003.
84 p. : il.

Orientador: César Augusto Brasil Pereira Pinto.

Tese (doutorado) – UFLA.

Bibliografia.

1. Milho. 2. Melhoramento genético vegetal. 3. Característica agronômica. 4.
Marcador molecular. 5. AFLP. 6. Germoplasma. I. Universidade Federal de
Lavras. II. Título.

CDD-633.1523

DÉA ALECIA MARTINS NETTO

**VALIDAÇÃO E VALORAÇÃO DA COLEÇÃO NÚCLEO DE MILHO,
SUBGRUPO ENDOSPERMA DURO**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do programa de Agronomia, área de concentração em Genética e Melhoramento de Plantas, para obtenção do título de “Doutor”.

APROVADA em 11 de junho de 2003.

Ph.D. Isabel Regina Prazeres de Souza	Embrapa Milho e Sorgo
Dr. Antônio Carlos de Oliveira	Embrapa Milho e Sorgo
Ph.D. Paulo Evaristo de Oliveira Guimarães	Embrapa Milho e Sorgo
Prof. Dr. João Bosco dos Santos	UFLA

Prof. Ph.D. César Augusto Brasil Pereira Pinto
UFLA
(Orientador)

**LAVRAS
MINAS GERAIS – BRASIL**

O único que é digno de receber a honra e a glória, a força e o poder
Ao Rei Eterno imortal, invisível, mas real
Deus, o Pai, reconheço e O louvo,

Quando nos momentos mais difíceis tu falaste a mim:
“tu és o meu servo, eu te escolhi e não te rejeitei, não temas, porque eu
sou contigo; não te assombres, porque eu sou o teu deus; eu te fortaleço,
e te ajudo, e te sustento com a minha destra fiel.” isaías 41. 9b-10.

Ao meu pai, Paulo da Silva Netto (*in memoriam*), a minha mãe, lanny
e a minha filha Raquel

DEDICO

Aos meus irmãos Paulo Fernando e Eduardo, a toda minha família,
pelo incentivo

AGRADEÇO

Ao Paulo Roberto, pela compreensão e amor,

OFEREÇO

AGRADECIMENTOS

Ao grande autor da vida, Deus, o Pai e Senhor.

À Universidade Federal de Lavras pela oportunidade de realização do curso de doutorado.

À Embrapa Milho e Sorgo pelo apoio material e financeiro concedido durante o curso.

Ao professor César Augusto Brasil Pereira Pinto pelos ensinamentos transmitidos, apoio e confiança demonstrados.

Aos professores Magno e João Bosco pelos ensinamentos transmitidos e colaborações durante o curso.

À pesquisadora, colega e amiga da Embrapa, Isabel Regina Prazeres de Souza pela confiança, apoio, ensinamentos e amizade transmitidos.

Ao pesquisador, colega e amigo da Embrapa, Antônio Carlos de Oliveira, pelo incentivo, inúmeras discussões e sugestões na realização deste trabalho, e grande amparo nas horas mais difíceis.

Aos colegas, Ubiraci, Miguel, Wellerson, Eliane e LÍlian nos procedimentos laboratoriais e de semi-automatização.

Ao pesquisador Ramiro Vilela Andrade e técnico agrícola José da Silva e equipe do BAG pelo apoio, incentivo e ajuda na coleta de dados experimentais.

À amiga Juciane Motta e ao doutorando e amigo Sandro Barbosa pela amizade e colaboração demonstrados.

À minha família, em especial, a minha mãe e minha filha, por todo sustento físico e espiritual, e ao Paulo Roberto, que deu novo sentido à minha vida.

A todos que colaboraram pelo êxito deste trabalho.

SUMÁRIO

	Página
RESUMO.....	i
ABSTRACT	ii
1 INTRODUÇÃO.....	1
2 REFERENCIAL TEÓRICO	3
2.1 Recursos genéticos e coleção de germoplasma.....	3
2.2 A espécie <i>Zea mays</i> e sua diversidade genética	6
2.3 Coleção núcleo brasileira de milho.....	9
2.4 Marcadores morfológicos e moleculares.....	11
2.5 Técnica de AFLP	13
2.6 Avaliação da diversidade genética	13
2.7 Validação da coleção de germoplasma	16
3 MATERIAL E MÉTODOS	19
3.1 Material genético.....	19
3.2 Caracterização morfo-agronômica	22
3.2.1 Dados morfo-agronômicos dos acessos da amostra da coleção base .	22
3.2.2 Dados morfo-agronômicos dos acessos da coleção núcleo	22
3.3 Análise dos dados morfo-agronômicos	25
3.3.1 Estatísticas descritivas.....	25
3.3.2 Dissimilaridade genética entre acessos e análise de agrupamento	25
3.4 Caracterização molecular	27
3.5 Análise dos dados moleculares	32
3.5.1 Similaridade genética entre acessos e análise de agrupamento.....	32
3.5.2 Diversidade genética.....	37
3.5.3 Análise de variância molecular (AMOVA).....	41

4 RESULTADOS E DISCUSSÃO	43
4.1 Caracterização morfo-agronômica	43
4.1.1 Análise descritiva.....	43
4.1.2 Dissimilaridade genética e agrupamento.....	51
4.2 Caracterização molecular	57
4.2.1 Similaridade genética e agrupamento	60
4.2.2 Diversidade genética.....	68
4.3 Considerações sobre a validação da coleção núcleo.....	72
4.4 Considerações sobre a valoração da coleção núcleo	74
5 CONCLUSÕES	76
REFERÊNCIAS BIBLIOGRÁFICAS.....	77

RESUMO

NETTO, D.A.M. **Validação e valoração da coleção núcleo de milho, subgrupo endosperma duro.** 2003. 84 p. Tese (Doutorado em Agronomia) – Universidade Federal de Lavras, Lavras, MG.*

A eficiência de utilização dos acessos de germoplasma de milho que compõem a coleção núcleo pode ser incrementada ao se agregar valores de caracterização molecular e morfo-agronômica, fornecendo informações mais detalhadas sobre esses acessos. Este trabalho objetivou validar e valorar a coleção núcleo de milho, subgrupo endosperma duro, da Embrapa Milho e Sorgo, através do emprego de descritores morfo-agronômicos e marcadores moleculares AFLP. Para isso, utilizaram-se 58 acessos da coleção núcleo e uma amostra de 21 acessos da coleção base, para os quais foram avaliados 22 caracteres morfo-agronômicos e seis combinações de *primers* AFLP para obtenção dos perfis genéticos. Apesar dos dados morfológicos referentes à amostra da coleção base terem sido obtidos em diferentes anos e locais, através de consulta ao banco de dados, foram feitos testes comparativos entre as duas coleções. Houve diferenças significativas para médias, variâncias e distribuição de frequência de acessos das duas coleções em relação à maioria dos descritores. O estudo da validação da coleção núcleo utilizando dados morfo-agronômicos não pode ser conclusivo, possivelmente, em razão da influência do efeito de ambiente. Os dendrogramas da coleção núcleo e da amostra da coleção base apresentaram estrutura geral semelhante e indicaram um bom ajuste entre as matrizes de distâncias genéticas e cofenéticas. A diversidade genética da amostra da coleção base e da coleção núcleo, baseada nos marcadores AFLP, foi atribuída em 94,45% à diversidade presente dentro de coleções, e apenas 5,55% devido à diferença entre coleções. Com relação às frequências alélicas entre as duas coleções, a identidade genética foi alta, significando que as coleções possuem frequências alélicas semelhantes. A distribuição de frequência alélica demonstrou que não houve diferença significativa para 79% dos locos. A análise de variância molecular não mostrou diferença significativa entre a variabilidade genética da coleção núcleo e a amostra da coleção base. A caracterização molecular dos acessos permitiu verificar que a estratégia de amostragem estratificada em regiões ecogeográficas empregada no desenvolvimento da coleção núcleo de milho, tipo de endosperma duro, foi adequada, indicando que esta representa a variabilidade genética da coleção base.

*Comitê Orientador: Prof. Ph.D. César Augusto Brasil Pereira Pinto – UFLA (Orientador), Ph.D. Isabel Regina Prazeres de Souza – Embrapa Milho e Sorgo, Dr. Antônio Carlos de Oliveira - Embrapa Milho e Sorgo.

ABSTRACT

NETTO, D.A.M. **Validation and valorization of maize core collection, sub-group flint endosperm.** 2003. 84 p. Thesis (Doctorate in Agronomy) – Federal University of Lavras, Lavras, MG.*

The utilization efficiency of a maize core collection can be incremented by using molecular and morpho-agronomical characteristics, providing more detailed descriptions about these accessions. The objectives of this work were to validate and valorize maize core collection, sub-group flint endosperm of the Brazilian Agricultural Research Corporation (Embrapa), through the use of morpho-agronomical descriptors and AFLP molecular markers. Fifty eight accessions from the core collection and a sample of 21 accessions of the whole collection were used. Twenty two morpho-agronomical descriptors and six AFLP primers combinations were tested to obtain the genetic patterns. Although morpho-agronomical data of the whole collection sample had been obtained from different years and locals, through data base consulting, comparative tests were made between the two collections. There were significant differences for means, variances, and frequency distribution of the accessions in the two collections in relation to the majority of descriptors. The core collection validation study using morpho-agronomical data was not conclusive due the existence of genotype-environment interaction. Dendrograms of core collection and whole collection sample showed similar general structure, indicating a good fit among the cophenetic and genetic distance matrices. The genetic diversity of whole collection sample and core collection, based on AFLP markers, was 94,45% attributed to the diversity existing inside the collections and only 5,55% to the difference between the collections. Nei's genetic identity was high, showing similar allelic frequencies between collections. The allelic frequency distribution did not present any significant difference for 79% of loci. The analysis of molecular variance did not show significant difference of genetic variability between the core collection and whole collection sample. Molecular characterization of accessions allowed to verify that the stratified sampling strategy using ecogeographic regions in the development of the maize core collection, subgroup flint, was adequate, indicating that the core collection genetic variability is representative of the whole collection.

*Guidance Committee: Ph.D. César Augusto Brasil Pereira Pinto – UFPA (Major Professor), Ph.D. Isabel Regina Prazeres de Souza – Brazilian Agricultural Research Corporation (Embrapa), Dr. Antônio Carlos de Oliveira - Brazilian Agricultural Research Corporation (Embrapa).

1 INTRODUÇÃO

Os recursos genéticos, conjunto de plantas, animais ou microrganismos mantidos com o objetivo de tornar disponíveis características genéticas úteis com valor atual ou potencial, constituem-se na integração entre a biodiversidade e o desenvolvimento sustentável da agricultura (Vilela-Morales et al., 1997). Por isso os recursos genéticos devem ser investigados, manejados, conservados e usados adequadamente. As informações sobre os recursos genéticos, no Brasil e em todo o mundo, são organizadas e disponíveis em coleções ou bancos de germoplasma.

Especificamente no caso do milho (*Zea mays* L.) no Brasil, a coleção de germoplasma está armazenada na Embrapa Milho e Sorgo, Sete Lagoas, MG, é chamada de coleção ativa, pois fornece material aos melhoristas e demais usuários, de universidades ou empresas de pesquisa. Em Brasília, DF, se encontra a coleção base ou de reserva mantida na Embrapa Recursos Genéticos e Biotecnologia, com o objetivo de preservar o germoplasma como estratégia de segurança nacional (Valois et al., 1996; Faiad et al., 1998). Existe ainda uma duplicata da coleção base no *Centro Internacional de Mejoramiento de Maiz y Trigo* (CIMMYT), México.

Para facilitar o acesso à variabilidade genética disponível e estimular a utilização de germoplasma, estabeleceu-se uma coleção nuclear de milho na Embrapa Milho e Sorgo. Por definição, coleção nuclear é um sub-conjunto considerado representativo da coleção base desenvolvido com 10 – 20% dos acessos, mas preservando 70 – 80% de diversidade genética (Brown, 1995). A coleção núcleo de milho foi criada em 1997, através de ações conjuntas de melhoristas, curadores e estatísticos, voltada, principalmente, para ser mais uma coleção de uso de germoplasma do que para a sua conservação simplesmente. É

constituída por 300 acessos, correspondendo a 13% da coleção base, classificados em diferentes categorias, raças e características morfo-agronômicas.

Vários procedimentos estatísticos são utilizados para a criação de coleções núcleo como as técnicas multivariadas de classificação (agrupamento) e os de ordenação (análises de componentes principais) dos acessos (Crossa et al., 1995b). Para verificar se a coleção núcleo contempla a variabilidade genética de todo germoplasma presente na coleção ativa, ou seja, para validar a coleção núcleo de milho é importante avaliar a sua representatividade comparando-se as características morfológicas e moleculares dos acessos de ambas coleções.

Além disso, é de fundamental importância a valoração da coleção núcleo, ou seja, devem-se agregar o máximo de informações sobre os acessos que a compõem para um melhor conhecimento de suas características, diversidade genética e potencial de uso em programas de melhoramento.

Este trabalho teve por objetivo valorar e validar a coleção núcleo de milho, subgrupo endosperma duro, do Banco Ativo de Germoplasma de Milho da Embrapa utilizando-se descritores morfo-agronômicos e marcadores moleculares AFLP.

2 REFERENCIAL TEÓRICO

2.1 Recursos genéticos e coleção de germoplasma

Entende-se como recursos genéticos, toda a variabilidade de espécies de plantas, animais e microrganismos integrantes da biodiversidade, de interesse sócio-econômico atual e potencial, para utilização em programas de melhoramento genético, biotecnologia e em outras ciências afins. A coleção de germoplasma é definida como o conjunto de genótipos de uma espécie com origens geográfica e ambiental variadas, sendo matéria-prima para a pesquisa (Valois et al, 1996). Essa coleção também é chamada de banco de genes, banco de germoplasma ou banco genético e constitui-se na base física onde é feita a conservação de todo o conjunto de materiais hereditários de uma espécie (Valois et al, 1996; Vilela-Morales et al., 1997). No caso de espécies vegetais, são armazenadas, prioritariamente no banco de germoplasma, as sementes, partes da planta, ou até células, consideradas úteis para o propósito de estudar, manejar ou utilizar a informação genética que possuem (Vilela-Morales et al., 1997).

Para facilitar o manejo e a obtenção de informações sobre o germoplasma, este é organizado em agrupamentos de gêneros, famílias ou produtos, ou grupos de produtos como forrageiras, frutíferas, etc. O termo acesso é utilizado para qualificar toda amostra de germoplasma representativa de uma variação genética de uma população ou de um indivíduo. Praticamente todo acesso propagado adequadamente reproduz as características genéticas da população de onde foi obtido (Vilela-Morales et al., 1997).

Os melhoristas, de uma maneira geral, reconhecem a importância dos recursos genéticos armazenados em bancos de germoplasma. Apesar disso, utilizam as coleções próprias como fonte de variabilidade em seus programas de melhoramento. Existem alguns aspectos que impedem a maior utilização dos

recursos genéticos fazendo com que a taxa de uso de germoplasma conservado em bancos genéticos, de uma maneira global, esteja em torno de 2% (Morales e Valois, 1996). No Brasil, Nass et al. (1993) verificaram que 14,7% dos pesquisadores de milho e soja utilizam regularmente os acessos do banco ativo de germoplasma (BAG). Estes autores citaram ainda que a maioria das coleções de trabalho de milho, que é a coleção particular do melhorista, tem um tamanho aproximado de 100 acessos, e que, apesar da qualidade das informações dos acessos nos BAGs ser considerada boa (24,1% para o milho; 43,8% para a soja), a quantidade de informações, segundo os usuários pesquisados, foi considerada insuficiente para a cultura do milho (66,7%) e suficiente para soja (68,8%). Além disso, as principais fontes de acessos visando resistência a pragas, doenças e outros estresses utilizados nos programas de melhoramento de milho foram provenientes da coleção de trabalho do próprio melhorista e, para a soja, a troca de materiais entre pesquisadores. Nass (2001) menciona que as causas da baixa utilização dos bancos de germoplasma são a falta de conhecimento sobre os fatores bióticos e abióticos que afetam o germoplasma, a falta de divulgação das informações dos acessos que compõem as coleções, além de documentação inadequada. Vilela-Morales et al. (1997) também descrevem que as dificuldades em se adquirir o germoplasma podem estar nos impedimentos burocráticos e nos procedimentos de quarentena inadequados. Para aumentar a utilização dos acessos mantidos em bancos de germoplasma, alternativas como programas de pré-melhoramento (*pre-breeding*) e a concentração de esforços em caracterização e avaliação dos acessos que compõem as coleções nucleares (*core collections*) devem merecer maior atenção (Nass, 2001).

O número e a abrangência dos acessos de um banco de germoplasma, a informação sobre as características e a fácil obtenção desses acessos são fatores freqüentemente mencionados como cruciais para o uso do germoplasma em melhoramento de plantas. Dada a necessidade de otimizar a relação entre custos

e manutenção de uma coleção de germoplasma, Frankel (1984), citado por Brown (1995), propôs o termo coleção núcleo que deveria representar, com um mínimo de repetitividade, a diversidade genética de uma espécie e seus parentes. Essa coleção núcleo (CN) consiste em ser o conjunto mais importante de acessos de toda a coleção base (CB) (Brown, 1989). O tamanho da coleção núcleo, segundo a teoria de amostragem de alelos neutros em populações finitas, deve ser de cerca de 10% de acessos tomados aleatoriamente da coleção base tendo a eficiência de reter a variação genética total em aproximadamente 70% (Brown, 1995). A coleção núcleo vem facilitar e incrementar a acessibilidade de usuários desde melhoristas de plantas até geneticistas fundamentais à diversidade genética conservada (Hamon et al., 1995).

Os acessos da coleção núcleo são escolhidos por ser primariamente representativos, ecologicamente e geneticamente distintos entre si, permitindo que se maximize a diversidade genética e minimize a similaridade entre os acessos. diferentes e perceptíveis vantagens são apontadas por Brown (1995) ao se referir às ações de gerenciamento do germoplasma conservado na coleção núcleo: a) conta com um conjunto de referência para decidir se o novo acesso acrescentará características desejáveis à coleção; b) os materiais que a compõem são de alta prioridade para conservação, monitoração da viabilidade, regeneração e multiplicação de sementes; c) a coleção núcleo possui materiais adequados para elaboração de uma lista de descritores distintos e eficientes; d) a avaliação dos acessos que compõem a coleção núcleo cobre uma ampla variação de toda a coleção base, permitindo desenvolver estudos multivariados dos dados e suas interrelações entre os caracteres; e) o enriquecimento do germoplasma é realizado por meio de melhoramento das características desejáveis encontradas nos acessos da coleção núcleo; f) a distribuição de germoplasma, é feita de maneira acelerada, pois as entradas da coleção núcleo poderão ser multiplicadas previamente.

A cultura do milho, por ocupar lugar de destaque na economia brasileira, desponta como um referencial para o desenvolvimento de uma coleção núcleo e assim facilitar o acesso de melhoristas e usuários à ampla variabilidade genética existente.

2.2 A espécie *Zea mays* e sua diversidade genética

Gramínea, da família Poaceae, o milho pertence a tribo Maydeae e compreende sete gêneros, dos quais dois são nativos do hemisfério ocidental – *Zea* e *Tripsacum*. o gênero *Zea* compreende cinco espécies: *Z. mays*, *Z. mexicana*, *Z. luxurians*, *Z. diploperennis*, e *Z. perennis*. O gênero *Zea*, em sua maioria diplóide, apresenta $2n = 20$ cromossomos com exceção de *Z. perennis*, que é tetraplóide. Aproximadamente, 300 raças de milho já foram identificadas e dentro de cada raça, milhares de variedades. O milho é cultivado praticamente em todos os continentes, sob as mais diversas condições. daí observa-se a sua grande variabilidade genética estendendo-se na maioria dos caracteres da planta, dos grãos e da espiga (Paterniani et al., 2000).

O milho é originário do México e a teoria mais aceita hoje é que tenha resultado da evolução do teosinte. A domesticação do milho aconteceu com os povos primitivos que habitavam a América Central, que por seleção, produziram um grande número de raças. Quando o europeu chegou à América, já encontrou vários tipos de milho domesticados e cultivados pelos indígenas.

Na América do Sul, os milhos indígenas foram adotados pelo homem ocidental e desenvolveram-se nas chamadas raças comerciais antigas. As raças comerciais recentes são resultantes dos milhos introduzidos da América Central e do Norte (Paterniani et al., 2000).

Vários são os milhos indígenas que tiveram grande importância no melhoramento genético atual. Entre eles citam-se o Avatí morotí com grãos amarelos e o Avatí morotí branco com todas as estruturas brancas, cultivados pelos índios Guaranis do Paraguai, Bolívia e regiões adjacentes do Brasil. Outro milho cultivado era o dos Kaingang, grupo indígena que vivia desde o Uruguai até o estado de São Paulo. Este milho apresenta grãos amiláceos, dentados e normalmente brancos, ocorrendo variações na coloração do grão, como vermelho, variegado e roxo. Há também o milho entrelaçado, cultivado numa extensa área amazônica, com espigas longas e as fileiras de grãos entrelaçadas, ao invés de retas, como ocorre no milho em geral. Os milhos entrelaçados são amiláceos e com grande variação na coloração dos grãos. No estado do Rio Grande do Sul, a raça Lenha foi encontrada com grãos brancos amiláceos, espiga curta com elevado número de fileiras (Paterniani et al., 2000).

O milho pipoca foi desenvolvido, principalmente, pelos Guaranis, no Paraguai. Nesse milho, predominam os grãos redondos e também os pontudos. Os índios Guaranis também cultivavam um milho de grãos duros, chamado *flint*, de cor branca. Este foi adotado pelo homem civilizado passando a ser cultivado para fins especiais, como a variedade Cristal que é usada para fazer canjica. Os índios Tupis viviam em toda a costa atlântica e cultivavam milhos de grãos duros laranja escuros, os quais originaram o milho Cateto. Esses foram os principais milhos desenvolvidos pelos índios na América do Sul e tiveram grande importância no melhoramento genético atual, principalmente o milho Cateto e o Cristal (Paterniani et al., 2000).

Algumas raças foram desenvolvidas a partir de cruzamentos ocorridos entre as introduções de milhos dos Estados Unidos da América e o Cateto, originando variedades que apresentavam germoplasma Cateto em combinação com milhos dentados. Novas introduções de milhos dentados americanos foram feitas por volta de 1915 pelo pesquisador Benjamin Hunnicut, e, passando a ser

cultivados, naturalmente se cruzaram com o Cateto, resultando em inúmeras variedades. Muitas delas ficaram conhecidas por nomes regionais: Itaiçi, Caiano, Argentino, Dente-de-Cavalo, etc., porém apresentavam o tipo de grão dentado amarelo mais produtivo que o Cateto. As diversas variedades encontradas em São Paulo foram denominadas de milho Dente Paulista. Aquelas encontradas no Rio Grande do Sul foram chamadas de Dentados Riograndenses. Na Bahia, foi identificada uma raça de milho, originada de uma variedade da América Central chamada Tusón. Esta raça é dentada, com espigas grandes de grãos amarelos e com freqüência aparecem grãos avermelhados (Paterniani et al., 2000).

As ações de organização do banco de germoplasma de milho iniciaram-se no México, pela Fundação Rockefeller, a partir de 1943. Foi feito um levantamento dos milhos existentes, e sua classificação resultou na descrição de 25 raças. Verificou-se um potencial de germoplasma bastante diversificado que poderia ser útil em programas de melhoramento. Surgiu, então, o interesse de se verificar a diversidade de germoplasma existente em outros países latino-americanos. O estudo das raças de milho existentes no Brasil começou em 1952, pelo Departamento de Genética da Escola Superior de Agricultura Luiz de Queiroz – ESALQ. Detectaram-se centenas de variedades dos milhos Avatí Morotí, Pipoca, Catetos, Dentados e Semi-dentados, bem como o dos Kaingang e milho entrelaçado. Inicialmente o banco de germoplasma do Departamento de genética da ESALQ contou com mais de 3 mil amostras de sementes, que foram caracterizadas e avaliadas morfológica e agronomicamente (Paterniani et al., 2000). Uma duplicata dessa coleção foi enviada para o *National Seed Storage Laboratory* (NSSL) nos Estados Unidos da América, e posteriormente enviada para o CIMMYT para manutenção (Taba, 1994). Quando a Embrapa (Empresa Brasileira de Pesquisa Agropecuária) desenvolveu um banco de germoplasma em 1975, amostras de sementes de 283 acessos foram enviadas para a Embrapa Milho e Sorgo em Sete Lagoas, MG. Parte da coleção enviada ao NSSL (593

acessos) foi repatriada para o Brasil em 1984, e o restante (1.371 acessos) foi repatriada em 1999, onde encontram preservados os acessos na coleção base da Embrapa Recursos Genéticos e Biotecnologia em Brasília, DF e no Banco Ativo de Germoplasma (BAG) de Milho na Embrapa Milho e Sorgo (Ramiro Andrade, comunicação pessoal).

A coleção de germoplasma de milho é mantida em banco ativo, em Sete Lagoas, MG, cujas sementes são armazenadas em câmaras de conservação a médio prazo, e têm como objetivo atender aos programas de pesquisa. A coleção base, que conserva a longo prazo todo o germoplasma de milho, considerado de segurança nacional, está armazenada em Brasília, DF. A coleção de germoplasma de milho é uma das maiores do Brasil, possuindo mais de 3.000 acessos.

2.3 Coleção núcleo brasileira de milho

O desenvolvimento de uma coleção núcleo é basicamente um exercício de amostragem que tenta assegurar a máxima conservação dos alelos presentes na coleção base. A amostragem estratificada aleatória é um procedimento recomendado por vários autores para se obter a conservação dos alelos comumente dispersos, raramente dispersos e comumente localizados (Abadie et al., 2000). Esses últimos são particularmente importantes, porque incluem os alelos que têm sido submetidos a grande pressão de seleção, conferindo adaptação a condições específicas ambientais. Quatro passos podem ser adotados para seleção de uma coleção núcleo. São eles: a) definição da coleção base; b) divisão da coleção base em grupos geneticamente distintos; c) alocação das entradas por grupos; d) escolha das entradas de cada grupo que farão parte da coleção núcleo (van Hintum, 1999).

Só foi possível elaborar a coleção núcleo de milho devido à organização dos dados de passaporte e de caracterização morfológica e à avaliação agrônômica dos acessos. A coleção núcleo de milho foi desenvolvida utilizando a estratégia de estratificação da coleção, uma vez que este procedimento possibilita a amostragem dos alelos localizados e comuns (Brown, 1995; Crossa et al., 1995b; Spagnoletti Zeuli e Qualset, 1995; Abadie et al., 2000). Foram utilizadas as informações de origem do germoplasma, origem ecogeográfica e tipo de grão das variedades autóctones, e para cada grupo, a representação na coleção núcleo foi relativa ao logaritmo do número de acessos da coleção base (Tabela 1) (Abadie, 1997; Abadie et al., 2000). Assim, pode-se estabelecer uma coleção núcleo com 235 variedades autóctones, 35 materiais melhorados e 30 materiais introduzidos resultando em uma amostra de 300 acessos, considerada de tamanho adequado para ser manejada pelos curadores a baixo custo.

Abadie et al. (1997, 2000) enfatizam a necessidade de se proceder a avaliação e caracterização da coleção núcleo brasileira de milho para outros caracteres visando aumentar o conhecimento de cada material e também utilizar técnicas moleculares que permitirão refinamentos e adição de valor agregado à coleção base. Outros autores (Hamon, 1995; Brown, 1995; Brown e Spillane, 1999; Nass, 2000) enfatizam o uso de marcadores moleculares, juntamente com os morfológicos e agrônômicos, na caracterização de germoplasma como forma de aumentar o conhecimento sobre os acessos armazenados em coleções.

Tabela 1. Distribuição das variedades autóctones brasileiras na coleção núcleo (CN) a partir do tipo de grão e região ecogeográfica de origem dos acessos da coleção base (CB). Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

Região Ecogeográfica	Tipo de Grão							
	Pipoca		Duro		Farináceo		Dentado	
	CB	CN	CB	CN	CB	CN	CB	CN
Sul	29	10	23	9	5	5	279	17
Cerrados	26	10	77	13	50	12	321	19
Cerrados Norte	12	8	9	7	6	5	110	14
Amazônia	35	12	94	15	19	8	121	14
Caatinga	17	8	38	11	1	1	169	16
Agreste e litoral	1	1	14	8	0	0	62	12
Não classificado	4	0	5	0	7	0	10	0
Total	124	49	260	63	88	31	1072	92

Total das variedades autóctones na coleção núcleo = 235

Fonte: Tabela 1 (Abadie et al., 2000) com modificações.

2.4 Marcadores morfológicos e moleculares

Desde os primórdios da agricultura, os marcadores morfológicos foram utilizados pelo homem para seleção de plantas com fenótipo de seu interesse. Seguramente, os marcadores morfológicos contribuíram de forma significativa em estudos de genética e melhoramento de plantas, especificamente no estabelecimento dos princípios teóricos do mapeamento genético e das análises de ligação gênica (Ferreira e Grattapaglia, 1998). No entanto, por apresentarem algumas desvantagens como a necessidade de um grande número de descritores a serem identificados em diferentes partes da planta e no estágio adulto, tempo e necessidade de espaço físico para avaliação do material e passíveis de serem influenciados pelo ambiente, tornam sua aplicação limitada. E apenas

ocasionalmente é que os marcadores morfológicos estão ligados a alelos de importância econômica identificados, limitando seu emprego em programas de melhoramento, além do nível de polimorfismo gerado ser baixo (Ferreira e Grattapaglia, 1998). Isto ocorre devido ao pequeno número de marcadores distintos disponíveis para um mesmo material. Além disso, o marcador morfológico deve possuir uma alta herdabilidade e deve estar intimamente ligado ao alelo de interesse (Ramalho et al., 1999).

Com o rápido desenvolvimento da biotecnologia, foram otimizados diversos métodos de detecção de polimorfismo diretamente em nível de DNA. Os marcadores moleculares que utilizam o DNA apresentam elevado nível de polimorfismo, herança mendeliana, podem conter grande informação genética por loco e podem ser detectados em amostras de células ou de tecidos em qualquer estágio de desenvolvimento da planta (Ferreira e Grattapaglia, 1998; Guimarães e Moreira, 1999).

Os marcadores isoenzimáticos foram inicialmente utilizados, porém com a descoberta e uso de enzimas de restrição aliadas ao processo de amplificação em cadeia utilizando uma DNA polimerase (PCR), o desenvolvimento de diversas classes de marcadores moleculares pode ser incrementado. Dentre as técnicas mais utilizadas atualmente, a de AFLP (*Amplified Fragment Length Polymorphism*) ou polimorfismo de comprimento de fragmentos amplificados, destaca-se por ser precisa e interessante para o estudo da variabilidade genética. Marcadores AFLP amostram de maneira ampla o genoma gerando um grande número de bandas (alelos) por gel, aumentando o rendimento das informações que podem ser resolvidas em um único gel. Porém, uma das limitações do AFLP é o baixo conteúdo de informação genética por loco, por ser uma técnica essencialmente dominante (Ferreira e Grattapaglia, 1998).

2.5 Técnica de AFLP

A técnica de AFLP é baseada na amplificação seletiva, via PCR, de fragmentos de restrição após a digestão do DNA genômico, sendo composta, basicamente, por três etapas: (i) clivagem do DNA com enzimas de restrição e ligação de adaptadores oligonucleotídeos; (ii) amplificação seletiva dos conjuntos de fragmentos de restrição; (iii) análise em gel dos fragmentos amplificados. O AFLP alia vantagens de usar DNA de qualquer origem, ou seja, de qualquer espécie ou complexidade, à detecção de grandes números de fragmentos em uma única reação. Além disso, é uma técnica precisa, porque condições de reação adstringentes são usadas no anelamento do *primer* (Vos et al., 1995; Ferreira e Grattapaglia, 1998).

2.6 Avaliação da diversidade genética

A diversidade genética tem uma estrutura complexa multidimensional. Agrupamentos baseados em dados de progênie, marcadores genéticos, caracteres qualitativos ou quantitativos podem diferir, porque todos os tipos de dados requerem suas próprias medidas de divergência genética. A diversidade genética não é distribuída ao acaso nas populações de plantas. Como um resultado do processo natural de evolução, domesticação e melhoramento, ela tem uma estrutura que, geralmente, pode ser resumida num modelo hierárquico de uma árvore. Para descrever a estrutura, a árvore da diversidade genética pode ser construída por ramificações (baseadas em conhecimentos da evolução natural, domesticação, distribuição e utilização) ou por agrupamentos. Para o propósito

de agrupamento, os dados morfológicos, moleculares e outros marcadores representam a sustentação da diversidade genética total (van Hintum, 1995).

Tanto em bancos de germoplasma como em programas de melhoramento genético de plantas são avaliados diversos caracteres morfológicos e fisiológicos muitas vezes de maneira visual. Uma forma mais adequada de análise dos caracteres é a utilização de técnicas de análise multivariada para agrupar e/ou descrever um grupo de indivíduos considerando todo o conjunto de descritores. Análises multidimensionais como os coeficientes de distância genética têm proporcionado importantes contribuições ao melhoramento genético, à conservação de recursos genéticos e aos estudos de evolução das espécies (Crossa et al., 1995b; Dias, 1998, Pinto-Coelho, 2000). No melhoramento, tem-se obtido sucesso na classificação de linhagens em grupos heteróticos a partir de cruzamento de materiais com maior grau de divergência genética; os coeficientes de distâncias genéticas têm auxiliado na identificação de populações e locais de coleta de germoplasma com elevada diversidade (Dias, 1998).

A análise de componentes principais, análise de agrupamento, distância de Mahalanobis, e outras, são técnicas multivariadas de grande utilização no melhoramento de plantas. A proposta de análise da divergência genética considerando a divisão dos acessos em sub-grupos, constituiu-se em uma estratégia eficiente, com manipulação mais racional e objetiva da variabilidade encontrada no Banco Ativo de Germoplasma de Mandioca (Pereira, 1989).

A utilização de métodos de ordenação em estudos de divergência genética também vem sendo aplicada em várias espécies de plantas, tais como: guandu (Santos et al., 1995); mandioca (Cordeiro et al., 1995); trigo (DeLacy et al., 2000); amendoim silvestre (Veiga et al., 2001); coco (Zibumbo e Colunga, 2001); *Sesbania* spp (Veasey et al., 2001). Esses estudos têm sido de grande utilidade no reconhecimento de genitores para cruzamentos, na manipulação de

acessos em bancos de germoplasma, estabelecimento da relação entre diversidade genética e geográfica, na avaliação da variabilidade total disponível em grupos geneticamente relacionados e validação da coleção núcleo. O método de Tocher é citado como um dos preferidos pelos melhoristas na avaliação da divergência genética (Cruz e Regazzi, 1997).

As medidas feitas entre pares de acessos são chamadas de associação por considerarem que há uma distância que pode ser quantificada entre eles (Crossa et al., 1995b; Cruz e Regazzi, 1997). Essas medidas podem ser de similaridade, como o coeficiente de Jaccard, que mede o quanto os acessos são similares num conjunto de atributos, ou de dissimilaridade, como a distância euclidiana, que verifica o quanto os acessos diferem entre si.

Vários são os coeficientes de similaridades aplicados no agrupamento de acessos de germoplasma. Oito deles foram estudados no agrupamento de 27 cultivares de feijão (*Phaseolus vulgaris* L.) analisados por marcadores RAPD (Duarte et al., 1999b). Esses autores encontraram poucas alterações no agrupamento dos cultivares. Verificaram que as correlações entre as distâncias genéticas foram sempre maiores que 0,86, e que o coeficiente de similaridade de Jaccard foi o de mais simples interpretação, devendo ser preferencialmente empregado. A divergência genética dessas 27 cultivares de feijão foi avaliada com base na similaridade entre os caracteres morfológicos e os marcadores RAPD. A correlação de 0,89 entre as distâncias genéticas obtidas por RAPD, e as distâncias generalizadas de Mahalanobis obtidas por dados morfológicos, indicou que estes marcadores forneceram classificações semelhantes de divergência genética (Duarte et al., 1999a).

A similaridade genética estimada pelo coeficiente de Dice, utilizando dados de RAPD e agrupamento UPGMA, foi eficiente em avaliar a variabilidade genética liberada por diferentes métodos de condução da população segregante de feijão, segundo Castanheira (2001).

Outras medidas da diversidade genética também têm sido utilizadas como o índice de diversidade de Nei e o índice de informação de Shannon, bem como a diferenciação genética entre populações como o G_{st} (van Hintum, 1995; Galwey, 1995, Robinson, 1998). Van Hintum (1995), Skinner et al. (1999) e van Hintum et al. (2000) mencionaram que para analisar a diversidade da cultura com dados quantitativos dentro de populações, o desvio padrão comum e o coeficiente de variação podem ser usados, e a diferenciação entre populações é usualmente quantificada usando médias ou componentes de variância.

2.7 Validação da coleção de germoplasma

Quando uma coleção núcleo é definida, sua representatividade pode ser avaliada por comparação de sua diversidade com a da coleção base, em cada descritor individual (Galwey, 1995; van Hintum et al., 2000). Assim, pode-se quantificar quão próximo é a variação presente na coleção núcleo em relação à coleção base e se os padrões de variação da núcleo se assemelham aos da base. Galwey (1995) propõe os mesmos índices de diversidade citados anteriormente. O termo verificação pode ser usado para indicar a avaliação da representatividade relativa à coleção base, e o termo validação, para indicar a avaliação relativa a critérios externos aos acessos como dados de passaporte, estudos taxonômicos e ecogeográficos da planta, caracteres externos ou não contemplados na coleção de germoplasma (Galwey, 1995). No presente trabalho, o termo validação será usado com o significado de verificação.

As aplicações de marcadores moleculares no melhoramento de plantas abrangem desde a identificação do genótipo até construção de mapas genéticos passando pela quantificação da variabilidade genética. Uma das aplicações de curto prazo do marcador molecular é o estudo da diversidade e distância

genética. Cada marcador é analisado como sendo um caráter fenotípico distinto e independente dos demais cuja interpretação é feita da seguinte forma: alelos (bandas - fragmentos de DNA) em comum entre acessos representam similaridades genéticas, enquanto os não comuns representam diferenças genéticas. A informação molecular de diversidade genética pode auxiliar na avaliação de redundância e deficiências das coleções de germoplasma, bem como no estabelecimento e comparações de coleções nucleares (Ferreira e Grattapaglia, 1998).

Um ensaio utilizando marcadores RAPD, para validar a coleção núcleo de feijão no México, foi realizado em uma amostra de 90 acessos tanto para a coleção núcleo quanto para a coleção base. Marcadores RAPDs foram gerados usando 39 *primers* que produziram 224 bandas polimórficas. A variabilidade relativa foi avaliada nas duas amostras. Essencialmente, não puderam ser detectadas diferenças nas frequências alélicas (bandas) ou variabilidade em cada amostra, ou seja, as amostras foram virtualmente idênticas demonstrando que as análises da diversidade dos marcadores moleculares sugerem que a coleção núcleo foi representativa da coleção base (Skroch et al., 1998).

Para validar a coleção núcleo de mandioca, Cordeiro et al. (1995) utilizaram dados de caracterização e avaliação de acessos. Esses autores usaram técnicas multivariadas como análises de correspondência e escala multidimensional para validar o agrupamento e a classificação ecogeográfica. Concluíram que a estratificação dos acessos foi bem sucedida, pois as informações agroecológicas estavam ligadas às morfológicas. Para Cordeiro et al. (1995), Crossa et al. (1995b), Tohme et al. (1995) e Schoen e Brown (1995), os marcadores genéticos devem ser usados para auxiliar a validação da coleção núcleo.

Numerosos são os trabalhos utilizando técnicas moleculares na caracterização da diversidade genética de inúmeras espécies. Tohme et al.

(1996) analisaram a diversidade genética em coleção núcleo de feijão utilizando AFLP e dados morfológicos. Um total de 110 bandas foi encontrado em 114 genótipos utilizando 16 combinações de *primers*. A análise com marcadores AFLP produziu grande quantidade de informações em curto período de tempo, permitindo discernimento sobre a estrutura genética da coleção núcleo de feijão silvestre (Tohme et al., 1996).

Barrett et al. (1998) compararam a diversidade genética obtida por marcadores AFLP e a obtida com informações de pedigree em 43 cultivares de trigo. A diversidade genética foi calculada pelo complemento do coeficiente de similaridade de Jaccard e a análise de agrupamento pelo método UPGMA. A diversidade genética baseada no pedigree e a baseada nos marcadores tiveram padrões hierárquicos semelhantes entre as 43 cultivares.

Avaliando a diversidade genética entre 61 acessos chineses de soja utilizando oito pares de *primers* de AFLP que geraram 569 fragmentos, Zhu et al. (1999) observaram que a diversidade genética da coleção chinesa de germoplasma é maior que aquela encontrada na coleção americana.

Usando duas combinações de *primers* AFLP, Cervera et al. (1998) encontraram 108 locos polimórficos, em 67 acessos de *Vitis vinifera* L.. Esses resultados permitiram identificar, com elevada segurança, cultivares que pertenciam à mesma variedade dentro da coleção, o que reduziu tempo e custo na sua conservação.

3 MATERIAL E MÉTODOS

A pesquisa foi conduzida na Embrapa Milho e Sorgo, Sete Lagoas, MG, localizada no km 65 da Rodovia MG 424, que liga Belo Horizonte a Sete Lagoas. As coordenadas geográficas são 19° 28' latitude sul e longitude 44° 15' 08" W GrW. A altitude, em sua estação meteorológica, é de 732 m. Ocupa uma área de 1.932,80 ha, que tem como solos predominantes os latossolos vermelho-escuro e vermelho-amarelo, ocorrendo ainda, em menor escala, cambissolos, aluviais e hidromórficos. O clima da região, segundo Koopen, é do tipo AW (clima de savana com inverno seco) (<http://www.cnpms.embrapa.br>).

3.1 Material genético

Os materiais genéticos utilizados foram os acessos de milho da coleção núcleo do tipo de endosperma duro e uma amostra da coleção base. Esse grupo foi selecionado por se constituir de materiais de grande importância para o melhoramento genético do milho.

Apesar da coleção núcleo ser constituída por 63 acessos do tipo de endosperma duro, apenas 58 apresentavam quantidade suficiente de sementes que pudessem ser disponibilizadas. Portanto, foram caracterizados morfo-agronomicamente e molecularmente 58 acessos constituídos de variedades autóctones provenientes de várias regiões ecogeográficas do Brasil (Tabela 2).

A amostragem dos materiais genéticos da coleção base para caracterização morfo-agronômica e molecular seguiu a estratégia logarítmica estratificada. Essa estratégia é baseada na proporção do logaritmo do número de acessos de cada estrato eco-geográfico em relação a soma dos logaritmos de todos os estratos, multiplicados pela amostragem de 10% do total de acessos da

Tabela 2. Identificação e região ecogeográfica dos acessos de milho endosperma duro da coleção núcleo e da amostra da coleção base. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

Coleção Núcleo		Coleção Base			
Nome*	Região	Nome **	Região	Nome***	Região
PR013	1	RS001	1	RS 093	1
PR034	1	RS099	1	RS 095	1
RS194	1	SC038	1	RS 100	1
RS160	1	MG001	2	RS 076	1
RS162	1	MG003	2	RS 081	1
RS165	1	MG004	2	RS 165	1
SC005	1	MG022	2	RS 001	1
MG010	2	MG145	2	RS 099	1
MG089	2	SP014	2	SP 008	2
MG104	2	RR131	3	SP 015	2
MS019	2	AC017	4	SP 021	2
BA178	2	AC018	4	SP 132	2
Cateto SL	2	AC033	4	SP 186	2
GO001	2	AC034	4	SP 187	2
MG084	2	PA086	4	SP 170	2
PR056	2	AL005	5	MG 007	2
SP015	2	BA052	5	RJ 005	2
SP050	2	PB007	5	SP 014	2
SP087	2	SE026	5	MG 001	2
SP145	2	BA070	6	MG 003	2
RR001	3	BA103	6	SP 173	2
RR106	3			RR 033	3
RR012	3			RR 131	3
RR088	3			RR 009	3
RR090	3			AM 002	4
RR003	3			PA 023	4
AC014	4			PA 039	4
AC036	4			PA 082	4
PA003	4			PA 111	4
PA022	4			AC 017	4
PA032	4			AC 018	4
RR200	4			AC 034	4
AC022	4			PA 086	4
AM001	4			PA 050	4

Continua ...

Tabela 2. Continuação.

Coleção Núcleo		Coleção Base			
Nome*	Região	Nome **	Região	Nome***	Região
MT038	4			PA 033	4
PA047	4			PA 070	4
RO009	4			AC 044	4
RR201	4			PA 097	4
AM003	4			PA 105	4
AM006	4			PA 107	4
PA089	4			PA 031	4
AL009	5			PA 040	4
BA020	5			PA 054	4
BA226	5			BA 012	5
PB020	5			CE 014	5
BA003	5			CE 015	5
BA042	5			PB 018	5
BA061	5			PE 040	5
CE039	5			AL 005	5
RN012	5			BA 052	5
SE014	5			PB 007	5
BA066	6			SE 026	5
BA117	6			BA 009	5
BA201	6			BA 041	5
BA224	6			BA 237	5
PE029	6			BA 233	5
PE001	6			RJ 003	6
BA132	6			BA 070	6

Região: 1. Sul; 2. Cerrados; 3. Cerrados-Norte; 4. Amazônia; 5. Caatinga; 6. Agreste-Litoral

*Acessos caracterizados morfo-agronomicamente e molecularmente.

** Acessos caracterizados molecularmente.

*** Acessos caracterizados morfo-agronomicamente.

coleção base, após retirados aqueles que compõem a coleção núcleo, conforme Abadie et al. (2000). Assim, a caracterização molecular foi realizada em 21 acessos do tipo de endosperma duro, amostrados da coleção base para comparar com os dados moleculares da coleção núcleo.

Para obtenção dos dados morfo-agronômicos dos acessos da coleção base, uma amostra de 58 acessos constituída com base também na amostragem estratificada, alocação logarítmica conforme Abadie et al. (2000), foi considerada (Tabela 2).

3.2 Caracterização morfo-agronômica

3.2.1 Dados morfo-agronômicos dos acessos da amostra da coleção base

Foram realizadas consultas às bases de dados existentes a respeito das informações sobre as características morfo-agronômicas dos 58 acessos de germoplasma de milho endosperma duro. Foram utilizados dados médios de 22 caracteres morfo-agronômicos dos acessos de milho avaliados nos locais de Sete Lagoas e Janaúba, no período de 1991 a 1998 (Tabela 3).

3.2.2 Dados morfo-agronômicos dos acessos da coleção núcleo

O experimento de caracterização morfológica dos 58 acessos da coleção núcleo foi realizado na área experimental da Embrapa Milho e Sorgo em Sete Lagoas, MG, com época de semeadura em outubro de 2001. Para a semeadura utilizaram-se 20 gramas de sementes de cada acesso distribuídas em duas linhas de cinco metros com espaçamento de 0,90 m entre linhas. O desbaste foi realizado deixando cinco plantas por metro linear, perfazendo um estande de 50 plantas. A irrigação suplementar foi realizada quando necessária, bem como o controle de pragas, doenças e plantas invasoras de acordo com o manejo utilizado para a cultura. foram tomados os dados de 10 plantas ao acaso de cada acesso.

Tabela 3. Caracteres morfo-agronômicos avaliados nas coleções de milho. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

Abreviação	Caracter	Descrição do Caracter
FM	Floração masculina	Número de dias após semeadura com 50% de espiguetas abertas
FF	Floração feminina	Número de dias após semeadura com 50% dos estigmas liberados
RP	Ramificação do pendão	Número de ramificações primárias e secundárias do pendão
AP	Altura da planta	Medida em cm do nível do solo à ponta do pendão
AE	Altura da espiga	Medida em cm do nível do solo à base da 1 ^a espiga
FA	Folhas acima da espiga	Número de folhas acima da 1 ^a espiga
TF	Ttotal de folhas	Número total de folhas da planta
DC	Diâmetro do colmo	Medida em mm do colmo tomada a 10 cm do solo
C5F	Comprimento da folha	Medida em cm da 5 ^a folha
CH	Comprimento da haste	Medida em cm da haste principal do pendão
EP	Número de espiga por planta	Número de espigas por planta
CP	Comprimento da espiga	Medida em cm da base à ponta da espiga despilhada
DE	Diâmetro da espiga	Medida em mm na porção intermediária da espiga
NF	Número de fileiras de grãos	Número de fileiras de grãos na porção intermediária da espiga
NG	Número de grãos por fileira	Número de grãos por fileira, tomada ao acaso, por espiga
PE	Peso de espiga	Peso em gramas da espiga
PG	Peso de grãos por espiga	Peso em gramas de grãos por espiga
DS	Diâmetro do sabugo	Medida em mm na porção intermediária do sabugo
NC	Número de entrenós	Número de entrenós da planta
AF	Ângulo da folha/colmo	Ângulo tomado da 5 ^a folha em relação ao colmo

Continua ...

Tabela 3. Continuação.

Abreviação	Caracter	Descrição do Caracter
PMG	Peso de mil grãos	Peso em gramas de mil grãos tomados ao acaso
CF	Cor da folha	Observação visual da cor predominante da folha
QE	Qualidade da espiga	Observação visual numa escala de 1 a 9 onde 1= péssima, 9= excelente
SE	Sanidade da espiga	Observação visual numa escala de 1 a 9 onde 1= completamente doente, 9=completamente sadia
TE	Tipo de espiga	Observação visual: cônica, cilíndrica ou cilíndrica cônica
CE	Cor do endosperma	Observação visual: branca, amarela, laranja, vermelha
CC	Cor da coroa	Observação visual: branca, amarela, laranja, vermelha
PQ	Plantas quebradas	Porcentagem de plantas quebradas (estande de 50 plantas)
PA	Plantas acamadas	Porcentagem de plantas acamadas (estande de 50 plantas)
CPF	Comportamento da folha	Observação visual da 5ª folha (reto, recurvado, fortemente recurvado)
F1F	Forma da primeira folha	Observação visual no 10º dia de plantio (pontiaguda, pontiaguda arredondada, arredondada para espatulada, espatulada)
AG	Arranjo dos grãos	Observação visual do arranjo dos grãos (reto, levemente curvado, espiral, irregular)

Além dos descritores considerados para a coleção base, foram aplicados mais oito caracteres valorando assim os acessos de milho da coleção núcleo descritos também na Tabela 3. São eles (IBPGR, 1991; Brasil, 1999):

1. Número de entrenós;
2. Cor do endosperma;
3. Ângulo entre a lâmina da folha, logo acima da espiga superior, e o caule;
4. Comportamento da 5ª lâmina foliar acima da espiga superior;
5. Comprimento da folha;

6. Cor da folha;
7. Comprimento da haste principal do pendão;
8. Cor da coroa do grão na espiga.

Um paquímetro digital foi usado para medir os caracteres: diâmetro do colmo, da espiga e do sabugo. Para os outros caracteres foram utilizadas a régua de campo graduada de 5 em 5 cm, régua graduada em mm e balança digital.

3.3 Análise dos dados morfo-agronômicos

3.3.1 Estatísticas descritivas

Foram obtidas estimativas de médias, variâncias, desvios-padrão, coeficientes de variação, máximo e mínimo referentes às características consideradas utilizando-se o programa Genes vs 2001.0.0 (Cruz, 2001). As médias dos descritores da coleção núcleo e da amostra da coleção base foram comparadas pelo teste de t e as variâncias pelo teste de Bartlett.

3.3.2 Dissimilaridade genética entre acessos e análise de agrupamento

A dissimilaridade dos dados quantitativos teve por finalidade estudar a divergência genética entre acessos. A dissimilaridade genética foi estimada para os dados morfo-agronômicos pelo procedimento de distância Euclidiana média. O cálculo das distâncias Euclidianas foi feito a partir dos dados padronizados por meio do desvio-padrão da variável x. a distância Euclidiana média ($d_{ii'}$) (Cruz & Regazzi, 1997) foi estimada por:

$$d_{ii'} = \sqrt{\sum_j (x_{ij} - x_{i'j})^2}$$

em que :

$d_{ii'}$ é a distância euclidiana média entre o i -ésimo e o i' -ésimo acesso;

$$x_{ij} = \frac{X_{ij}}{s(X_j)}$$

sendo:

X_{ij} é a observação no i -ésimo acesso ($i = 1, 2, \dots, m$) ; em referência ao j -ésimo descritor estudado ($j = 1, 2, \dots, n$);

$s(X_j)$ é o desvio-padrão dos dados do j -ésimo descritor.

O método de agrupamento utilizado foi o upgma (*unweighted pair group method with arithmetic average*). Este método adota o critério de formação dos grupos utilizando média das distâncias entre todos os pares de acessos que formam cada grupo. A distância intergrupo é a média das distâncias dos pares dos membros dos dois grupos (Dias, 1998).

Estas análises foram executadas no programa NTSyspc 2.02k (Rohlf, 1998).

A análise dos componentes principais consiste em transformar um conjunto de v variáveis em um novo conjunto, que são funções lineares das variáveis originais e independentes entre si. Esse novo conjunto de variáveis tem a função de reter o máximo da informação da variação total em poucos componentes, tendo-se demonstrado como uma técnica viável no estudo da divergência genética. Análises de componentes principais foram conduzidas para os caracteres morfológicos e agrônômicos conjuntamente utilizando o programa Genes.

Para descartar as variáveis (descritores) que contribuíram pouco para a discriminação dos acessos utilizou-se o critério de contribuição relativa para divergência genética, segundo Singh (1981) conforme programa Genes.

Para avaliar o método de agrupamento verificou-se o grau de concordância entre o dendrograma e a matriz de distâncias Euclidianas. Esta avaliação é dada pelo coeficiente de correlação cofenética. A matriz cofenética é a matriz de distâncias entre os pares de acessos, constituída de elementos que representam os níveis de fusão verificados no dendrograma. Os valores da correlação cofenética acima de 0,80 indicam um bom ajuste entre as duas matrizes. Para realizar esta análise, utilizou-se o módulo *Clustering/ Coph* do programa NTSyspc 2.02k.

3.4 Caracterização molecular

Para a caracterização molecular dos acessos de milho da coleção núcleo e base foi utilizada a técnica de AFLP.

Etapa 1: Obtenção do material vegetal

O plantio das sementes dos acessos foi realizado em canteiros e amostrados 100 indivíduos de cada acesso, sendo a terceira folha de cada plântula coletada 20 dias após plantio. O conjunto de 100 folhas foi embalado em gaze e papel alumínio e durante o período de coleta, os materiais foram mantidos no gelo, em caixas de isopor, e imediatamente após imersos em nitrogênio líquido e armazenados em ultra freezer -80° C, até que o DNA fosse extraído.

Etapa 2: Isolamento e quantificação de DNA genômico

O DNA foi extraído de um *bulk* de 100 folhas de cada acesso utilizando-se o método de Saghai-Marooft et al. (1984), com modificações.

Foram usadas cerca de cinco gramas de folhas jovens, maceradas em nitrogênio líquido, com o auxílio de pistilo e almofariz, até obtenção de um pó fino. Foram adicionados 10ml de tampão CTAB (CTAB; Tris-HCl 1M pH 7,5; NaCl 5M; EDTA 0,5M pH 8,0; β-mercaptoetanol) e a mistura foi incubada a 65° C por 90 minutos com homogeneização a cada 15 minutos. Após a incubação, adicionou-se 5 ml de clorofórmio/octanol (24:1), homogeneizou-se por 10 minutos e centrifugou-se a 3000 rpm por 10 minutos. Removeu-se o sobrenadante à nova lavagem com clorofórmio. Os ácidos nucleicos foram precipitados adicionando-se 6 ml de isopropanol. Após remoção do isopropanol, adicionou-se ao precipitado 3 ml de TE pH 8,0 (Tris HCl e EDTA) e 20 µl de RNase 10 mg/ml, o qual ficou incubado a 37° C por uma hora. O DNA genômico foi precipitado adicionando-se 6 ml de etanol e transferido para 2 ml de etanol 70% por 10 minutos. O DNA precipitado foi ressuspenso em 300 µl de TE pH 8,0.

A quantificação da concentração de DNA foi realizada em espectrofotômetro através de leituras a 260 nm, empregando-se a fórmula:

$$[DNA] = \frac{A_{260} \times 50 \times 100}{1000}$$

em que:

A_{260} é a leitura da absorbância a 260 nm, cujo comprimento de onda é da região ultra-violeta;

50 é o fator de OD (valor de leitura ótica) para o DNA de fita dupla;

100 é o fator de diluição;

1000 é o fator de correção das unidades para o volume dado em µl.

Posteriormente, a solução estoque de DNA foi armazenada a -20°C e a solução de DNA diluída para a concentração de trabalho de 25 ng/µl e mantida a 4° C.

Etapa 3: Marcadores AFLP

Para o desenvolvimento desta técnica foram empregados *kits* da GIBCO (Rockville, MD, EUA) para a etapa de restrição do DNA genômico com as enzimas *EcoRI* e *MseI* e ligação dos adaptadores (Figura 1). Para as etapas de amplificação pré-seletiva e seletiva utilizaram-se os *kits* da Invitrogen (Carlsbad, Califórnia, EUA). Na amplificação seletiva, os *primers* referentes à região de corte da *EcoRI* foram marcados com os fluorocromos 6-FAM, HEX e NED, para que durante a eletroforese os fragmentos emitissem a fluorescência azul, verde e amarela, respectivamente.

A reação de AFLP foi preparada de acordo com o manual de AFLP (AFLP, 1997) com modificações e utilizadas seis combinações de conjuntos de *primers*: *MseI* CAG e *EcoRI* ACA (6-FAM), *MseI* CAG e *EcoRI* ACG (HEX), *MseI* CAG e *EcoRI* ACC (NED); *MseI* CAG e *EcoRI* ACT (6-FAM), *MseI* CAG e *EcoRI* AGG (HEX), *MseI* CAG e *EcoRI* AGC (NED).

Etapa 3.1: Restrição e ligação dos adaptadores: Adicionaram-se 10 µl da solução de trabalho de DNA na concentração de 25ng/µl aos reagentes constituintes do *kit* GIBCO: 5 µl do tampão de reação 5X, 2 µl *EcoRI/MseI* e 8 µl de água destilada autoclavada para completar um volume final de 25 µl. Após misturar suavemente e centrifugar brevemente, incubou-se a mistura a 37° C por duas horas e depois a 65° C por 20 minutos para inativar as enzimas de restrição. Colocou-se o tubo no gelo para resfriar, fez-se uma breve centrifugação e adicionaram-se 24 µl da solução adaptador/ligação e 1 µl de T₄ DNA ligase. Misturou-se suavemente, centrifugou-se e incubou-se a 20° C por duas horas. Para uma diluição 1:10 da mistura de ligação adicionaram-se 90 µl de tampão TE e 10 µl da mistura da reação. Essa mistura de ligação diluída foi usada para a reação de pré-amplificação. A porção não usada da mistura de reação foi armazenada a -20° C.



Figura 1. Representação esquemática da técnica de AFLP (*Amplified Fragment Length Polymorphism*, Polimorfismo de Comprimento de Fragmentos Amplificados) (Fonte: Perkin Elmer, 1996). Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

Etapa 3.2: Amplificação pré-seletiva das seqüências específicas: Para a reação de PCR adicionaram-se 15 µl de AFLP *Core Mix*, 1 µl de AFLP *EcoRI* +A e *MseI* +C *primer* de pré-amplificação e 4 µl da solução de DNA diluído preparados na etapa de restrição e ligação dos adaptadores, perfazendo-se um volume final de 20 µl. As misturas foram submetidas ao termociclador, em programa de 72° C por 2 min, mais 20 ciclos sendo cada um de: 94° C por 20 seg, 56° C por 30 seg, 72° C por 2 min, seguido de 60° C por 30 min e 4° C por tempo indeterminado.

Preparou-se a diluição do produto da reação da amplificação pré-seletiva usando-se 95 µl de tampão TE_{0,1} e 5 µl do produto da reação de DNA .

Etapa 3.3: Amplificação seletiva dos fragmentos pré-amplificados: Para a reação de PCR adicionaram-se 2,85 µl de água destilada e autoclavada, 1 µl de tampão PCR 10X sem magnésio, 0,3 µl de cloreto de magnésio (MgCl₂) 50 mM, 0,8 µl de dNTPs (2,5 mM cada) , 0,5 µl de *primer MseI* CAG a 5 µM, 0,5 µl de *primer Eco RI* ACA a 5 µM, 0,5 µl de *primer Eco RI* ACC a 5 µM, 0,5 µl de *primer Eco RI* ACG, 0,05 µl da enzima Taq polimerase e 4 µl da solução de DNA, perfazendo um volume final de 10 µl. Nesta etapa os ciclos para amplificação dos fragmentos foram: dez ciclos de 94° C por 2 min, 94° C por 20 seg, 66° C por 30 seg, 72° C por 2 min, 20 ciclos de 94° C por 20 seg, 66° C por 30 seg, 72° C por 2 min, com uma extensão final a 60° C por 30 min e 4° C por tempo indeterminado.

Etapa 3.4: Preparo das amostras para carregamento no gel: Adicionou-se 1 µl de formamida, 0,1 µl de Rox 500 (marcador padrão de peso molecular), 0,4 µl de corante *Blue* Dextran, 0,8 µl do produto da amplificação seletiva com o *primer* marcado com 6-FAM, 1 µl do HEX e 1,4 µl do NED. Centrifugou-se, desnaturou-se a amostra a 95° C por três minutos e colocou-a em gelo. Foram carregados 1,5 µl da amostra por canaleta em gel de

poliacrilamida 5% e a eletroforese dos fragmentos de DNA realizada a 1.388 volts por sete horas. O equipamento utilizado foi o seqüenciador automático ABI PRISM 377 da Perkin Elmer. A extração dos dados obtidos na corrida eletroforética foi automática pelos programas *GeneScan 2.1* e *GenoTyper 2.0* em que foram analisados os tamanhos de fragmentos. O padrão de bandas AFLP utilizando combinações de *primers* com as diferentes fluorescências pode ser visualizado no exemplo dado na Figura 2.

Os perfis das bandas de AFLP em variedades de milho com seus tamanhos, em pb, são apresentados em eletroferogramas (Figura 3) gerados pelo programa *GenoTyper 2.0*. A montagem das tabelas binárias para cada conjunto de *primer* foi semi-automática e conferência com os tamanhos de fragmentos amplificados no gel foi manual.

3.5 Análise dos dados moleculares

3.5.1 Similaridade genética entre os acessos e análise de agrupamento

A análise dos marcadores AFLP foi realizada tendo como base uma matriz considerando os dados binários, onde o valor 1 (um) foi atribuído para a presença do fragmento (banda) polimórfico e o valor 0 (zero) para a sua ausência.

As estimativas de similaridades genéticas entre cada par de acessos foram determinadas por sete coeficientes diferentes (CS: Coincidência Simples; RT: Rogers e Tanimoto; RR: Russel e Rao; J: Jaccard; D: Dice; O: Ochiai; OII: Ochiai II). Esses coeficientes comparam dois acessos quaisquer possibilitando a construção de uma tabela de contingência 2 x 2, utilizando-se das frequências de bandas observadas no gel, conforme o esquema (Dias, 1998):

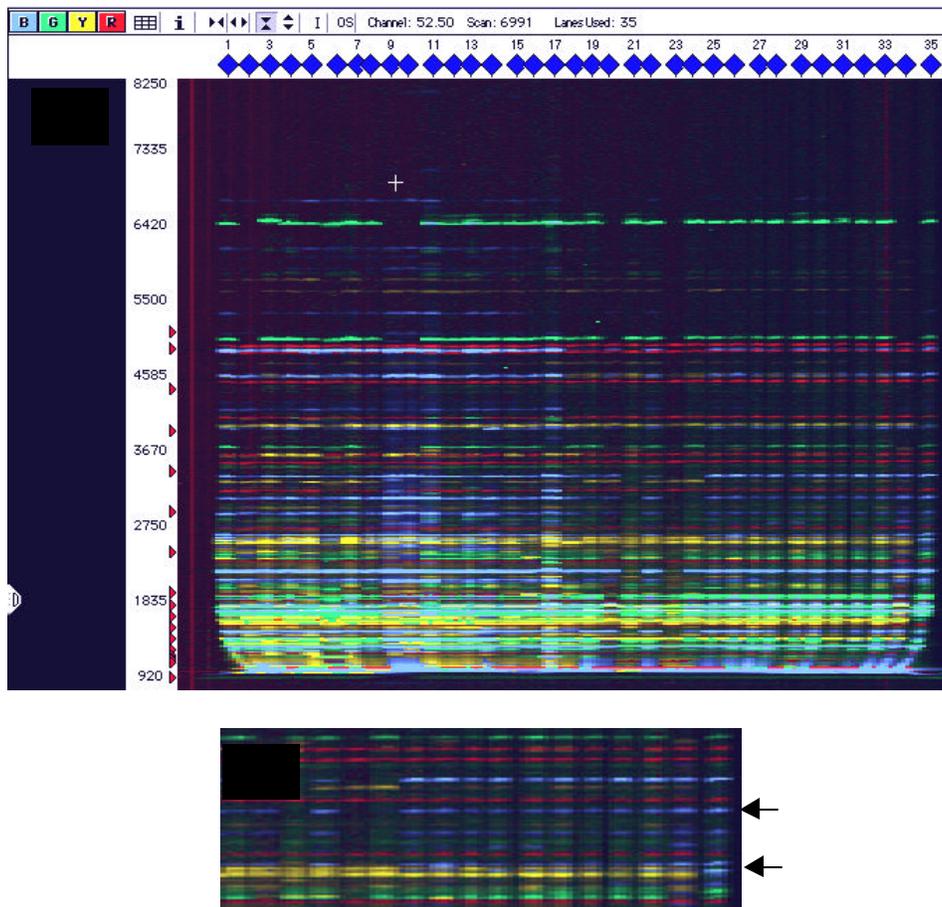


Figura 2. Imagem de gel de poliacrilamida obtida utilizando-se sequenciador de DNA ABI 377. Os fragmentos em azul, verde e amarelo foram gerados através da amplificação seletiva de AFLP com conjuntos de *primers* marcados (A). Detalhe da amplificação dos fragmentos mostrando o polimorfismo gerado pelos diferentes conjuntos de *primers*, nos acessos de milho. As setas indicam algumas das bandas polimórficas. (B). Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

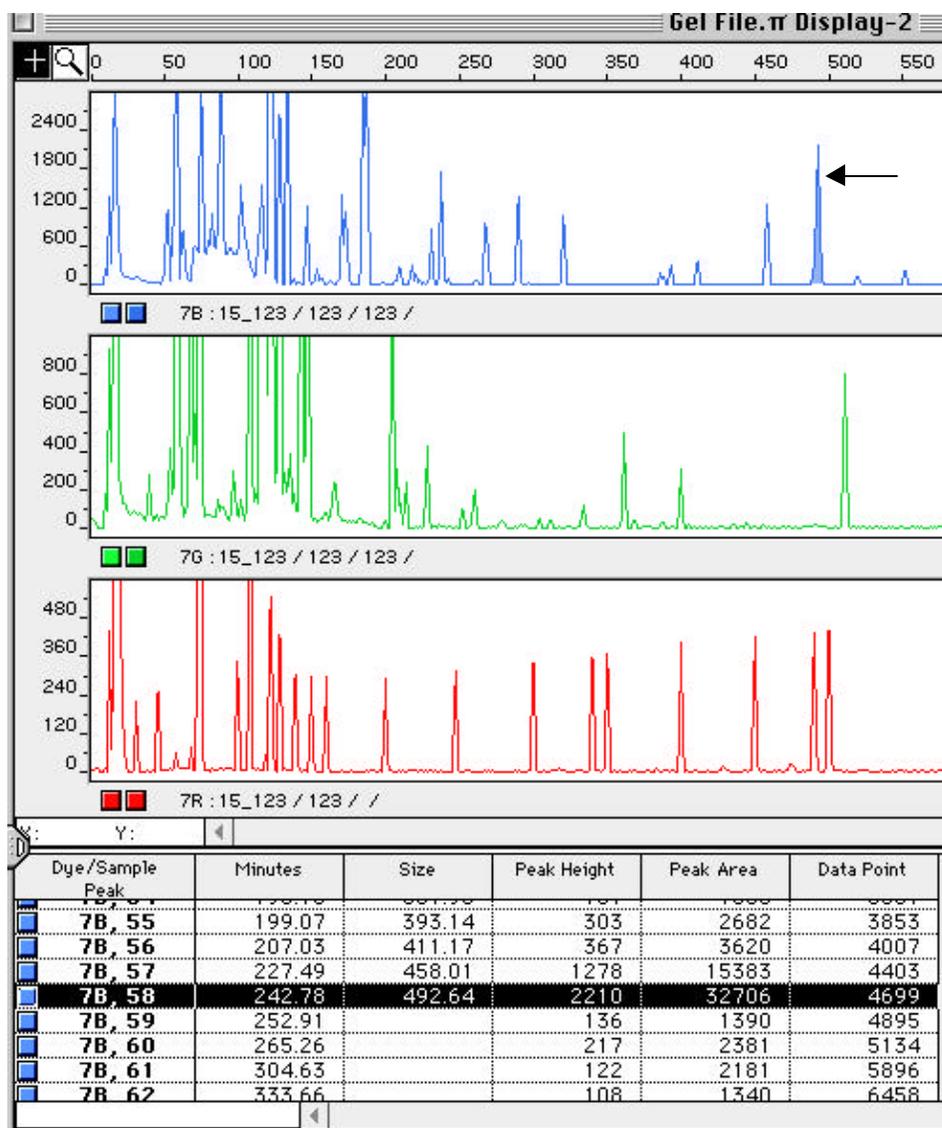


Figura 3. Eletroferogramas gerados pelo programa *GeneScan* para uma mesma amostra de DNA genômico, com duas combinações de *primers* seletivos de AFLP. Eletroforese realizada no sequenciador de DNA ABI 377. A seta indica o pico preenchido em azul que corresponde ao fragmento amplificado de tamanho de 492,64 pb (Ver na tabela, a linha destacada em preto). O tamanho do fragmento foi calculado pelo programa que gera uma curva de calibração, baseada nos padrões de peso molecular do Rox 500, eletroferograma em vermelho. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

		Acesso 1	
		Presença (1)	Ausência (0)
Acesso 2	Presença (1)	a	b
	Ausência (0)	c	d

sendo: a = presença da banda em ambos os acessos;

b = presença da banda no acesso 2 e ausência em 1;

c = ausência da banda no acesso 2 e presença em 1;

d = ausência da banda em ambos os acessos.

As estimativas de similaridades genéticas foram obtidas com a utilização do programa NTSyspc 2.02k (Rohlf, 1998).

A comparação entre os coeficientes de similaridade foi realizada pela correlação de Spearman. Optou-se pelo coeficiente de Jaccard por ser um coeficiente simples, grandemente utilizado e preferido em estudos de biologia na comparação de populações dentro da mesma espécie (Dias, 1998). A expressão

do coeficiente de Jaccard (S_j) é:
$$S_j = \frac{a}{(a+b+c)}$$

Para a construção dos dendrogramas foi utilizado o método da média aritmética entre pares não ponderados (UPGMA), utilizando a distância dada pelo coeficiente de similaridade de Jaccard, sendo cada acesso denominado uma unidade taxonômica operacional. A comparação dos dendrogramas foi feito pelo índice de ramificação de consenso (CIc) obtido a partir de dendrogramas comparados dois a dois. Este índice fornece uma estimativa relativa da semelhança dos dendrogramas e foi usada a opção *Clustering/ Consen* com a interpretação de acordo com Rohlf (1998).

Os erros associados a cada similaridade (s_{sg}) foram estimados de acordo com a expressão (Skroch et al., 1992):

$$s_{sg} = [sg_{ij}(1 - sg_{ij}) / (n - 1)]^{1/2}$$

em que:

sg_{ij} é a similaridade genética entre os acessos i e j ;

n é o número total de combinações a , b e c utilizadas na estimativa da similaridade genética.

Os acessos geneticamente diferentes foram identificados no dendrograma a partir da estimativa do valor mínimo de similaridade, acima do qual os acessos são semelhantes, ou o valor máximo significativo de similaridade (sg_{max}). O sg_{max} foi estimado por meio do teste de t , em nível de 1% de probabilidade pela expressão (Hagiwara et al., 2001):

$$sg_{max} = 1 - (t \cdot \bar{s}_{sg})$$

em que:

t é o valor tabelado de t com $n-2$ graus de liberdade;

\bar{s}_{sg} é o erro médio das similaridades genéticas consideradas no dendrograma.

O índice de eficiência da técnica de AFLP (ET_{AFLP}) foi calculado conforme Pejic et al. (1998):

$$ET_{AFLP} = \frac{nlocos \cdot ne}{cp}$$

em que:

$nlocos$ é o número total de locos;

ne é o número efetivo de alelos;

cp é o número de combinações de *primers* AFLP utilizado.

3.5.2 Diversidade genética

i) Índice de diversidade genética e frequência alélica

A partir dos dados tabulados foi possível calcular a frequência de cada alelo para cada um dos locos, sendo a ausência da banda identificada por 0 e a sua presença por 1. Cada acesso foi considerado como um indivíduo. As frequências alélicas foram utilizadas para avaliar o conteúdo de informação polimórfica do loco calculando-se o índice de diversidade genética, também chamada de heterozigosidade esperada (h) ou *Polymorphic Information Content* (PIC) admitindo-se que a população está em Equilíbrio de Hardy-Weinberg (Hedrick, 1999), segundo a fórmula:

$$h = 1 - \sum_{i=1}^m (p_i)^2$$

em que:

m = número de alelos;

p_i = frequência relativa do i -ésimo alelo em um loco.

O intervalo de variação da heterozigosidade para dois alelos é de 0 a 0,5.

ii) Diversidade genética de Nei

Nei (1987) menciona a partição hierárquica da diversidade genética em seus componentes entre e dentro das populações. A diversidade total será H_T , e é constituída por:

$$H_T = H_S + D_{ST}$$

em que:

$$H_T = 1 - \sum_{i=1}^m \bar{p}_i^2 \quad \text{em que:}$$

$$\bar{p}_i = (N_1 p_{i1} + N_2 p_{i2}) / (N_1 + N_2)$$

H_S é o componente de diversidade dentro da população ou coleção de acessos dado pela fórmula:

$$H_S = \frac{[(N_1 \cdot H_1) + (N_2 \cdot H_2)]}{(N_1 + N_2)}$$

sendo:

N_1, N_2 o número de acessos da população (coleção) 1 e 2;

H_1, H_2 a diversidade estimada em cada população;

D_{ST} é o componente entre coleções, dentro da espécie estudada, dado pela diferença entre H_T e H_S .

A proporção da diversidade genética, que é atribuída ao componente entre coleções, é estimada por G_{ST} :

$$G_{ST} = \frac{D_{ST}}{H_T}$$

Para o marcador AFLP, cada fragmento de DNA foi considerado como um loco (banda) independente no genoma do milho.

iii) Número efetivo de alelos (N_e)

Esta é uma medida de diversidade que representa o número médio de alelos por loco polimórfico (Nei, 1987; Hedrick, 1999), que é estimada por:

$$N_e = \frac{1}{\sum_{i=1}^m (p_i)^2}$$

em que:

m = número de alelos;

p_i = frequência relativa do i -ésimo alelo em um loco.

O intervalo de variação do número efetivo de alelos para dois alelos é de 1 a 2 e muda conforme o número de alelos considerados.

iv) Distância genética de Nei e Identidade de Nei

A distância genética de Nei (D_N) é a mais conhecida das medidas de distância genética e leva em conta tanto os locos polimórficos como monomórficos. Os valores de D_N são calculados a partir de uma amostra da população, portanto eles variam entre amostras. Alguns fatores podem influenciar D_N como o tamanho de amostras, número de locos estudados e a quantidade de divergência genética entre as populações. É definida como:

$$D_N = -\ln(I_N)$$

em que:

I_N é a identidade genética de Nei, que expressa a probabilidade de um dado alelo de um loco, tomado ao acaso em duas diferentes populações, seja idêntico em relação à probabilidade de que dois alelos do mesmo loco, tomados também ao acaso em cada população, sejam também idênticos. É usada para quantificar a semelhança genética entre duas populações e mede a correlação entre frequências alélicas em cada loco em duas populações. É definida como:

$$I_N = \frac{J_{12}}{\sqrt{(J_1 J_2)}}$$

em que:

$$J_{12} = \sum_{i=1}^m p_{i1} p_{i2} \quad ; \quad J_1 = \sum_{i=1}^m p_{i1}^2 \quad ; \quad J_2 = \sum_{i=1}^m p_{i2}^2$$

em que:

p_{i1}^2 e p_{i2}^2 são as frequências do alelo i nas coleções 1 e 2, respectivamente.

I_N varia de 0 a 1, enquanto que D_N varia de 0 a infinito.

v) Índice de Shannon

O índice de diversidade de Shannon-Weaver foi calculado para comparar a diversidade genotípica entre as coleções. Esse índice é usado em estudos de recursos genéticos como uma medida conveniente de riqueza e uniformidade alélica quando se usam dados genéticos. É definido por:

$$H' = - \sum_{i=1}^n P_i \log(P_i)$$

em que:

n é o número de classes genotípicas;

P_i é a frequência de acessos na classe i .

Cada valor de H' foi normalizado pela divisão por $\log(n)$ para enquadrar o valor do índice no intervalo $0 \leq H' \leq 1$. Um baixo valor de H' indica menor diversidade genética (Dias, 1998).

O índice de diversidade genética de Nei, frequência alélica, número efetivo de alelos, diversidade genética total, dentro e entre as coleções, distância

genética de Nei e identidade de Nei e índice de Shannon foram estimados com a utilização do programa PopGene versão 1.31 (Yeh, 1999).

3.5.3 Análise de variância molecular (AMOVA)

A análise de variância molecular (AMOVA) foi realizada de modo a se determinar a variância entre e dentro das coleções, conforme o método proposto por Excoffier et al. (1992) e utilizando-se o programa Arlequin versão 2.000 (Schneider et al., 2000). Este método utiliza a soma dos quadrados dos desvios entre todos os pares de indivíduos para construir uma análise hierárquica da variância molecular. A análise de variância molecular é feita diretamente da matriz das distâncias Euclidianas entre todos os pares de acessos. A soma dos quadrados totais correspondentes às distâncias genéticas pode ser particionada em componentes que representam as somas dos quadrados dentro e entre coleções de germoplasma (Tabela 4).

Procedimentos permutacionais foram usados na matriz original das distâncias para subsidiar testes de significância para cada componente de variância e relacionados com a estatística F de Wright. O coeficiente F_{ST} representa a correlação de acessos entre os marcadores AFLP tomados ao acaso dentro de uma coleção de germoplasma, em relação a acessos tomados ao acaso em toda população. Esta estatística pode ter valores entre 0 e 1; altos valores indicam grandes particionamentos da população em subgrupos.

Tabela 4. Esquema da análise de variância molecular (AMOVA). Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

Fonte de variação	GL	SQ	E(QM)
Entre coleções	c - 1	SQ _{Entre coleções}	$n\mathbf{s}_a^2 + \mathbf{s}_b^2$
Dentro de coleções	a - c	SQ _{Dentro de coleções}	\mathbf{s}_b^2
Total	a - 1	SQ _{Total}	\mathbf{s}_T^2

c : número de coleções (c = 2);

a : número de acessos;

n : valor ponderado em função do número diferente de acessos de cada coleção (Excoffier et al., 1992), sendo:

$$n = \frac{N - \sum \frac{N_p^2}{N}}{c - 1}$$

\mathbf{s}_a^2 : componente de variância entre as distâncias genéticas das coleções;

\mathbf{s}_b^2 : componente de variância entre as distâncias genéticas dos acessos dentro das coleções.

4 RESULTADOS E DISCUSSÃO

4.1 Caracterização morfo-agronômica

4.1.1 Análise descritiva

As análises descritivas dos caracteres morfo-agronômicos quantitativos mostram a variabilidade genética encontrada nos acessos de milho tipo de endosperma duro da amostra da coleção base e coleção núcleo (Tabela 5). Apesar dos caracteres avaliados serem muito diferentes e com diversas unidades de medida, a variabilidade fenotípica foi semelhante observando-se os descritores individualmente. O maior valor para o coeficiente de variação foi de 26,95% e 25,14% para peso de grãos por espiga nas duas coleções base e núcleo, respectivamente, e o menor valor foi de 8,53% e 6,09% para folhas acima da espiga (FA) na amostra da coleção base e coleção núcleo, respectivamente.

Observando os valores de coeficiente de variação, verifica-se que as maiores variações dos 17 caracteres avaliados na amostra da coleção base foram peso de grãos por espiga (PG), peso de espiga (PE), peso de mil grãos (PMG), altura da espiga (AE) e número de grãos por fileira (NG).

Para os acessos da coleção núcleo, os caracteres morfo-agronômicos de maior variação foram peso de grãos por espiga (PG), peso de espiga (PE), ramificação do pendão (RP) e altura da espiga (AE). As menores variações foram apresentadas pelas variáveis número de folhas acima da espiga (FA), diâmetro do colmo (DC), altura da planta (AP).

Tabela 5. Estatísticas descritivas dos caracteres morfo-agronômicos avaliados na amostra da coleção base e coleção núcleo de milho. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

Caracter	Estimador	Amostra da Coleção Base	Coleção Núcleo	Teste de t	Teste de Bartlett
Floração masculina (número de dias)	Média	70,60	64,08	4,39**	
	Mínimo	49,00	54,00		
	Máximo	94,00	79,00		
	Des. Padrão	9,71	5,80		
	C. Variação	13,73	9,05		
	Variância	93,99	33,66		
Floração feminina (número de dias)	Média	74,74	68,24	4,06*	
	Mínimo	55,00	58,00		
	Máximo	104,0	82,00		
	Des. Padrão	10,42	6,33		
	C. Variação	13,94	9,28		
	Variância	108,62	40,12		
Ramificação do pendão	Média	23,47	24,22	1,09 ^{NS}	
	Mínimo	14,00	16,60		
	Máximo	32,00	34,70		
	Des. Padrão	3,65	3,91		
	C. Variação	15,54	16,15		
	Variância	13,31	15,33		
Altura da planta (cm)	Média	291,50	323,60	5,91**	
	Mínimo	209,00	263,50		
	Máximo	344,00	362,50		
	Des. Padrão	33,64	24,02		
	C. Variação	11,54	7,42		
	Variância	1131,80	577,04		
Altura da primeira espiga (cm)	Média	168,86	193,09	4,32**	
	Mínimo	100,00	132,00		
	Máximo	221,00	237,50		
	Des. Padrão	30,96	29,43		
	C. Variação	18,34	15,24		
	Variância	958,65	866,20		

Continua ...

Tabela 5. Continuação

Caracter	Estimador	Amostra da Coleção Base	Coleção Núcleo	Teste de t	Teste de Bartlett
Folhas acima da primeira espiga	Média	5,74	5,78	0,76 ^{NS}	*
	Mínimo	5,00	5,10		
	Máximo	7,00	6,50		
	Des. Padrão	0,49	0,35		
	C. Variação	8,53	6,09		
	Variância	0,24	0,12		
Número total de folhas	Média	14,55	15,49	3,31 ^{**}	NS
	Mínimo	12,00	12,90		
	Máximo	18,00	24,70		
	Des. Padrão	1,34	1,66		
	C. Variação	9,19	10,76		
	Variância	1,79	2,78		
Diâmetro do colmo (mm)	Média	22,83	22,95	0,30 ^{NS}	**
	Mínimo	15,00	18,80		
	Máximo	29,00	27,30		
	Des. Padrão	2,75	1,58		
	C. Variação	12,06	6,88		
	Variância	7,58	2,50		
Número de espigas por planta	Média	1,06	1,05	0,31 ^{NS}	NS
	Mínimo	1,00	1,00		
	Máximo	1,60	1,60		
	Des. Padrão	0,12	0,12		
	C. Variação	11,40	11,52		
	Variância	0,017	0,015		
Comprimento da espiga (cm)	Média	12,67	13,49	2,62*	**
	Mínimo	8,00	9,90		
	Máximo	17,00	17,10		
	Des. Padrão	1,87	1,38		
	C. Variação	14,77	10,29		
	Variância	3,50	1,93		
Diâmetro espiga (mm)	Média	42,02	38,77	4,74 ^{**}	*
	Mínimo	31,0	29,50		
	Máximo	50,0	46,80		
	Des. Padrão	4,23	3,03		
	C. Variação	10,07	7,82		
	Variância	17,91	9,20		

Continua ...

Tabela 5. Continuação.

Caracter	Estimador	Amostra da Coleção Base	Coleção Núcleo	Teste de t	Teste de Bartlett
Número de fileiras de grãos na espiga	Média	13,51	12,99	2,03*	
	Mínimo	10,00	10,60		
	Máximo	18,00	16,60		
	Des. Padrão	1,46	1,27		
	C. Variação	10,84	9,81		
	Variância	2,15	1,63		
Número de grãos por fileira	Média	30,81	32,53	1,98 ^{NS}	
	Mínimo	20,00	21,30		
	Máximo	44,00	41,10		
	Des. Padrão	5,08	4,23		
	C. Variação	16,50	13,03		
	Variância	25,84	17,97		
Peso de espiga (g)	Média	127,95	109,62	3,32**	
	Mínimo	39,00	51,30		
	Máximo	191,00	175,50		
	Des. Padrão	33,34	25,60		
	C. Variação	26,05	23,35		
	Variância	1111,31	655,58		
Peso de grãos por espiga (g)	Média	108,05	91,13	3,32**	
	Mínimo	51,00	38,80		
	Máximo	165,00	150,50		
	Des. Padrão	29,12	22,91		
	C. Variação	26,95	25,14		
	Variância	848,08	525,11		
Diâmetro do sabugo (mm)	Média	24,36	23,76	1,26 ^{NS}	
	Mínimo	20,00	18,60		
	Máximo	30,00	31,70		
	Des. Padrão	2,45	2,61		
	C. Variação	10,05	11,00		
	Variância	5,99	6,84		
Peso de mil grãos (g)	Média	276,24	242,34	3,61**	
	Mínimo	159,00	168,80		
	Máximo	425,00	319,30		
	Des. Padrão	62,23	35,15		
	C. Variação	22,53	14,50		
	Variância	3872,99	1235,96		

* Significativo a 5%; ** Significativo a 1%; ^{NS} Não significativo.

O teste F, aplicado para testar as variâncias das duas coleções em relação a cada descritor detectou diferenças significativas para nove dos 17 descritores morfo-agronômicos.

O teste de t não mostrou diferenças significativas entre as médias da coleção núcleo e base para os seguintes descritores: ramificação do pendão, folhas acima da primeira espiga, diâmetro do colmo, número de grãos por fileira, diâmetro do sabugo e número de espigas por planta. Para os outros 11 descritores em um total de 17, houve diferenças significativas a 5% e 1% de probabilidade.

As variâncias de cada descritor da amostra da coleção base e coleção núcleo comparadas pelo teste de Bartlett mostraram-se homogêneas, ou seja, não houve diferenças significativas para oito caracteres morfo-agronômicos em um total de 17: ramificação do pendão, altura da primeira espiga, número total de folhas, número de espigas por planta, número de fileiras de grãos na espiga, número de grãos por fileira, peso de grãos por espiga e diâmetro do sabugo.

Quando se estudou a distribuição de frequência dos acessos das duas coleções em relação a cada descritor, verificou-se pelo teste de χ^2 diferenças significativas em 16 caracteres morfo-agronômicos.

Hu et al. (2000) estabeleceram os seguintes critérios para considerar representativa a coleção núcleo em relação à coleção base: a) não mais que 20% dos descritores com diferenças entre as médias das coleções, significativas a 5% de probabilidade, e b) distribuições em classes dos acessos em relação aos descritores apresentadas pela coleção núcleo não for menor que 80%. Adotando o critério de Hu et al. (2000) e apoiando-se nos resultados dos descritores morfo-agronômicos avaliados, verifica-se que a coleção núcleo não poderia ser considerada como uma amostra representativa da coleção base. É importante ressaltar, no entanto, que os dados morfo-agronômicos da amostra da coleção

base foram obtidos em anos e épocas diferentes, tendo sido possivelmente afetados pelo efeito de ambiente e ocorrência da interação genótipo-ambiente.

Adotando-se o critério de Singh (1981) usado pelo programa Genes para se estudar a contribuição relativa dos caracteres para divergência na coleção base e núcleo, observou-se que os caracteres número de espigas por planta, número de folhas acima da espiga, total de folhas e número de fileiras de grãos apresentaram quase nenhuma contribuição para a divergência genética da coleção base variando entre 0,0002% e 0,0251%.

Quanto as maiores contribuições relativas dos caracteres para a divergência genética dos acessos da amostra da coleção base, o peso de mil grãos foi o de maior importância (45,3%), seguido pela altura de planta (13,24%), peso de espiga (13,0%), altura de espiga (11,21%) e peso de grãos por espiga (9,92%). As menores variações foram em folhas acima da espiga (FA), total de folhas (TF) e diâmetro do sabugo (DS).

Os caracteres que mais contribuíram para a divergência genética da coleção núcleo, segundo o critério de Singh (1981), foram peso de mil grãos (31,0%), altura da espiga (21,7%) e peso de espiga (16,4%). Os caracteres da coleção núcleo número de espigas por planta, número de folhas acima da espiga, número total de folhas, número de fileiras de grãos, número de colmos, comprimento da espiga e diâmetro do colmo contribuíram com menos de 0,04% para a divergência genética, portanto foram descartados. Assim, foram descartadas as variáveis número de espiga por planta (EP), folhas acima da espiga (FA), número total de folhas (TF) e número de fileiras de grãos (NF), pois tiveram uma contribuição pequena para a divergência genética dos acessos da amostra da coleção base.

Ao se utilizar técnicas multivariadas no estudo de divergência genética, Cruz e Regazzi (1997) e Ribeiro Júnior (2001) mencionaram a possibilidade de se descartarem as variáveis, que contribuíram pouco para a discriminação de

materiais, tornando-se caracteres dispensáveis. Ferreira (1993) avaliando a utilização de técnicas multivariadas para estimar a divergência genética em 28 variedades de milho e 19 caracteres morfo-agronômicos concluiu que três variáveis foram redundantes (comprimento do entrenó, largura da folha e número de grãos por fileira) e, portanto descartadas. A prática de descarte de variáveis também foi usada por Santos et al. (1995) num estudo de seleção de descritores de germoplasma de guandu e por Veasey et al. (2001) ao caracterizar acessos de *Sesbania*.

Quanto aos caracteres morfo-agronômicos qualitativos, alguns não tiveram variação na avaliação das plantas dos acessos da coleção núcleo. Por exemplo, a arquitetura da planta foi normal em todos os acessos não se encontrando planta anã; a cor da folha foi considerada verde médio para todos os acessos; o comportamento da folha foi sempre recurvado; a forma da primeira folha foi sempre reta; e o arranjo dos grãos foi reto, levemente curvo para todas as espigas. O caráter cor da coroa do grão foi avaliado apenas nos acessos da coleção núcleo, sendo que 40% apresentou cor laranja, 43% amarela e 17% branca.

De maneira geral, os resultados da avaliação dos caracteres qualitativos de sanidade e tipo de espiga, porcentagem de plantas quebradas e acamadas foram semelhantes entre os acessos das coleções base e núcleo (Tabela 6). Observa-se que a qualidade e sanidade da espiga foram elevadas com predominância de notas 7 a 9 que correspondem a espigas boas, bem formadas e sadias, com poucos grãos doentes a completamente sadias. Nota-se que houve concordância de resultados quanto ao tipo de espiga, sendo que a maioria das espigas apresentavam o formato cilíndrico. As maiores porcentagens de plantas quebradas e acamadas apareceram na classe até 25 plantas. Desses caracteres qualitativos, os que mais variaram entre as duas coleções foram qualidade e sanidade da espiga evidenciando que estas variáveis são altamente subjetivas,

Tabela 6. Caracteres morfo-agronômicos qualitativos dos acessos da amostra da coleção base e coleção núcleo. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

Descritor	Escala*	Coleção Base (%)	Coleção Núcleo (%)
Qualidade da espiga	9	46	13
	8	46	53
	7	8	16
	6	0	15
Sanidade da espiga	9	18	19
	8	68	40
	7	14	29
	6	0	12
Tipo de espiga	Cônica	0	0
	Cilíndrica	64	66
	Côn/Cil	36	34
Plantas Quebradas (%)	0-25	87	78
	26-50	13	3
	51-75	0	12
	76-100	0	7
Plantas Acamadas (%)	0-25	87	64
	26-50	13	14
	51-75	0	14
	76-100	0	8

*9: completamente bem formada e sadia; 8: 85% bem formada e sadia; 7: 70% bem formada e sadia; 6: 55% bem formada e sadia

principalmente devido a amplitude das escalas (1 a 9), difícil classificação e dependentes das influências dos genótipos, ambientes e suas interações, o que infere em maior variação.

4.1.2 Dissimilaridade genética e agrupamento

A análise de componentes principais utilizada para avaliar a divergência genética entre os acessos foi realizada com 13 caracteres morfo-agronômicos da amostra da coleção base. As variáveis número de espiga por planta, folhas acima da espiga, número total de folhas e número de fileiras de grãos foram descartadas por contribuírem pouco para a variação total.

A variância total dos dados morfo-agronômicos da coleção base foi decomposta em três componentes principais. O primeiro componente principal concentrou 33,99% do total da variância, o segundo 23,95% e o terceiro 16,35% (Tabela 7). A variância acumulada dos três primeiros componentes principais foi de 74,29%, resultado que mostra relativamente alta porcentagem da variação total das variáveis originais.

A análise dos componentes principais dos acessos da coleção núcleo foi realizada com 14 variáveis morfo-agronômicas e demonstrou que o primeiro componente concentrou 33,65% da variância total, o segundo 27,57% e o terceiro 9,35%, explicando assim 70,57% da variância total.

Tabela 7. Estimativas da variância (% acumulada) dos três primeiros componentes principais para os caracteres morfo-agronômicos avaliados nos acessos da amostra da coleção base e coleção núcleo, respectivamente. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

CP	Amostra da Coleção Base (CB)	Coleção Núcleo (CN)	CB e CN
CP1	33,99	33,65	29,16
CP2	57,94	61,22	48,34
CP3	74,29	70,57	66,75

Estes resultados aproximam-se dos valores citados por Dias (1998), onde 75% a 85% da variância total deve ser explicada pelos dois ou três primeiros componentes principais. A estimativa acumulada de variância de 74,29% encontrada, também aproxima-se do valor considerado por Cruz e Regazzi (1997), de 80%, como indicativo da variância total em estudos de divergência genética. Veasey et al. (2001) encontraram uma variância acumulada nos dois primeiros componentes principais de 74,38% utilizando 17 caracteres morfo-agronômicos em 19 acessos de *Sesbania* spp. Também, Zibumbo e Colunga (2001) em um estudo com 18 populações de coco e 19 descritores morfo-agronômicos aplicados encontraram 76% de variância total para os dois primeiros componentes principais.

Embora os três primeiros componentes principais não tenham absorvido alta proporção da variância total de CB e CN juntas resultando em 66,75%, pode-se notar que grande parte da variabilidade encontrada na amostra da coleção base é contemplada também na coleção núcleo de milho tipo de endosperma duro (Figura 4). Nota-se que a maioria dos acessos estão dispersos de maneira aleatória, ou seja, sem a formação de grupos definidos mostrando que as duas coleções não formam agrupamentos independentes sendo, portanto, partes de um mesmo conjunto.

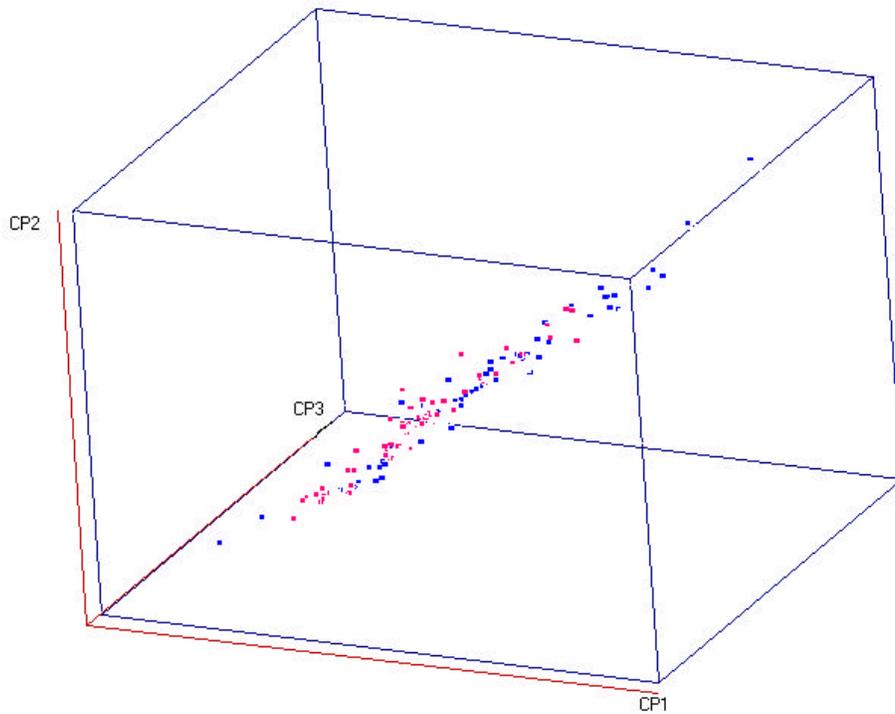


Figura 4. Dispersão dos acessos da coleção base (■) e da coleção núcleo (■) considerando os três primeiros componentes principais. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

Os dendrogramas construídos a partir da medida de dissimilaridade pelo coeficiente de distância euclidiana média com padronização dos dados morfo-agronômicos, e o método de agrupamento UPGMA estão apresentados nas Figuras 5 e 6 para a coleção base e coleção núcleo, respectivamente. Pode-se observar uma estrutura geral semelhante para as duas coleções, em ambos os casos, sem a formação de grupos homogêneos até uma distância correspondente

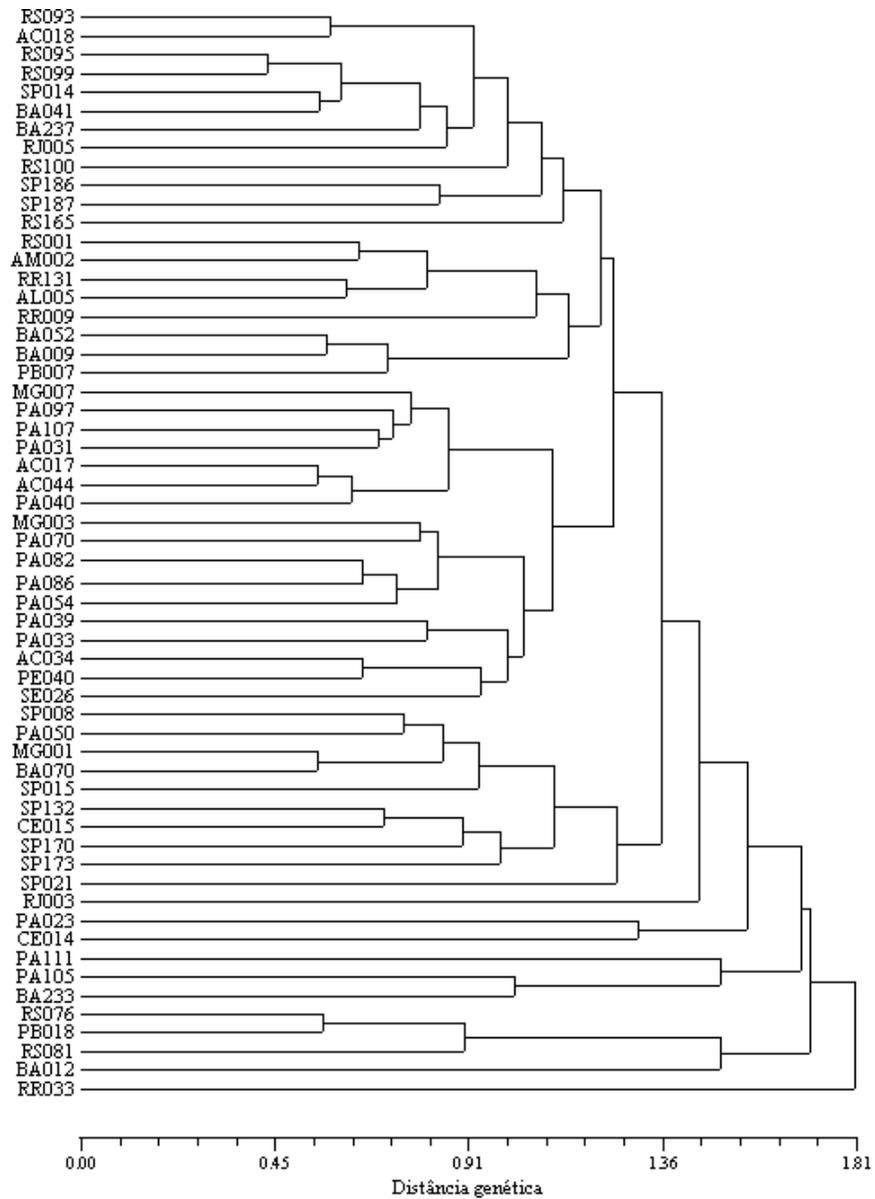


Figura 5. Dendrograma das distâncias genéticas entre os acessos da amostra da coleção base utilizando a distância euclidiana média para os dados morfo-agronômicos e o método de agrupamento de UPGMA. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

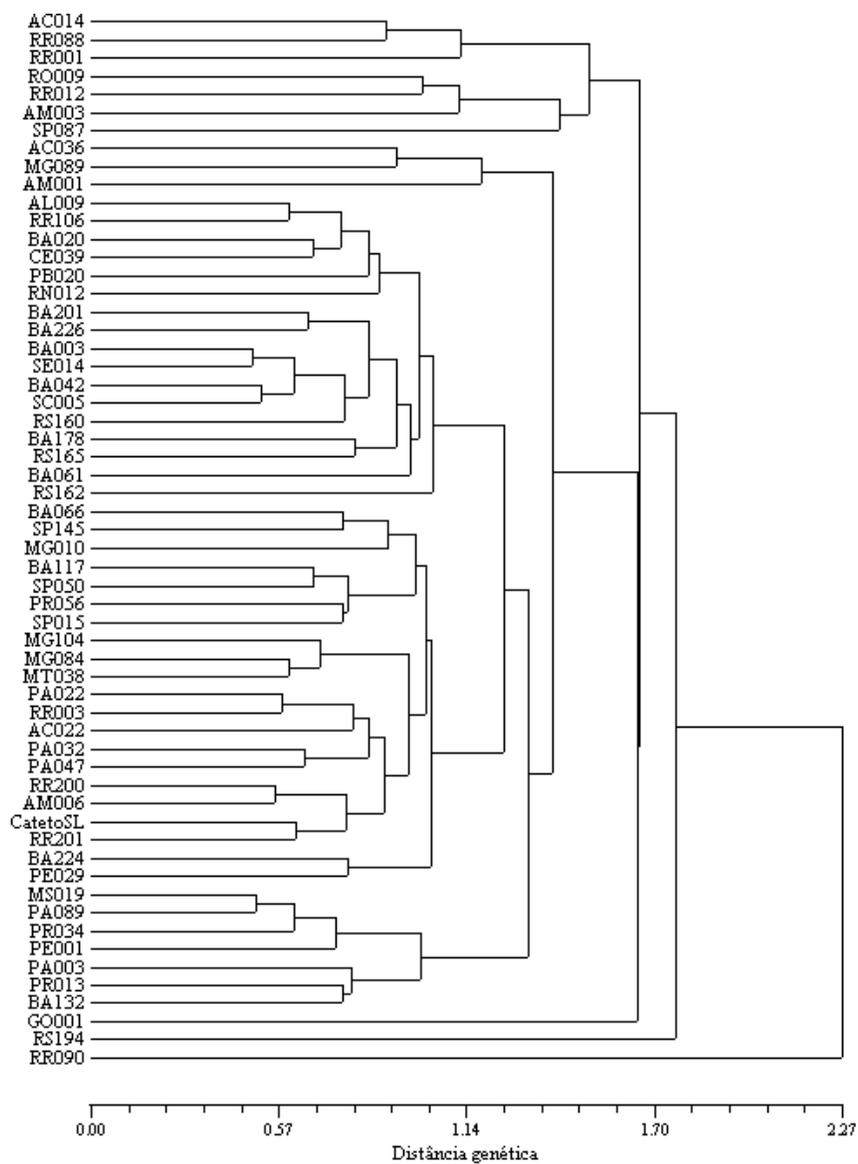


Figura 6. Dendrograma das distâncias genéticas entre os acessos da coleção núcleo utilizando a distância euclidiana média para os dados morfo-agronômicos e o método de agrupamento de UPGMA. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

a 25% da maior distância entre grupos de acessos, evidenciando a não ocorrência de redundância de acessos.

Em relação à amostra da coleção base, as distâncias genéticas entre os acessos variaram de 0,44 a 1,81 sendo a menor distância observada entre os acessos RS095 e RS099 (Figura 5). Observa-se que o acesso RR033 ficou isolado dos demais.

As distâncias genéticas detectadas para a coleção núcleo foram de 0,45 a 2,27. As menores distâncias verificadas foram para os pares de acessos BA003 e SE014, e MS019 e PA089 (Figura 6).

Observa-se pelos dendrogramas que os acessos iniciam o agrupamento a partir de distâncias bem próximas, 0,44 para a coleção base e 0,45 para a coleção núcleo.

Para verificar o grau de concordância entre o dendrograma e a matriz de distâncias euclidiana média foram realizadas as análises de correlação cofenética. A matriz cofenética foi comparada com a matriz das distâncias originais. A correlação entre as matrizes normalizada pela estatística de Mantel foi de 0,71. Este valor, segundo a escala proposta por Mantel (Rohlf, 1998), significa que o grau de ajustamento entre as matrizes foi considerado regular para os acessos da amostra da coleção base. O mesmo foi realizado para os acessos da coleção núcleo, e verificou-se uma correlação de 0,78, aproximando-se do valor de um bom ajustamento. O valor da correlação cofenética de 0,83 indicou um bom ajuste entre as duas matrizes em um estudo feito por Dias (1998) utilizando o método de agrupamento UPGMA para sete populações de arbustos tropicais.

Este tipo de validação entre matrizes também foi utilizado por Mumm et al. (1994). Esses autores chamaram de validação interna correlacionando dados de RFLP em linhagens de milho. O coeficiente da correlação cofenética foi de

0,93 sugerindo que o dendrograma representava as estimativas da relação genética corretamente.

4.2 Caracterização molecular

Os marcadores AFLP demonstraram ser muito informativos sobre a diversidade genética presente nos 58 acessos da coleção núcleo e 21 acessos da coleção base, ambas constituídas de variedades autóctones. Os tamanhos de fragmentos AFLP variaram de 52,43 a 498,79 pares de base (pb) (Tabela 8). As análises dos 58 acessos com seis combinações de *primers* revelaram um total de 207 fragmentos ou bandas, com média de 32,8 bandas polimórficas por combinação de *primer* para a coleção núcleo, correspondendo a 98% de polimorfismo. Para a amostra da coleção base verificou-se média de 25 bandas polimórficas, ou seja, 84,5% de polimorfismo. Dos 207 marcadores AFLP, 150 (76%) ocorreram em ambas as coleções.

O número de fragmentos variou de 22 para a combinação *MseI CAG + Eco RI ACG* para acessos da amostra da coleção base até 50 para a combinação *MseI CAG + Eco RI ACT* para acessos da coleção núcleo. Esta última combinação de *primers* foi particularmente interessante já que conseguiu distinguir 45 fragmentos polimórficos que representam a característica de diferenciação em 58 acessos da coleção núcleo e 34 fragmentos polimórficos em 21 acessos da coleção de base. A combinação de *primers MseI CAG + Eco RI ACG* gerou o menor número de fragmentos nas duas coleções, porém com alta porcentagem de polimorfismo. As combinações *MseI CAG + Eco RI ACA* e *MseI CAG + Eco RI ACC* apresentaram 100% de polimorfismo nos acessos da coleção núcleo. Padilha (2002) estudou em 35 linhagens de milho, 11

Tabela 8. Combinação de *primers* utilizados na técnica de AFLP e número de fragmentos encontrados na coleção núcleo (CN) e amostra da coleção base (CB). Embrapa Milho e Sorgo, Sete Lagoas, MG, 2002.

Combinação de <i>Primers</i>	Amplitude Alélica (pb)	Nº. de Bandas CN			Nº. de Bandas CB		
		Total	Poli- mórfica	%	Total	Poli- mórfica	%
1) <i>MseI</i> CAG + <i>EcoRI</i> ACA	59,76 – 488,69	29	29	100	29	22	75,9
2) <i>MseI</i> CAG + <i>EcoRI</i> ACG	57,53 – 426,06	24	23	95,8	22	20	90,9
3) <i>MseI</i> CAG + <i>EcoRI</i> ACC	81,02 – 460,15	31	31	100	31	24	77,4
4) <i>MseI</i> CAG + <i>EcoRI</i> ACT	52,43 – 498,79	50	45	90,0	44	34	77,3
5) <i>MseI</i> CAG + <i>EcoRI</i> AGG	57,76 – 444,30	35	32	91,4	26	24	92,3
6) <i>MseI</i> CAG + <i>EcoRI</i> AGC	68,92 – 474,48	38	37	97,4	28	26	92,9
Total	52,43 – 498,79	207	197	-	180	150	-
Média	62,90 – 465,41	34,5	32,8	95,8	30	25	84,5

combinações de *primers* AFLP gerando 260 fragmentos, sendo 158 polimórficos e média de 14 bandas polimórficas por combinação de *primers*. Em 114 genótipos da coleção núcleo de feijão e duas combinações de *primers*, Tohme et al. (1996) encontraram em média 102 bandas sendo que 90% eram polimórficas.

O nível de polimorfismo encontrado no presente trabalho também foi superior ao relatado por Pejic et al. (1998), os quais obtiveram 38,7 bandas polimórficas de AFLP utilizando seis combinações de *primers* em linhagens de milho. Já Barrett et al. (1998) encontraram 229 bandas polimórficas de AFLP em 43 cultivares de trigo e Zhu et al. (1999) encontraram um total de 569 fragmentos com oito pares de *primers*, com média de 71 fragmentos por *primer* em 61 acessos de soja cultivada e silvestre, sendo que 48% dos fragmentos eram polimórficos.

O procedimento de simulações sobre o número de marcadores usados em estudos de diversidade genética tem sido executado para se determinar a confiabilidade e a precisão do agrupamento produzido (Pejic et al., 1998). Da mesma forma, foram realizadas simulações utilizando matrizes de distâncias calculadas pelo coeficiente de Jaccard com base em um número crescente de marcadores AFLP, verificando-se uma diminuição no desvio padrão e um aumento na correlação entre as matrizes de distâncias genéticas (Figura 7). A partir de 100 marcadores de AFLP, o desvio padrão foi inferior a 0,05 e a correlação superior a 0,95 para os acessos da amostra da coleção base. Já para a coleção núcleo, a partir de 80 marcadores AFLP, o desvio padrão e a correlação foram constantes em 0,3 e 0,90, respectivamente. Estes valores corresponderiam à utilização de quatro combinações de *primers* mais polimórficos para o AFLP, baseados nos resultados da amostra da coleção base. Estes resultados indicam que as 150 e 197 bandas polimórficas avaliadas na amostra da coleção base e coleção núcleo, respectivamente, foram suficientes para definir, com confiabilidade, a divergência genética entre os acessos. Isto pôde ser verificado por Johns et al. (1997) que encontraram o mesmo agrupamento utilizando 50 e 106 bandas com 69 genótipos de feijão do Chile. Dias (1998) comentou que a utilização de, pelo menos, 30 locos polimórficos são suficientes para a construção de dendrogramas de elevada precisão.

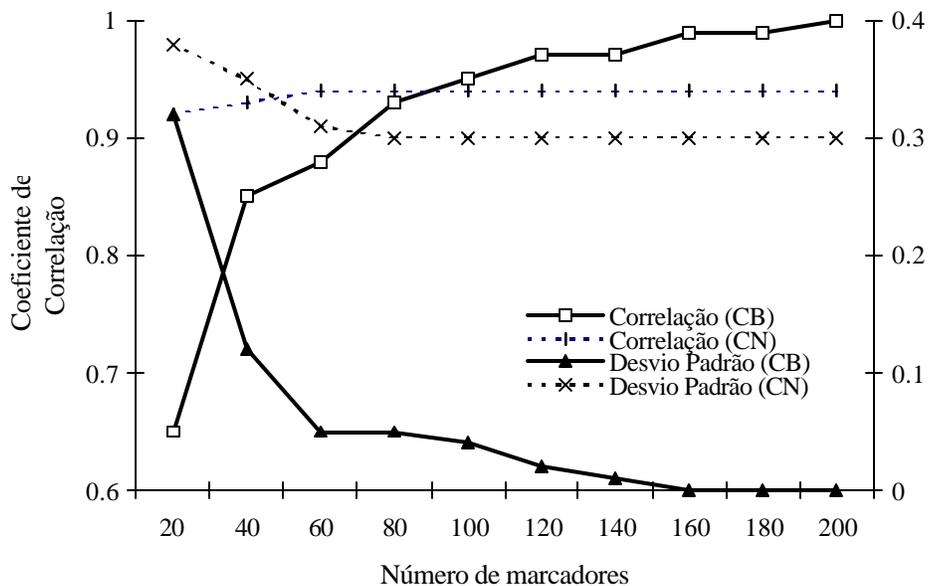


Figura 7. Coeficiente de correlação entre matrizes de distâncias de CN e CB e desvio padrão obtidos por simulações de número de marcadores AFLP para os acessos de milho da amostra da coleção base e da núcleo. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

4.2.1 Similaridade genética e agrupamento

O estudo da similaridade genética entre os acessos foi realizado com sete coeficientes. A distância genética foi gerada pelo complemento destes coeficientes e a comparação entre os coeficientes por meio da correlação de Spearman. As correlações entre as diferentes distâncias genéticas da amostra da

coleção base estão apresentadas na Tabela 9 e para a coleção núcleo na Tabela 10.

Tabela 9. Correlação de Spearman entre as distâncias genéticas geradas a partir do complemento dos coeficientes de similaridades entre os acessos da amostra da coleção base. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

Coeficiente	CS	RT	RR	J	D	O	OII
CS	1,00						
RT	0,94	1,00					
RR	0,52	0,57	1,00				
J	0,92	0,98	0,70	1,00			
D	0,92	0,98	0,70	1,00	1,00		
O	0,93	0,98	0,69	1,00	1,00	1,00	
OII	0,94	1,00	0,59	0,99	0,99	0,99	1,00

CS: Coincidência Simples; RT: Rogers e Tanimoto; RR: Russel e Rao; J: Jaccard; D: Dice; O: Ochiai; OII: Ochiai II.

Tabela 10. Correlação de Spearman entre as distâncias genéticas geradas a partir do complemento dos coeficientes de similaridades entre os acessos da coleção núcleo. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

Coeficiente	CS	RT	RR	J	D	O	OII
CS	1,00						
RT	1,00	1,00					
RR	0,51	0,51	1,00				
J	0,90	0,90	0,82	1,00			
D	0,90	0,90	0,82	1,00	1,00		
O	0,89	0,89	0,82	1,00	1,00	1,00	
OII	0,99	0,99	0,47	0,87	0,87	0,87	1,00

CS: Coincidência Simples; RT: Rogers e Tanimoto; RR: Russel e Rao; J: Jaccard; D: Dice; O: Ochiai; OII: Ochiai II.

Verifica-se pela Tabela 9 que a maioria das correlações são altas, mostrando uma grande associação entre as distâncias estimadas. Observa-se ainda que, as correlações para o coeficiente Russel e Rao apresentaram valores inferiores em relação aos outros coeficientes. As mesmas observações podem ser feitas para a Tabela 10.

Altas correlações entre as distâncias são freqüentes quando se avaliam diferentes coeficientes em variáveis binárias. Duarte (1999a), trabalhando com marcadores RAPD em feijão, encontrou correlações de 0,96 entre o coeficiente Coincidência Simples e os coeficientes de Jaccard, Dice e Ochiai; e também entre o coeficiente de Rogers e Tanimoto e os de Jaccard, Dice, Ochiai e Ochiai II. Este autor encontrou correlações baixas para o coeficiente de Russel e Rao em relação aos demais coeficientes. Estes resultados estão de acordo com os encontrados no presente trabalho. Johns et al. (1997) igualmente trabalhando com marcadores RAPD em feijão encontraram correlações de 0,99; 0,97 e 0,98 entre o coeficiente Coincidência Simples e os coeficientes de Jaccard, Dice e distância modificada de Rogers, respectivamente. Os resultados encontrados também estão de acordo com os de Skroch et al., (1998) que encontraram o valor de 0,99 para a correlação de distância genética estimada pelo coeficiente de Jaccard e Coincidência Simples para dados de RAPD.

Como os resultados das correlações de Spearman talvez não expressassem claramente as diferenças entre os coeficientes testados, utilizou-se o procedimento chamado de índice de ramificação de consenso comparando-se os dendrogramas gerados pelos coeficientes de similaridade (Tabela 11). Por este índice, cuja amplitude varia de 0 a 1, dois dendrogramas são considerados idênticos quando seu valor corresponde à unidade. Verificou-se, portanto, que os coeficientes de Jaccard e Dice conduzem a dendrogramas idênticos. Levando-se em consideração a base genética binária dos marcadores AFLP observa-se que a

Tabela 11. Índices de ramificação de consenso (CIc) na comparação dos dendrogramas gerados pelos coeficientes de similaridade. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

Coeficientes	CS	RT	RR	J	D	O
CS	1,00	0,80	0,07	0,59	0,59	0,59
RT		1,00	0,07	0,57	0,57	0,57
RR			1,00	0,09	0,09	0,09
J				1,00	1,00	0,95
D					1,00	0,95
O						1,00

CS: Coincidência Simples; RT: Rogers e Tanimoto; RR: Russel e Rao; J: Jaccard; D: Dice; O: Ochiai; OII: Ochiai II.

ausência de amplificação de uma determinada banda em dois acessos não representa, necessariamente, semelhança genética entre eles, o que faz com que aqueles coeficientes que excluem as co-ocorrências negativas de sua expressão de similaridade (Jaccard, Dice, Ochiai), a princípio, sejam mais adequados para o uso com este tipo de marcador (Williams et al, 1990). Optou-se, portanto, em usar o coeficiente de Jaccard por ser um coeficiente simples, largamente utilizado e preferido em estudos de biologia e, segundo Dias (1998), é melhor empregado para comparar materiais dentro da mesma espécie, em que concordâncias são mais frequentes.

Os dendrogramas obtidos pelo método UPGMA estão apresentados na Figura 8 para a amostra da coleção base e Figura 9 para a coleção núcleo. Pode-se observar que esses dendrogramas possuem estrutura semelhante, apesar da Figura 8 ter sido construída com 21 acessos da coleção base. Os distintos acessos não formaram grupos homogêneos definidos tanto na amostra da coleção base quanto na coleção núcleo até uma distância correspondente a

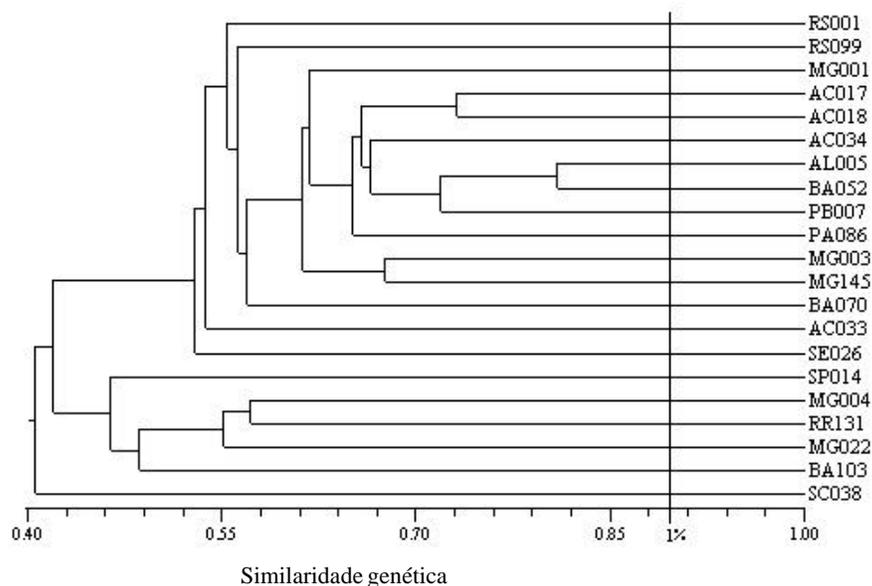


Figura 8. Dendrograma das similaridades genéticas entre os acessos da amostra da coleção base pelo coeficiente de Jaccard e o agrupamento por UPGMA. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

aproximadamente 25% da maior distância, reafirmando os resultados obtidos para os dados morfo-agronômicos. As linhas de corte nestas figuras representam o valor máximo significativo de similaridade (sg_{max}), determinado pelo teste de t em nível de 1% de probabilidade, sendo considerado como ponto de referência para discriminação ou agrupamento dos acessos.

O grau de concordância entre o dendrograma e a matriz de similaridades pelo coeficiente de Jaccard foi verificada pela análise de correlação cofenética. A correlação de matrizes normalizada pela estatística de Mantel (Rohlf, 1998) foi de 0,86 significando que o grau de ajustamento entre as matrizes foi considerado bom para os acessos da coleção base. Para a coleção núcleo, o valor da correlação de matrizes foi de 0,76 aproximando-se do valor de um bom

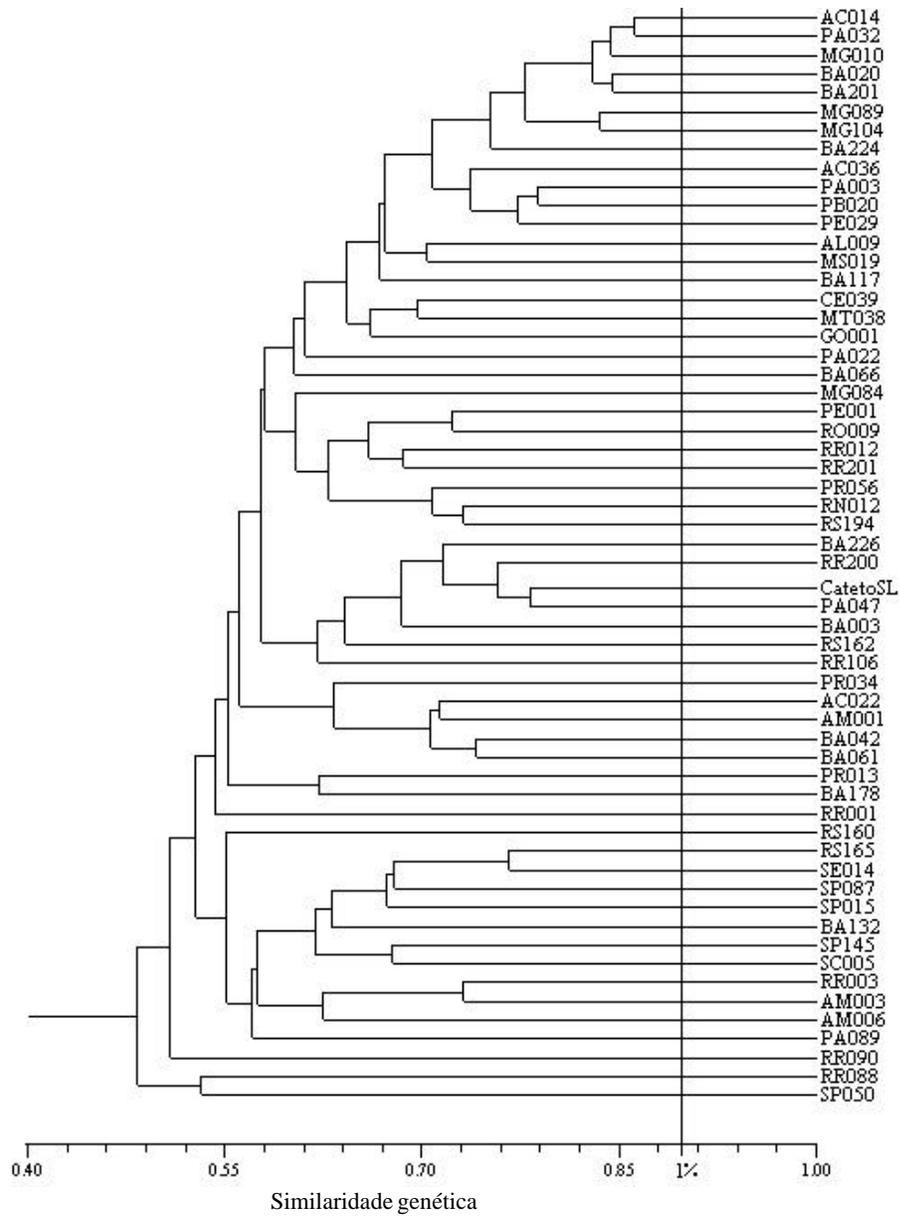


Figura 9. Dendrograma das similaridades genéticas entre os acessos da coleção núcleo pelo coeficiente de Jaccard e o agrupamento por UPGMA. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

ajustamento. Os valores dos coeficientes de correlação cofenética encontrados por Pejic et al. (1998) trabalhando com milho, foram de 0,83 para os dados de AFLP, 0,80 para SSR e 0,72 para RAPD, os quais consideraram o grau de ajustamento das matrizes como regular para bom.

As distribuições dos marcadores classificados por frequência de amplificação de fragmentos AFLP em cada coleção de acessos foram comparadas e pareceram bem semelhantes por possuírem a mesma forma do gráfico (Figura 10). Os gráficos sugerem que a diversidade genética encontrada na amostra da coleção base também aparece na coleção núcleo. Skroch et al. (1998) encontraram uma distribuição de marcadores quase idêntica quando compararam as distribuições dos marcadores classificados por frequência de amplificação de fragmentos RAPD em 90 acessos de cada coleção, base e núcleo pertencentes ao CIAT.

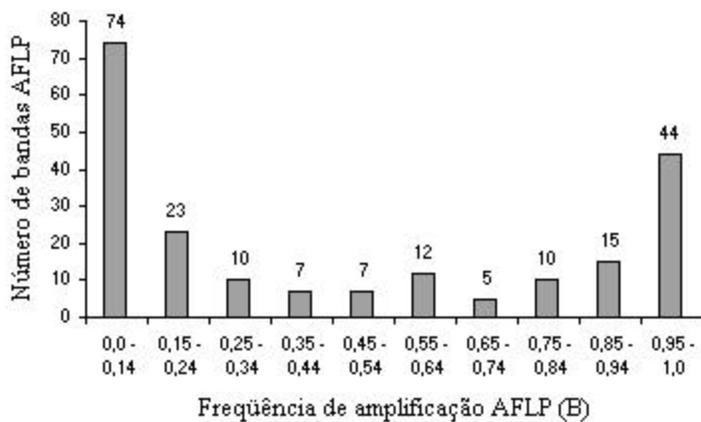
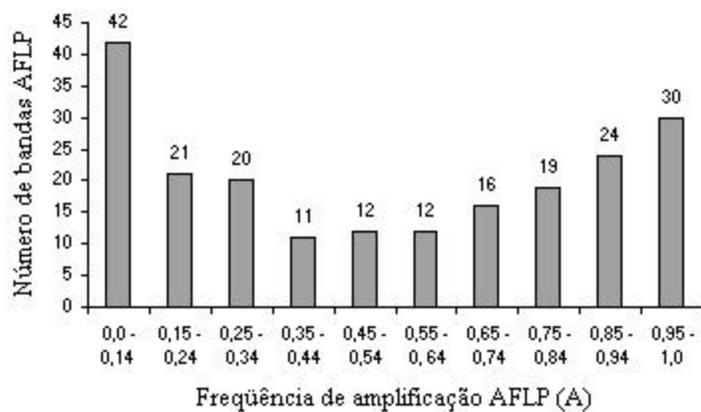


Figura 10. Distribuição de marcadores AFLP por frequência de amplificação dos fragmentos entre os acessos da coleção núcleo (A) e amostra da coleção base (B). Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

4.2.2 Diversidade genética

Os resultados da análise da variabilidade e estrutura genética da amostra da coleção base e coleção núcleo usando os marcadores AFLP quanto a frequência alélica, número efetivo de alelos, diversidade genética de Nei e informação sobre o índice de Shannon estão apresentados na Tabela 12. O número médio de alelos observado foi de 1,9662 e o número médio efetivo de alelos foi de 1,4420 para as duas coleções. O número efetivo de alelos por loco encontrado por Pejic et al. (1998) trabalhando com AFLP foi de 1,6 usando seis combinações de *primers* em 33 linhagens de milho e comentaram que este valor reflete uma baixa heterozigosidade esperada. Já o índice de eficiência da técnica encontrado para o AFLP foi de 61,9 e o encontrado no presente trabalho foi de 49,7. Estes valores são considerados altos quando comparados com índices de eficiência de RFLP, RAPD e SSR que foram de 3,2, 5,8, e 4,4, respectivamente, encontrados por Pejic et al. (1998). O alto índice de eficiência do AFLP é devido à detecção simultânea de vários locos em apenas uma reação de amplificação (Pejic et al., 1998).

A heterozigosidade esperada (h), também chamada de índice de diversidade genética de Nei, foi usada para avaliar o conteúdo polimórfico de cada loco. O valor obtido de 0,1988 foi calculado quando se consideraram todos os locos para a amostra da coleção base e 0,2807 para a coleção núcleo, utilizando-se seis combinações de *primers* de AFLP. Considerando que a amplitude de variação de h vai de 0 a 0,5, os valores encontrados no presente trabalho foram baixos. Padilha (2002) obteve o valor de h de 0,39 para marcadores AFLP em milho usando 11 combinações de *primers*. Os resultados do presente estudo também foram inferiores aos encontrados por Pejic et al. (1998) quando obtiveram o valor de h de 0,34 para marcadores AFLP usando

Tabela 12. Estimativas do número de alelos e diversidade genética para os 207 marcadores AFLP na amostra da coleção base e coleção núcleo. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

Parâmetro	Coleção Base (CB)		Coleção Núcleo (CN)		CB e CN	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
Número de alelos observados	1,7246	0,4478	1,9517	0,2149	1,9662	0,1812
Número de alelos efetivos (N_e)	1,3247	0,3430	1,4677	0,3410	1,4420	0,3312
Heterozigosidade (h)	0,1988	0,1823	0,2807	0,1668	0,2698	0,1629
Índice de Shannon (H')	0,3094	0,2561	0,4303	0,2184	0,4183	0,2120
Diversidade total das duas coleções (H_T)	-	-	-	-	0,2698	0,0265
Diversidade dentro das coleções (H_s)	-	-	-	-	0,2589	0,0241
Diversidade entre as coleções (D_{ST})	-	-	-	-	0,0109	-
Proporção da Diversidade entre as coleções (G_{st})	-	-	-	-	0,0404	-

seis combinações de *primers* em linhagens de milho. Estes valores indicam baixo conteúdo polimórfico, quando comparados com valores de h para RFLP de 0,63 e 0,72 para SSR (Pejic et al., 1998). O baixo conteúdo de informação genética por loco, ou heterozigosidade, é uma das limitações do AFLP, por ser essencialmente dominante.

A diversidade genética total (H_T) apresentada pelas duas coleções foi de 0,2698 quando se consideraram todos os 207 locos para os 79 indivíduos e frequências médias ponderadas de alelos. A diversidade genética dentro (H_s) das

coleções foi de 0,2589 e a diversidade entre coleções (D_{ST}) foi de 0,0109, que é a diferença entre H_T e H_S .

A proporção da diversidade genética devido à diferença entre coleções, em relação à diversidade total (G_{ST}) foi 0,0404. De acordo com os resultados apresentados, conclui-se portanto que apenas 4,04% da variabilidade genética total é atribuível à diferença entre as duas coleções, enquanto que o restante, 95,96% ocorre em virtude da diversidade presente dentro destas.

O índice de Shannon (H') também é empregado para medir a diversidade genotípica em populações. Esse índice é usado em estudos de recursos genéticos como uma medida de riqueza e uniformidade alélica, cujo baixo H' indica freqüência de classes desbalanceadas para uma determinada característica e uma falta de diversidade genética (Perry e McIntosh, 1991). No presente trabalho o índice de Shannon normalizado médio foi de 0,31 para a coleção base, 0,43 para a coleção núcleo e 0,42 quando se estimou o H' para as duas coleções juntas. O valor de H' para cada loco da coleção base variou de 0,1136 a 0,6929. O valor de H' para cada loco da coleção núcleo variou de 0,0877 a 0,6930. Estes resultados demonstram que a coleção núcleo apresentou maior diversidade genética em relação à amostra da coleção base. Isto pode ter ocorrido devido à quantidade menor de acessos amostrados na coleção base.

A identidade genética de Nei (I) obtida a partir das freqüências dos alelos nas duas coleções base e núcleo foi de 0,9643 e tem sido usada para medir a extensão da similaridade genética entre populações (Nei, 1987). A distância genética D de Nei foi de 0,0363, caracterizando, em média, 3,6 substituições alélicas a cada 100 locos ou marcadores analisados conforme Dias (1998). Considerando que I expressa a probabilidade de que um dado alelo de um loco, tomado ao acaso em duas diferentes populações seja idêntico em relação à probabilidade de que dois alelos do mesmo loco tomados também ao acaso em cada população sejam também idênticos e que I varia de 0 a 1, verifica-se que a

identidade genética de Nei foi alta, aproximando-se de 1, significando que as duas coleções possuem frequências alélicas semelhantes. Considerando que a distância genética D de Nei varia de 0 ao infinito, conclui-se que a distância genética verificada neste trabalho é muito pequena, confirmando que as frequências alélicas da amostra da coleção base e da coleção núcleo são semelhantes.

As frequências alélicas foram calculadas para cada loco de ambas coleções base e núcleo e a sua distribuição de frequência foi testada para homogeneidade pelo teste de χ^2 . A distribuição de frequência alélica foi estatisticamente homogênea para 164 locos ou 79,23%, significando que em 79% dos locos não houve diferença na distribuição da frequência alélica entre a amostra da coleção base e coleção núcleo. Este resultado foi semelhante ao encontrado por Huaman et al. (2000), no qual mencionam que em 77,78% dos locos isoenzimáticos analisados não houve diferença significativa entre acessos das coleções base e núcleo de batata.

A análise de variância molecular (AMOVA) foi usada como um outro critério para se determinar o grau de diferenciação genética entre as duas coleções. Pelos resultados da AMOVA (Tabela 13), observa-se que a maior porcentagem de variação ocorreu dentro das coleções (88,37%), evidenciando a alta variabilidade genética entre os acessos.

Os resultados confirmam que a fração da variância exposta pela variação dentro das coleções é, aproximadamente, oito vezes maior que a variação devido a diferenciação entre coleções, não havendo diferença significativa para a variância devido a diferenças entre as duas coleções com um valor de P menor que 0,0001 pelo teste permutacional.

Tabela 13. Análise de variância molecular dos 79 acessos de milho analisados por marcadores AFLP. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

Fontes de Variação	Graus de Liberdade	Soma dos Quadrados	Componentes de Variância	Porcentagem de Variação	P
Entre coleções	1	130,155	3,386	11,63	0,0000
Dentro de coleções	77	1981,785	25,738	88,37	
Total	78	2111,949	29,124		

Índice de Fixação de Alelos (F_{ST}): 0,1163

A porcentagem de variação atribuída às diferenças entre coleções foi de 11,63%, o que corresponde ao valor de F_{ST} de 0,1163.

4.3 Considerações sobre a validação da coleção núcleo

Conforme Brown (1989) a coleção núcleo deve ter 10% do tamanho da coleção original e conter 70% da variação genética. Uma estratégia adequada para se obter essa representatividade é o estabelecimento da coleção núcleo de uma maneira estratificada (Morales et al., 1995). Uma vez organizada, a validação ou a representatividade da coleção núcleo deve ser avaliada comparando sua diversidade com a da coleção base em cada descritor individual (Galwey, 1995). No presente trabalho, apesar dos dados morfo-agronômicos referentes à amostra da coleção base terem sido obtidos nos anos de 1991 a 1998 em Sete Lagoas e Janaúba, através de consulta ao banco de dados, foram feitos testes comparativos entre as duas coleções. Os resultados mostraram que houve diferenças significativas pelo teste de t para médias, teste de Bartlett para variâncias e teste de χ^2 para a distribuição de frequência de acessos das duas

coleções em relação à maioria dos descritores. O estudo da validação da coleção núcleo utilizando dados morfo-agronômicos não pode ser conclusivo em razão da ocorrência da interação genótipo-ambiente e do efeito de ambiente.

Os resultados da análise de componentes principais estimaram em aproximadamente 70%, a variação fenotípica considerando os três componentes principais. Os dendrogramas da coleção núcleo e da amostra da coleção base apresentaram estrutura geral semelhante, resultando em um bom ajuste entre as matrizes de distâncias genéticas e a cofenética.

Quanto à caracterização molecular, o nível de polimorfismo encontrado para a coleção núcleo e a amostra da coleção base foi bastante informativo e confiável. Através dos estudos de diversidade genética, apenas 4,04% (G_{ST}) da variabilidade genética total é atribuível à diferença entre a amostra da coleção base e coleção núcleo, enquanto que 95,96% é devido à diversidade presente dentro das coleções.

Com relação às frequências alélicas entre as duas coleções, a identidade genética de Nei foi alta, significando que as coleções possuem frequências alélicas semelhantes, e a distância genética de Nei foi pequena. Quanto a distribuição de frequência alélica verificou-se que não houve diferença significativa pelo teste de χ^2 para 79% dos locos. Finalmente, a análise de variância molecular veio confirmar os resultados anteriormente obtidos, mostrando diferença não significativa entre a variabilidade genética das coleções base e núcleo.

Portanto, os resultados obtidos nas avaliações moleculares, permitiram concluir que a formação da coleção núcleo com base na estratificação dos acessos, de acordo com as regiões agroecológicas, apresentou-se adequada, evidenciando a representatividade da coleção núcleo em relação à coleção base.

4.4 Considerações sobre a valoração da coleção núcleo

Como se sabe, a coleção núcleo deve ser um subconjunto representativo da variabilidade genética da coleção base. Além disso, os acessos que a compõem devem ser amplamente caracterizados de modo a fornecer o máximo de informação aos seus usuários. O incremento do conhecimento dos acessos que compõem a coleção núcleo adicionará valor estratégico à coleção base (Abadie et al., 2000), sendo que a caracterização morfo-agronômica e molecular é essencial e de demanda prioritária (Paterniani et al., 2000).

A obtenção de dados de mais oito descritores morfo-agronômicos: número de entrenós, comprimento da 5^a folha, comprimento da haste do pendão, ângulo da folha em relação ao caule, comportamento da folha acima da espiga, cor da folha, cor da coroa do grão e cor do endosperma permitiram valorar os acessos que compõem a coleção núcleo de milho, subgrupo endosperma duro (Tabela 14).

Quanto aos caracteres morfo-agronômicos qualitativos, a maioria não apresentou variação entre os acessos. Por exemplo, a cor da folha foi considerada verde médio para todos os acessos; o comportamento da folha foi sempre recurvado.

Quanto à cor da coroa do grão, 40% dos acessos apresentou cor laranja, 43% amarela e 17% branca. A cor do endosperma laranja foi a predominante (81%), seguida pela cor branca (16%), amarela (2%) e vermelha (1%).

Tabela 14. Estatísticas descritivas dos caracteres morfo-agronômicos avaliados nos acessos da coleção núcleo. Embrapa Milho e Sorgo, Sete Lagoas, MG, 2003.

Estimador	Descritor			
	Número de entrenós	Comprimento da 5ª folha (cm)	Comprimento da haste do pendão (cm)	Ângulo da folha/ caule (°)
Média	15,75	103,21	40,00	46,16
Mínimo	12,60	88,30	31,50	30,10
Máximo	19,50	132,00	77,90	60,60
Des. Padrão	1,31	7,93	5,60	6,58
C. Variação	8,33	7,69	14,02	14,25
Variância	1,73	63,03	31,45	43,30

A caracterização molecular permitiu a valoração de cada acesso, proporcionando uma nova informação acerca da estrutura genética da coleção núcleo, obtendo-se estimativas de frequências alélicas e índices de diversidade genética. Dessa maneira, a valoração dos acessos da coleção núcleo de milho, subgrupo endosperma duro, poderá incrementar o uso dos recursos genéticos, mantidos em banco de germoplasma pelos melhoristas e usuários em geral.

5 CONCLUSÕES

O uso do marcador molecular AFLP foi um instrumento valioso na validação da coleção núcleo de milho, tipo de endosperma duro, sendo eficiente em detectar o polimorfismo e útil nos estudos de diversidade genética dos acessos.

A caracterização morfo-agronômica com mais oito descritores e a caracterização molecular adicionaram valores agregados a cada acesso da coleção núcleo de milho.

A caracterização molecular dos acessos permitiu verificar que a estratégia de amostragem estratificada em regiões ecogeográficas empregada no desenvolvimento da coleção núcleo de milho, tipo de endosperma duro, foi adequada, demonstrando que esta representa a variabilidade genética da coleção base.

REFERÊNCIAS BIBLIOGRÁFICAS

ABADIE, T. **Obtenção e tratamento analítico de dados para organizar coleções nucleares**. Brasília: EMBRAPA-CENARGEN, 1997. 8 p. Relatório Final de Consultoria.

ABADIE, T.; CORDEIRO, C. M.; ANDRADE, R. V. de. A MAGALHÃES, J. R.; PARENTONI, S. N. A coleção nuclear de germoplasma de milho no Brasil. In: UDRY, C. V.; DUARTE, W. **Uma história brasileira do milho** – o valor dos recursos genéticos. Brasília: Paralelo 15, 2000. p. 65-78.

ABADIE, T.; MAGALHÃES, J. R.; CORDEIRO, C. M. T.; PARENTONI, S. N.; ANDRADE, R. V. **Obtenção e tratamento analítico de dados para organizar coleção nuclear de milho**. Brasília: EMBRAPA-CENARGEN, 1997. 8 p. (EMBRAPA-CENARGEN. Comunicado Técnico, 20).

AFLP. **Plant mapping: protocol**. The Perkin-Elmer Corporation, California, 1997.

BARRETT, B. A.; KIDWELL, K. K.; FOX, P. N. Comparison of AFLP and pedigree-based genetic diversity assessment methods using wheat cultivars from the Pacific Northwest. **Crop Science**, Madison, v. 38, n. 5, p. 1271- 1278, Sept./Oct. 1998.

BRASIL. Ministério da Agricultura e do Abastecimento. Secretaria de Desenvolvimento Rural, Serviço Nacional de Proteção de Cultivares. **Descritores mínimos do milho (*Zea mays* L.)** Brasília, 1999.

BROWN, A. H. D. The case for core collections. In: BROWN, A. H. D.; FRANKEL, O. H.; MARSHALL, R. D.; WILLIAMS, J. T. (Ed.) **The use of plant genetic resources**. Cambridge, UK: Cambridge University, 1989. p. 136-156.

BROWN, A. H. D. The core collections at the crossroads. In: HODGKIN, T.; BROWN, A. H. D.; VAN HINTUM, TH. J. L.; MORALES, E. A. V. (Ed.). **Core collections of plant genetic resources**. Chichester: J. Wiley, 1995. p. 3-19.

BROWN, A. H. D.; SPILLANE, C. Implementing core collections – principles, procedures, progress, problems and promise. In: JOHNSON, R. C. ; HODGKIN, T. **Core collections for today and tomorrow**. Rome: IPGRI, 1999. p. 1-9.

CASTANHEIRA, A. L. M. **Marcadores RAPD na avaliação do potencial de métodos de condução de população segregante de feijão**. 2001. 76 p.
Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras, MG.

CERVERA, M. T.; CABEZAS, J. A.; SANCHÁ, J. C.; TODA, F. M.; MARTINEZ-ZAPATER, J. M. Application of AFLPs to the characterization of grapevine *Vitis vinifera* L. genetic resources. A case study with accessions from Rioja (Spain). **Theoretical and Applied Genetics**, Berlin, v. 97, n. 1/2, p. 51-59, July 1998.

CORDEIRO, C. M. T.; MORALES, E. A. V.; FERREIRA, P.; ROCHA, D. M. S.; COSTA, I. R. S.; VALOIS, A. C. C.; SILVA, S. Towards a Brazilian core collection of cassava. In: HODGKIN, T.; BROWN, A. H. D.; VAN HINTUM, TH. J. L. ; MORALES, E. A. V. (Ed.). **Core collections of plant genetic resources**. Chichester: J. Wiley, 1995. p. 155-168.

CROSSA, J.; BASFORD, K.; TABA, S.; DELACY, I.; SILVA, E. Three- mode analyses of maize using morphological and agronomic attributes measures in multilocational trials. **Crop Science**, Madison, v. 35, n. 5, p. 1483-1491, Sept./Oct. 1995a.

CROSSA, J.; DeLACY, I. H.; TABA, S. The use of multivariate methods in developing a core collection. In: HODGKIN, T.; BROWN, A. H. D.; VAN HINTUM, TH. J. L. ; MORALES, E. A. V. (ed.). **Core collections of plant genetic resources**. Chichester: J. Wiley, 1995b. p. 77-92.

CRUZ, C. D. **Programa GENES- versão windows: aplicativo computacional em genética e estatística - Versão 2001**. 0. 0. Viçosa, MG: UFV, 2001. 642 p. 2001.

CRUZ, C. D.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: Universidade Federal de Viçosa, 1997. 390 p.

DeLACY, I. H.; SKOVMAND, B.; HUERTA, J. Characterization of México wheat landraces using agronomically useful attributes. **Genetic Resources and Crop Evolution**, Dordrecht, v. 47, n. 6, p. 591- 602, Dec. 2000.

DIAS, L. A. S. Análises multidimensionais. In: ALFENAS, A. C. **Eletroforese de isoenzimas e proteínas afins**. Viçosa: UFV, 1998. p. 405-475.

DUARTE, J. M.; SANTOS, J. B. DOS; MELO, L. C. Comparison of similarity coefficients based on RAPD markers in the common bean. **Genetics and Molecular Biology**, Ribeirão Preto, v. 22, n. 3, p. 427-432, Sept. 1999a.

DUARTE, J. M.; SANTOS, J. B. DOS; MELO, L. C. Genetic divergence among common bean cultivars from different races based on RAPD markers. **Genetics and Molecular Biology**, Ribeirão Preto, v. 22, n. 3, p. 419-426, Sept. 1999b.

EXCOFFIER, L.; SMOUSE, P. E.; QUATTRO, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. **Genetics**, Baltimore, v. 131, n. 2, p. 479-491, June 1992.

FAIAD, M. G. R.; SALOMÃO, A. N.; FERREIRA, F. R.; GONDIM, M. T. P.; WETZEL, M. M. V. da S.; MENDES, R. A.; GOES, M. de; MIRANDA, A. R. de. **Manual de procedimentos para conservação de germoplasma-semente a longo prazo na Embrapa**. Brasília: Embrapa – Cenargen, 1998. 21 p. (Embrapa – Cenargen. Documentos, 30).

FERREIRA, D. F. **Métodos de avaliação da divergência genética em milho e suas relações com os cruzamentos dialélicos**. 1993. 72 p. Dissertação (Mestrado) – Universidade Federal de Lavras, Lavras, MG

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3. ed. Brasília: EMBRAPA-CENARGEN, 1998. 220 p.

GALWEY, N. W. Verifying and validating the representativeness of a core collection. In: HODGKIN, T.; BROWN, A. H. D.; VAN HINTUM, TH. J. L. ; MORALES, E. A. V. (Ed.). **Core collections of plant genetic resources**. Chichester: J. Wiley, 1995. p. 187-198.

GUIMARÃES, C. T.; MOREIRA, M. A. Genética molecular aplicada ao melhoramento de plantas. In: BORÉM, A. **Melhoramento de espécies cultivadas**. Viçosa: UFV, 1999. p. 715-740.

HAGIWARA, W. E.; SANTOS, J. B. dos; CARMO, S. L. M. do. Use of RAPD to aid selection in common bean backcross breeding programs. **Crop Breeding and Applied Biotechnology**, Londrina, v. 1, n. 4, p. 355-362, Dec. 2001.

HAMON, S.; DUSSERT, S.; NOIROT, M.; ANTHONY, F.; HODGKIN, T. Core collections - accomplishments and challenges. **Plant Breeding Abstracts**, Cambridge, v. 65, n. 8, p. 1125- 1133, Aug. 1995.

HEDRICK, P. W. **Genetics of populations** . 2. ed. Massachusetts: Jones and Bartlett Publishers, 1999. 553 p.

HU, J.; ZHU, J.; XU, H. M. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. **Theoretical & Applied Genetics**, Berlin, v. 101, n. 1/2, p. 264-268, July 2000.

HUAMAN, Z.; ORTIZ, R.; ZHANG, D.; RODRÍGUEZ, F. Isozyme análisis of entire and core collections of *Solanum tuberosum* subsp. *andigena* potato cultivars. **Crop Science**, Madison, v. 40, n. 1, p. 273-282, Jan./Feb. 2000.

IBPGR. (Rome, Italy) **Descriptors for Maize** . Mexico: CIMMYT/ Rome: IBPGR, 1991. 88 p.

JOHNS, M. A.; SKROCH, P. N.; NIENHUIS, J.; HINRICHSEN, P. BASCUR, G.; MUÑOZ-SCHICK, C. Gene pool classification of common bean landraces from Chile based on RAPD and morphological data. **Crop Science**, Madison, v. 37, n. 2, p. 605-613, Mar./Apr. 1997.

MORALES, E. A. V.; VALOIS, A. C. C. Princípios para conservação e uso de recursos genéticos. In: PUIGNAV, A. C. C. (Ed.). **Conservacion de germoplasma vegetal**. Montevideo: IICA, 1996. p. 13-34. (IICA-PROCISUR. Dialogo, 45).

MORALES, E. A. V.; VALOIS, A. C. C.; COSTA, I. R. S. Core collections for gene banks with limited resources. In: HODGKIN, T.; BROWN, A. H. D.; VAN HINTUM, TH. J. L. ; MORALES, E. A. V. (Ed.). **Core collections of plant genetic resources**. Chichester: J. Wiley, 1995. p. 241-249.

MUMM, R. H.; HUBERT, L. J.; DUDLEY, J. W. A classification of 148 U. S. maize inbreds: II. Validation of cluster analysis based on RFLPs. **Crop Science**, Madison, v. 34, n. 4, p. 852-865, July/Aug. 1994.

NASS, L. L. Utilização de recursos genéticos vegetais no melhoramento. In: NASS, L. L.; VALOIS, A. C. C.; MELO, I. S. de; VALADARES-INGLIS, M.

C. **Recursos genéticos e melhoramento/ plantas** . Rondonópolis: Fundação MT, 2001. p. 30- 55.

NASS, L. L.; PELLICANO, I. J.; VALOIS, A. C. C. Utilization of genetics resources for maize and soybean in Brasil. **Brazilian Journal of Genetics** , Ribeirão Preto, v. 16, n. 4, p. 983-988, Dec. 1993.

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

PADILHA, L. **Marcadores moleculares semi-automatizados na caracterização e determinação da diversidade genética entre linhagens de milho tropical**. 2002. 85 p. Tese (Doutorado) – Universidade Federal de Lavras, Lavras, MG.

PATERNIANI, E.; NASS, L. L.; SANTOS, M. X. dos. O valor dos recursos genéticos de milho para o Brasil: uma abordagem histórica da utilização do germoplasma. In: UDRY, C. V.; DUARTE, W. (Org.). **Uma história brasileira do milho – o valor dos recursos genéticos**. Brasília: Paralelo 15, 2000. p. 11-42.

PEJIC, I.; AJMONE-MARSAN, P.; MORGANTE, M.; KOZUMPLICK, V.; CASTIGLIONI, P.; TARAMINO, G.; MOTTO, M. Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs and AFLPs. **Theoretical Applied Genetics** , Berlin, v. 97, n. 8, p. 1248-1255, Dec. 1998.

PEREIRA, A. V. **Utilização de análise multivariada na caracterização de germoplasma de mandioca (*Manihot esculenta* Crantz)**. 1989. 180 f. Tese (Doutorado em Genética e Melhoramento de Plantas)- Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.

PERKIN ELMER. **AFLP™ plant mapping kit**. The PCR marker of choice for plant mapping. California, The Perkin Elmer Corporation, 1996.

PERRY, M. C. , MCINTOSH, M. S. Geographical patterns of variation in the USA soybean germplasm collection: I. Morphological traits. **Crop Science** , Madison, v. 31, n. 5, p. 1350-1355, Sept./Oct. 1991.

PINTO-COELHO, R. M. **Fundamentos em ecologia**. Porto Alegre: Artmed Editora, 2000. 252 p.

RAMALHO, M. A. P.; SANTOS, J. B. dos; PINTO, C. A. B. P. **Genética na agropecuária**. Lavras: UFLA, 1999. 472 p.

RIBEIRO JÚNIOR, J. I. **Análises estatísticas no SAEG**. Viçosa: UFV, 2001. 301 p.

ROBINSON, I. P. Aloenzimas na genética de populações de plantas. In: ALFENAS, A. C. **Eletroforese de isoenzimas e proteínas afins**. Viçosa: Editora UFV, 1998. p. 359-376.

ROHLF, F. J. **Numerical taxonomy and multivariate analysis system: version2. 02k**. New York, 1998.

SAGHAI-MAROOF, M. A.; SOLIMAN, K. M.; JORGENSEN, R. A.; ALLARD, R. W. Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. **Proceedings of the National Academy of Sciences**, Washington, v. 81, n. 24, p. 8014-8019, 1984.

SANTOS, C. A. F.; OLIVEIRA, C. A. V.; MENEZES, E. A. Seleção de descritores na caracterização e avaliação preliminar de germoplasma de guandu. **Pesquisa Agropecuária Brasileira**, Brasília, v. 30, n. 7, p. 971- 975, jul. 1995.

SCHNEIDER, S.; ROESSLI, D.; EXCOFFIER, L. **Arlequin ver. 2. 000: A software for population genetics data analysis**. Genetics and Biometry Laboratory, University of Geneva, Switzerland. 2000.

SCHOEN, D. J.; BROWN, A. H. D. Maximizing genetic diversity in core collections of wild relatives of crop species. In: HODGKIN, T.; BROWN, A. H. D.; VAN HINTUM, TH. J. L.; MORALES, E. A. V. (Ed.). **Core collections of plant genetic resources**. Chichester: J. Wiley, 1995. p. 55-76.

SINGH, D. The relative importance of characters affecting genetic divergence. **The Indian J. of Genet. and Plant Breed.**, v. 41, p.237-245, 1981.

SKINNER, D. Z.; BAUCHAN, G. R.; AURICHT, G.; HUGHES, S. Developing a core collection from a large annual *Medicago* germplasm collection. In: JOHNSON, R. C.; HODGKIN, T. **Core collections for today and tomorrow**. IPGRI, Rome, Italy. 1999. p. 61-67.

SKROCH, P. W.; NIENHUIS, J. BEEBE, S.; TOHME, J.; PEDRAZA, F. Comparison of Mexican common bean (*Phaseolus vulgaris* L.) core and reserve

germplasm collections. **Crop Science**, Madison, v. 38, n. 2, p. 488- 496, Mar./Apr. 1998.

SKROCH, P. W.; TIVANG, J.; NIENHUIS, J. Analysis of genetic relationships using RAPD marker data. In: APPLICATIONS OF RAPD TECHNOLOGY TO PLANT BRREDING, Minneapolis, 1992. **Proceedings...** Minneapolis: Crop Science Society of America, 1992. p. 26-30.

SPAGNOLETTI ZEULI, P. L.; QUALSET, C. O. The durum wheat core collection and the plant breeder. In: HODGKIN, T.; BROWN, A. H. D.; VAN HINTUM, TH. J. L. ; MORALES, E. A. V. (ed.). **Core collections of plant genetic resources**. Chichester: J. Wiley, 1995. p. 213-228.

TABA, S. The conservation of maize genetic resources in the western hemisphere. In: TABA, S. (Coord.). **Cooperative project to regenerate maize germplasm collections** - Review. Mexico: CIMMYT, 1994.

TOHME, J.; GONZALEZ, D. O.; BEEBE, S.; DUQUE, M. C. AFLP analysis of gene pools of a wild bean core collection. **Crop Science**, Madison, v. 36, n. 5, p. 1375-1384, Sept./Oct. 1996.

TOHME, J.; JONES, P.; BEEBE, S.; IWANAGA, M. The combined use of agroecological and characterization data to establish the CIAT *Phaseolus vulgaris* core collection. In: HODGKIN, T.; BROWN, A. H. D.; VAN HINTUM, TH. J. L. ; MORALES, E. A. V. (Ed.). **Core collections of plant genetic resources**. Chichester: J. Wiley, 1995. p. 95-107.

VALOIS, A. C. C.; SALOMÃO, A. N.; ALLEN, A. C. **Glossário de recursos genéticos vegetais**. Brasília: Embrapa-SPI, 1996. 62p (Embrapa- Cenargen. Documentos ; n. 22).

VAN HINTUM, T. J. L. The general methodology for creating a core collection. In: JOHNSON, R. C.; HODGKIN, T. **Core collections for today and tomorrow**. Rome: IPGRI, 1999. p. 10-17.

VAN HINTUM, T. J. L. Hierarchical approaches to the analysis of genetic diversity in crop plants. In: HODGKIN, T.; BROWN, A. H. D.; VAN HINTUM, T. J. L. ; MORALES, E. A. V. (Ed.). **Core collections of plant genetic resources**. Chichester: J. Wiley, 1995. p. 23-34.

VAN HINTUM, T. J. L.; BROWN, A. H. D.; SPILLANE, C.; HODGKIN, T.
Core collections of plant genetic resources. Rome: IPGRI, 2000. 48p. (IPGRI
Technical Bulletin, 3).

VEASEY, E. A.; SCHAMMASS, E. A.; VENCOVSKY, R.; MARTINS, P. S.;
BANDEL, G. Germplasm characterization of *Sesbania* accessions based on
multivariate analyses. **Genetic Resources and Crop Evolution**, Dordrecht, v.
48, n. 1, p. 79-90, Feb. 2001.

VEIGA, R. F. DE A.; QUEIROZ-VOLTAN, R. B.; VALLS, J. F. M.;
FÁVERO, A. P.; BARBOSA, W. Caracterização morfológica de acessos de
germoplasma de quatro espécies brasileiras de amendoim silvestre. **Bragantia**,
Campinas, v. 60, n. 3, p. 167- 176, 2001.

VILELA-MORALES, E. A.; VALOIS, A. C. C.; NASS, L. L. **Recursos
genéticos vegetales.** Brasília: Embrapa-SPI/Embrapa- Cenargen, 1997. 78 p.

VOS, P.; HOGERS, R.; BLEEKER, M.; REIJANS, M.; VAN DE LEE, T.;
HORNES, M.; FRIJTERS, A.; POT, J.; PELEMAN, J.; KUIPER, M.;
ZABEAU, M. AFLP: a new technique for DNA fingerprinting. **Nucleic Acids
Research**, Oxford, v. 23, n. 21, p. 4407-4414, Nov. 1995.

WILLIAMS, J. G. K.; KUBELELIK, A. R.; LIVAK, K. J.; RAFALSKI, J. A.;
TINGEY, S. V. DNA polymorphisms amplified by arbitrary primers are useful
as genetic markers. **Nucleic Acids Research**, Oxford, v. 18, n. 22, p. 6531- 6535,
Nov. 1990.

YEH, F. C. **Popgene version 1.31.** Microsoft window-based freeware for
population genetic analysis. Dept. of Renewable Resources, University of
Alberta. 1999.

ZHU, S. L.; MONTI, L. M.; AVITABILE, A.; RAO, R. Evaluation of genetic
diversity in Chinese soyabean germplasm by AFLP. **Plant Genetic Resources
Newsletter**, Rome, n. 119, p. 10-14, 1999. Supplement.

ZIBUMBO, D. V.; COLUNGA, P. G. Morpho-physiological variation and
phenotypic plasticity in Mexican populations of coconut (*Cocos nucifera* L.).
Genetic Resources and Crop Evolution, Dordrecht, v. 48, n. 6, p. 547-554,
Dec. 2001.