

Erro experimental em marcadores AFLP: consequências e estimativas

Experimental error in AFLP markers: consequence and estimative

Jefferson Luís Meirelles Coimbra^{1*}; Altamir Frederico Guidolin¹; Maurício Marini Kopp²; Juliano Garcia Bertoldo³; Bruna Arruda³; Jussara Cristina Stingenhen³

Resumo

O objetivo deste trabalho foi estimar e comparar os coeficientes de dissimilaridades de DICE por meio da estimativa do erro experimental, proveniente da utilização de repetições, que foram compostas por *bulks* de plantas e, comparar tanto estas estimativas entre si e com a estimativa da dissimilaridade genética baseada em caracteres agronômicos avaliados por meio da experimentação a campo. A estimativa deste coeficiente pode se tornar mais confiável e robusta com a utilização de repetições nas análises moleculares. O uso de repetições em análises moleculares favorece tanto os testes de hipóteses quanto a avaliação dos contrastes de interesse. A associação entre as medidas avaliadas (genética e morfológica), com base na utilização de dados moleculares *versus* morfológicos evidenciou 80% de concordância nos valores estimados, ($P < 0,05$) quando foram utilizadas as repetições, na análise molecular. Desta forma, fica evidente que o emprego de repetições em análises moleculares, além de fornecer uma estimativa do erro experimental, propicia um ganho imediato por meio dos testes de hipóteses; podendo incrementar desta maneira tanto a confiabilidade das estimativas dos parâmetros genéticos, como por exemplo, a repetibilidade, quanto à concordância dos valores de laboratório e de experimentos de campo.

Palavras-chave: *Oryza sativa* L., coeficiente de dissimilaridade, repetibilidade Mahalanobis

Abstract

The objective of this study was to estimate and to compare the Dice coefficients of dissimilarity through experimental error estimation, from the use of replications composed by plant bulks. The estimate of this coefficient can become more reliable and robust with the use of replications in the molecular analyses. This procedure favors either the tests of hypotheses or the evaluation of the contrasts of interest, according to the objective of the experiment. Also, to compare these estimates with each other and with the genetic dissimilarity estimate based on agronomical traits measured in the field. Among the evaluated variables (genetic and morphological), the distances estimated with base in the use of molecular *versus* morphological data evidenced 80% of agreement in the estimated values, compared to 5% ($P < 0.05$) when the replications were used, in the molecular analysis. This way, it is evident that the use of replications in molecular analyses, besides supplying an estimate of the experimental error, enables an immediate gain through the tests of hypotheses, increasing both the reliability of the estimates of the genetic parameters (i.e., repeatability), and the agreement between laboratory and field data.

Key words: *Oryza sativa* L., genetic dissimilarity, repeatability Mahalanobis

¹ Prof. da Universidade do Estado de Santa Catarina, UDESC, Deptº de Agronomia, Programa de Pós Graduação em Ciências Agrárias, Instituto de Melhoramento e Genética Molecular da UDESC, IMEGEM. Lages, SC. E-mail: coimbrajefferson@cav.udesc.br; a2afg@cav.udesc.br

² Engº Agrº D.Sc. Embrapa Pecuária Sul, Bagé, RS. E-mail: mauricio.kopp@cppsul.embrapa.br

³ Discentes da Universidade do Estado de Santa Catarina, UDESC, Deptº de Agronomia, Programa de Pós Graduação em Ciências Agrárias, Instituto de Melhoramento e Genética Molecular da UDESC, IMEGEM. Lages, SC. E-mail: jgbertoldo@gmail.com; bru_agro@hotmail.com; jcstingenhen@hotmail.com

* Autor para correspondência

Introdução

Todo e qualquer experimento tem como objetivo testar uma hipótese. Portanto, *a priori*, o pesquisador precisa obter uma estimativa para o erro experimental e, *a posteriori*, testar a significância dos contrastes. É importante a estimação correta do erro experimental, pois pode ser utilizado em qualquer das etapas posteriores (CARDELLINO; SIEWERDT, 1992). O erro experimental, por sua vez, apenas pode ser predito se os tratamentos forem repetidos. Os testes de hipóteses podem ser tanto mais precisos quanto menor for a estimativa do erro experimental, sendo que quanto maior o número de repetições, menor o erro padrão da média (RAMALHO; FERREIRA; OLIVEIRA, 2000). Agrupamento de erros em categorias discretas de acordo com suas causas é um desafio devido suas diferentes causas, que por vezes, interagem para gerar um erro. Os mesmos autores agruparam os erros de laboratório em quatro categorias: i) os erros que estão ligados a própria sequência de DNA; ii) os erros que são devido à baixa qualidade ou quantidade de DNA; iii) reagentes bioquímicos e; iv) fatores humanos.

Técnicas de marcadores moleculares para estimar a distância genética vem sendo amplamente utilizadas, principalmente a partir da última década. Este progresso tem por base o fato de que estas técnicas permitem a obtenção da informação genética contida no genoma de forma direta. Dentre as técnicas de marcadores moleculares, a de AFLP se destaca pelo número de marcadores analisados num único gel, grande poder de detecção de variabilidade genética e maior confiabilidade do ensaio (VOS et al., 1995).

Comparativamente aos caracteres fenotípicos, os marcadores moleculares expressam a vantagem de não sofrerem influência do ambiente (BARBOSA NETO, 1998). O termo ambiente, num sentido específico, pode ser conceituado molecularmente como toda a variação, intrínseca e/ou extrínseca, suscetível de ocorrer durante a realização da técnica

de AFLP. Assim, a técnica de AFLP pode sofrer alterações intrínsecas de maior ou menor intensidade na digestão com enzimas de endonucleases (corte raro e corte frequente) na amplificação seletiva e na separação dos fragmentos no gel. Ainda, no entanto, por acessarem o genoma como um todo e não somente as regiões responsáveis pela expressão dos caracteres de interesse, os marcadores moleculares podem gerar informações não associadas aos caracteres morfológicos (VIEIRA et al., 2005; MILACH, 1998; BARBOSA NETO, 1998; OLIVEIRA, 1998).

No emprego tanto de técnicas moleculares quanto morfológicas, é prudente que o pesquisador possua um número de repetições adequado, com o propósito de testar as hipóteses de seu interesse. Além disso, é importante que os laboratórios possam demonstrar que seus resultados são confiáveis, garantindo a repetitividade de testes por meio da validação de metodologias (OLIVEIRA et al., 2010). Ainda, é essencial que as atividades laboratoriais, como as diferentes técnicas de análise utilizadas, possam ser padronizadas no sentido de serem reproduzíveis entre diferentes laboratórios (JONES et al., 1997). Em geral, no emprego das técnicas moleculares, como, por exemplo, AFLP (*Amplified Fragment Length Polymorphism*), não se utilizam repetições das amostras. Ainda assim, estas técnicas podem ser consideradas úteis, porém menos consistentes. Bensch e Akesson (2005) verificaram 2223 referências no ano de 2004, somente no Editor ISI Web of Science, relacionadas a aplicação de AFLP. Tal fato revela a grande utilização desta técnica. Apesar da relevância das metodologias moleculares, as estimativas dos coeficientes de dissimilaridade computados por técnicas moleculares podem não fornecer uma qualidade completamente satisfatória na discriminação entre os genótipos testados; de fato, essa informação incompleta pode ter origem nas causas de variação intrínseca e perturbadora, por exemplo: *i)* nível de endogamia entre os genótipos estudados; *ii)* pureza varietal; e *iii)* heterose residual. Na metodologia de AFLP um possível problema

pode surgir caso algumas bandas não amplifiquem ou não apareçam de forma consistente na corrida em gel, e, se as populações são analisadas em lotes, isto poderia conduzir a estimativas significativas da estrutura populacional apenas resultante de um viés da metodologia (BENSCH; AKESSON, 2005). Ainda, segundo os mesmos autores, esse problema é evitado se as amostras forem analisadas de forma aleatória em relação aos grupos e as populações que são comparados, e, salientam que, as análises repetidas de DNA re-extraído dos mesmos indivíduos representam importantes controles de qualidade. Nesse caso, é possível verificar que pode ter ocorrido um erro de genotipagem. Erros de genotipagem ocorrem quando um determinado genótipo, após a análise molecular, não corresponder ao real genótipo do indivíduo em questão (BONIN et al., 2004).

A utilização de técnicas de marcadores moleculares para estimar a distância genética vem crescendo rapidamente, principalmente na última década. Este progresso tem por base o fato de que estas técnicas permitem a obtenção da informação genética contida no genoma de forma direta. Dentre as técnicas de marcadores moleculares, a de AFLP se destaca pelo número de marcadores analisados num único gel, grande poder de detecção de variabilidade genética e maior confiabilidade do ensaio (VOS et al., 1995). Comumente na análise molecular, os resultados são interpretados por meio de dendrogramas, principalmente quando o objetivo de estudo for o de diversidade e distância genética. A análise da distância é uma ferramenta auxiliar de grande importância em programas de melhoramento e um importante elo entre a conservação e a utilização dos recursos genéticos disponíveis (SILVA et al., 2009).

Tanto a confecção quanto a interpretação do dendrograma, resultantes tanto da técnica molecular quanto da morfológica, são bastante insuficientes, mesmo utilizando adicionalmente, por exemplo, a técnica de reamostragem por *bootstrapping*. A técnica de dendrogramas é subjetiva, portanto não oferece

uma adequada probabilidade de aceitar ou rejeitar a hipótese básica a ser testada, ou seja, não deve ser considerada como estatística, pois não testa parâmetros.

Os dendrogramas são especialmente úteis na visualização de semelhanças entre um grande número de genótipos ou acessos. No entanto, os dendrogramas podem sofrer alterações expressivas simplesmente por meio da modificação da escala de aferimento, levando o pesquisador a uma interpretação ambígua. Além disso, a identificação de agrupamentos pode ser feita por diversos algoritmos que podem produzir resultados diferentes entre si (MOITA NETO; MOITA, 1997).

Desta forma, este trabalho foi realizado com o intuito de: *i*) estimar o erro experimental em técnicas moleculares (AFLP); *ii*) comparar os resultados com os dados morfológicos (Mahalanobis) e; *iii*) testar os coeficientes de dissimilaridade genética por meio do teste t; utilizando para tanto dois *bulks* de plantas de arroz como repetições para testar as hipóteses e os contrastes de interesse.

Material e Métodos

Avaliação fenotípica

Para realização do experimento foram testados seis genótipos de arroz, amplamente utilizados pelos agricultores no Rio Grande do Sul: *i*) BRS 7 ‘Taim’; *ii*) BRS Pelota; *iii*) BRS Ligeirinho; *iv*) BRS Atalanta; *v*) BRS 6 ‘Chuí’; e *vi*) BRS Firmeza. Os genótipos foram semeados a campo, por meio do delineamento blocos casualizados com três repetições por tratamento, onde foram medidos 12 caracteres de importância agrônômica nos cultivares citados anteriormente: 1) número de afilhos férteis (NAF); 2) comprimento da panícula principal em centímetros (CPP); 3) peso da panícula principal em gramas (PPP); 4) número de grãos por panícula (NGP); 5) peso de grãos por panícula em gramas (PGP); 6) peso médio de grãos em gramas (PMG); 7) peso de grãos por planta em gramas (PGPL); 8)

largura da bainha da folha bandeira em centímetros (LFB); 9) comprimento da bainha da folha bandeira (CFB); 10) estatura de planta do solo até o final da panícula (EFP); 11) estatura da planta do solo até o final da folha bandeira (EFFB); 12) número de dias entre a emergência a floração (DEF).

O coeficiente de dissimilaridade (distância fenotípica) foi estimado por meio da estatística de Mahalanobis (D^2), por meio do software SAS (Learning Edition, 2002), via procedimento *CANDISC* (*Canonical discriminant analysis*).

Todos os seis cultivares foram avaliados molecular e morfológicamente no presente trabalho.

Avaliação molecular

ODNA utilizado nas análises de AFLP foi extraído de acordo com o protocolo descrito por Castelo-Branco (2005). Foram semeadas 15 sementes de cada genótipo, aproximadamente, sendo utilizados dois *bulks* de cinco plantas, para extração de DNA, denominados de *bulk-1* e *bulk-2*. Esses *bulks* foram efetivados com o intuito de considerar como duas repetições do mesmo genótipo, com o objetivo de contemplar as estimativas do erro experimental dos coeficientes de dissimilaridades, via análise de variância, realizada pelo *PROC MIXED* do SAS, considerando estas repetições como fator de efeito aleatório, segundo a metodologia descrita por Littell et al. (1996).

A análise de marcadores AFLP foi realizada de acordo com a metodologia sugerida por Vos et al. (1995), por meio do protocolo fornecido pelo fabricante (*AFLP Analysis System I, Gibco/BRL*). Foram empregadas seis combinações de iniciadores: C_1) *M-CTG/E-AGC*; C_2) *M-CTC/E-AAC*; C_3) *M-CAG/E-ACG*; C_4) *M-CAA/E-ACT*; C_5) *M-CTT/E-ACA*; C_6) *M-CTA/E-AAG*; e C_7) *M-CAC/E-AGG* em C_i : i -ésima combinação de iniciadores; E: *EcoRI* e M: *MseI*. Para visualizar os fragmentos amplificados e separados eletroforeticamente em gel desnaturante de poliacrilamida (6%), foi utilizado o protocolo de

coloração a base de nitrato de prata, descrito por Creste, Tulmann-Neto e Figueira (2001).

Análise estatística

Os dados de presença (1) ou ausência (0) de bandas obtidas na análise de AFLP dos seis genótipos avaliados permitiram o cálculo da similaridade genotípica entre todos os pares de genótipos, com o auxílio do software SAS versão *Learning Edition* (2002). Para o cálculo da similaridade genotípica (S_{ij}), foi empregado o coeficiente de Dice (DICE, 1945), conforme a equação: $S_{ij} = 2N_{ij} \div N_i + N_j$, em que N expressa o número de bandas presentes nos genótipos i e j . Tanto a similaridade genética quanto a dissimilaridade genética foram estimadas por meio do software SAS com auxílio de macros específicos, de acesso livre, denominado de *distnew*.

Após a obtenção dos valores estimados dos coeficientes de dissimilaridades genéticas de Dice (DICE, 1945), foi determinado tipo de distribuição assumida por estes coeficientes, sendo possível verificar que a distribuição normal representava de modo apropriado o tipo de distribuição desses valores de forma significativa (5% de significância), por meio das estatísticas de Shapiro-Wilk e Kolmogorov-Smirnov, que testam à normalidade dos dados.

Considerando o caráter dissimilaridade genética (y), o modelo empregado para as observações moleculares foi:

$$y = \mu + p_j + b_k + e_{jk},$$

onde: μ : média fenotípica ou geral (intercepto); p : é o efeito do j -ésimo de todas as combinações de genótipos; b_k : é o efeito da k -ésima repetição; e_{jk} : é o efeito do resíduo.

As pressuposições para os efeitos fixos, média geral e das combinações dos genótipos, são independentes entre si, para os efeitos aleatórios

das repetições (b_k) e do efeito do resíduo (e_{jk}) assume-se que possuem uma distribuição normal com média 0 e variâncias σ_b^2 e σ^2 , respectivamente. Adicionalmente, a todos os efeitos aleatórios pode ser assumida a independência.

Os componentes de variâncias foram estimados por meio do método da máxima verossimilhança restrita (REML), pois se tornou o método padrão para a estimativa de componentes de variância e parâmetros genéticos a partir de dados desbalanceados (LITTELL et al., 1996). A preferência por este método decorre de suas propriedades estatísticas, os quais são superiores as propriedades dos estimadores de quadrados mínimos e de máxima verossimilhança (SEARLE; CASELA; McCULLOCH, 1992). As estimativas dos componentes da variância foram preditas pelo método proposto por Patterson e Thompson (1971) com o intuito de se estimar a repetibilidade.

Resultados e Discussão

Ao analisarmos uma variável qualquer (principalmente uma variável com distribuição desconhecida) temos em primeiro lugar que verificar a condição de normalidade da distribuição dos dados, no entanto, não há necessidade de um rigor extremo a esse respeito (SILVA, 1993). A normalidade dos dados foi verificada por meio dos testes estatísticos de Shapiro-Wilk e Kolmogorov-Smirnov, sendo que os valores dos testes e das significâncias observados foram 0,92 ($P=0,02$) e 0,92 ($P=0,11$), respectivamente. Por meio desses testes de significância utilizados, pode ser concluído que a distribuição dos dados esta suficientemente próxima da curva normal.

As seis combinações de iniciadores usados geraram um total de 195 bandas, das quais 172 (88%) foram polimórficas, e 23 (12%) bandas foram monomórficas entre os seis genótipos de arroz considerados neste estudo. Este fato confirma que a técnica de AFLP é um poderoso método de análise molecular que pode ser utilizado eficientemente para

detectar variabilidade genética na cultura do arroz, em especial. Um dos grandes atributos da técnica de AFLP é a sua capacidade *multiplex*, ou seja, um grande número de locos pode ser amostrado em único gel AFLP (FERREIRA; GRATTAPAGLIA, 1998). Ainda de acordo com estes autores, o número médio de locos polimórficos também e alto nestas amostras, corroborando os dados encontrados neste trabalho. Onde o grau de polimorfismo encontrado no presente trabalho indica que existe um grau de variabilidade genética bastante ampla entre estas combinações genóticas avaliadas.

De modo contrário, alguns autores afirmam que a base genética do melhoramento da cultura é estreita (GUIDOLIN et al., 1994). Por exemplo, em 1992, na América Latina, o número de cultivares de arroz utilizado em programas de melhoramento estava reduzido a doze cultivares (CUEVAS-PÉREZ et al., 1992).

Atualmente cerca de 81% da produção nacional de arroz é oriunda dos Estados do Rio Grande do Sul (7.844,781 t), Santa Catarina (1.039.720 t), Mato Grosso (803.542 t) e Maranhão (605.789 t) (ICEPA, 2009), sendo que, a base genética dos principais programas de melhoramento genético pode ser resumida em apenas sete cultivares (RANGEL; GUIMARÃES; NEVES, 1996).

Provavelmente, o alto polimorfismo, encontrado neste trabalho, não esteja associado a caracteres adaptativos, como por exemplo, a características principais do grão e da planta como: *i*) classe longo fino; *ii*) aspecto vítreo; *iii*) temperatura de gelatinização; e *iv*) estatura de planta, pois todos os caracteres de importância agrônômica, entre outros, são bastante semelhantes entre estes genótipos estudados.

Desta forma, todo o conhecimento, molecular e morfológico, que possa contribuir de maneira efetiva para facilitar a obtenção de genótipos com maior potencial de produtividade e alta adaptabilidade é de suma importância para programas de melhoramento genético da cultura. Esta importância aumenta,

fundamentalmente, na presença de variabilidade genética restrita, pois tal fato pode ofuscar o êxito na seleção tanto de plantas individuais (autógamas ou as de propagação vegetativa) quanto de populações heterozigóticas (alógamias).

Dentre os genótipos caracterizados molecularmente (Tabela 1), foi observada uma elevada variação (0 a 1) em termos de coeficiente de dissimilaridade genética tanto entre quanto dentro dos *bulks* de plantas. Por outro lado, pode ser verificado que apenas três pares (20% do número total) de genótipos (BRS 7 ‘Taim’ x BRS 6 ‘Chuí’, BRS Pelota x BRS Ligeirinho e BRS Atalanta x BRS 6 ‘Chuí’) apresentaram coincidência perfeita

nas estimativas dos valores para o índice de dissimilaridade genética. Na mesma tabela, pode ser verificado ainda que o coeficiente de dissimilaridade médio variou de 0,42 a 0,52, entre as repetições (*bulks*). Provavelmente esta técnica acessou os mesmos locos nos dois *bulks* de plantas para estas três combinações de genótipos. Houve coincidência em termos de significância nas estimativas destes coeficientes de dissimilaridade genética ($P < 0,05$), comparativamente aos valores da dissimilaridade morfológica, estimadas por Mahalanobis para as combinações genotípicas que coincidiram os valores nos diferentes *bulks* avaliados: *i*) BRS 7 ‘Taim’ x BRS 6 ‘Chuí’; *ii*) BRS Pelota x BRS Ligeirinho; e *iii*) BRS Atalanta x BRS 6 ‘Chuí’ (Tabela 2).

Tabela 1. Estimativa da dissimilaridade genética baseada em dados moleculares por meio da técnica de AFLP (D_{mo}) entre seis genótipos de arroz desenvolvido pela Embrapa Clima Temperado (CPACT). Pelotas, RS, 2005.

Genótipos (i/j)		<i>Bulk</i> -1	<i>Bulk</i> -2
BRS 7 ‘Taim’	BRS Pelota	0,25	0,50
	BRS Ligeirinho	0,33	0,43
	BRS Atalanta	0,33	0,50
	BRS 6 ‘Chuí’	1,00	1,00
	BRS Firmeza	0,50	0,43
BRS Pelota	BRS Ligeirinho	0,10	0,10
	BRS Atalanta	0,10	0,00
	BRS 6 ‘Chuí’	0,67	0,71
	BRS Firmeza	0,67	0,10
BRS Ligeirinho	BRS Atalanta	0,00	0,10
	BRS 6 ‘Chuí’	0,71	0,67
	BRS Firmeza	0,71	0,20
BRS Atalanta	BRS 6 ‘Chuí’	0,71	0,71
	BRS Firmeza	0,71	0,10
BRS 6 ‘Chuí’	BRS Firmeza	1,00	0,67
Dissimilaridade Média		0,52	0,42

Tabela 2. Estimativas dos coeficientes de dissimilaridade para dados moleculares (D_{mo}) e morfológico (D_{mf}), baseadas em AFLP, graus de liberdade (GL) e desvio padrão (SD), utilizando como repetições dois *bulks* de cinco plantas cada um, separadamente para cada genótipo arroz desenvolvido pela Embrapa Clima Temperado (CPACT). Pelotas, RS, 2005.

Genótipos		D_{mo}		D_{mf}	
I	j	¹ Estimativa	² Pr > t	Pr>F	
BRS 7 'Taim'	BRS 6 'Chuí'	1,00	**	A ¹	14.910*
BRS 6 'Chuí'	BRS Firmeza	0,84	**	AB	1.120
BRS Atalanta	BRS 6 'Chuí'	0,71	**	ABC	183.837*
BRS Pelota	BRS 6 'Chuí'	0,69	**	ABC	19.776*
BRS Ligeirinho	BRS 6 'Chuí'	0,69	**	ABC	28.410*
BRS 7 'Taim'	BRS Firmeza	0,47	**	BCD	20.250*
BRS Ligeirinho	BRS Firmeza	0,46	**	BCDE	40.284*
BRS 7 'Taim'	BRS Atalanta	0,42	*	CDE	133.383*
BRS Atalanta	BRS Firmeza	0,41	*	CDE	212.639*
BRS 7 'Taim'	BRS Ligeirinho	0,38	*	CDE	10.671*
BRS Pelota	BRS Firmeza	0,39	*	CDE	29.655*
BRS 7 'Taim'	BRS Pelota	0,38	*	CDE	5.372
BRS Pelota	BRS Ligeirinho	0,10	ns	DE	1.183
BRS Pelota	BRS Atalanta	0,05	ns	E	86.401*
BRS Ligeirinho	BRS Atalanta	0,05	ns	E	6.810

¹: Estimativas seguidas pelas mesmas letras não diferem significativamente pelo teste dms ($P>0,05$).

²: $H_0: D_{mo}=0$, pelo teste t.

³: $H_0: D_{mf}=0$, pelo teste F ($P<0,05$).

Opar de genótipos (BRS Pelota x BRS Ligeirinho) similar molecularmente (0,10 significativamente igual a zero, pelo teste t, a 5% de probabilidade) mostrou um valor de 1.183 (significativamente igual a zero, pelo teste F, a 5% de probabilidade) para o coeficiente de dissimilaridade morfológica, já os pares de genótipos BRS Atalanta x BRS 6 'Chuí' molecularmente (0,71, significativamente diferente de zero, pelo teste t, a 5% de probabilidade) e morfológicamente (183.837 significativamente diferente de zero, pelo teste F, a 5% de probabilidade) apontaram um valor significativo pelo teste F e BRS 7 'Taim' x BRS 6 'Chuí' molecularmente (1,00 significativamente diferente de zero, pelo teste t, a 5% de probabilidade) e morfológicamente (14.910 significativamente diferente de zero, pelo teste F, a 5% de probabilidade), respectivamente. Este fato revela que a consistência encontrada entre os *bulks* nos valores estimados dos coeficientes de dissimilaridades genético coincidiu perfeitamente com os resultados encontrados morfológicamente

para os coeficientes de dissimilaridades, especialmente para esses três pares de genótipos. Por exemplo, o par de genótipos BRS 7 'Taim' x BRS 6 'Chuí' apontou um valor significativamente diferente de zero tanto pelo teste t quanto pelo teste F (5% de probabilidade), o que indica que este par de genótipos pode ser considerado dissimilar tanto molecular quanto morfológicamente.

A exatidão das técnicas moleculares, em especial AFLP dependem muito de alguns fatores a considerar: *i*) heterozigose residual; *ii*) imprecisão humana; *iii*) qualidade DNA; *iv*) qualidade dos reagentes; *v*) aferição dos equipamentos, como por exemplo, termociclador; e *vi*) qualidade da revelação; e *vii*) avaliação da presença e ausência, etc. Mesmo assim, estas técnicas têm se mostrado eficientes de acordo com autores como Graham, McNilcil e McNicol (1996). Por exemplo, o coeficiente de similaridade estimado entre as cultivares a partir de dados originários do estudo

do pedigree, foi menor que o grau de similaridade estimado por marcadores RAPD. Assim estes autores concluíram que nos dados de pedigree há grandes diferenças entre os cultivares, que na verdade são semelhantes quando o estudo utiliza dados de marcadores moleculares. Outro fato que deve ser mencionado é que o resultado positivo da técnica de AFLP depende essencialmente da qualidade e da quantidade do DNA. Ou seja, um DNA com alto nível de pureza é necessário para garantir uma digestão completa pelas enzimas de restrição em todas as amostras de DNA (VOS et al., 1995; FERREIRA; GRATTAPAGLIA, 1998).

Durante a extração do DNA de um grande número genótipos, este pode variar indiscriminadamente tanto em qualidade quanto em quantidade. Por outro lado, a ausência de correlação entre coeficientes de dissimilaridade genética (D_{mo}) com a dissimilaridade morfológica (D_{mf}) pode estar relacionada direta ou indiretamente aos seguintes fatores: *i*) número limitado de caracteres morfológicos avaliados; *ii*) pequena variabilidade genética entre estes caracteres; *iii*) limitado número de genes responsáveis pela expressão desses caracteres morfológicos; e *iv*) possibilidade real de genes epistáticos (SCHUT; QI; STAM, 1997). Além da técnica de AFLP acessar regiões que determinam a expressão de caracteres não adaptativos, esta técnica pode acessar regiões genômicas não funcionais.

A técnica de AFLP, polimorfismo de comprimento de fragmentos amplificados, envolve basicamente nove etapas que incluem: *i*) extração do DNA; *ii*) digestão do DNA com enzimas de restrição, uma de corte raro (*EcoRI*) e outra de corte freqüente (*MseI*); *iii*) ligação de adaptadores específicos; *iv*) pré amplificação; *v*) amplificação seletiva de fragmentos com *primers* específicos; *vi*) eletroforese; *vii*) amplificação; *viii*) revelação; *ix*) avaliação; e *x*) testar as hipótese do experimento. O polimorfismo obtido com AFLP está baseado em diferenças entre genótipos, no que se refere a distribuição dos sítios de restrição e na amplificação diferencial de fragmentos (MILACH, 1998).

De modo que, as diferenças encontradas entre os pares de genótipos podem ser em virtude do erro experimental e não do efeito genotípico em si. O erro experimental, denominado também de resíduo, pode ser conceituado como sendo a variação entre as unidades experimentais atribuíveis as fontes de variação que não são controladas nem por controle local ou por controle estatístico. Sendo que a comparação entre tratamentos num experimento consiste em comparar a variação entre unidades experimentais com diferentes tratamentos com a variação entre unidades experimentais com um mesmo tratamento, caso a primeira fonte de variação seja superior à segunda fonte de variação pode ser atribuída tal desempenho superior a diferenças reais entre efeitos de tratamentos com um grau de probabilidade (significância).

Cada uma das etapas da técnica de AFLP, que são compostas por passos intermediários, além de passos subjetivos e visuais, como por exemplo, a digestão do DNA (devido a sua qualidade), torna esta técnica sensível a erro (s). Deste modo, a variação extrínseca pode ocorrer por: *i*) contaminação genética dos genótipos, ou seja, pureza genética das sementes ou das plantas que serão utilizadas para extração do DNA?; *ii*) tamanho da amostra ideal para extração do DNA, uma planta ou um conjunto (*bulk*) de plantas?; *iii*) facilidade da extração de DNA, normalmente as plantas da família das leguminosas (ex. feijão) são bem mais simples comparativamente a família das gramíneas (ex. aveia), provavelmente em virtude da relação C/N, tornando-se mais acessível à ruptura das fibras, com reflexo, direto na qualidade geral do DNA (integridade e contaminação)?.

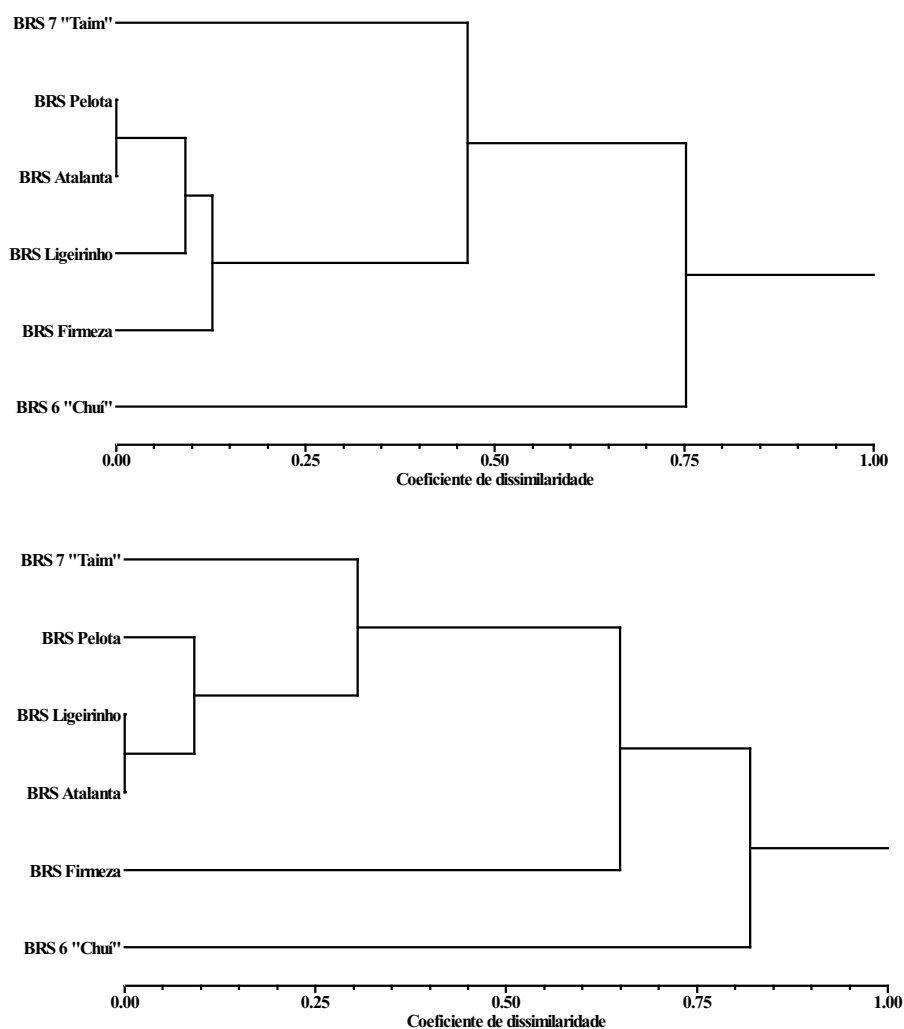
De modo geral, pode ser verificada certa similaridade entre os dendrogramas, ou seja, os dendrogramas apresentaram estruturas de agrupamento similares (Figura 1). Os resultados comparados demonstraram que a disposição dos genótipos de arroz nas figuras não alterou o agrupamento mais divergente, ou seja, o par de genótipos BRS 6 ‘Chuí’ x BRS 7 ‘Taim’ foi sempre

o mais dissimilar, diferindo somente no grau de dissimilaridade entre os agrupamentos formados. Os dendrogramas obtidos por meio dos *bulk-1* e *bulk-2* foram capazes de separar os genótipos BRS 7 'Taim' x BRS 6 'Chuí'. No entanto, se for predefinido um ponto de corte (coeficiente de dissimilaridade médio), em ambos o dendrogramas, podemos observar uma discrepância no coeficiente de dissimilaridade entre os agrupamentos formados. Por exemplo, o dendrograma referente ao primeiro *bulk* (Figura 1-A) mostrou um coeficiente de dissimilaridade médio de 0,52, aproximadamente, indicando o par de genótipos BRS Firmeza x BRS 6 'Chuí' como dissimilar. Para o dendrograma referente ao segundo *bulk* (Figura 1-B), a dissimilaridade média foi de 0,42 e, neste caso, para este mesmo par de genótipos (BRS Ligeirinho x BRS Firmeza) é indicado que estes pertençam ao mesmo agrupamento (mesmo *cluster*) (Figura 1). De fato este resultado pode ser corroborado nos coeficientes de dissimilaridade obtidos nas duas repetições executadas, onde na primeira e na segunda repetição os valores encontrados foram de 0,71 e 0,20, respectivamente, apontando assim, uma variação percentual de -72%, aproximadamente. Ainda, a partir da Tabela 2, pode ser observado dois resultados relevantes, especificamente para este par de genótipos (BRS Ligeirinho x BRS Firmeza): *i*) o coeficiente de dissimilaridade estimados por meio das repetições (0,46) é significativamente diferente de zero, ou seja, este par de genótipos revelou uma divergência genética significativa pelo teste t ($P=0,0065$); *ii*) o par de genótipos BRS Ligeirinho x BRS Firmeza não diferem significativamente do par de genótipos BRS Firmeza x BRS 6 'Chuí'. Esse resultado reforça a importância de testar os contrastes de interesse, refletindo diretamente na possibilidade de distinguir, eficientemente, os agrupamentos formados (variabilidade genética

acessada). Este resultado, referente ao teste de comparação de médias evidencia um poder de discriminar por completo a variabilidade genética acessada por marcadores moleculares.

Ainda na tabela 2 pode ser constatado que apenas três combinações de genótipos apontaram um coeficiente de dissimilaridade igual à zero BRS Pelota x BRS Ligeirinho (0,10), BRS Pelota x BRS Atalanta (0,05) e BRS Ligeirinho x BRS Atalanta (0,05), ou seja, essas três combinações genotípicas não diferem estatisticamente (similares) pelo teste t. Podemos concluir que os cinco primeiros pares de combinações genotípicas testadas mostraram um grau de divergência genética significativamente diferente de zero, pelo teste t. No entanto, quando estas cinco combinações são avaliadas comparativamente não revelam diferença significativa pelo teste dms, ou seja, o grau de divergência genética estimado entre o par de genótipos BRS 7 'Taim' x BRS 6 'Chuí' (1,00) não difere estatisticamente do grau de divergência genética entre os genótipos BRS Ligeirinho x BRS 6 'Chuí' (0,69). Evidentemente que esta diferença entre os valores estimados deve ser atribuída ao erro experimental. É importante destacar que o fato desse tipo de análise por meio de dendrograma não apresentar um critério objetivo para identificação dos grupos, fato que dificulta muito a interpretação dos resultados obtidos (MEYER, 2002; CONTI; MINAMI; TAVARES, 2002). Todas estas causas de variações perturbadoras potenciais são passíveis de redução e, até de controle num grau aceitável, mas nem sempre são realizadas como rotinas em laboratórios de análise molecular, fato que pode involuntariamente reduzir tanto a precisão quanto a exatidão de parâmetros genéticos estimados molecularmente, como por exemplo, coeficiente de repetibilidade (correlação entre as medidas feitas no mesmo indivíduo).

Figura 1. Dendrograma de seis genótipos de arroz obtido com análise de AFLP, provenientes do primeiro e do segundo *bulk* (primeira e segunda repetição, respectivamente) de cinco plantas, utilizando o índice de dissimilaridade de DICE (1945) e o método de agrupamento UPGMA. O valor do coeficiente de correlação cofenética (r) foi de 0,64 (primeiro *bulk*) e 0,72 (segundo *bulk*). FAEM/UFPel, Pelotas, 2005.



Nas análises de variância para o caráter dissimilaridade genética, tanto para dados moleculares quanto morfológicos, utilizando as duas repetições para estimar o erro experimental, pode ser observada diferença significativa ($P < 0,05$), independentemente do tipo de distância estimada. De fato, analisando conjuntamente os dois tipos de dissimilaridade genética estimada, pode ser observada uma concordância em todos os pares de combinações avaliados, exceto para os pares BRS 6 'Chuí' x BRS Firmeza, BRS 7 'Taim' x BRS Pelota e BRS Pelota x BRS Atalanta. De forma

prática, os genótipos BRS 6 'Chuí' e BRS Firmeza são do tipo moderno filipino e moderno americano, respectivamente, sendo assim, estes genótipos deveriam mostrar uma diferença significativa em relação ao seu grau de dissimilaridade genética, fato que foi corroborado pela análise de AFLP.

Esta coincidência de aproximadamente 80% fortalece a utilização das repetições em análises moleculares, sendo nestes casos, uma estratégia para minimizar distorções provocadas por causa de variações não genéticas, ou seja, integralmente

atribuídas ao ambiente (erro experimental). Além de fornecer uma probabilidade que possibilita testar as hipóteses básicas em qualquer tipo de experimento conduzido, seja a campo ou em laboratório.

Esta proposta alternativa pode ser uma estratégia para minimizar as distorções observadas nos valores de dissimilaridades genéticas, principalmente quando se trabalha no campo da biologia, em que as variâncias preditas são os resultados da ação simultânea de inúmeros fatores, muito deles não mensuráveis. A utilização de repetições em análise de AFLP pode ser um procedimento efetivo e deve ser empregado pelos melhoristas na avaliação da distância genética. Contudo estes índices de dissimilaridade genética devem ser submetidos à análise de variância e não apenas a apresentação dos índices, sem nenhum teste adicional, o que inviabiliza muitas inferências que são feitas. De fato este procedimento possibilita testar as hipóteses básicas, com certa confiabilidade nos resultados obtidos quanto à repetibilidade. Além disso, sempre que possível, as pressuposições básicas da análise de variância devem ser checadas para maior segurança nas inferências a serem feitas a partir dos testes estatísticos realizados.

Observando os dados inseridos na Tabela 3, pode ser notado o valor dos componentes de variância em arroz avaliada para o caráter (índice de dissimilaridade). O objetivo mais relevante para a utilização do modelo aleatório é a estimação destes componentes de variância que estão ligados diretamente às estimativas dos

parâmetros genéticos. Ainda na tabela 3, pode ser observado que os componentes de variância para genótipos (0,06) e resíduos (0,04) apontaram os maiores valores preditos; de modo que, o valor predito para o componente de variância dos genótipos superou em 1,6 vezes o valor do resíduo, aproximadamente. Assim o valor estimado para o coeficiente de repetibilidade (estimado por $\rho = \sigma^2_g \div \sigma^2_e + \sigma^2_g$), apontou um valor de aproximadamente 62%, onde o coeficiente de repetibilidade expressa a proporção da variância total, que pode ser explicada pelas variações proporcionadas pelo genótipo e pelas alterações permanentes atribuídas ao resíduo comum ou ambiente comum (CRUZ; REGAZZI, 1997). Isto revela a importância de se estimar o resíduo mesmo em ensaios moleculares. Pois todas as etapas que são subjetivas, como as citadas anteriormente, são passíveis de variação aleatória, reduzindo assim, consideravelmente sua verossimilhança genética. Esse valor estimado para o coeficiente de repetibilidade, relativamente baixo, indica que não houve regularidade na repetição do coeficiente de dissimilaridade genética de um *bulk* de plantas para outro. A utilização de quatro *bulks* de plantas pode, entretanto selecionar indivíduos com 85%, aproximadamente, de certeza da predição de seu valor real, com base nesta estimativa do erro experimental. Ou seja, quanto mais preciso for o experimento, menor é o erro experimental, sendo que este fato possibilita a redução do número de repetições.

Tabela 3. Componentes de variância dos efeitos aleatórios estimados pelo método da máxima verossimilhança restrita (REML). Pelotas, 2005.

Componentes de variância	Valores preditos	Erro Padrão predito
Genótipos	0,06031	0,03070
Repetições	0,00324	0,00816
Resíduo	0,03748	0,01416

O melhorista quando determina um conjunto de genótipos em *pools* heteróticos tem como objetivo principal a ampliação da variabilidade genética em populações segregantes. De maneira evidente, se espera que o bom desempenho manifestado em certos caracteres que compõem o indivíduo ideal reflita o potencial do genótipo a ser selecionado pelo melhorista como um todo, sendo que a veracidade desta expectativa pode ser confirmada pelo coeficiente de repetibilidade da distância genética estimada entre genótipos avaliados. Esta proposta possui uma relevância de apresentar um critério objetivo para identificação tanto entre grupos quanto dentro de grupos heteróticos, podendo se tornar assim, uma ferramenta importante para este tipo de análise molecular. Deve ser ressaltado, que neste trabalho como eram conhecidos *a priori* os genótipos avaliados, a interpretação dos dendrogramas tornou-se mais simples, porém, nem sempre este fato ocorre na prática.

Conclusões

O valor estimado para o erro experimental, em análises moleculares, proporciona uma confiabilidade maior em relação às hipóteses testadas.

Os testes de médias associados ao teste t para os coeficientes de dissimilaridades permite uma interpretação mais consistente dos resultados obtidos, tornando as interpretações dos dados por meio de dendrogramas mais informativas.

Agradecimentos

Agradecemos à UDESC, CNPq, CAPES e FAPESC, pela concessão de bolsa e apoio financeiro.

Referências

BARBOSA NETO, J. F. Seleção assistida por marcadores moleculares. In: MILLACH, S. C. K. (Ed.). *Marcadores de DNA em plantas*. Porto Alegre: UFRGS, 1998. p. 75-

88.

BENSCH, S.; AKESSON, M. Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology*, Vancouver, v. 14, p. 2899-2914, 2005.

BONIN, A.; BELLEMAIN, E.; BRONKEN EIDENSEN, P.; POMPANON, F.; BROCHMANN, C.; TABERLET, P. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, Vancouver, n. 13, p. 3261-3273, 2004.

CARDELLINO, R. A.; SIEWERDT, F. Utilização adequada e inadequada dos testes de comparação de médias. *Revista da Sociedade Brasileira de Zootecnia*, Viçosa, v. 21, n. 6, p. 985-995, 1992.

CASTELO-BRANCO, J. S. Extração de DNA de vegetais. In: ZIMMER, P. D.; OLIVEIRA, A. C.; MALONE, G. *Ferramentas da biotecnologia no melhoramento genético vegetal*. Pelotas: Editora e Gráfica Universitária, p. 17-39, 2005.

CONTI, J. H.; MINAMI, F.; TAVARES, F. C. A. Comparação de caracteres morfológicos e agrônômicos com moleculares em morangueiro cultivados no Brasil. *Horticultura Brasileira*, Brasília, v. 20, n. 3, p. 419-423, 2002.

CRESTE, S.; TULMANN-NETO, A.; FIGUEIRA, A. Detection of single sequence repeat polymorphism in denaturing polyacrylamide sequencing gels by silver staining. *Plant Molecular Biology Reporter*, Canadá, v. 19, n. 4, p. 1-8, 2001.

CRUZ, C. D.; REGAZZI, A. J. *Modelos biométricos aplicados ao melhoramento genético*. 2. ed. Viçosa: UFV, 1997. 390 p.

CUEVAS-PÉREZ, F. E.; GUIMARÃES, E. P.; BERRIO, L. E.; GONZÁLES, D. I. Genetic base of irrigated rice in Latin America and the Caribbean, 1971 to 1989. *Crop Science*, Madison, v. 32, n. 4, p. 1054-1059, 1992.

DICE, L. R. Measures of the amount of ecological association between species. *Ecology*, Washington, v. 26, n. 3, p. 297-307, 1945.

FERREIRA, M. E.; GRATTAPAGLIA, D. *Introdução ao uso de marcadores RAPD e RFLP em análise genética*. 3. ed. Brasília: EMBRAPA: Embrapa-Cenargen, 1998. 220 p.

GETTING STARTED WITH THE SAS® *Learning Edition*. Cary, NC: SAS Institute. 2002. 83 p.

GRAHAM, J.; McNICIL, R. J.; McNICOL, J. W. A comparison of methods for the estimation of genetic diversity in strawberry cultivars. *Theoretical and Applied Genetics*, Berlin, v. 93, n. 3, p. 402-406, 1996.

- GUIDOLIN, A. F.; OLIVEIRA, A. C. de; TERRES, A. L.; COSTA, F. L. C. da. Caracterização eletroforética das cultivares de arroz irrigado em uso no RS. *Lavoura Arrozeira*, Porto Alegre, v. 47, n. 414, p. 3-5, 1994.
- ICEPA. *Síntese anual da agricultura de Santa Catarina*. 2009. Disponível em: <http://www.icepa.com.br>. Acesso em: 05 abr. 2011.
- JONES, C. J.; EDWARD, K. J.; CASTAGLIONE, S.; WINFIELD, M. O.; SALA, F.; WIEL, C. van de; BREDEMEIJER, G.; VOSMAN, B.; MATTHE, M.; DALY, A.; BRETTSCHEIDER, R.; BETTINI, P.; BUIATTI, M.; MAESTRI, E.; MALCEVSCHI, A.; MARMIROLI, N.; AERT, R.; VOLCKAERT, G.; RUEDA, J.; LINACERO, R.; VAZQUEZ, A.; KARP, A. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding*, Spain, v. 3, p. n. 5, 381-390, 1997.
- LITTELL, R. C.; MILLIKEN, G. A.; STROUP, W. W.; WOLFINGER, R. D. *Sas system for mixed models*. Cary, NC. Sas Institute Inc. 1996. 633 p.
- MEYER, A. S. *Comparação de coeficientes de similaridade usados em análise de agrupamento com dados de marcadores moleculares dominantes*. 2002. Dissertação (Mestrado em Agronomia) – Universidade de São Paulo, São Paulo.
- MILACH, S. C. K. Principais tipos de marcadores e suas características. In: MILLACH, S. C. K. *Marcadores de DNA em plantas*. Porto Alegre: UFRGS, 1998. p. 17-28.
- MOITA NETO, J. M.; MOITA, G. C. Uma introdução à análise exploratória de dados multivariados. *Química Nova*, Florianópolis, v. 21, n. 4, p. 467-469, 1997.
- OLIVEIRA, A. C. Mapeamento comparativo. In: MILLACH, S. C. K. *Marcadores de DNA em plantas*. Porto Alegre: UFRGS, 1998. p. 57-66.
- OLIVEIRA, T. S.; HOFFMANN, L. V.; ALVES, P. F.; LUCENA, V. F.; FILHO, J. L. S. Validação de métodos laboratoriais aplicadas a análises com marcadores microssatélites. *Revista Ciência Agronômica*, Fortaleza, v. 41, n. 2, p. 279-284, 2010.
- PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. *Biometrika*, Oxford, v. 58, p. 545-554, 1971.
- RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. *Experimentação em genética e melhoramento de plantas*. São Paulo: UFLA, 2000, 293 p.
- RANGEL, P. H. N.; GUIMARÃES, E. P.; NEVES, P. C. F. Base genética das cultivares de arroz (*Oryza sativa*) irrigado do Brasil. *Pesquisa Agropecuária Brasileira*, Brasília, v. 31, n. 5, p. 349-357, 1996.
- SCHUT, J. W.; QI, X.; STAM, P. Association between relationship measures based on AFLP markers, pedigree data and morphological traits in barley. *Theoretical and Applied Genetics*, Berlin, v. 95, p. 1161-1168, 1997.
- SEARLE, S.; CASELA, G.; McCULLOCH, C. E. *Variance components*. New York: John Wiley, 1992, 501 p.
- SILVA, G. O. da; PEREIRA, A. da S.; SOUZA, V. Q. de; CASTRO, C. M.; CARVALHO, F. I. F. de; VIEIRA, E. A. Distâncias genéticas entre genótipos de batata a partir de dados morfológicos, moleculares e genealógicos. *Semina: Ciências Agrárias*, Londrina, v. 30, p. 983-992, 2009. Suplemento 1.
- SILVA, R. G. da. *Manual de procedimentos em análise por quadrados mínimos*. Jaboticabal: FUNEP, 1993, 169 p.
- VIEIRA, E. A.; CARVALHO, F. I. F.; OLIVEIRA, A. C.; BENIN, G.; ZIMMER, P. D.; SILVA, J. A. G.; MARTINS, A. F.; BERTAN, I.; SILVA, G. O.; SCHMIDT, D. A. M. Comparação entre medidas de distancia genealógica, morfológica e molecular em aveia em experimentos com e sem a aplicação de fungicida. *Bragantia*, Campinas, v. 64, n. 1, p. 51-60, 2005.
- VOS, P.; HOGERS, R.; BLEEKER, M.; REIJANS, M.; LEE, T. van de; HORNES, M.; FRITERS, A.; POT, J.; PALEMAN, J.; KUIPER, M.; ZABEAU, M. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, Oxford, v. 23, n. 21, p. 4407-4414, 1995.

