DIVERGÊNCIA GENÉTICA E FENÉTICA MULTIVARIADA NA
PREDIÇÃO DE HÍBRIDOS E PRESERVAÇÃO DE GERMOPLASMA
DE CACAU (*Theobroma cacao* L.)

LUIZ ANTÔNIO DOS SANTOS DIAS
Engenheiro Agrônomo

Orientador: Prof. Dr. PAULO YOSHIO KAGEYAMA

Tese apresentada à Escola Superior de
Agricultura "Luiz de Queiroz", da
Universidade de São Paulo, para obtenção
do título de Doutor em Agronomia. Área de
concentração: Genética e Melhoramento de
Plantas.

PIRACICABA
Estado de São Paulo - Brasil
Março - 1994

Ficha catalográfica preparada pela Seção de Livros da
Divisão de Biblioteca e Documentação - PCLQ/USP

Dias, Luiz Antônio dos Santos

D541d Divergência genética e fenética multivariada na
predição de híbridos e preservação de germoplasma
de cacau (*Theobroma cacao* L.). Piracicaba, 1994.
94p.

Tese - ESALQ
Bibliografia.

1. Análise multivariada 2. Cacau - Divergência gené-
tica 3. Cacau - Germoplasma 4. Híbrido de cacau I. Es-
cola Superior de Agricultura Luiz de Queiroz, Piracica-
ba

CDD. 633.74

DIVERGÊNCIA GENÉTICA E FENÉTICA MULTIVARIADA NA
PREDIÇÃO DE HÍBRIDOS E PRESERVAÇÃO DE GERMOPLASMA
DE CACAU (*Theobroma cacao* L.)

LUIZ ANTÔNIO DOS SANTOS DIAS

Aprovada em: 05.04.1994

Comissão julgadora:

Prof. Dr. Paulo Yoshio Kageyama

ESALQ/USP

Prof. Dr. Roland Vencovsky

ESALQ/USP

Prof. Dr. Cassio Roberto de Melo Godoi

ESALQ/USP

Prof. Dr. Cosme Damião Cruz

UFV

Prof. Dr. Mário Luiz Teixeira de Moraes

FEIS/UNESP

Prof. Dr. Paulo Yoshio Kageyama
Orientador

"Chove ouro baço, mas não no lá-fora... É em mim... Sou a hora,
E a hora é de assombros e toda ela escombros dela..."

Fernando Pessoa

"Hora Absurda" - 1913.

Numa deferência especial à colega e esposa Denise Cunha Fernandes dos Santos Dias,

Aos nossos filhos Raquel e Pedro,

Aos meus pais e irmãos.

AGRADECIMENTOS

À Comissão Executiva do Plano da Lavoura Cacaueira - CEPLAC, à Escola Superior de Agricultura "Luiz de Queiroz" - ESALQ/USP e, em especial, o seu Departamento de Genética por possibilitarem a realização do curso;

Ao estimado Professor Dr. Paulo Yoshio Kageyama pela orientação, ensinamentos e estímulos transmitidos ao longo dos cursos de Mestrado e Doutorado;

À Coordenadoria de Aperfeiçoamento de Pessoal de Ensino Superior - CAPES pela concessão da bolsa de estudos, sem a qual o treinamento não seria possível;

Ao Professor Dr. Roland Vencovsky do Departamento de Genética e ao Professor Dr. Cassio Roberto de Melo Godoi do Departamento de Matemática da ESALQ pelo estímulo e discussões referentes ao tema da tese. Ao Pesquisador Marcos Deon Vilela de Resende da EMBRAPA/CNPFloresta pela leitura crítica da redação final;

Aos colegas da Estação Experimental "Filogônio Peixoto" - ESFIP, em especial a Carlos Alberto Spaggiari Souza e Sebastião Geraldo Augusto, pelo apoio logístico prestado durante todo o curso;

Ao amigo e colega Dr. Jairo Cunha pelo apoio constante;

Aos colegas do curso de pós-graduação em Genética e Melhoramento de Plantas do Departamento de Genética da ESALQ, com os quais tive o prazer de desfrutar do convívio. Tempos de profunda amizade que recordarei com orgulho;

As funcionárias e aos colegas do Laboratório de Biologia Reprodutiva e Genética de Espécies Arbóreas;

Aos Professores do Departamento de Genética pelos seus ensinamentos e aos funcionários pela cordialidade e presteza.

SUMÁRIO

	Página
RESUMO-----	VII
SUMMARY-----	IX
1. INTRODUÇÃO -----	01
2. REVISÃO DE LITERATURA -----	05
2.1. A seleção de genitores e a hibridação -----	05
2.2. Divergência genética por análises multivariadas-----	07
2.2.1. Análise da variância multivariada -----	09
2.2.2. Análise de cluster -----	10
2.2.3. Análise por componentes principais -----	13
2.2.4. Análise por variáveis canônicas -----	17
2.3. Importância relativa dos caracteres para divergência -----	20
2.4. Estabilidade temporal da divergência expressa por D^2 -----	22
3. MATERIAL E MÉTODOS -----	24
3.1. Estudo I: Predição da performance de híbridos por divergência dos genitores -----	24
3.1.1. Material de estudo -----	24
3.1.2. Metodologias-----	25
3.1.2.1. Análise de cluster -----	27
3.1.2.2. Análise por componentes principais -----	29
3.1.2.3. Análise por variáveis canônicas -----	30
3.2. Estudo II: Divergência fenética entre clones de cacauero -----	32
3.2.1. Material de estudo -----	32
3.2.2. Metodologias-----	34
4. RESULTADOS E DISCUSSÃO -----	36
4.1. Estudo I: Predição da performance de híbridos por divergência dos genitores -----	36
4.1.1. Análises da variância uni e multivariada-----	36
4.1.2. A divergência genética por diferentes estatísticas multivariadas-----	37
4.1.3. A importância dos caracteres para a divergência-----	40
4.1.4. Eficiência da predição da performance de híbridos -----	43
4.1.5. Relação entre distância geográfica e divergência genética -----	52
4.1.6. Estabilidade temporal da divergência-----	54

4.2. Estudo II: Divergência fenética entre clones de cacaueiro -----	58
4.2.1. A distância Euclidiana média -----	59
4.2.2. Distâncias Euclidianas intra e intergrupos. -----	63
4.2.3. A divergência fenética por meio dos componentes principais -----	64
4.2.4. Relação entre distância geográfica e divergência fenética -----	71
4.2.5. Importância e descarte de caracteres -----	73
4.3. Considerações finais -----	77
5. CONCLUSÕES -----	80
6. REFERÊNCIAS BIBLIOGRÁFICAS -----	82

DIVERGÊNCIA GENÉTICA E FENÉTICA MULTIVARIADA NA PREDIÇÃO DE HÍBRIDOS E PRESERVAÇÃO DE GERMOPLASMA DE CACAU (*Theobroma cacao* L.)

Autor: LUIZ ANTÔNIO DOS SANTOS DIAS

Orientador: Prof. Dr. PAULO YOSHIO KAGEYAMA

RESUMO

No presente trabalho, dois estudos foram desenvolvidos abordando o tema comum da divergência entre genótipos de cacau. Análises multivariadas foram implementadas ao programa de melhoramento genético visando aumento de produção e junto ao programa de preservação de recursos genéticos de cacau. O estudo I objetivou a predição da performance de híbridos por divergência dos genitores S_1 . No estudo II, a divergência fenética entre clones de cacauzeiros foi quantificada.

O estudo I envolveu cultivares e híbridos avaliados em sistemas de cruzamentos dialélicos. Assim, os cultivares autocompatíveis CC 41, SIAL 169, CEPEC 1, ICS 1 e SIC 19 e seus 20 híbridos foram ensaiados no Centro de Pesquisas de Cacau, CEPEC, em Itabuna, Bahia. O ensaio foi instalado no ano de 1975, em blocos casualizados com quatro repetições e parcelas de 16 plantas. As sementes dos cruzamentos foram obtidas por polinizações manuais controladas nos cultivares clonais. Por sua vez, as sementes dos cultivares foram obtidas por autofecundações dos clones correspondentes. Dados de componentes de produção dos anos agrícolas de 1986 a 1990 foram utilizados para análise. Para avaliação dos cultivares e híbridos foram empregados o número de frutos sadios (NFSP) e colhidos (NFCP), o peso de sementes úmidas por planta (PSUP) e por fruto (PSUF) e a percentagem de frutos doentes por planta (PFDP).

As análises da variância uni e multivariada detectaram diferenças altamente significativas entre os cultivares. A divergência entre eles foi quantificada pelas distâncias de Mahalanobis (D^2), Euclidiana (d_e), Euclidiana estimada de componentes principais (d_{cp}) e Euclidiana estimada de variáveis canônicas (d_{vc}). As estatísticas que consideram a matriz de dispersão residual, no caso D^2 e d_{vc} , apresentaram estreita associação ($r = 0,95$). O mesmo ocorreu entre as estatísticas d_e e d_{cp} ($r = 0,97$). Fraca concordância, no entanto, foi observada entre esses dois pares de estatísticas. A aplicação

dos métodos de Tocher, do vizinho mais próximo e do gráfico de variáveis canônicas evidenciou dois grupos divergentes entre si. O primeiro envolvendo os cultivares CC 41, SIAL 169, CEPEC 1 e SIC 19 e o segundo formado pelo ICS 1.

Componentes principais e variáveis canônicas identificaram o caráter PSUP entre os caracteres de menor importância para a divergência. A distância geográfica entre os cultivares esteve relacionada à distância genética, com base no agrupamento promovido pelo método de Tocher. O padrão estável da divergência genética entre os cultivares, nos cinco anos, possibilitou o emprego de um simples ano, de ótimas condições climáticas, para a estimação da divergência.

A performance média e heterótica dos híbridos foi satisfatoriamente predita pela divergência entre os genitores S_1 , expressa por D^2 . Correlações de Spearman moderadas e altas foram obtidas entre D^2 e PSUP ($r = 0,68$) e entre D^2 e PSUF ($r = 0,76$), para a performance média. Correlação alta de 0,65 foi obtida entre D^2 e PSUP, com relação à performance heterótica dos híbridos. A heterose foi medida como desvio do híbrido F_1 em relação ao correspondente genitor S_1 superior. Os melhores híbridos, com referência a PSUP e PSUF, surgiram dos cruzamentos envolvendo o cultivar de maior divergência (ICS 1). Também o melhor híbrido, resultou do cruzamento envolvendo o par com a segunda maior divergência, no caso SIAL 169 e ICS 1. Por fim, o híbrido de pior performance média surgiu do cruzamento entre o par de cultivares com a segunda maior similaridade, ou seja, CEPEC 1 e SIC 19.

No estudo II, a divergência fenética entre 26 clones de cacau do germoplasma do CEPEC foi quantificada por análise de cluster e de componentes principais. Nos 26 clones das séries SIC e SIAL, selecionados na população de cacauzeiros forasteiros amazônicos do sul da Bahia, quatro grupos de similaridade fenética foram discriminados. Esses grupos foram detectados pelo método de Tocher. A presença de divergência entre esses clones locais contrariou o paradigma da uniformidade das populações cacaueiras baianas e sugeriu a viabilidade das estratégias de seleção e hibridação entre essas populações.

Dentre os 13 caracteres utilizados na avaliação dos 26 clones, quatro foram descartados, por serem redundantes e/ou pouco variáveis. A metodologia de descarte adotada produziu resultados satisfatórios. Os métodos multivariados provaram ser úteis na manipulação das coleções de germoplasma, otimizando o conjunto de descritores e identificando acessos em duplicata.

MULTIVARIATE GENETIC AND PHENETIC DIVERGENCE IN
HYBRID PREDICTION AND CACAO (*Theobroma cacao* L.)
GERMPLASM PRESERVATION

Author: LUIZ ANTÔNIO DOS SANTOS DIAS

Adviser: Prof. Dr. PAULO YOSHIO KAGEYAMA

SUMMARY

Two studies involving the common theme of divergence among cacao genotypes were developed. Multivariate analyses were applied to the genetic improvement program with a view to increasing yield in association with the preservation of cacao genetic resource program. The objective of study I was to predict the performance of hybrids by parents S_1 divergence. In study II the phenetic divergence among cacao clones was quantified.

Study I involved cultivars and hybrids evaluated in diallel crossing systems. Thus, the auto-compatible cultivars CC 41, SIAL 169, CEPEC 1, ICS 1, and SIC 19, and 20 possible hybrids were tested at the Centro de Pesquisa de Cacau (CEPEC), in Itabuna, Bahia. The trial was set up in 1975, in randomized block with four replications and plots containing 16 plants. The seeds for the crossings were obtained by controlled manual pollination in clonal cultivars. In turn, the cultivar seeds were obtained by selfing of the corresponding clones. Yield component data obtained in the agricultural years from 1986 to 1990 were used in the analysis. The traits studied were number of healthy and of collected fruits per plants NHFP and NCFP, weight of humid seed per plant and per fruit WHSP and WHSF, and percentage of diseased fruit per plant PDFP.

The uni- and multivariate variance analyses detected highly significant differences among cultivars. The divergence among them was quantified by Mahalanobis' D^2 and Euclidian distance (d_e), Euclidian distance estimated from principal components (d_{pc}), and from canonical variates (d_{cv}). The statistics that consider the residual dispersion matrix - in the case, D^2 and d_{cv} - showed a close association ($r = 0.95$). This was also true between statistics d_e and d_{pc} ($r = 0.97$). However, these two pairs of statistics revealed a slight concordance. The application of Tocher's, single linkage methods, and the graphic representation from canonical variates confirmed two groups

diverging between themselves. The first involved cultivars CC 41, SIAL 169, CEPEC 1, and SIC 19 and the second was formed by ICS 1.

Principal components and canonical variates discriminated the trait WHSP as those of least importance for the divergence. The geographic distance among cultivars was related to genetic distance, based on groups formed by Tocher's method. Due to the stable pattern of genetic divergence among cultivars in five years of the study, it was possible to use one single year, of optimal climatic conditions, for estimating divergence.

The average and heterotic performance of the hybrids were satisfactorily predicted by the divergence among S_1 generation parents, expressed by D^2 . Moderate and high Spearman's correlations were obtained between D^2 and WHSP (0.68) and between D^2 and WHSF (0.76) for average performance. A high correlation (0.65) between D^2 and WHSP was obtained with relation to the heterotic performance of the hybrids. Heterosis as measured by difference between F_1 's and average performance of corresponding superior parental S_1 . The best hybrids in relation the WHSP and WHSF resulted from crossings involving the cultivar of highest divergence (ICS 1). Also, the best hybrid resulted from the crossing between the second pair of highest divergence - in this case SIAL 169 and ICS 1. Finally, the hybrid showing the worst average performance resulted from the crossing between the pair of cultivars showing the second highest similarity, this is, CEPEC 1 and SIC 19.

In study II the phenetic divergence among 26 cacao clones of CEPEC's germplasm was quantified by cluster and principal components analyses. In the 26 clones SIC and SIAL, selected from the cacao Amazon populations from southern Bahia, four groups of phenetic similarity were detected. Such groups were identified by Tocher's method. The presence of divergence among these local clones contradicted the uniformity paradigm of cacao populations from Bahia and suggested that selection and breeding strategies between such populations are feasible.

Among 13 traits used in evaluating the 26 clones, four were discarded due to redundancy and/or low variability. The methodology adopted for discarding produced satisfactory results. The multivariate methods proved to be useful for handling germplasm collections, optimizing the group of descriptors and identifying duplicate accessions.

1 - INTRODUÇÃO

Na quase totalidade das vezes o melhorista vegetal é levado a obter múltiplas mensurações no seu material experimental. O fato decorre da necessidade que o profissional tem de observar vários caracteres no mesmo indivíduo. Entretanto, de posse dessas múltiplas mensurações, o melhorista recorre a procedimentos analíticos univariados, em que cada caráter é analisado em separado, de forma individualizada. Optando por procedimentos multivariados o melhorista estaria avaliando o indivíduo na sua multidimensionalidade. Por essa razão, esses últimos procedimentos proporcionam uma visão holística dos indivíduos. Na verdade, são procedimentos mais equilibrados e estão em harmonia com o sistema biológico.

A seleção de tipos parentais, baseada em um número de caracteres importantes coletivamente, pode ser mais vantajosa que a seleção baseada em caracteres individuais. Essa postulação torna-se relevante em programas de melhoramento que objetivam incrementos nos caracteres quantitativos complexos, a exemplo da produção de grãos. É que, frequentemente, a produção de grãos é formada por múltiplos componentes. E, segundo GRAFIUS (1964), a produção não pode ser alterada sem alterar-se um ou mais de seus componentes.

Em cacau, o desenvolvimento de híbridos representou um substancial avanço no programa de melhoramento genético e na produção brasileira de cacau. O híbrido foi, em grande parte, o responsável pelo fato do Brasil ter sido, no passado recente, o país de mais alta produtividade mundial de cacau, cerca de 750 kg/ha/ano de amêndoas secas. Não obstante, produtividade ao redor de 3.500 kg/ha/ano tem sido obtida em condições experimentais. A lavoura cacaueira ocupa área de 550 mil hectares, distribuída em 25 mil propriedades, empregando diretamente 200 mil trabalhadores. Esse panorama se verifica no Estado da Bahia, responsável por 90% da produção nacional. Além disso, a lavoura tem papel estratégico na preservação da Mata Atlântica. Economicamente o cacau destacou-se pela geração de sete bilhões de dólares de divisas ao Brasil no período 1977/87. Portanto, o potencial produtivo da espécie e sua importância sócio-econômica evidenciam o quanto a cultura pode se beneficiar de processos preditivos de híbridos superiores.

A literatura contém relatos de utilização da diversidade geográfica como índice sólido para se avaliar a divergência genética nos programas de hibridação. A hibridação em cacau não constitui exceção a essa prática. De modo que a hibridação, envolvendo cruzamentos entre clones parentais de origem geográfica distinta é comum. A estratégia de cruzamento objetiva a complementação da expressão fenotípica de caracteres. Em alguns casos, cruzamentos são conduzidos de modo aleatório, sobretudo quando não se dispõe de informações sobre os clones parentais. Portanto, o programa de híbridos em cacau baseia-se em uma combinação de metodologias dos tipos convencional ou de complementação da expressão fenotípica de caracteres, de acordo com a diversidade geográfica e, até mesmo, na aleatoriedade. As técnicas multivariadas, todavia, ainda não foram aplicadas em cacau para se estimar a divergência genética e prever combinações heteróticas.

Certamente que as técnicas multidimensionais não são as únicas ferramentas preditivas de híbridos heteróticos, à disposição do melhorista de cacau. Entretanto, tais técnicas têm se mostrado muito adequadas para estimar divergência e prever cruzamentos promissores em diversos cultivos. A aplicação da análise multivariada não representa custos adicionais, uma vez que exige sistemas computacionais já existentes. Além disso, pode ser aplicada sobre caracteres tradicionalmente avaliados em experimentos de cacau.

As técnicas multivariadas, a exemplo das análises de cluster, análises de componentes principais e variáveis canônicas, podem ser uma opção favorável para predição de combinações heteróticas. Com a aplicação de tais técnicas, centenas de cruzamentos indesejáveis podem ser evitados. Afinal é melhor, mais prático e mais econômico realizar alguns cruzamentos bem planejados do que centenas deles, com base em critérios subjetivos. Evitar cruzamentos indesejáveis em cacau significa economizar, em média, nove anos necessários para obtenção e avaliação de centenas de progêies híbridas. Percebe-se o quanto a predição ganha em importância. É bom reforçar que a subjetividade no processo de hibridação é inevitável e, frequentemente, é resultante da experiência do melhorista na manipulação dos materiais. Por mais sólidos que sejam os critérios adotados na identificação de híbridos heteróticos, algum grau de subjetividade sempre existirá. Entretanto, a subjetividade não deve prevalecer sobre critérios científicos.

A associação de heterose com divergência está bem estabelecida na teoria. Mais recentemente, alguns estudos estão confirmando, na prática, o que a teoria já demonstrou. No entanto, quantificar o grau de associação entre heterose nos híbridos e

divergência genética nos genitores, constitui-se em um grande desafio do melhoramento vegetal. Se essa associação ocorre em um grau elevado, então é possível utilizar-se da estimativa de divergência como critério preliminar para seleção de genitores e síntese de híbridos heteróticos.

As técnicas multivariadas, contudo, não se prestam apenas para estimar a divergência genética com objetivos de hibridação. Elas podem e devem ser empregadas para otimizar coleções de germoplasmas. Em cacau, os germoplasmas são mantidos, principalmente, na forma de coleções clonais de campo. As coleções de germoplasmas da Bahia, por exemplo, detêm mais de 540 acessos. A Amazônia, por sua vez, detêm mais de 1.300 acessos, constituídos, majoritariamente, de materiais regionais silvestres, representados, em média, por dez clones cada. Além da enorme área que ocupam e do elevado custo de instalação e manejo que representam, tais coleções são passíveis de deter germoplasmas em duplicata. A aplicação de técnicas multivariadas permite quantificar a divergência existente nas coleções e identificar grupos de similaridade, que podem se constituir em duplicatas. Informam, ainda, sobre os caracteres mais importantes para a divergência.

Depreende-se, então, que a análise multivariada pode tornar-se uma poderosa ferramenta a ser manipulada pelo melhorista de cacau, em particular, já que a eficiência dessas técnicas tem sido confirmada em diversas culturas. O fato demonstra a necessidade do programa de pesquisas de cacau ser suficientemente flexível para incorporar inovações tecnológicas desenvolvidas em outras culturas.

A presente tese envolveu dois estudos em separado, porém conectados por um tema comum, no caso a divergência entre genótipos. Os objetivos propostos foram divididos em objetivos gerais e específicos. Os últimos delineados para atender a cada um dos estudos em particular. Os objetivos gerais foram:

- a. Implementar a aplicação de análises multivariadas no programa de melhoramento genético do cacaueiro (*Theobroma cacao* L.) para aumento de produção de amêndoas;
- b. Implementar a aplicação de análises multivariadas no programa de preservação de recursos genéticos de cacau;

No estudo I os objetivos foram:

- a. Quantificar e visualizar a divergência genética entre cinco cultivares de cacau, por diferentes estatísticas multivariadas;
- b. Comparar quatro diferentes estatísticas utilizadas na estimação da divergência entre os cinco cultivares de cacau;
- c. Identificar os caracteres de menor importância para a divergência genética;

- d.** Quantificar o grau de associação entre performance média e heterose realizada nos híbridos com a divergência genética manifestada entre os cultivares parentais;
- e.** Relacionar a diversidade ecogeográfica com a divergência genética dos cultivares;
- f.** Avaliar a estabilidade temporal da estimativa de divergência genética expressa pela estatística D^2 de Mahalanobis;

Para o estudo II, os seguintes objetivos foram propostos:

- a.** Quantificar e visualizar a divergência fenética entre 26 clones de cacau das séries SIC e SIAL, pertencentes às seleções conduzidas na população cacaueira da Bahia, por duas técnicas multivariadas;
- b.** Relacionar a diversidade ecogeográfica com a divergência fenética dos clones;
- c.** Selecionar e descartar caracteres de pouca importância para a divergência fenética entre os clones.

2. REVISÃO DE LITERATURA

2.1. A seleção de genitores e a hibridação

A postulação de critérios racionais e eficientes para identificação de tipos parentais visando a síntese de híbridos e populações é uma questão aberta no melhoramento de plantas. Existem duas possibilidades para solucionar a questão. A primeira é testar os tipos parentais quanto a capacidade combinatória. Nesse caso, os sistemas de cruzamentos dialélicos são conduzidos para avaliar a capacidade geral e específica de combinação. Genótipos bons combinadores, ou seja com alta frequência de alelos favoráveis, são esperados produzir os cruzamentos superiores. A segunda possibilidade baseia-se na premissa de que genótipos parentais altamente divergentes e com alta performance média podem gerar híbridos superiores.

A segunda possibilidade é, comparativamente, mais vantajosa que a primeira. A seleção de combinações parentais promissoras baseada na performance dos genótipos parentais *per se* tem natureza preditiva, significando economia de recursos financeiros, de tempo e mão de obra. Em outras palavras, a seleção de genitores pela performance *per se* evita que grande número de cruzamentos sejam realizados, em contraste à primeira possibilidade que tem uma natureza realizada. A dificuldade, no entanto, está em estimar a diversidade entre os futuros genitores, como relatam VELLO & PIRES (1992).

É interessante notar que desde longo tempo os melhoristas têm se utilizado da diversidade ecogeográfica como um índice sólido de diversidade genética em diversas culturas (MOLL et alii, 1962; MOLL et alii, 1965; TIMOTHY, 1963 e RAM & PANWAR, 1970). Entretanto, algumas forças como deriva genética, seleção em diferentes ambientes (MURTY & ARUNACHALAM, 1966), intercâmbio de materiais e seleção artificial (ADAMS, 1977; KATIHAR & SINGH, 1979; JOSHI & VASHI, 1992) podem gerar maior diversidade que distância geográfica. Se tais forças atuam mais intensamente que o isolamento geográfico, então, a distância geográfica pode não expressar a verdadeira diversidade genética.

BHATT (1973) comparou experimentalmente várias metodologias de seleção de parentais em trigo, denominadas de i) convencionais, ii) aleatórias, iii) multivariadas e iv) com base na diversidade geográfica. Na estratégia convencional um genitor tem performance testada em uma dada área, enquanto o outro genitor é escolhido por complementar um caráter específico, pouco expressivo no primeiro. A estratégia aleatória se aplica quando não se dispõe de informações sobre os tipos parentais que são, então, escolhidos ao acaso. A seleção baseada em técnicas multivariadas objetiva selecionar os parentais pela diversidade, com base em um conjunto de caracteres. Por último, os parentais são selecionados pela diversidade ecogeográfica, considerando-a um índice de diversidade genética. Os resultados proporcionados por BHATT (1973) revelaram que as técnicas multivariadas foram mais eficientes que o critério da diversidade ecogeográfica para predição das combinações parentais superiores. Além disso, a variância genética para produção de grãos entre linhagens F₅ foi proporcional à divergência dos genitores, expressa por D^2 , envolvidos nos cruzamentos (correlação de Spearman de 0,864 entre as duas estimativas). Os resultados demonstraram claramente a validade da predição da performance de cruzamentos em gerações avançadas. E mais, cruzamentos entre genitores altamente divergentes geraram maior número de segregantes transgressivos para a produção de grãos. As metodologias convencionais e aleatórias foram menos eficientes.

Em cacau, a seleção de genótipos parentais para hibridação é de grande importância já que o cultivo é feito com híbridos. Assim, sob uma perspectiva histórica o melhoramento genético do cacaueiro pode ser dividido em antes e depois do híbrido (DIAS, 1993). Na seleção dos clones parentais dos híbridos adota-se uma combinação de três das estratégias abordadas por BHATT (1973). Pratica-se uma hibridação para complementação da expressão fenotípica de caracteres, a partir de cruzamentos entre genitores de origem geográfica distinta. A estratégia aleatória é também utilizada, num processo de tentativa e erro (DIAS, 1991), quando se dispõe de poucas informações sobre os clones parentais em cruzamento. Por fim, a seleção de clones parentais baseada em técnicas multivariadas é a única estratégia ainda não aplicada ao melhoramento do cacaueiro.

É interessante abordar os grupos raciais existentes em cacau, já que os híbridos são, frequentemente, o resultado do cruzamento de clones pertencentes às diferentes raças. O primeiro grupo, denominado crioulo, foi cultivado há dois mil anos por Maias e Astecas no sul do México (CUATRECASAS, 1964). O grupo forasteiro tem sido cultivado há apenas três séculos na bacia Amazônica, embora exista em estado

silvestre na região (SORIA, 1965; 1970). Um terceiro grupo denominado trinitário é resultante da hibridação natural entre os dois primeiros grupos e teve origem em Trinidad, no século XVIII (CHEESMAN, 1944).

As características de fruto e semente são usadas para definir os grupos raciais em cacau (TOXOPEUS, 1985a). Os crioulos têm frutos longos, verdes imaturos e vermelhos quando maduros. Suas sementes arredondadas têm coloração rósea ou branca. Já os forasteiros apresentam sementes achatadas, violeta escuro, contidas em frutos verdes imaturos e amarelos amadurecidos. Os forasteiros dividem-se em amelonados, oriundos do baixo Amazonas, cujos frutos são pequenos e com formato de melão. Da região do alto Amazonas deriva o segundo subgrupo dos forasteiros conhecido pelo nome da região de origem. Os alto Amazonas exibem grande variabilidade, o que levou CHEESMAN (1944) a propor o centro de diversidade do cacau como sendo na região de origem desses alto Amazonas, ou seja, no leste do Equador. Por último, trinitários têm sementes de cor rósea intermediária.

Heterose para crescimento vegetativo, produção de amêndoas de cacau, precocidade de produção, resistência a doenças e outros caracteres tem sido encontrada em cruzamentos envolvendo clones de grupos raciais distintos (RUSSEL, 1952; SORIA, 1964; OPEKE & JACOB, 1969; ATANDA & TOXOPEUS, 1971; REYES & PEREZ, 1971; ATANDA, 1972; 1973). Comercialmente, os cacauzeiros híbridos são multiplicados por sementes. Entretanto, os genitores desses híbridos são clones heterozigóticos (S_0), o que provoca intensa segregação para produção no campo. Todavia, a propagação clonal de híbridos biclonais não têm sido realizada comercialmente.

2.2. Divergência genética por análises multivariadas

A hibridação envolvendo tipos parentais divergentes é passível de produzir segregantes transgressivos favoráveis em gerações avançadas. As técnicas multivariadas, a exemplo da análise de cluster, análise de componentes principais e análises canônicas têm contribuído para a estimação da divergência genética entre populações. A quantidade de heterose, segundo FALCONER (1989), é uma função linear do quadrado das diferenças das frequências gênicas entre as populações, nos locos com dominância. A linealidade se verifica no campo experimental. Todavia, essa linealidade desaparece considerando-se populações altamente diferenciadas (CRESS, 1966), em função de um desbalanço gênico. Por outro lado, as populações que têm grande número de alelos em comum, para um caráter considerado, são similares e inadequadas para

produzir cruzamentos superiores. Já populações divergentes, com diferenças de frequências nos locos considerados, podem se apresentar apropriadas para hibridação.

A associação da heterose nos híbridos com divergência genética dos genitores está estabelecida em teoria (FALCONER, 1989; CRESS, 1966). O grau de associação entre heterose para produção e seus componentes e divergência genética, por sua vez, tem sido quantificado, experimentalmente, em triticales (SRIVASTAVA & ARUNACHALAM, 1977), em amendoim (ARUNACHALAM et alii, 1984), e em amendoim e colza (ARUNACHALAM & BANDYOPADHYAY, 1984). Esses estudos têm revelado a existência de um nível ótimo de divergência parental, estimada por distância de Mahalanobis, que garante heterose nos híbridos. Assim, maior frequência de cruzamentos heteróticos para produção de grãos e seus componentes, com heterose em alta magnitude, ocorre quando os tipos parentais exibem divergência em níveis intermediários. Portanto, existem suficientes argumentos científicos que justificam o uso da divergência como um critério preliminar de seleção de parentais para produzir híbridos heteróticos.

As técnicas multivariadas consideram um grande número de caracteres simultaneamente. A definição dos mais potentes para expressar divergência entre as populações, segundo MURTY & ARUNACHALAM (1966); ARUNACHALAM & RAM (1967) e JOSHI & VASHI (1992), são os caracteres de importância para a adaptação sob condições naturais. Os autores advertem que os caracteres selecionados pelo homem têm importância limitada na discriminação das populações. É que, em dado momento, seleção artificial e seleção natural operam em direções opostas. Em sorgo, por exemplo, materiais de ciclo avançado de melhoramento exibem floração sincrônica e altura de planta similar. Essa é a condição necessária para se produzir sementes híbridas comerciais. Então, a possibilidade de se obter tipos parentais fenotipicamente similares é grande, a despeito de acumularem complexos gênicos diversos. Nesses casos, segundo JOSHI & VASHI (1992), as análises multivariadas empregadas para acessar a divergência genética têm utilidade limitada por se basearem em fenótipos.

Em princípio, caracteres pouco influenciados pelo ambiente, como os bioquímicos e moleculares são os mais adequados para estimação de divergência genética por técnicas multivariadas (SMITH, 1984; 1988; SMITH et alii, 1985a; 1985b; SMITH & SMITH, 1987; 1988; 1989b e SMITH et alii, 1990). Entretanto, ambos têm limitações sérias. Dados bioquímicos, a exemplo de isoenzimas e proteínas, refletem em alguns casos, uma amostragem insuficiente do genoma. Por sua vez, dados moleculares,

como marcadores RFLP, geram estimativas de divergência confiáveis, segundo MELCHINGER et alii (1991), apenas a partir do emprego de um grande número de combinações sonda-enzima.

A divergência genética inferida de coeficientes de parentesco, por seu turno, reflete relações genéticas baseadas em maior número de locos que aqueles trabalhados por isoenzimas e proteínas. Todavia, tais coeficientes têm como limitação a admissão da contribuição equitativa de genes dos tipos parentais e a ausência de seleção. A seleção natural opera sobre caracteres morfológicos ligados à adaptação. Assim, a divergência genética estimada de caracteres morfológicos quantitativos, conforme CAMUSSI et alii (1985), permitem inferências sobre padrões de adaptação e coadaptação das populações.

Dados bioquímicos, moleculares e morfológicos em combinação parecem mais apropriados para estimação da divergência genética entre populações, não obstante divergência morfológica e bioquímica apresentarem-se discrepantes, algumas vezes. O fato sugere padrões evolutivos morfológicos e bioquímicos distintos. Entretanto, na ausência de dados bioquímicos e moleculares, a divergência fenotípica e a divergência genética podem guardar estreita correspondência entre si. Isso é verdadeiro quando a estimativa de divergência baseia-se em caracteres morfológicos quantitativos, avaliados em experimentos com repetições, repetidos em diversos locais e épocas. Desse modo, se o caráter a ser predito é a produção heterótica de grãos, então a produção e seus componentes parecem apropriados para o estudo da divergência.

2.2.1. Análise da variância multivariada

A análise multivariada é uma extensão da análise univariada. Enquanto a última testa a igualdade entre as médias dos tratamentos, a multivariada testa a igualdade dos vetores de médias dos tratamentos. Então, a multivariada é o procedimento mais adequado para se avaliar um grupo de k tratamentos ou amostras a partir do conjunto de p variáveis respostas. A grande vantagem de se utilizar um procedimento multivariado é a visão holística que ele proporciona. Assim, a discriminação entre tratamentos no caso multivariado, ao contrário do univariado, é feita considerando-se as múltiplas variáveis conjuntamente, sendo que nenhuma delas assume a posição de variável independente.

Ao proceder-se à análise da variância multivariada, assume-se que todas as variáveis sejam normalmente distribuídas. Portanto, a multinormalidade dos dados deve ser uma condição a ser atendida. Nesse tipo de análise, as variáveis devem ser

preferencialmente correlacionadas, pois caso contrário, a análise univariada apresenta-se como a opção mais adequada. Particularmente importante é a correlação residual entre as variáveis, mais que as correlações lineares simples. Segundo KENDALL (1963) as variáveis precisam ser consideradas no conjunto. Sendo p as variáveis, a multivariada considera as p médias, as p variâncias e, ainda, as $p(p-1)/2$ diferentes covariâncias entre as variáveis. Para COOLEY & LOHNES (1986) esses são os parâmetros naturais da distribuição normal multivariada.

Na prática, o procedimento multivariado difere do univariado pela maior complexidade matemática e também pela maior dificuldade de interpretação dos resultados (KENDALL, 1963). Entretanto, tais aspectos não devem desencorajar a aplicação da análise multidimensional, sobretudo em biologia, onde os organismos precisam ser avaliados com base no conjunto de suas principais características, simultaneamente. A propósito, foi com o objetivo de encorajar novos aplicadores que VAN LAAR (1987) revisou alguns dos métodos multivariados listando, ao final, suas aplicabilidades. "Análise multivariada: Um caminho para o melhor entendimento da complexidade" foi o título do seu artigo.

Admitir a análise multivariada como sendo uma generalização da univariada, aplicável a muitas variáveis, significa atribuir a WILKS (1932) o desenvolvimento dessa técnica. De fato foi WILKS quem propôs a estatística Λ - lâmbda maiúscula - seguindo o método da razão de verossimilhança e que se constituiu, mais tarde, em um teste básico para as análises multivariadas. Para avaliar a significância de Λ é comum promover-se a transformação dessa estatística em um valor aproximado de F de Snedecor. Segundo VAN LAAR (1991), Λ e F são relacionados ao inverso, ou seja, um baixo valor de Λ revela diferenças significativas entre as populações multivariadas.

2.2.2. Análise de cluster

A análise de cluster é uma técnica multivariada que objetiva agrupar itens diversos como populações, clones, variedades, etc, após a estimação da dissimilaridade entre os itens. Ou seja, em um primeiro momento estima-se uma matriz de dissimilaridade entre os itens e, com base nessa matriz, identificam-se os grupos pela similaridade. Os métodos de análise de cluster classificam-se, basicamente, em hierárquicos e de otimização. Dentre os hierárquicos destaca-se o método do vizinho mais próximo, que permite a construção de um diagrama de árvore denominado dendrograma. Por sua vez, o método de otimização é representado pelo método de Tocher. Percebe-se a utilidade dessa técnica ao empregá-la sobre um grande conjunto de

dados, de modo a transformá-lo em pequenos subconjuntos de mais fácil interpretação e manipulação. Ressalte-se que a técnica desconhece *a priori* o número e a composição dos diferentes grupos ou clusters a serem formados.

Na estruturação da matriz de dissimilaridade empregam-se a distância Euclidiana e a distância generalizada de Mahalanobis, entre outras medidas de distância. VAN LAAR (1991) relatou que a última é uma medida de distância mais favorável que a primeira. Na realidade, a distância de Mahalanobis é também mais precisa pois considera a covariância residual entre as variáveis em estudo. Assim, todas as p variâncias e as $p(p-1)/2$ covariâncias residuais envolvidas são levadas em conta, no cálculo da distância. Na aplicação da distância de Mahalanobis assume-se multinormalidade dos caracteres e matriz de covariância comum. GOODMAN (1967) admite a multinormalidade essencial para testes de significância, mas não para estimação da distância em si. A matriz de covariância não é, frequentemente, homogênea, impossibilitando que a transformação das variáveis elimine a correlação entre elas. Logo, a mesma razão que cria vantagens para D^2 cria problemas teóricos. Na prática, contudo, essa escassez de homogeneidade é ignorada. A matriz de dispersão do erro, comum a todos os genótipos avaliados por múltiplos caracteres em experimentos com repetições, é adequada para estimação da distância genética por Mahalanobis (ARUNACHALAM, 1981).

A distância Euclidiana é uma distância baseada no teorema de Pitágoras, porém, aplicada a múltiplos eixos ortogonais. A ortogonalidade dos eixos admite a inexistência da correlação entre caracteres, o que frequentemente não ocorre (SNEATH & SOKAL, 1973). Essa dificuldade pode ser contornada empregando-se a distância de Mahalanobis ou então a própria distância Euclidiana obtida de escores de componentes principais, esses sim ortogonais entre si. Dois cultivares sendo geneticamente próximos ocorrem na mesma região do espaço multidimensional e exibem pequena distância entre si. A distância Euclidiana torna-se mais eficaz quando obtida de variáveis padronizadas ou estandardizadas, ponderada pelo número de variáveis envolvidas. Com esse procedimento a distância não é alterada pelas diferentes escalas de mensuração das variáveis e nem pelo número delas. Nesse caso, obtém-se a distância Euclidiana padronizada média.

Os métodos de agrupamento existentes na literatura são inúmeros e podem conduzir a diferentes padrões de agrupamento. Entretanto, serão abordados nessa revisão apenas os métodos do vizinho mais próximo e o de Tocher, classificados como hierárquicos aglomerativos e de otimização, respectivamente. A aplicação do método do vizinho mais próximo resulta em um dendrograma que permite identificar grupos

homogêneos. A delimitação de grupos nos dendrogramas obedece aos chamados pontos de alta mudança de nível. O outro método, desenvolvido a partir de uma sugestão de Tocher, conforme RAO (1952), estabelece grupos cujas médias das distâncias intragrupos são sempre menores que as distâncias médias intergrupos. Difere do método do vizinho mais próximo por levar à formação de grupos homogêneos mutuamente exclusivos. O método de Tocher, segundo ARUNACHALAM (1981), é bem exato dado que trabalha com o espaço multidimensional completo.

O agrupamento é promovido em virtude das distâncias serem obtidas de n indivíduos aos pares, nas $n(n-1)/2$ combinações possíveis. Se n for um número grande, a identificação de grupos homogêneos pela observação da matriz de distâncias tornar-se-á, obviamente, impraticável. O agrupamento, contudo, leva à perda de informações ao nível de indivíduos, restando apenas informações sobre grupos de similaridade. Por esse motivo, o estudo da divergência é conduzido também e, por vezes simultaneamente, por componentes principais e variáveis canônicas, que serão vistos nos itens seguintes. Essas duas últimas técnicas permitem recuperar as informações da divergência ao nível de indivíduos, a partir da representação gráfica com eixos formados pelos escores obtidos dos primeiros componentes e variáveis canônicas.

A versatilidade e o poder discriminatório da análise de cluster tem permitido a sua aplicação em diferentes áreas do melhoramento de plantas. Frequentemente, a análise de cluster utiliza-se da distância generalizada de Mahalanobis, ou seja, da estatística D^2 , como medida de diversidade genética, e do método de Tocher como método de agrupamento. Esse esquema tem sido utilizado por BHATT (1976) em trigo, RAMANUJAM et alii (1974) em *Phaseolus aureus*, SINGH et alii (1981) em *Pennisetum americanum*, SINGH & GUPTA (1979) em *Pennisetum typhoides*, DAS & DAS GUPTA (1984) em *Vigna mungo*, JAIN et alii (1981) e SHERIFF (1992) em *Eleusine coracana*, SINGH & SINGH (1976) em *Capsicum annum*, SOMAYAJULU et alii (1970) e SHAMSUDDIN (1985) em trigo.

Esquema idêntico foi utilizado por SINGH et alii (1979) em arroz do tipo anão, KATIYAR & SINGH (1979) em *Cicer arietinum*, KALLOO & SIDHU (1982) em melão, AMALRAJ (1982) em *Gossypium barbadense* e SINGH & GILL (1984) em *Gossypium hirsutum*, CHAUDHARY & SINGH (1975) e VARMA & GULATI (1982) em cevada, JOSHI & VASHI (1992) em sorgo, JAGADEV et alii (1991) em *Brassica campestris* e GUPTA et alii (1991) em *Brassica juncea*, KHARE & SINGH (1992) em *Vicia faba*, JAGADEV & SAMAL (1991) em *Guizotia abyssinica*, SHERIFF & SHIVASHANKAR (1992) em *Setaria italica*, JESWANI et alii (1970) em

Linum usitatissimum, RESENDE & HIGA (1991) e PULCINELLI et alii (1993) em eucalipto e BAGCHI (1992) em *Acacia nilotica*.

2.2.3. Análise por componentes principais

O princípio da técnica dos componentes principais é a análise simultânea de múltiplas variáveis condensadas em umas poucas variáveis, sem perda substancial de informação. Isso é feito a partir da estimação de componentes ditos principais, que correspondem a combinações lineares de todas as variáveis estudadas. Obviamente que a passagem de um espaço multidimensional, proporcionado pelas p variáveis, para outro bi ou tridimensional facilita a interpretação e a visualização dos resultados. Essa metodologia foi desenvolvida por K. Pearson (1901), segundo MORRISON (1976).

A técnica envolve a matriz de covariância H ou a matriz de correlação R , sendo a segunda de uso mais frequente (MORRISON, 1976). É que, frequentemente, as variáveis têm escalas distintas devendo serem padronizadas, o que torna idênticas as matrizes H e R . A importância de padronizar as variáveis antes do cálculo dos componentes principais foi demonstrada por VAN LAAR (1991). O resultado da ponderação das variáveis padronizadas pelos respectivos elementos dos autovetores, estimados de H ou R , constitui os chamados componentes principais.

O primeiro componente é uma variável adicional, assim como os demais, e dotado da maior variância. Essa variância corresponde ao autovalor λ_1 da matriz R . De modo análogo, o segundo componente é uma variável não-correlacionada ao primeiro e que detém a segunda maior variância, e assim por diante. KENDALL (1963) demonstra que a escolha do maior autovalor (λ_1) gera uma linha, cuja a soma de quadrados das distâncias dos seus pontos é mínima e a variável medida ao longo dela tem a máxima variância. Geometricamente, a estimação do segundo componente λ_2 significa encontrar a linha perpendicular ao primeiro autovalor λ_1 , cuja soma de quadrados das distâncias seja mínima. A importância de cada componente é dada pela percentagem de variância total que ele absorve.

É certo admitir que quanto maior o grau de correlação entre as variáveis, tanto maior será a proporção da variância total explicada por um número reduzido dessas variáveis e, logicamente, pelos primeiros componentes principais. Segundo VAN LAAR (1991), essa alta correlação é comum em estudos biológicos. MORRISON (1976) sustenta que é melhor sumarizar o complexo multivariado logo nos primeiros componentes, desde que esses respondam por 75% ou mais da variância total.

Nesses casos é possível construir gráficos bi ou tridimensionais, com os eixos cartesianos formados pelos respectivos escores dos dois ou três primeiros componentes principais, e visualizar grupos de similaridade.

Análises por componentes principais também vêm sendo aplicadas numa gama de espécies vegetais com os mais diversos propósitos. A identificação de genitores divergentes para hibridação e geração de populações segregantes, a avaliação da divergência genética entre acessos em germoplasmas e a avaliação da uniformidade em cultivos são algumas das muitas aplicações da técnica de componentes principais. A seguir serão abordados algumas pesquisas envolvendo essa técnica.

LANAUD (1986) utilizou-se da técnica para reconstituir a história das introduções de cacauzeiros na África. A análise por componentes principais foi aplicada a 157 indivíduos de distintos grupos raciais pertencentes a coleção de cacau da Costa do Marfim, caracterizados por sete locos isoenzimáticos. Os dois primeiros componentes absorveram apenas 35% da variação total disponível. Mesmo assim, segundo o autor, a representação gráfica, com os eixos formados por esses dois componentes permitiu uma boa separação dos vários grupos raciais. O grupo racial baixo Amazonas, que constitui a população de cacauzeiros cultivados da África, mostrou a menor diversidade nos eixos. Os cacauzeiros do alto Amazonas, por outro lado, distinguiram-se pela grande diversidade. O grupo dos trinitários africanos ocupou posição intermediária entre trinitários americanos e baixo Amazonas, sugerindo introgressão entre trinitários e os primeiros baixo Amazonas introduzidos na Costa do Marfim. A história do cacau africano registra que sua origem deriva de poucas plântulas introduzidas do Brasil no século XIX justificando, portanto, a homogeneidade detectada pelo estudo. Apesar da limitação do número de locos estudados, do tamanho reduzido da amostra e da reduzida absorção da variação total pelos primeiros componentes, o autor pôde acessar a diversidade genética das populações cacaueiras africanas. Informações sobre a diversidade foram complementadas pela frequência de alelos e pelo número médio de alelos por locos. Portanto, a análise de componentes principais aplicada a dados isoenzimáticos acusou a necessidade de novas introduções de cacauzeiros na África, para obtenção de maiores progressos no melhoramento genético local.

A aplicação de componentes principais na quantificação da similaridade genética no cultivo do feijoeiro foi proposta por ADAMS (1977). O estudo envolveu 27 cultivares, em nove Estados produtores dos EUA, avaliados por 36 caracteres químico-agronômicos obtidos de sementes e plantas. Estimou-se a distância Euclidiana entre pares de cultivares, dentro de uma dada região produtora, a partir de

escores dos seis primeiros componentes principais. Na etapa seguinte, essa distância foi ponderada pela área ocupada de cada cultivar e utilizada como um índice de "homogeneidade genética" para o cultivo, naquela região. Cultivares relacionados por pedigree foram menos distantes que aqueles não relacionados, demonstrando a natureza genética dessa distância Euclidiana ponderada. Admitiu-se, então, que os valores de distância refletiram, numa relação direta, a diversidade de alelos, incluindo aqueles alelos que afetam a taxa de infecção a patógenos. Ora, a homogeneidade genética é a base da vulnerabilidade a epidemias. Logo, os Estados produtores de feijão que apresentaram valores altos para a tal distância Euclidiana ponderada foram classificados como menos vulneráveis. Em decorrência, o cultivo do feijoeiro no Estado da Califórnia apresentou a menor vulnerabilidade, enquanto a maior vulnerabilidade foi detectada no Colorado.

Componentes principais também vêm sendo utilizados para estudar a diversidade genética entre híbridos comerciais de milho. SMITH (1984) avaliou 111 híbridos amplamente cultivados nos EUA. Análise de componentes principais foi aplicada a dados isoenzimáticos, referentes a 21 locos. A representação gráfica estudada pelo autor envolveu os quatro primeiros componentes, o 1º com o 2º e o 3º com o 4º, os quais absorveram 66,4% da variação total. Assim, os componentes principais permitiram visualizar agrupamentos contrastantes de híbridos, proporcionando ao agricultor a escolha de híbridos verdadeiramente diferentes. Finalmente, foi verificado que 95% dos híbridos avaliados detinham frequências alélicas diferentes. Todavia, muitas dessas diferenças foram devidas a desvios das frequências alélicas em poucos locos.

SMITH et alii (1985a; 1985b); SMITH (1988) e SMITH & SMITH (1987; 1988) têm aplicado análise de componentes principais a dados cromatográficos, isoenzimáticos e de pedigree em linhagens e híbridos de milho dos E.U.A.. O objetivo tem sido a estimação da divergência genética relativa e da diversidade dentro do germoplasma da espécie. BROWN (1991), entretanto, utilizou-se de caracteres agronômicos e de fibra para quantificar a variabilidade dos cultivares de algodão dos E.U.A.. A análise por componentes principais resultou em divergência com sentido agronômico e genético.

Recentemente, análise de componentes principais tem sido aplicada a dados de RFLP para acessar a similaridade genética entre linhagens do cinturão de milho dos E.U.A.. Os resultados revelaram que os marcadores RFLP discriminaram linhagens em grupos heteróticos, em concordância com o pedigree (MELCHINGER et alii, 1991).

Na pesquisa com espécies arbóreas, a técnica de componentes principais vem sendo utilizada no estudo da origem e evolução de *Ulmus* sp. (RICHENS & JEFFERS, 1978); na diferenciação de populações de *Pinus brutia* (CALAMASSI et alii, 1988) e de *Pinus kesiya* (BURLEY & BURROWS, 1972); na taxonomia de híbridos naturais entre *Eucalyptus melanophloia* e *E. crebra*, entre *E. populnea* e *E. crebra* (DRAKE, 1980), *Eucalyptus obliqua* x *E. pulchella* (POTTS & REID, 1983), *Eucalyptus risdonii* e *E. amygdalina* (POTTS & REID, 1985a) e entre *Abies amabilis* e *A. lasiocarpa* (MAZE et alii, 1981) e na pesquisa de "sites" potenciais similares em florestas (BURTON et alii, 1991). Não obstante, os resultados dessas pesquisas foram, no geral, obtidos de sementes ou plântulas. Em decorrência, tais resultados não necessariamente correspondem àqueles com árvores adultas. Ademais, as avaliações realizadas em folhas, sementes e frutos não incluíram caracteres de maior interesse silvicultural. Em função dessas restrições esses estudos não serão objeto de maior aprofundamento.

Entre as várias aplicações de componentes principais a mais comum é a estimação da divergência genética entre genótipos ou populações, com base em caracteres quantitativos. WILCHES (1983) avaliou a diversidade entre 34 genótipos de amendoim, com base em 13 caracteres de produção selecionados, utilizando os quatro primeiros componentes (absorção de 74% da variação total). O estudo revelou a estreita associação entre o caráter rendimento e o primeiro autovetor. Em feijão-vagem, MALUF & FERREIRA (1983) empregaram gráficos com os três primeiros componentes principais para estudar a divergência entre 10 cultivares, avaliados por sete caracteres. Os referidos componentes explicaram mais de 90% da variação observada.

Em eucalipto, a obtenção de híbridos tanto intra como interespecífico vem sendo pesquisada. RESENDE & HIGA (1991) estudaram a divergência entre sete procedências de *Eucalyptus viminalis* por componentes principais. Os dois primeiros componentes explicaram 85,36% da variação total. PULCINELLI et alii (1993) investigaram oito diferentes procedências, sendo seis de *Eucalyptus grandis* e duas de *E. saligna*, quanto à divergência genética. Os dois primeiros componentes principais absorveram 95,7% da variação observada. Então, no gráfico de dispersão de escores três grupos de similaridade foram identificados. CRUZ (1990) praticamente esgotou toda a variação observada ao utilizar os dois primeiros componentes principais, para estudar a divergência entre cinco cultivares de milho, avaliados por oito caracteres. Esses dois componentes explicaram 94% da variação total.

HUSSAINI et alii (1977) analisaram 640 acessos da coleção mundial de *Eleusine coracana* por componentes principais, com base em 18 caracteres selecionados. 12 grupos foram subjetivamente formados no gráfico de dispersão estabelecido pelos dois primeiros componentes que, juntos, explicaram 60% da variação. O trabalho demonstrou então a grande diversidade presente na coleção. A divergência entre seis cultivares de pimentão foi estudada por MIRANDA et alii (1988), por análise de componentes principais. A visualização dos cultivares em gráfico, proporcionado pelos dois primeiros componentes (89% da variação total), resultou na identificação de alguns cultivares com maior divergência entre si. Os componentes, neste estudo, representaram combinação linear de 11 caracteres agronômicos.

A técnica de componentes principais é a mais difundida entre os aplicadores de análises multidimensionais. Não obstante essa técnica considera apenas a matriz de covariância entre médias de tratamentos H ou mais comumente a matriz de correlação R. Todavia, a correlação residual entre caracteres não é considerada. A análise da divergência por variáveis canônicas, por sua vez, considera ambas as matrizes de covariâncias, apresentando maior consistência nas estimativas de divergência. A análise por variáveis canônicas será discutida a seguir.

2.2.4. Análise por variáveis canônicas

A análise por variáveis canônicas constitui uma técnica multivariada que, a semelhança de componentes principais, busca condensar um complexo de múltiplas variáveis em uma ou poucas variáveis canônicas que são combinações lineares das primeiras. Variáveis canônicas produzem, então, a máxima separação entre grupos, pela maximização das diferenças entre eles. Ainda, variáveis canônicas ou funções discriminantes foram desenvolvidas por FISHER (1936), quando investigava caracteres morfológicos visando classificar espécies do gênero *Iris*.

Para aplicação de análise por variáveis canônicas necessita-se da matriz de covariâncias entre tratamentos H e da matriz de covariância residual E. A seguir são calculados os autovalores e autovetores associados referentes à matriz produto $E^{-1}H$. A combinação linear das variáveis estudadas, cujos coeficientes de ponderação são os elementos do autovetor normalizado, ou seja de comprimento unitário, associado ao maior autovalor, corresponde à primeira variável canônica. O segundo maior autovalor corresponde à segunda variável canônica e assim sucessivamente. Como em componentes principais, a importância relativa das variáveis canônicas é determinada com base na percentagem da variância total atribuída a cada uma delas.

Ainda à semelhança de componentes principais, a análise por variáveis canônicas visa reduzir a dimensionalidade dos dados. Desse modo, ao plotarem-se os escores das primeiras variáveis canônicas em um gráfico bidimensional, obtém-se uma aproximação da configuração multidimensional dos dados originais. Óbvio que é melhor sumarizar o complexo multivariado logo nas primeiras variáveis, desde que essas respondam por 70% ou mais da variância total. Nos gráficos, a distância relativa entre os genótipos pode ser visualizada, assim como a identificação subjetiva de grupos de similaridade pode ser promovida.

Análise por variáveis canônicas tem sido aplicada em espécies arbóreas com diferentes objetivos. No estudo sobre taxonomia de híbridos entre *Eucalyptus cinerea* x *E. maculosa* (MERGEN et alii, 1966), híbridos intra e interespecíficos usando *Eucalyptus nitens* (TIBBITS, 1988) e híbridos de *Humulus lupulus* (SMALL, 1981); na diferenciação de populações de *Picea sitchensis* (FALKENHAGEN & NASH, 1978), *Pinus brutia* (CALAMASSI et alii, 1988), *Pinus kesiya* (BURLEY & BURROWS, 1972), *Abies amabilis* e *A. lasiocarpa* (MAZE et alii, 1981) e diferenciação no complexo *Eucalyptus gunnii-archeri* (POTTS & REID, 1985b). Outra importante aplicação das variáveis canônicas é a identificação de áreas próprias para produção sustentada de madeira e áreas destinadas à preservação em florestas naturais (SOUZA et alii, 1990).

Em eucalipto, PULCINELLI et alii (1993) identificaram três grupos de similaridade ao estudar seis populações de *Eucalyptus grandis* e duas de *E. saligna*. Os grupos foram identificados no gráfico cujas coordenadas foram as duas primeiras variáveis canônicas. A primeira e a segunda variáveis acumularam 95,6% da variabilidade observada. Nessa análise apenas três caracteres silviculturais foram considerados e resultaram suficientes para revelar divergência entre as populações. RESENDE & HIGA (1991) quantificaram a divergência genética entre sete procedências de *E. viminalis* objetivando a hibridação intraespecífica. As duas primeiras variáveis canônicas absorveram 86,74% da variação total. Sete variáveis foram empregadas nesse estudo.

RAMANUJAM et alii (1974) aplicaram análise por variáveis canônicas em 35 populações, sendo 10 genitores e 25 F₁'s, de *Phaseolus aureus*. No gráfico de dispersão, estabelecido com as duas primeiras variáveis canônicas, identificaram-se 19 grupos de similaridade. O resultado sugeriu a presença de substancial diversidade na espécie, em desacordo com a escassez de variabilidade frequentemente relatada na literatura de *Phaseolus aureus*. A diversidade genética entre 23 variedades de *Vigna mungo*, coletadas de diferentes partes da Índia, foi também analisada por variáveis

canônicas (DAS & DAS GUPTA, 1984). As duas primeiras variáveis canônicas, formadas pela combinação linear de sete características, absorveram 70% da variabilidade total.

Em *Eleusine coracana*, 30 populações avaliadas por 12 caracteres foram investigadas quanto à divergência genética por JAIN et alii (1981). Análise conduzida por variáveis canônicas revelou que 87,46% da variabilidade total foi explicada pelas duas primeiras raízes canônicas. HUSSAINI et alii (1977) plotaram as duas primeiras raízes canônicas, que acumularam 57,7% da variação total, em um gráfico bidimensional e identificaram 12 grupos de similaridade entre 640 acessos de *Eleusine coracana*. CRUZ (1990) detectou acúmulo de 97,39% da variação nas duas primeiras variáveis canônicas, ao investigar a divergência entre cinco cultivares de milho. FERREIRA (1993), estudando a divergência entre 28 variedades de milho avaliadas por 19 caracteres, encontrou absorção de 73,3% da variação total pelas quatro primeiras variáveis.

A coleção mundial de *Cicer arietinum* foi avaliada para oito caracteres quanto a divergência genética. As duas primeiras raízes canônicas absorveram 84,16% da variação total, permitindo a visualização das posições relativas das variedades (NARAYAN & MACEFIELD, 1976). RESENDE (1991) quantificou a divergência fenética entre 11 espécies de orquídeas do gênero *Laelia* seção *Parviflorae*, com base nas duas primeiras variáveis canônicas. Juntas essas variáveis acumularam 80,31% da variação total. Divergência fenética, no caso, é um termo empregado por SNEATH & SOKAL (1973) para caracterizar relações entre unidades taxonômicas que se baseiam em expressões fenotípicas e não propriamente genéticas.

Os estudos de RAMANUJAM et alii (1974); NARAYAN & MACEFIELD (1976); JAIN et alii (1981); DAS & DAS GUPTA (1984); RESENDE (1991) e RESENDE & HIGA (1991) apresentaram resultados concordantes com referência ao padrão de divergência obtido por D^2 e variáveis canônicas. Alta associação entre D^2 e distância Euclidiana foi detectada por FERREIRA (1993). Por sua vez, HUSSAINI et alii (1977) e CALAMASSI et alii (1988) encontraram concordância entre as análises de divergência conduzidas por componentes principais e variáveis canônicas. PULCINELLI et alii (1993) apresentaram resultados concordantes entre as análises por D^2 , componentes principais e variáveis canônicas. E, finalmente, CRUZ (1990) e PIRES (1993) relataram elevada concordância entre D^2 , distância Euclidiana, componentes principais e variáveis canônicas. Esses resultados são de fundamental importância, na

medida que demonstram a consistência das técnicas multivariadas e validam a aplicação de quaisquer delas, em estudos de estimação da divergência genética entre populações.

SPAGNOLETTI & QUALSET (1987) avaliaram 3.000 acessos de trigo, originários de 26 países, por variáveis canônicas. Utilizando oito caracteres quantitativos medidos na espiga os autores detectaram a formação de cinco grupos. As duas primeiras variáveis canônicas acumularam 73% da variação total. Em arroz, 67 germoplasmas, sendo 59 brasileiros e 8 japoneses, foram analisados por MONTALVAN DEL AGUILA (1990) através de variáveis canônicas, aplicadas a padrões eletroforéticos protéicos de grãos. O gráfico de dispersão de variáveis confirmou a diferenciação entre as variedades de diferentes origens.

2.3. Importância relativa dos caracteres para divergência

A finalidade última da análise multidimensional é extrair um subconjunto de variáveis ou caracteres as quais contenham as principais informações da amostra completa. JOLLIFFE (1972; 1973) comentou que, frequentemente, o resultado final é pouco alterado se um subconjunto de variáveis é utilizado. A razão é que algumas variáveis presentes na análise, sendo pouco informativas, contribuem para complicar os dados, ao invés de proporcionarem informação adicional. Desse ponto de vista, as variáveis não redundantes e de expressiva variabilidade são as mais adequadas para análise multidimensional.

A importância relativa das variáveis ou caracteres em estudos de divergência pode ser deduzida tanto de componentes principais, quanto de variáveis canônicas. Todavia, os procedimentos desenvolvidos a partir das técnicas comentadas referem-se à seleção ou descarte de variáveis, sendo as remanescentes as variáveis de maior importância. JOLLIFFE (1972; 1973) examinou diferentes métodos de rejeição de variáveis utilizando inclusive a técnica de componentes principais. Estudando variáveis simuladas e reais o autor deduziu que o número de variáveis rejeitadas devia ser igual ao número de componentes principais com variância menor que 0,70, obtidos de matriz de correlação. Assim, as variáveis associadas aos maiores coeficientes de ponderação nos últimos componentes, até o limite de variância de 0,70, seriam passíveis de descarte. MORRISON (1976); MARDIA et alii (1979) e VAN LAAR (1991) relataram o mesmo procedimento.

Da mesma maneira as variáveis descartáveis podem ser identificadas com o emprego de variáveis canônicas. Descartáveis seriam então aquelas variáveis associadas aos maiores produtos entre o desvio-padrão residual da variável e seus

respectivos coeficientes de ponderação das últimas variáveis canônicas (SINGH, 1981). VAN LAAR (1991) comenta o procedimento. SINGH (1981), no entanto, avança na questão da contribuição relativa dos caracteres para a divergência. Antes dele tais caracteres eram identificados a partir das variáveis transformadas não correlacionadas.

PEREIRA (1989) empregou a recomendação de descarte de variáveis, a partir de componentes principais, descrita por JOLLIFFE (1972; 1973), em banco de germoplasma de mandioca. Concluiu o autor ser possível o descarte de 14 dos 28 descritores utilizados no estudo (50% de descarte). No entanto, CURI (1993), também em mandioca, descartou apenas 31% das variáveis (6 em 20) ao adotar parcialmente a recomendação de JOLLIFFE (1972; 1973). A estratégia previu a reanálise do conjunto de variáveis remanescentes a cada novo descarte e a observação da matriz de correlação fenotípica entre variáveis, para auxiliar a eliminação das redundantes.

CRUZ (1990) empregou tanto componentes principais quanto variáveis canônicas, para detectar caracteres de menor importância no estudo de divergência entre cinco cultivares de milho. Muito embora não tenha promovido o descarte de caracteres, uma vez que contava com apenas oito deles, o autor demonstrou satisfatória concordância entre as metodologias de JOLLIFFE (1972; 1973) e SINGH (1981).

A seleção de variáveis abordada até o momento se processa com a aplicação das técnicas multivariadas. Todavia, a literatura contém estudos em que ocorre uma seleção prévia de caracteres, antes mesmo de aplicarem-se as técnicas multidimensionais. Tal seleção baseia-se em análises univariadas apenas. WILCHES (1983), por exemplo, empregou o coeficiente de repetibilidade para selecionar 13 dentre 31 caracteres avaliados em 34 variedades de amendoim, para serem submetidos a análises multivariadas. Idêntico critério foi utilizado por SMITH & SMITH (1989a) para selecionar entre 57 caracteres morfológicos avaliados em 31 linhagens de milho, ensaiadas por dois anos em três locais. Em *Eleusine coracana*, HUSSAINI et alii (1977) usaram como critério de seleção de caracteres a razão entre o componente de variância entre médias de acessos e a soma dos componentes de variância do erro e das repetições. Caracteres com valores da razão inferiores a 1,00 (11 em 29 caracteres) não foram submetidos a análises multivariadas. Claramente os trabalhos anteriores preocuparam-se em eliminar da análise multidimensional os caracteres muito influenciados pelo ambiente.

2.4. Estabilidade temporal da divergência expressa por D^2

A literatura contém vários exemplos de emprego de genitores divergentes para gerar segregantes transgressivos nas mais diversas culturas. Todavia, é desejável que a estimativa da divergência genética entre os futuros genitores dos segregantes transgressivos manifeste-se de modo consistente no tempo e no espaço. Essa consistência é traduzida pela estabilidade temporal e espacial do padrão de agrupamento dos genitores, determinado, tanto pelos métodos de otimização e hierárquicos quanto pelos gráficos de dispersão. A estabilidade pode, também, ser acessada pelo grau de correlação entre as estatísticas de divergência, notadamente D^2 , nos diferentes locais e/ou anos. Entretanto, a literatura pouco tem tratado dessa última possibilidade.

BHATT (1976) avaliou 12 genótipos de trigo em quatro diferentes ambientes, envolvendo locais e anos. A estatística D^2 de Mahalanobis foi estimada com base em seis caracteres de qualidade para o conjunto dos ambientes e para cada ambiente separadamente. O padrão de agrupamento por Tocher, contudo, manteve-se satisfatoriamente estável nos diferentes ambientes. O resultado sugeriu então a possibilidade de agrupar um grande número de genótipos, avaliados pela qualidade, com base na divergência estimada em um simples ambiente. Aplicando o método de estabilidade de Finlay e Wilkinson a cada caráter o autor detectou independência entre essa estabilidade e aquela dos agrupamentos por D^2 . Portanto, a estabilidade dos grupos pareceu ser função da expressão balanceada dos diferentes caracteres sobre os vários ambientes.

Dois ambientes envolvendo diferentes níveis de fertilidade foram utilizados para cultivo de 18 variedades de arroz do tipo anão. Métodos de taxonomia numérica foram empregados para a formação de clusters. Os resultados demonstraram que o padrão de agrupamento, no diagrama obtido da matriz de distâncias fenéticas (fenograma), foi concordante com o pedigree e satisfatoriamente estável em ambos os ambientes (JANORIA et alii, 1976). SOMAYAJULU et alii (1970) detectaram estabilidade no número e na composição de grupos em linhagens de trigo, cultivadas sob três níveis de fertilidade. Enfatizaram que, sobretudo, os grupos amplamente divergentes não foram influenciados pelos diferentes ambientes. A preocupação com a estabilidade também foi demonstrada por GHADERI et alii (1979), ao aplicar análise de cluster a 16 genótipos de *Vigna radiata*, submetidos a 18 ambientes distintos (três níveis de fertilidade, três densidades e duas épocas de plantio).

A divergência entre 20 variedades de *Eleusine coracana* foi estimada por D^2 , seguida do agrupamento pelo método de Tocher. Quatro grupos foram

formados sob ambiente não irrigado e 11 grupos sob condições irrigadas (SHERIFF, 1992). A inconsistência dos clusters nos dois ambientes foi evidente e o melhor ambiente proporcionou maior discriminação entre as variedades. Também SINGH & GUPTA (1979) relataram que a diversidade entre 34 linhagens de *Pennisetum typhoides* foi melhor expressa em ambientes fertilizados, em comparação àqueles ambientes sem fertilização. Na coleção mundial da mesma espécie, UPADHYAY & MURTY (1970) também relataram consideráveis diferenças na composição dos grupos formados entre os 360 acessos avaliados em duas regiões, com e sem déficit hídrico.

Experimentos envolvendo 36 variedades de cevada, instalados sob condições irrigadas e não irrigadas, apresentaram padrões distintos de agrupamentos por Tocher aplicado a D^2 . A inconsistência do padrão de agrupamento baseado em D^2 também tem sido relatada em *Gossypium hirsutum* cultivado sob quatro diferentes ambientes agronômicos (SINGH & GILL, 1984). Desses últimos estudos citados presume-se que a resposta diferencial dos diferentes caracteres, exibida nos vários ambientes, reflete a inconsistência dos agrupamentos de similaridade. Portanto, na ausência de informações sobre interação genótipo por ambiente é prudente conduzir estudos sobre divergência em vários locais e anos. Melhor ainda que esse estudo seja conduzido sobre ambientes mais favoráveis.

3. MATERIAL E MÉTODOS

3.1. Estudo I: Predição da performance de híbridos por divergência dos genitores

3.1.1. Material de estudo

O presente estudo utilizou dados obtidos de cultivares e híbridos de cacau. Estes materiais foram avaliados em sistemas de cruzamentos dialélicos, em estudo apresentado anteriormente por DIAS (1991). Assim, cinco cultivares S_1 de cacau não comerciais foram ensaiados em conjunto com seus 20 híbridos. Desse modo, foi possível prever a performance dos híbridos com base nos seus genitores e testar a eficiência da predição pela comparação com a performance realizada. Os cultivares, suas origens e grupos raciais estão descritos abaixo:

- 1) CC 41 - grupo trinitário originário da Costa Rica, do Centro de Cacao;
- 2) SIAL 169 - forasteiro amazônico originário do sul da Bahia e selecionado pelo extinto Instituto Agrônômico do Leste;
- 3) CEPEC 1 - forasteiro amazônico do sul da Bahia selecionado como mutante espontâneo de amêndoas brancas. Nomeado pelo Centro de Pesquisas de Cacau, CEPEC;
- 4) ICS 1 - trinitário selecionado em Trinidad pelo Imperial College of Tropical Agriculture;
- 5) SIC 19 - forasteiro amazônico selecionado em um único município do sul da Bahia pelo extinto Instituto de Cacau da Bahia.

Esses cultivares e seus 20 híbridos, incluindo os híbridos recíprocos, foram conduzidos em blocos casualizados com quatro repetições e parcelas de 16 plantas. O sombreamento provisório envolveu o plantio de bananeiras no espaçamento de 3 x 3 m, igual ao dos cacaueiros. O sombreamento definitivo foi realizado com *Erythrina* sp. a cada 24 metros. As sementes dos cruzamentos foram obtidas por polinizações manuais controladas nos genitores clonais. Por sua vez, as sementes dos cultivares foram obtidas por autofecundação dos clones correspondentes. Todos os cultivares envolvidos nesse estudo são autocompatíveis.

O experimento foi instalado no Centro de Pesquisas de Cacau - CEPEC, em Itabuna, Ba, no ano de 1975, com produção efetiva considerada a partir de 1980. Para efeito de análise do presente estudo utilizou-se de dados de colheitas mensais obtidos de 1986 a 1990, considerando-se a produção acumulada do ano agrícola (abril do ano 0 a março do ano 1). Para obter informações detalhadas de clima e solo do local e mesmo sobre o manejo do experimento o leitor deve remeter-se ao estudo de DIAS (1991).

Os dados coletados por parcela foram o número de frutos sadios e doentes, o peso de sementes úmidas, o número real de plantas ou estande e o número de plantas produtivas. Entretanto, na avaliação dos cultivares e cruzamentos foram considerados cinco caracteres: i) número médio de frutos sadios (NFSP), ii) número médio de frutos colhidos (NFCP), iii) peso médio de sementes úmidas (PSUP), iv) percentagem média de frutos doentes (PFDP) por planta e v) peso de sementes úmidas por fruto (PSUF).

NFSP e NFCP representam a produção real e potencial, respectivamente. PSUP, em kg, é a razão entre o peso de sementes úmidas por parcela e o estande. PSUF, em gramas, refere-se a razão entre o peso de sementes úmidas e o número de frutos sadios, obtidos por parcela. Já PFDP representa a razão entre o número de frutos doentes e o número de frutos colhidos na parcela. Este último caráter sofreu transformação do tipo $\sqrt{x + 0,5}$.

3.1.2. Metodologias

Inicialmente procedeu-se à análise conjunta da variância univariada, visando detectar a existência de variabilidade entre os cultivares avaliados no período 1986-90. Com relação a efeitos, anos e resíduo foram considerados aleatórios, com cultivares previamente definidos como de efeitos fixos. A seguir, realizou-se a análise da variância multivariada, aplicada a todas as cinco variáveis simultaneamente. Este procedimento favorece a análise da correlação que possa existir entre tais variáveis. Para tanto, foram obtidas as matrizes de somas de quadrados e de produtos de tratamentos (H), e de resíduo (E) da análise conjunta da variância univariada. Para o teste de significância das diferenças entre os vetores de médias dos tratamentos se utilizou a estatística Λ de Wilks (WILKS, 1932), assim definida:

$$\Lambda = \frac{\det E}{\det (H + E)}$$

Na avaliação da significância da estatística Λ optou-se por transformar o valor de Λ em um valor correspondente de F , conforme Harris (1975) citado por PIMENTEL GOMES (1990, p. 345). A fórmula geral empregada foi:

$$F[pk_1; ab - c] = \left[(1/\Lambda)^{1/b} - 1 \right] [ab - c/pk_1]$$

onde,

$$a = k_2 - 0,5(p - k_1 + 1) \quad b = \sqrt{(p^2 k_1^2 - 4)/(p^2 + k_1^2 - 5)} \text{ com } p^2 + k_1^2 \neq 5$$

$$c = 0,5(pk_1 - 2)$$

p = número de variáveis;

k_1 = número de graus de liberdade associado a tratamentos;

k_2 = número de graus de liberdade do resíduo;

Λ = razão entre o determinante da matriz de somas de quadrados e de produtos residuais e o determinante da matriz resultante da soma dos elementos das matrizes de somas de quadrados e de produtos residuais e de tratamentos.

A quantificação da divergência genética entre os genitores foi realizada através da distância Euclidiana média calculada de dados originais padronizados, dos escores dos dois primeiros componentes principais e dos escores das duas primeiras variáveis canônicas. Ambos os escores foram também obtidos de dados originais padronizados. Ainda, na quantificação da divergência utilizou-se a distância generalizada de Mahalanobis. A visualização da divergência no plano cartesiano foi possível pela dispersão dos escores das duas primeiras variáveis canônicas nos eixos perpendiculares.

A concordância entre si das estimativas de divergência, avaliada pelas quatro estatísticas multivariadas mencionadas acima, foi quantificada com base no relacionamento estatístico entre elas; no teste de igualdade entre a matriz de correlação residual e a matriz identidade (GODOI, 1985) e; pela magnitude dos coeficientes de correlação de Spearman, obtidos entre os pares dessas estatísticas. A seguir, a eficiência da predição da performance dos híbridos, pela divergência entre os genitores S_1 , foi também quantificada pela correlação de Spearman, entre a distância generalizada de Mahalanobis e a performance média e heterótica realizada desses híbridos. Para maior eficácia no cálculo dessa correlação, os 20 híbridos foram combinados pela média dos F_1 's e F_1 's recíprocos correspondentes, em apenas 10 híbridos. A não significância de efeitos recíprocos detectada por DIAS (1991), no mesmo dialelo e para os mesmos caracteres e período, possibilitou esse tipo de ajuste.

A contribuição relativa dos caracteres para a divergência foi avaliada segundo o método de SINGH (1981), aplicado às estimativas de distância de Mahalanobis. Por outro lado, a importância desses caracteres foi acessada também com base na metodologia de SINGH (1981) aplicada às variáveis canônicas e na recomendação de JOLLIFFE (1972; 1973) aplicada a componentes principais. Outra informação relevante acessada nesse primeiro estudo referiu-se à estabilidade temporal da divergência entre os genitores. Para tanto, foram comparadas as estimativas de distância de Mahalanobis obtidas com base no período 1986/90, com as distâncias obtidas para cada ano, isoladamente. O coeficiente de correlação de Spearman foi utilizado para essa comparação. Também a relação entre divergência genética e diversidade ecológica e geográfica foi acessada, com base no padrão de agrupamento pelo método de Tocher.

Os métodos multivariados de predição da divergência genética empregados nesse estudo foram a análise de cluster ou de conglomerado, a análise por componentes principais e a análise por variáveis canônicas. Considerações teóricas sobre esses métodos são apresentadas a seguir. No interesse da simplicidade essas considerações resumem-se a uma abordagem suficiente para aplicação das técnicas multivariadas nas questões do melhoramento. O leitor interessado nos fundamentos teórico-estatísticos de tais técnicas deve remeter-se à literatura citada.

3.1.2.1. Análise de cluster

A análise de cluster ou de conglomerado compõe-se de duas fases. Na primeira, estima-se uma medida de dissimilaridade entre os genitores. A segunda, desenvolvida a partir da primeira, emprega uma técnica de identificação e agrupamento dos genitores pela similaridade. As medidas de dissimilaridade utilizadas nesse trabalho foram a distância Euclidiana média e a distância generalizada de Mahalanobis. Como métodos de agrupamento empregaram-se o do vizinho mais próximo (single linkage method, também conhecido como nearest neighbour method) e o método da distância mínima generalizada de Tocher. A análise de cluster é tratada em detalhes por VAN LAAR (1991).

a. Distância Euclidiana (d_{ij})

Seja X_{ij} a observação no genitor i ($i = 1, 2, \dots, g$) referente ao caráter j ($j = 1, 2, \dots, p$) estudado. Define-se a distância Euclidiana (d_{ij}) entre dois genitores i e i' pela seguinte expressão:

$$d_{ii'} = \sqrt{\sum_{j=1}^p (X_{ij} - X_{i'j})^2}$$

A distância Euclidiana, entretanto, altera-se com a mudança de escala de mensuração dos caracteres e com o número desses caracteres em estudo. Esses dois inconvenientes foram superados pelo emprego da distância Euclidiana média estimada de dados padronizados. Pela padronização a observação X_{ij} feita no caráter j do genitor i foi subtraída à média geral e o resultado dividido pelo desvio padrão do caráter j , como expresso abaixo:

$$x_{ij} = \frac{X_{ij} - \bar{X}_j}{s(X_j)}$$

após a padronização a distância Euclidiana média foi calculada pela nova expressão:

$$d_{ii'} = \sqrt{\left[\sum_{j=1}^p (x_{ij} - x_{i'j})^2 \right] / p}$$

onde p = número de caracteres em estudo.

b. Distância generalizada de Mahalanobis ($D_{ii'}^2$)

Ao contrário da distância Euclidiana, a distância generalizada de Mahalanobis (RAO, 1952) considera a correlação residual existente entre os caracteres. Por esse motivo, tem sido amplamente utilizada nos estudo sobre divergência.

Sejam X_{ij} as estimativas de médias das X_{ijk} observações feitas no genitor i ($i = 1, 2, \dots, g$), no caráter j ($j = 1, 2, \dots, p$) avaliadas na repetição k . Sejam, ainda, E a matriz de variâncias e covariâncias residuais ou matriz de dispersão residual e δ o vetor de elementos d_j representando a diferença entre médias de dois genitores i e i' , para cada j . Então, por definição, tem-se:

$$D_{ii'}^2 = \delta E^{-1} \delta$$

c. Método do vizinho mais próximo

É um método adequado quando trata-se de identificar grupos homogêneos de genitores. O dendrograma obtido por essa metodologia compõe-se de grupos formados a diversos níveis, por genitores similares. Assim, o grupo inicial foi formado pelos genitores mais próximos, determinados pela estimativa de distância de Mahalanobis. A seguir, calculou-se a distância desse grupo com relação aos demais genitores e, em outra etapa, em relação a outros grupos já formados.

d. Método de Tocher

No método de Tocher, apresentado por RAO (1952), são estabelecidos grupos cujas médias das distâncias intragrupos são sempre inferiores às distâncias médias intergrupos. Formam-se, então, grupos mutuamente exclusivos. O passo inicial para implementação desse método foi a identificação dos genitores mais próximos na matriz de distâncias de Mahalanobis, para formação do primeiro grupo. Ainda nessa mesma matriz foram identificadas as distâncias mínimas referentes a cada genitor. O maior elemento do conjunto das mínimas distâncias foi estabelecido como limite máximo para inclusão de novos genitores no grupo. É que, a inclusão de um novo genitor implica em aumento na média da distância intragrupo. Superado o limite máximo permitido para entrada de novos candidatos ao grupo, formou-se novo grupo e, de modo análogo, reiniciou-se o processo, até a inclusão do último candidato.

3.1.2.2. Análise por componentes principais

Essa técnica consiste em transformar um conjunto original de variáveis, oriundo de um espaço multidimensional, em outro conjunto equivalente, porém no espaço bi ou tridimensional. Percebe-se que essa técnica ganha importância quando o estudo envolve um número relativamente elevado de variáveis. Nesse caso, a utilização de componentes principais representa uma simplificação na interpretação dos resultados. Diversas obras tratam da análise de componentes principais, destacando-se ANDERSON (1958); KENDALL (1963); MORRISON (1976) e VAN LAAR (1991).

Seja X_{ij} a média do caráter j ($j = 1, 2, \dots, n$) avaliada no genitor i ($i = 1, 2, \dots, g$) formando o conjunto de variáveis $X_{i1}, X_{i2}, \dots, X_{in}$. Seja, ainda, R a matriz de correlações fenotípicas entre caracteres ou matriz de covariâncias entre caracteres cujos dados foram padronizados. Desse modo, obter-se-á um novo conjunto de variáveis $Y_{i1}, Y_{i2}, \dots, Y_{in}$, onde cada Y_{ij} é um componente principal que representa uma combinação linear de variáveis do primeiro conjunto. Supondo dois componentes principais, como abaixo:

$$Y_{ij} = a_1 x_{i1} + a_2 x_{i2} + \dots + a_n x_{in}$$

$$Y_{ij} = b_1 x_{i1} + b_2 x_{i2} + \dots + b_n x_{in}$$

onde as seguintes propriedades se verificam

$$\sum_{j=1}^n a_j^2 = \sum_{j=1}^n b_j^2 = 1$$

$$\sum_{j=1}^n a_j b_j = 0$$

ou seja, os coeficientes de ponderação (a e b) são normalizados e os componentes principais são não-correlacionados. Por definição, o primeiro componente é o de maior variância (λ_1) seguido pela segunda maior variância (λ_2), dada pelo segundo componente e assim sucessivamente. Para maximizá-los estimam-se as raízes características ou autovalores (λ_j) da matriz R e os vetores característicos ou autovetores (α_j) associados. Esses últimos contém os coeficientes de ponderação dos caracteres. Estimativas de λ_j e de α_j são obtidas pela solução do sistema

$$(R - \lambda_j I)\alpha_j = \Phi \text{ o que equivale a}$$

$$\det(R - \lambda_j I) = 0$$

sendo

R = matriz de correlações fenotípicas entre variáveis padronizadas;

λ_j = autovalores da matriz R, um para cada componente principal;

I = matriz identidade;

α_j = autovetores cujos elementos são os coeficientes de ponderação das variáveis originais;

Φ = vetor coluna nulo.

Ainda a importância relativa dos caracteres na divergência é informada pelos componentes principais. Assim, a variável ou caráter que apresentar o maior coeficiente de ponderação no componente principal de menor autovalor, será a de menor importância. Naturalmente, que essa informação presta-se ao descarte de variáveis. Também a importância relativa de um componente é avaliada pela percentagem da variação total observada que ele explica. Desse modo, a importância de Y_j é expressa por:

$$Y_j = \frac{\lambda_j}{\sum \lambda_j} = \frac{\lambda_j}{\text{Traço}R}$$

Na etapa seguinte, a divergência entre os genitores foi avaliada pela distância Euclidiana média aplicada aos escores dos primeiros componentes principais. Para tanto, o critério adotado foi a explicação mínima de 70% da variação total, por parte desses primeiros componentes.

3.1.2.3. Análise por variáveis canônicas

As variáveis canônicas ou funções discriminantes proporcionam uma grande redução da dimensionalidade do espaço, à semelhança dos componentes principais. Aqui também, um conjunto de variáveis originais é convertido em um outro conjunto de variáveis canônicas, com o último conjunto menor que o primeiro. Os livros-

textos de RAO (1952); ANDERSON (1958); KENDALL (1963); MORRISON (1976) e VAN LAAR (1991) abordam a técnica.

Seja X_{ij} a média do caráter j ($j = 1, 2, \dots, n$) avaliado no genitor i ($i = 1, 2, \dots, g$), H a matriz de covariâncias entre as médias dos genitores e E a matriz de covariâncias residuais. Pode-se estimar, então, a raiz característica ou autovalor (λ_j) e o vetor característico ou autovetor (α_j) associado a cada raiz. O autovetor compõe-se de elementos que se constituem nos coeficientes de ponderação. As estimativas de λ_j e α_j são obtidas pela solução do sistema:

$$(H - \lambda_j E)\alpha_j = \Phi \quad \text{a partir do sistema determinantal,}$$

$$|H - \lambda_j E| = 0$$

Cada raiz característica corresponde a uma variável canônica. Por sua vez, cada variável canônica é uma função linear das X_{ij} variáveis originais, como abaixo:

$$Y_{ij} = a_1 x_{i1} + a_2 x_{i2} + \dots + a_n x_{in}$$

sendo uma variável canônica e,

$$Y_{ij} = b_1 x_{i1} + b_2 x_{i2} + \dots + b_n x_{in}$$

a outra variável canônica, as seguintes propriedades se verificam:

$$\sum_{j=1}^n \sum_{j'=1}^n a_j a_{j'} \delta_{jj'} = \sum_{j=1}^n \sum_{j'=1}^n b_j b_{j'} \delta_{jj'} = 1$$

$$\sum_{j=1}^n \sum_{j'=1}^n a_j b_{j'} \delta_{jj'} = 0$$

$\delta_{jj'}$ = covariância residual entre os caracteres j e j' .

Na padronização dos dados originais para obtenção dos escores, tanto de componentes principais como de variáveis canônicas, foi adotado o procedimento de dividir o valor da observação pelo correspondente valor do desvio padrão do caráter. Na transformação dos dados originais em dados não correlacionados foi aplicado o processo de condensação pivotal (RAO, 1952). Pela condensação, os novos dados apresentam covariâncias residuais nulas e variâncias unitárias, igualando a matriz de dispersão residual à matriz identidade.

Como em componentes principais, o processo de estimação de variáveis canônicas prevê que a primeira variável Y_{i1} apresente a maior variância. Y_{i2} a segunda maior variância e assim sucessivamente, ou seja, $\lambda_1 > \lambda_2 > \dots > \lambda_n$. Também, a importância relativa das variáveis canônicas é determinada com base na percentagem da variação total que elas explicam na análise multidimensional. Por sua vez, a importância relativa dos caracteres para a divergência pode ser deduzida pela aplicação do método de

SINGH (1981). Por esse método, os caracteres de menor contribuição para a divergência são aqueles que exibem os maiores produtos entre o desvio-padrão residual do caráter e seus respectivos coeficientes de ponderação, em valores absolutos, nos menores autovalores.

Os escores das primeiras variáveis canônicas foram utilizados para quantificação da divergência entre os genitores, através da aplicação da distância Euclidiana média. E, para visualização dessa divergência, os mesmos escores foram dispostos em gráficos de dispersão. A definição do número de variáveis canônicas utilizadas para quantificação e visualização da divergência teve como critério uma absorção mínima de 70% da variação total observada. Todas as análises multivariadas foram processadas empregando-se os programas de computação GENES da Universidade Federal de Viçosa (UFV, Viçosa, MG), SISMUL do Departamento de Matemática (ESALQ, Piracicaba, SP) e CLUSTER de D. F. Ferreira (ESAL, Lavras, MG).

3.2. Estudo II: Divergência fenética entre clones de cacauero

3.2.1. Material de estudo

Dados obtidos por VELLO et alii (1969) e VELLO et alii (1972), e utilizados no presente estudo, demonstraram a existência de divergência entre clones locais, suficiente para a síntese de híbridos heteróticos. O estudo I demonstrou o potencial dos métodos multivariados na predição de híbridos pela quantificação da divergência entre genitores S_1 . Então, como um estudo adicional para complementar as conclusões do estudo I, dados obtidos de clones de cacau, publicados por CASTRO & BARTLEY (1983), foram utilizados na quantificação da divergência entre genitores clonais S_0 . Esses clones foram avaliados em 1981, na coleção de germoplasma do CEPEC, Itabuna, Ba. Os autores caracterizaram 15 clones das séries SIC (2, 17, 18, 19, 23, 24, 250, 433, 628, 644, 662, 680, 747, 765 e 823) e 11 das séries SIAL (70, 88, 105, 169, 244, 283, 325, 357, 557, 581 e 638) num total de 26 clones locais, baseados em descritores de folha, fruto e sementes.

Os clones das séries SIC e SIAL resultaram de seleção fenotípica individual praticada na população de cacaueros da Bahia. Tais seleções foram conduzidas pelos extintos Instituto de Cacau da Bahia (clones SIC) e Instituto Agrônomo do Leste (clones SIAL). Os SIC's com identificação menor que 100 são originários de seleção de plantas da antiga Estação Experimental de Uruçuca. SIC's entre

100 e 800 procedem de seleções feitas na população S₁ de clones com identificação inferior a 100. Finalmente, os SIC's acima de 800 são seleções praticadas no cultivar Catongo, um mutante de amêndoas brancas (VELLO et alii, 1972). Os clones SIAL, por sua vez, foram selecionados em numerosas fazendas do sul da Bahia, após um mínimo de sete anos e um máximo de 11 anos de colheitas controladas (PINTO et alii, 1969). Importante observar que todos os 26 clones envolvidos nesse estudo pertencem ao grupo racial forasteiro, introduzido do baixo Amazonas.

Na caracterização desses materiais foram seguidas as recomendações propostas por ENGELS et alii (1979) para os descritores de cacau. Cada descritor citado é acompanhado do tamanho (n) da amostra empregada na sua obtenção. Os descritores de folha por não serem suficientemente variáveis e sendo de menor importância agrônômica foram excluídos das análises. Assim, 13 descritores de fruto e semente foram utilizados no presente estudo:

A) Fruto

01. CF - comprimento em mm (n=35);
02. DF - diâmetro em mm (n=20);
03. C/D - relação comprimento/diâmetro do fruto;
04. PF - peso do fruto inteiro em gramas (n=35);
05. PC - peso da casca em gramas (n=10);
06. EG - espessura da casca no gomo em mm (n=35);
07. ES - espessura da casca no sulco em mm (n=35);
08. NSF - número de sementes por fruto referente apenas as sementes com desenvolvimento normal (n=35).

B) Sementes: Refere-se a uma amostra dos 10 primeiros frutos com retirada de 15 sementes por fruto.

09. PSU - peso das sementes úmidas sem a polpa e a testa em gramas;
10. PAS - peso das amêndoas secas em gramas, obtido após secagem em estufa a 100° C por 24 horas;

Os caracteres 11, 12 e 13 a seguir foram avaliados nos 20 primeiros frutos, tomando-se uma média de cinco amêndoas por fruto.

11. CA - comprimento da amêndoa em mm;
12. LA - largura da amêndoa em mm;
13. EA - espessura da amêndoa em mm;

É importante observar que no germoplasma em questão os clones são representados por, em média, 12 plantas estabelecidas em linha. Além disso, é

comum nessas coleções de germoplasmas de cacau encontrar-se clones de idades distintas. Entretanto, no caso das séries SIC e SIAL os clones apresentavam idades próximas, máxima de 18 anos, por ocasião do estudo conduzido em 1981.

Ainda em estudo de caracterização de germoplasmas de cacau não se adota, *a priori*, um número definido de plantas para serem avaliadas, quando se trata de caracteres referentes a frutos e sementes. É que a frutificação nas plantas de um mesmo clone é variável, encontrando-se desde plantas com grande carga de frutos, até aquelas onde a frutificação ainda não ocorreu. Então, em se tratando de avaliação de clones por frutos e sementes adotam-se amostras de tamanho compatível com a variabilidade apresentada por esses caracteres. POUND (1933), estudando a variabilidade de frutos e sementes, sugeriu que o tamanho amostral adequado não deveria ser inferior a 0,16 vezes o quadrado do coeficiente de variação do caráter, com 5% de probabilidade de erro. Esse tamanho amostral foi seguido por CASTRO & BARTLEY (1983) que verificaram que, para a maioria dos descritores utilizados, as amostras foram satisfatórias.

3.2.2. Metodologias

A quantificação da divergência fenética entre os clones foi realizada através da distância Euclidiana média (d_{ij}) calculada de dados originais padronizados e dos escores dos primeiros componentes principais. Para tanto, o critério adotado foi a explicação mínima de 70% da variação total disponível, por parte desses componentes. Os escores foram também obtidos de dados originais padronizados. A visualização da divergência no plano cartesiano foi obtida pela dispersão dos escores dos primeiros componentes principais nos eixos perpendiculares, desde que obedecido o critério anterior.

Poder-se-ia quantificar a divergência entre clones empregando-se a distância generalizada de Mahalanobis, em substituição à distância Euclidiana. A matriz de dispersão residual W seria obtida dentro de clones, entre plantas. Entretanto, essa matriz de covariância W é amostral e diferente da matriz E , obtida de estimativas de erros experimentais. Alguns trabalhos da literatura têm se utilizado dessa possibilidade, a exemplo de FALKENHAGEN & NASH (1978). Todavia, os dados disponíveis para o presente estudo foram apenas as médias dos clones, encontradas em CASTRO & BARTLEY (1983). Com dados médios somente os métodos que não consideram a matriz de dispersão residual são passíveis de aplicação.

Os métodos de agrupamento utilizados foram dos tipos hierárquico e de otimização. O método do vizinho mais próximo representando o primeiro tipo e o de

Tocher o segundo. Os métodos hierárquicos são muito informativos quanto à variabilidade existente em coleções de germoplasmas, por estabelecerem relações hierárquicas de afinidade entre acessos (PEETERS & MARTINELLI, 1989). Definidos os grupos, conforme o método de Tocher, as estimativas de distâncias Euclidianas intra e intergrupos foram obtidas. Distância Euclidiana intragrupo corresponde à média das distâncias entre os pares de clones que constituem o grupo. Por sua vez, as distâncias intergrupos correspondem às médias das distâncias entre os pares de clones pertencentes aos diferentes grupos. Também os valores médios dos grupos foram analisados.

A importância relativa dos caracteres na divergência foi determinada com base nos componentes principais, adotando-se a recomendação de JOLLIFFE (1972; 1973). Essa recomendação preconiza que as variáveis ou caracteres que apresentarem os maiores coeficientes de ponderação, nos componentes principais cujos autovalores forem inferiores a 0,7, sejam classificadas como de menor importância e passíveis de descarte. O descarte de variáveis, contudo, obedeceu à recomendação anterior aliada à análise das correlações fenotípicas entre os caracteres e à aplicação da estratégia utilizada por CURI (1993). Esse último autor promoveu o descarte como recomendado por JOLLIFFE (1972; 1973), porém reanalisando o conjunto de variáveis remanescentes com respeito a divergência, à cada novo descarte. A análise das correlações auxiliou o descarte pela eliminação das variáveis redundantes.

4. RESULTADOS E DISCUSSÃO

4.1. Estudo I: Predição da performance de híbridos por divergência dos genitores

4.1.1. Análises da variância uni e multivariada

A análise conjunta da variância univariada, envolvendo apenas os cultivares parentais, evidenciou diferenças significativas entre médias de cultivares e anos, a 1% de significância, para todos os caracteres avaliados (Tabela 1). A interação cultivares x anos, entretanto, não acusou significância, exceto para o caráter percentagem de frutos doentes por planta, PFDP. Assim, os cultivares tiveram uma performance relativa semelhante frente aos diferentes anos. Ou ainda, os diferentes anos influenciaram de modo relativamente semelhante os cinco cultivares em teste. Por sua vez, a análise da variância multivariada com base na estatística Λ de Wilks, transformada para um valor correspondente de F, detectou diferenças significativas ($P < 0,0001$) entre os vetores de médias dos cultivares, quando considerados todos os caracteres simultaneamente ($F = 19,37$ para 20 e 187 graus de liberdade). Convém ressaltar que a aplicação da análise multivariada foi precedida de um teste de normalidade das variáveis.

O presente experimento foi instalado na segunda metade da década de 1970. Nessa época, já se dispunha de informações sobre a heterose para produção em cacau. Sabia-se que a combinação de genitores de origem geográfica e grupos raciais distintos geravam híbridos heteróticos. Portanto, a diversidade verificada entre os cultivares era um fato previsível. Dos cinco cultivares em estudo, três resultaram de seleções conduzidas nas populações de cacaueiros forasteiros amazônicos da Bahia e dois tiveram origem nas populações de trinitários de Trinidad (ICS 1) e Costa Rica (CC 41). Essa mesma constatação de diversidade entre os cultivares foi detectada em trabalho anterior por DIAS (1991).

Tabela 1 - Resumo das análises conjunta da variância univariada, referentes a cinco características avaliadas em cinco cultivares de cacau, em blocos casualizados com quatro repetições, durante o período 1986-90. CEPEC, Itabuna, Ba.

FV	GL	QM				
		NFSP	NFCP	PSUP	PSUF	PFDP
Blocos/Anos	15	78,974	87,727	0,943	65,287	0,717
Cultivar (C)	4	1852,354**	2219,027**	13,164**	5206,731**	5,366**
Ano (A)	4	940,893**	874,154**	7,274**	408,131**	24,080**
C x A	16	71,305	61,863	0,784	96,108	0,742*
Resíduo	60	89,337	94,197	0,946	64,246	0,345
Média		28,44	30,02	2,72	98,81	2,08
CV (%)		33,23	32,33	35,71	8,11	28,20

* e ** valores de F significativos aos níveis de 5% e 1% de significância, respectivamente.

4.1.2. A divergência genética por diferentes estatísticas multivariadas

Neste estudo foram empregadas quatro diferentes estatísticas multivariadas para quantificar a divergência entre pares de cultivares de cacau. Na Tabela 2 estão as estimativas da distância generalizada de Mahalanobis (D^2); da distância Euclidiana média de dados padronizados (d_e) e da distância Euclidiana média estimada de escores dos dois primeiros componentes principais (d_{cp}) e das duas primeiras variáveis canônicas (d_{vc}). É que, os dois primeiros componentes principais e as duas primeiras variáveis canônicas praticamente esgotaram toda a informação contida na análise multidimensional. Ou, de outro modo, suas raízes características ou autovalores correspondentes acumularam 89,43% e 92,14% da variação total, como será visto adiante.

A concordância entre as estimativas de distâncias proporcionadas pelas diferentes metodologias foi acessada sob quatro formas. A primeira delas se resumiu a simples observação da Tabela 2. Assim, as estatísticas D^2 e d_{vc} identificaram os pares de cultivares 1 x 4 e 2 x 4 como os mais divergentes e, os pares 1 x 5 e 3 x 5, como os menos divergentes ou mais similares. Com base em d_e e d_{cp} , os cultivares mais divergentes foram 2 x 3 e 2 x 4, sendo 1 x 5 e 3 x 4, os mais similares. Tais resultados demonstraram uma certa discrepância entre as diferentes estatísticas aplicadas.

A razão entre $n \sum d_{cp}^2 / p \sum d_e^2$ constituiu uma outra forma de acessar o grau médio de concordância entre as estatísticas d_{cp} e d_e . Por sua vez, a

concordância entre as estatísticas d_{vc} e D^2 pôde ser quantificada pela razão $n\sum d_{vc}^2 / \sum D^2$. Ambas as razões foram facilmente calculadas da Tabela 2. Desse modo, sendo $n = 2$, o número de componentes principais ou variáveis canônicas usados no cálculo da distância Euclidiana e $p = 5$, o número de caracteres avaliados, a relação estatística entre d_{cp} e d_e proporcionou um grau de concordância de 89,43 %. Por outro lado, a razão entre a soma de quadrados das distâncias Euclidianas médias estimadas de variáveis canônicas e a soma das distâncias generalizadas de Mahalanobis foi, no caso, de 92,14 %, demonstrando maior grau de concordância entre essas duas últimas estatísticas.

Tabela 2 - Estimativas de medidas de dissimilaridade entre pares de cultivares de cacau obtidas pela distância generalizada de Mahalanobis (D^2), distância Euclidiana média padronizada (d_e) e distância Euclidiana média obtida com base nos escores dos dois primeiros componentes principais (d_{cp}) e escores das duas primeiras variáveis canônicas (d_{vc}). CEPEC, Itabuna, Ba.

Cultivares ^a	D^2	d_e	d_{cp}	d_{vc}
1 e 2	11,77	1,31	2,07	2,39
1 e 3	7,96	1,19	1,83	1,72
1 e 4	32,16	1,50	2,26	4,00
1 e 5	1,94	0,53	0,38	0,81
2 e 3	11,95	1,95	3,06	2,36
2 e 4	23,51	1,95	2,99	3,40
2 e 5	8,50	1,20	1,71	1,95
3 e 4	13,10	0,85	0,72	2,32
3 e 5	5,10	1,50	2,07	0,99
4 e 5	20,86	1,52	2,40	3, 21

^a 1 = CC 10; 2 = SIAL 169; 3 = CEPEC 1; 4 = ICS 1; 5 = SIC 19

A concordância entre as várias estimativas de distância foi também acessada, avaliando-se a correlação residual existente entre os caracteres em estudo. Convém lembrar que essa correlação é considerada apenas nas análises de distância generalizada de Mahalanobis e distância Euclidiana média obtida de variáveis canônicas. Logo, esse diferencial, permitiu comparar a distância de Mahalanobis com a distância Euclidiana média e a distância Euclidiana de componentes principais, com aquela obtida

de variáveis canônicas. Entretanto, a hipótese H_0 , empregada para testar a igualdade entre a matriz de correlação residual e a matriz identidade (GODOI, 1985), foi aceita ao nível de 5% de significância ($\chi^2 = 14,66$ com 10 graus de liberdade), a despeito dessa matriz (Tabela 3) ter apresentado 3 estimativas de correlações de elevadas magnitudes, no conjunto de 10 correlações.

Tabela 3 - Matriz de correlação residual entre cinco caracteres avaliados em cinco cultivares de cacau. CEPEC, Itabuna, Ba.

Caráter	NFSP	NFCP	PSUP	PSUF	PFDP
NFSP	1	0,9952	0,9720	-0,0304	-0,2636
NFCP		1	0,9704	-0,0297	-0,1994
PSUP			1	0,1770	-0,2249
PSUF				1	0,0800
PFDP					1

O grau de concordância entre si das diferentes estimativas de distâncias foi também avaliado pelo coeficiente de correlação de Spearman (Tabela 4). As correlações foram expressivas em se tratando da distância de Mahalanobis e distância Euclidiana obtida de variáveis canônicas ($r = 0,9515$) e entre a distância Euclidiana padronizada e a distância Euclidiana estimada de componentes principais ($r = 0,9757$). Entre D^2 e d_e , D^2 e d_{cp} , d_e e d_{vc} , d_{cp} e d_{vc} as correlações foram apenas moderadas, confirmando a discrepância observada entre as várias estatísticas utilizadas na quantificação da divergência entre os cultivares.

Tabela 4 - Estimativas dos coeficientes de correlação de Spearman entre as medidas de dissimilaridade obtidas entre pares de cultivares de cacau, através das estatísticas D^2 , d_e , d_{cp} e d_{vc} .

Estatísticas	D^2	d_e	d_{cp}	d_{vc}
D^2	1	0,5030	0,5879	0,9515
d_e		1	0,9757	0,5757
d_{cp}			1	0,6606
d_{vc}				1

Resultado semelhante foi relatado por MALUF & FERREIRA (1983). Esses autores, ao investigarem a divergência genética entre cultivares de feijão-vagem, encontraram uma baixa estimativa de correlação entre distância Euclidiana e de Mahalanobis ($r = 0,273$), minorando as distorções a partir da análise dos dendrogramas obtidos das distâncias citadas.

A discrepância verificada entre as estatísticas multivariadas em estudo pode vir a se constituir em um complicador adicional, sobretudo nos estudos de divergência entre acessos, dispostos em bancos germoplasmas ou mesmo em condições naturais. Nesses casos, entretanto, pode-se utilizar do artifício já anteriormente comentado e obter a matriz de dispersão residual, a partir da variação entre plantas dentro de acessos. Contudo, em casos onde apenas as médias dos acessos são disponíveis, a aplicação de distâncias Euclidianas e componentes principais é a única solução possível.

Entretanto, suficiente consistência entre as quatro estatísticas aqui estudadas tem sido relatada por CRUZ (1990) em milho, PIRES (1993) em cana-de-açúcar e, em menor escala, por PULCINELLI et alii (1993) em eucalipto. Também MALUF et alii (1983) e FERREIRA (1993) detectaram concordância entre as distâncias Euclidiana e de Mahalanobis, quando avaliaram a divergência entre cultivares de tomate e de milho, respectivamente. Desse modo, com base na literatura, quaisquer das quatro estatísticas prestam-se para aplicação em estudos de divergência genética. Todavia, o presente estudo, demonstrou que as estatísticas que consideram a correlação residual entre caracteres devem ser preferidas, sempre que for possível a sua estimação.

4.1.3. A importância dos caracteres para a divergência

A importância relativa dos componentes principais e das variáveis canônicas pode ser deduzida da percentagem da variância total contida em seus primeiros autovalores (λ_j) correspondentes. A Tabela 5 evidenciou que os dois primeiros componentes principais e as duas primeiras variáveis canônicas explicaram 89,43% e 92,14% da variação total disponível, respectivamente. No contexto da geometria, significa dizer que a distorção provocada pela passagem do espaço multidimensional, proporcionado pelos p caracteres ($p = 5$), para o espaço bidimensional dos dois primeiros componentes e variáveis canônicas, foi negligenciável.

Tabela 5 - Estimativas dos autovalores (λ_j), de suas importâncias relativas e dos seus correspondentes coeficientes (autovetores α_j) associados aos caracteres, constituindo os **componentes principais** e as variáveis canônicas, avaliados em cinco cultivares de cacau. CEPEC, Itabuna, Ba.

λ_j	$\lambda_j / \sum \lambda_j$ (%)	α_j				
		NFSP	NFCP	PSUP	PSUF	PFDP
3,4036	68,07	0,5326	0,5390	0,5010	-0,2941	0,2969
1,0680	89,43	-0,1225	-0,0554	0,2333	0,7200	0,6396
0,5275	99,98	-0,1861	-0,1205	-0,4059	-0,5365	0,7059
0,0008	99,99	-0,4323	-0,4193	0,7279	-0,3275	-0,0159
0,0001	100,00	-0,6926	0,7184	-0,0012	-0,0049	-0,0644
4,5914	67,10	0,1604	-0,0586	-0,4890	1,0254	-0,2376
1,7133	92,14	-5,8346	6,6403	-0,2983	0,3811	0,2856
0,4743	99,07	6,0774	0,1186	-5,7215	1,3199	-0,2899
0,0635	99,99	2,8568	3,7867	-7,2675	1,3221	0,1616
0,0001	100,00	12,1292	-11,7973	0,0212	0,0445	1,3031

Componentes principais

Variáveis Canônicas

Na avaliação da importância dos caracteres para a divergência também foram empregados os componentes principais e as variáveis canônicas. Para os primeiros foi adotada a recomendação de JOLLIFFE (1972; 1973), enquanto o critério definido por SINGH (1981) foi aplicado às variáveis canônicas. Nesse ponto, é preciso esclarecer que caracteres importantes para a divergência são aqueles que representam bem a estrutura do sistema biológico em estudo, apresentando-se suficientemente diversos. Inversamente, caracteres descartáveis para o estudo da divergência são os pouco variáveis e/ou os redundantes por correlação a outros caracteres.

Os critérios anteriormente citados estabelecem que os caracteres de menor contribuição para a divergência são aqueles que exibem os maiores coeficientes de ponderação, nos menores autovalores correspondentes. Isso significa que devam ser observados os coeficientes de ponderação dos últimos componentes e variáveis. Assim, em relação a componentes principais, os caracteres NFCP e PSUP foram identificados como de pouca importância na determinação da divergência. Em relação a variáveis

canônicas foram identificados os caracteres NFSP e PSUP, como os de menor importância (Tabela 5). Convém destacar que SINGH (1981) utiliza o produto entre o coeficiente de ponderação e o desvio-padrão residual do caráter correspondente (ver Tabela 5), como critério para identificar a importância, definida do mesmo modo que para componentes principais.

Os caracteres NFSP, NFSP e PSUP foram evidenciados como os menos importantes pelas análises da divergência por variáveis canônicas e componentes principais, devido a correlação com outros caracteres. Na realidade, supondo por exemplo o descarte dos caracteres NFSP e PSUP, esses ficariam representados pelo caráter NFSP, dado que os dois primeiros estão fortemente correlacionados a esse último (correlações fenotípicas de 0,9964 e 0,9310, respectivamente). Aliás, a matriz de Mahalanobis, estimada após descarte de PSUP, manteve-se inalterada. Essa mesma matriz também não exibiu distorção após o descarte simultâneo dos caracteres PSUP e NFSP, confirmando a afirmativa anterior de que estes ficaram representados pelo caráter NFSP.

Por outro lado, o teste de SINGH (1981) quando aplicado à matriz de distância de Mahalanobis detectou contribuições relativas para a divergência iguais a 38,57% (NFSP); 32,53% (PSUP); 16,91% (NFSP); 8,79% (PSUP) e 3,20% (PFDP). A maior contribuição dos caracteres número de frutos colhidos e peso de sementes úmidas por fruto pode ser devida à seleção humana sobre esses caracteres. Aqui também o componente PSUP demonstrou ser pouco importante para a divergência. De modo que, ao menos para o caráter PSUP, existiu concordância entre as três metodologias empregadas para acessar a importância dos caracteres. No geral, entretanto, as estatísticas multivariadas aqui aplicadas não discriminaram, de modo concordante, os caracteres pouco variáveis e redundantes, presentes no estudo.

Por último, o fato dos caracteres NFSP e PSUP terem sido apontados pela análise de variáveis canônicas como contribuintes de menor importância para a divergência não significa que, necessariamente, deveriam ser descartados. JOLLIFFE (1972) refere-se à aplicação da técnica de descarte para um conjunto formado por 10 ou mais variáveis. O estudo I, em particular, envolveu apenas um conjunto restrito de cinco componentes de produção, que guardam estreita relação entre si, e não teve como objetivo o descarte de variáveis. O propósito nesse estudo foi o de identificar aqueles caracteres de menor importância para a divergência genética.

4.1.4. Eficiência da predição da performance de híbridos por divergência dos genitores

A hibridação envolvendo genitores divergentes é esperada gerar segregantes transgressivos favoráveis. Logo, os estudos sobre divergência genética entre genitores podem ter grande aplicação no melhoramento. Assim, com base nas estimativas de diversidade genética é possível selecionar genitores favoráveis para hibridação podendo-se, a partir de seus cruzamentos, obter híbridos superiores.

A aplicação da técnica de agrupamento de Tocher (Tabela 6), com base nas estimativas de distância generalizada de Mahalanobis, evidenciou dois grupos divergentes entre si. O grupo I formado pelos cultivares 1, 2, 3 e 5 e o grupo II singular, constituído pelo cultivar 4. A visualização da divergência dos cultivares no dendrograma (Figura 1) obtido da matriz de Mahalanobis e no plano cartesiano, com os eixos formados pelas duas primeiras variáveis canônicas (Figura 2) também confirmou os cultivares 1, 2, 3 e 5 como os menos divergentes entre si e o cultivar 4 como o de maior divergência entre todos. Em decorrência, obteve-se uma concordância satisfatória entre os métodos de otimização, os aglomerativos e a dispersão gráfica subjetiva no plano cartesiano.

Tabela 6 - Agrupamentos de similaridade entre cinco cultivares de cacau estabelecido pelo método de Tocher, usando a distância generalizada de Mahalanobis. CEPEC, Itabuna, Ba.

Grupos	Cultivares			
I	CC 41(1)	SIC 19(5)	CEPEC 1(3)	SIAL 169(2)
II	ICS 1(4)			

A análise da divergência entre cultivares, com base em D^2 , não foi concordante com a classificação dos grupos raciais de cacau, em se tratando do par mais distante e do mais similar. Os cultivares 1 (CC 41) e 4 (ICS 1), estimados como os mais distantes entre si, constituem o grupo dos trinitários. Esse grupo racial tem uma natureza híbrida, resultado do cruzamento entre os grupos crioulos e forasteiros amazônicos. Por outro lado, o par mais similar (cultivares 1 e 5) constituído por CC 41 e SIC 19, respectivamente, tem origem distinta. O primeiro, como já visto, é um trinitário da Costa Rica e o segundo é um forasteiro amazônico selecionado nas populações da Bahia.

Por outro lado, a concordância da divergência com os grupos

raciais surgiu quando analisados o segundo par mais distante e o segundo mais similar, com base também em D^2 . Os cultivares SIAL 169 (2) e ICS 1 (4), os mais divergentes entre si, formam o grupo dos amazônicos e dos trinitários, respectivamente. Esses grupos raciais têm gerado híbridos heteróticos em cruzamento, como demonstrado exaustivamente pela literatura de cacau, reflexo de sua divergência. O par mais similar, no caso, CEPEC 1 (3) e SIC 19 (5) tem origem comum, sendo ambos resultantes de seleções conduzidas nas populações de forasteiros amazônicos da Bahia.

Figura 1 - Dendrograma representando a divergência genética entre cinco cultivares de cacau obtido pelo método do vizinho mais próximo, aplicado à matriz de distância de Mahalanobis. Códigos dos cultivares estão na Tabela 6.

Nesse ponto é necessário justificar a utilização de genitores S_1 no estudo da divergência, em lugar dos próprios genitores S_0 que originaram os híbridos. O fato decorreu da própria estratégia adotada no melhoramento do cacaueiro. O programa de melhoramento de cacau trabalha, basicamente, com germoplasma reproduzido clonalmente. Entretanto, os híbridos sintetizados desses clones são reproduzidos

seminalmente. Ocorre que clones e materiais reproduzidos seminalmente geram árvores com arquiteturas e hábitos produtivos muito distintos, a ponto de não serem postos, frequentemente, em competição, sob o mesmo experimento. Então, a estratégia adotada na cultura, para reproduzir integralmente os genótipos dos clones parentais e contornar as dificuldades fitotécnicas apresentadas por eles, é a autofecundação dos clones.

Figura 2 - Gráfico de dispersão de escores das duas primeiras variáveis canônicas (VC₁ e VC₂) referentes a cinco cultivares de cacau.

Na sequência, a divergência dos genitores S₁ dos híbridos foi avaliada. Entretanto, não é suficiente considerar a divergência entre cultivares como único critério para orientar os cruzamentos. A performance *per se* desses cultivares também deve ser considerada. Sobretudo, a performance com relação aos caracteres de maior importância econômica.

Em cacau, um genótipo superior é aquele que apresenta alta produção de amêndoas secas por unidade de área, tendo em vista que amêndoas secas constituem o produto comercial da cultura. Entretanto, essa alta produção deve ser combinada com elevado peso de amêndoas por fruto e menor percentagem de casca. Isso significa produzir mais a partir de um número não excessivo de frutos. Em outras palavras, um genótipo superior deve aliar alto rendimento com custos reduzidos de colheita. TOXOPEUS (1985b) recomenda que os cultivares comerciais devam apresentar amêndoas de peso médio uniforme, superior a 1 grama.

A performance média dos cultivares consta da Tabela 7. Observou-se que o cultivar 4, de maior divergência em relação aos demais, foi o que apresentou a maior média para o caráter PSUF (124,49 g/fruto), embora tenha sido apenas o quarto melhor produtor (NFSP = 18,80 e PSUP = 2,32 kg/planta). Assim, híbridos superiores devem ser originados dos cruzamentos envolvendo esse cultivar. Não obstante, o cultivar 2 exibiu o segundo melhor peso de sementes por fruto (99,62 g/fruto), combinado com a maior produção (NFSP = 37,86 e PSUP = 3,76 kg/planta). Como os cultivares 2 e 4 foram preditos entre os de maior divergência entre si, deve-se esperar que o cruzamento entre ambos tenha gerado o melhor híbrido.

Os 20 híbridos entre os cinco genitores S_0 , representados pelos correspondentes genitores S_1 , tiveram as suas médias combinadas em apenas 10 F_1 's (Tabela 8). Depreendeu-se da observação dessa Tabela que o híbrido de melhor performance média, em termos de PSUP e PSUF, resultou do cruzamento entre os genitores 2 e 4, em concordância com o predito, já que esse par de cultivares apresentou-se como o segundo mais divergente. De outro modo, os híbridos de melhor performance, com referência aos mesmos caracteres, foram aqueles que envolveram a participação do genitor 4 (ICS 1), o mais divergente dentre todos, a exemplo dos híbridos 1 x 4, 2 x 4, 3 x 4 e 4 x 5; como aliás foi também predito. Finalmente, os cruzamentos entre os genitores 3 e 5 propiciaram híbridos F_1 's de performances medíocres. Pela predição da divergência esses dois genitores formaram o segundo par mais similar.

Tabela 7 - Valores médios (1986-90) de cinco cultivares de cacau referentes a cinco características. CEPEC, Itabuna, Ba.

Cultivares	NFSP	NFCP	PSUP kg/planta	PSUF g/fruto	PFDP %
1- CC 41	32,25 a	33,62 b	2,62 bc	80,93 c	1,89 b
2- SIAL 169	37,86 a	41,61 a	3,76 a	99,62 b	2,97 a
3- CEPEC 1	17,49 b	18,40 c	1,67 d	98,01 b	2,07 b
4- ICS 1	18,80 b	19,42 c	2,32 cd	124,49 a	1,80 b
5- SIC 19	35,80 a	37,03 ab	3,25 ab	90,99 b	1,68 b

Médias seguidas pelas mesmas letras, na vertical, não diferem estatisticamente entre si, pelo teste de Tukey a 5% de significância.

A propósito, em dois experimentos instalados em 1981 no CEPEC o híbrido entre os clones SIAL 169 e ICS 1 vem se destacando no comportamento produtivo, em competição com outros 19 híbridos no Ensaio 34 e com outros 22 híbridos no Ensaio 35 (MARIANO, 1986). Por outro lado, os clones SIC 19 e SIC 24 em cruzamentos com o clone ICS 1 têm sido incluídos na mistura de híbridos distribuídos aos cacauicultores da Bahia (MARIANO & YAMADA, 1986) e Espírito Santo (PEREIRA et alii, 1987). Em anos recentes, o híbrido SIC 19 x ICS 1 vem sendo utilizado como testemunha nos experimentos de competição de híbridos.

Considerando-se ainda os híbridos de melhor performance para os caracteres PSUP e PSUF (Tabela 8), percebe-se que todos tiveram um genitor comum - o cultivar 4 (ICS 1). Esse cultivar, conforme avaliação feita por DIAS (1991), exibe elevada estimativa positiva de capacidade geral de combinação para os caracteres PSUF e PSUP. Por essa razão, o ICS 1 concorreu para elevar o peso das sementes úmidas por fruto e a produção por planta, em todos os cruzamentos em que participou. No presente estudo, o genitor ICS1 (4) foi o mais divergente dos demais. O fato sugere, então, uma relação entre efeitos de capacidade combinatória e divergência genética dos genitores. Na literatura, uma relação positiva entre distância parental e efeitos de capacidade combinatória nos híbridos tem sido encontrada por MOLL et alii (1962) em milho, MURTY & ANAND (1965) em linho, ANAND & RAWAT (1984) e GUPTA et alii (1991) em mostarda indiana, ARUNACHALAM et alii (1984) em amendoim e SHAMSUDDIN (1985) em trigo.

O grau de concordância entre a divergência dos genitores S_1 , expressa por D^2 , e a performance média realizada nos híbridos entre tais genitores pôde ser melhor apreciado pelo coeficiente de correlação de Spearman (Tabela 8). Tais coeficientes variaram de -0,31 para o caráter PFDP a 0,76 para PSUF. Convém destacar que as correlações foram moderadas e altas em se tratando dos componentes mais íntimos da produção, a exemplo de PSUP (0,68) e PSUF (0,76). Em conclusão, pode-se afirmar que coeficientes de correlação de magnitudes tão elevadas atestaram a eficiência da predição, ao menos para caracteres que melhor refletem a produção comercial de cacau.

Observando a Tabela 8 depara-se, ainda, com quatro estimativas de correlação positivas, referentes aos caracteres NFSP, NFCP, PSUP e PSUF. No entanto, para o caráter PFDP, a correlação entre D^2 e performance média dos híbridos foi negativa (-0,31). Matematicamente a correlação pode ser negativa, já que seu numerador é uma covariância. Entretanto, sendo tal estimativa de baixa magnitude pode-se interpretá-la como sendo zero.

Tabela 8 - Estimativas da divergência genética entre pares de genitores expressas pela distância generalizada de Mahalanobis (D^2), das médias dos híbridos F_1 's referentes a cinco características e suas correspondentes correlações de Spearman. CEPEC, Itabuna, Ba.

Genitores ^a	D^2	NFSP	NFCP	PSUP kg/planta	PSUF g/fruto	PFDP %
1 e 2	11,77	45,39	48,59	4,26	94,19	2,25
1 e 3	7,96	38,93	40,37	3,56	92,29	1,72
1 e 4	32,16	42,59	43,91	4,77	113,29	1,70
1 e 5	1,94	49,87	52,72	4,62	94,41	2,27
2 e 3	11,95	36,53	37,79	3,92	107,86	1,70
2 e 4	23,51	40,25	41,77	5,19	129,68	1,85
2 e 5	8,50	36,52	38,70	3,55	97,27	2,14
3 e 4	13,10	36,37	37,17	4,55	125,42	1,49
3 e 5	5,10	29,54	30,45	2,92	98,74	1,55
4 e 5	20,86	41,09	42,53	4,97	121,49	1,72
Correlação com D^2		0,11	0,09	0,68	0,76	-0,31

^a 1 = CC 10; 2 = SIAL 169; 3 = CEPEC 1; 4 = ICS 1; 5 = SIC 19

Convém enfatizar que, do ponto de vista agrônomo, o importante é a performance média do híbrido e, sob esse aspecto, a predição foi eficiente. A elevada concordância entre o predito e o realizado pode estar no emprego de componentes de produção do cacau, cujo controle genético supera o ambiental. Basta observar que as estimativas de repetibilidades encontradas por DIAS (1991), para esses mesmos componentes, referentes aos mesmos cinco cultivares e seus 20 híbridos no período, foram superiores a 0,84; com exceção apenas do componente PFDP (0,41). Por sua vez, os coeficientes de repetibilidade, envolvendo apenas os cultivares foram superiores a 0,76 para todos os componentes de produção, a exceção de PFDP, cuja repetibilidade foi de 0,55 (estimados da Tabela 1). Em decorrência, a divergência estimada entre os pares de genitores nesse estudo tem uma natureza predominantemente genética.

Outra razão para a eficiência da predição residiu nos componentes de produção aqui empregados. Tais componentes pareceram refletir bem a estrutura do sistema biológico em estudo. Destaque-se ainda, que o estudo envolveu dados referentes

a um período de cinco anos, coletados mensalmente. As plantas tinham 11 e 16 anos de campo, no início e no final do estudo, respectivamente. Significa dizer que o clímax produtivo do cacauzeiro, alcançado aos 10-12 anos de campo, foi atingido e que a competição de copas e parte radicular estava estabilizada. Aspectos da competição intergenotípica em cacau têm sido abordados por GLENDINNING & VERNON (1965).

Então fatores como esses podem ter contribuído significativamente para a elevada eficiência da predição. A respeito desse clímax produtivo, BARTLEY (1970) sugeriu que a seleção de cacauzeiros se processe sobre uma combinação de produções anuais, a partir do oitavo ano de plantio, para minimizar os efeitos de anos. Esse autor estudou o comportamento produtivo de clones ICS 1, híbridos entre eles e também híbridos entre ICS 1 e clones Escavinas, em condições experimentais.

A divergência entre os genitores S_1 's, expressa por D^2 , foi relacionada à performance heterótica dos híbridos. Dado que a heterose é uma medida relativa referente à duas gerações, no caso, gerações parental e filial, reforçou-se o critério de seleção de genitores para hibridação, baseado nas suas divergências e nas suas performances *per se*. Ressaltando que o termo heterose, aqui empregado, referiu-se à diferença entre a performance média dos F_1 's e a performance média dos correspondentes genitores S_1 's superiores.

A observação da Tabela 9 evidenciou que para o caráter PSUP a predição da heterose dos híbridos, com base na divergência dos genitores, foi eficiente ($r = 0,65$), repetindo um fato já ocorrido na predição da performance média dos híbridos ($r = 0,68$). Ora, o peso de sementes úmidas por planta é considerado o melhor estimador da produção de cacau comercial (POUND, 1933; ESQUIVEL & SORIA, 1967). E a literatura tem demonstrado que a performance média e a heterose têm sido mais eficientemente preditas pelos caracteres relativos à produção. Assim, SMITH & SMITH (1987) encontraram correlações entre divergência e performance média igual a 0,35 e de 0,34 para a heterose, referentes a produção de grãos em milho. Correlações de 0,79 e 0,56 foram obtidas por CRUZ (1990), para o mesmo caráter em milho, em estudos de predição da média e da heterose de híbridos, respectivamente. MOLL et alii (1962) e MOLL et alii (1965), também em milho, detectaram associação entre diversidade genética dos genitores e grande heterose nos cruzamentos entre eles, para produção de grãos.

Em trigo, a divergência dos genitores e os efeitos heteróticos e de capacidade específica nos correspondentes híbridos foi investigada. Correlações de 0,45** e de 0,44**, respectivamente, foram encontradas para a produção de grãos nos híbridos (SHAMSUDDIN, 1985). GHADERI et alii (1984), trabalhando com feijão,

encontraram correlação de 0,73** entre distância parental e efeitos heteróticos para produção de grãos nos híbridos e de -0,53** para baixo peso de grãos, em feijão-fava. Em tomate, correlação de 0,73**, entre heterose para produção de frutos nos híbridos e distância parental, foi encontrada por MALUF et alii (1983).

A maior eficiência da predição da performance média e da heterose nos híbridos constatada por diversos autores, nos caracteres relativos à produção de grãos é explicável. A produção é uma característica resultante da combinação de uma série de atributos de natureza fisiológica, anatômica, morfológica, etc, sugerindo a necessidade de múltiplos caracteres para avaliá-la. É exatamente esse o objetivo da análise multivariada. Ou seja, estimar a divergência a partir de múltiplos caracteres, simultaneamente. Geneticamente a produção é a resultante de um grande número de locos distribuídos aleatoriamente no genoma, sendo correlacionada a muitos outros caracteres. Por outro lado, a heterose manifestada nos híbridos F_1 é resultante, principalmente, das diferenças nas frequências gênicas entre populações, como será discutido adiante. Ao estimar-se a divergência entre populações estima-se, com certa aproximação, a diferença em tais frequências gênicas e, conseqüentemente, a heterose.

Em mostarda indiana, GUPTA et alii (1991) sugeriram associação entre distância parental e efeitos de heterose e de capacidade específica nos híbridos. Em feijão vigna, RAMANUJAM et alii (1974) detectaram considerável associação entre heterose para dias de florescimento, número de sementes por fruto e produção de grãos e divergência genética entre os genitores dos híbridos. A frequência de cruzamentos heteróticos e a magnitude da heterose para produção de grãos e seus componentes foi elevada, quando envolveu cruzamentos entre parentais de divergência intermediária. Esse resultado foi obtido por SRIVASTAVA & ARUNACHALAM (1977) em triticle, ARUNACHALAM & BANDYOPADHYAY (1984) em amendoim e colza e por ARUNACHALAM et alii (1984) em amendoim. Em amendoim, esse último autor encontrou 50% dos cruzamentos heteróticos com alta capacidade específica de combinação explicando a heterose. Todavia, a literatura também constata ausência de associação entre divergência de genitores e performance média para produção nos híbridos de milho (FERREIRA, 1993) e entre divergência e heterose nos híbridos como os de CHAUDHARY & SINGH (1975) em cevada, SINGH et alii (1981) em milheto e JOSHI & VASHI (1992) em sorgo.

Para os demais caracteres a predição não se confirmou. NFSP e NFPC não estiveram correlacionados com D^2 . Inclusive para os caracteres PSUF e PFDP as correlações com D^2 foram inversas (-0,11 e -0,22, respectivamente). Para interpretar

tais correlações negativas é preciso lembrar o conceito de heterose. Segundo FALCONER (1989), para um dado caráter herdado quantitativamente, a heterose é o resultado do produto dos desvios de dominância para um dado loco i , pelo quadrado da diferença de frequência alélica no loco i , verificada entre dois genótipos inter cruzados. Enquanto o primeiro termo da expressão biométrica da heterose admite sinal positivo ou negativo, a depender da direção da dominância, o segundo assume valores exclusivamente positivos. Logo, a associação entre D^2 e heterose pode ser positiva, nula ou até negativa.

Tabela 9 - Estimativas da divergência genética entre pares de genitores expressas pela distância generalizada de Mahalanobis (D^2), das heteroses¹ nos híbridos F_1 's referentes a cinco características e suas correspondentes correlações de Spearman. CEPEC, Itabuna, Ba.

Genitores ^a	D^2	NFSP	NFCP	PSUP kg/planta	PSUF g/fruto	PFDP %
1 e 2	11,77	7,53	6,98	0,50	-5,43	0,36
1 e 3	7,96	6,68	6,75	0,94	-5,72	-0,17
1 e 4	32,16	10,34	10,29	2,15*	-11,20*	-0,10
1 e 5	1,94	14,07	15,69*	1,37	3,42	0,59
2 e 3	11,95	-1,33	-3,82	0,16	8,24	-0,37
2 e 4	23,51	2,39	0,16	1,43	5,19	0,05
2 e 5	8,50	-1,34	-2,91	-0,21	-2,35	0,46
3 e 4	13,10	17,57*	17,75*	2,23*	0,93	-0,31
3 e 5	5,10	-6,26	-6,58	-0,33	0,73	-0,13
4 e 5	20,86	5,29	5,50	1,72*	-3,00	0,04
E. Padrão		6,96	7,04	0,70	4,93	0,46
Correlação com D^2		0,16	0,14	0,65	-0,11	-0,22

^a 1 = CC 10; 2 = SIAL 169; 3 = CEPEC 1; 4 = ICS 1; 5 = SIC 19

¹ heterose como desvio do híbrido F_1 em relação ao genitor S_1 superior.

* e ** significativo aos níveis de 5% e de 1% de significância, respectivamente.

A associação positiva entre D^2 e heterose prevê que a dominância unidirecional exista e que exista também diferenças de frequências alélicas para o caráter considerado. A associação nula ou ausência de correlação entre D^2 e heterose decorre da

ausência de dominância e similaridade gênica dos genitores. Também a correlação nula pode ser motivada pela expressão errática da heterose nos cruzamentos, promovida por cancelamentos dos desvios de dominância, conforme argumentam VENCOVSKY & BARRIGA (1992), para o caso de parâmetros lineares dependentes de médias.

Por último, a associação negativa ou inversa implica em desvios de dominância negativos. Para os caracteres PSUF e PFDP, aqui estudados, as correlações foram negativas porém de baixa magnitude, podendo serem interpretadas como nulas. Ora, no estudo anterior desenvolvido por DIAS (1991), com os mesmos genitores e híbridos e no mesmo período, apenas os caracteres PSUF e PFDP manifestaram-se sob o controle de efeitos gênicos de natureza predominantemente aditiva.

4.1.5. Relação entre distância geográfica e divergência genética

Considerações a respeito dessa possível relação foram feitas, embora tenham um caráter muito preliminar. O número reduzido de cultivares em estudo impediu considerações de ordem geral e conclusivas. Entretanto, algumas inferências sobre o tema demonstraram que a distância geográfica esteve relacionada à divergência genética.

O agrupamento dos cultivares pelo método de Tocher, com base em D^2 (Tabela 6), discriminou dois grupos de similaridade. No grupo I estão os cultivares pertencentes às seleções baianas (SIAL 169, CEPEC 1 e SIC 19). Como visto no material e métodos, esses cultivares formam o grupo racial dos forasteiros amazônicos, tendo sido selecionados nas populações locais do sul da Bahia. Por outro lado, no grupo II está o cultivar ICS 1, um trinitário originário de Trinidad (POUND, 1933). Não fosse a presença do cultivar CC 41, um trinitário segundo ENRIQUEZ & SORIA (1967) e ENGELS (1981; 1983) no grupo I, junto às seleções baianas, a concordância entre a distância geográfica e a divergência genética seria total.

A inclusão do cultivar CC 41, um trinitário, no grupo I formado pelas seleções de cacau baiano do baixo amazonas foi, em princípio, inesperada. Entretanto, a própria constituição genética desse cultivar pode explicar tal fato. Selecionado por A. L. Erickson em Turrialba, em 1954, dentro de uma progênie de polinização livre de UF 676, o CC 41 foi classificado como um tipo híbrido (ENRIQUEZ & SORIA, 1967). O clone UF 676 é verdadeiramente um trinitário. Aliás, uma grande proximidade entre o trinitário UF 676 e cultivares do baixo amazonas do oeste africano foi recentemente detectada por RABOIN et alii (1993), em estudos de divergência genética por D^2 em cacau. Ocorre que trinitários constituem um grupo racial resultante

da hibridação entre crioulos e amazônicos (CHEESMAN, 1944), combinando a qualidade do primeiros com a rusticidade dos últimos. Logo, o grupo trinitário inclui de fato uma mistura de tipos híbridos que varia desde tipos semelhantes a crioulos, passando por tipos intermediários, até aqueles tipos que são próximos dos amazônicos.

Na Tabela 8, o CC 41 apresentou-se como um cultivar próximo das seleções de amazônicos da Bahia, no caso SIAL 169, CEPEC 1 e SIC 19, com valores de D^2 iguais a 11,77; 7,96 e 1,94, respectivamente. Por outro lado, esse mesmo cultivar apresentou-se muito distante do ICS 1 (D^2 de 32,16), um clone também trinitário como o CC 41. Ou seja, o clone trinitário CC 41 esteve bem mais próximo das seleções baianas do que do clone trinitário ICS 1. A Observação da Tabela 7 evidenciou que CC 41, SIAL 169 e SIC 19 foram os cultivares de maior produção de cacau úmido, a partir de amêndoas de peso mediano e de elevado número de frutos por planta, com todos esses valores de produção em níveis muito próximos. O CEPEC 1, é bom que se repita, é um mutante espontâneo de amêndoas brancas surgido nas populações cacaueiras da Bahia e guardou a terceira maior distância do CC 41 (D^2 de 7,96). É possível então que o cultivar CC 41 tenha recebido maior contribuição de genes de amazônicos e, em função disso, se assemelhe mais com os clones desse grupo.

Assumindo então que a hipótese anterior seja verdadeira, confirma-se a postulação de CHEESMAN (1944) de que os trinitários compõem-se de populações altamente heterogêneas. A precariedade da classificação dos grupos raciais de cacau fica também evidente com a hipótese anterior. A esse respeito, BARTLEY¹ tem alertado para a necessidade de uma classificação racial que considere as relações genéticas entre as populações de cacau. Ao que parece nenhuma das características atribuídas aos crioulos são exclusivas deles. Por sua vez, os não-crioulos ou grupo dos forasteiros amazônicos compreendem populações heterogêneas como os trinitários. Com esses resultados, fica evidenciado o potencial das técnicas multivariadas nos estudos de evolução e sistemática da espécie.

A proximidade entre CC 41 e os clones amazônicos da Bahia pode ter sido ampliada por forças seletivas artificiais operando semelhantemente, embora de modo independente, nos locais de origem desses materiais. De fato, a prática de seleção humana, de objetivo comum a ambos os materiais, pode ter ocorrido. Ou seja, o processo seletivo pareceu ter sido direcionado para plantas de elevada produção de frutos, ainda que conduzido por Instituições diferentes, em países diferentes. No caso, o Centro de

¹ BARTLEY, B.G.D. (Centro de Pesquisas do Cacau, Itabuna, Ba.) Genetic variability in Theobroma cacao: its origin and distribution. (Não publicado)

Cacao na Costa Rica (CC 41) e os Institutos de Cacau e Instituto Agrônomo do Leste, no Brasil (seleções baianas). Ademais, a seleção para maior produção de frutos tem resultado na seleção inconsciente para menor peso de sementes. Esses dois caracteres são correlacionados geneticamente ao inverso ($r_g = -0,59$ no presente estudo). Evidências de que a seleção humana pode aumentar o grau de similitude entre os cultivares em diferentes espécies têm sido encontradas por ADAMS (1977); KATIYAR & SINGH (1979) e JOSHI & VASHI (1992).

Embora a literatura seja conflitante nesse particular, alguns trabalhos têm sugerido que a diversidade genética pode estar relacionada à diversidade geográfica. Sugestão como essa foi feita por RAM & PANWAR (1970) e MONTALVAN DEL AGUILA (1990) em arroz, MOLL et alii (1962); MOLL et alii (1965) e TIMOTHY (1963) em milho, RUSSEL (1952); SORIA & ESQUIVEL (1969) e ATANDA (1973) em cacau. É necessário contudo, enfatizar que a comparação de resultados de diferentes espécies a respeito dessa relação não faz muito sentido. É que cada espécie vegetal tem sua própria história de vida. E, nesse sentido, fatores como seleção, deriva genética, fluxo gênico, segregação e outros podem atuar conjunta ou isoladamente e a diferentes intensidades sobre uma dada espécie e, com isso, determinar a existência ou não de algum paralelismo entre diversidade geográfica e divergência genética.

4.1.6. Estabilidade temporal da divergência

O conhecimento a respeito da estabilidade da divergência genética é de fundamental importância. Com referência a estabilidade temporal, o que se deseja saber é como se comporta a divergência, expressa no caso pela estatística D^2 , frente a variação climática própria de uma série de anos. A utilidade de D^2 , como estimativa de divergência genética, depende da consistência da classificação dos genótipos sob diferentes anos. Caso D^2 se mantenha estável no tempo poder-se-á avaliar a divergência entre genótipos, com base em um ou poucos anos de avaliação.

Nesse estudo, a estabilidade dita temporal foi acessada pela comparação entre a estatística D^2 de Mahalanobis, obtida no período 1986-90, considerada como padrão, com cada D^2 obtido de um simples ano. A comparação foi conduzida pela estimação do coeficiente de correlação de Spearman entre os pares de D^2 , referentes aos diferentes períodos, conforme a Tabela 10. Observaram-se correlações moderadas e altas, exceto entre D_{CONJ}^2 e D_3^2 (0,57), ou seja, entre a divergência estimada no período 1986-90 e a divergência no ano de 1988. Portanto, os

resultados sugerem que um simples ano, no presente estudo, pode ser utilizado para avaliação da divergência genética entre os genitores S₁ de cacau. Todavia, esse simples ano deve apresentar condições climáticas ótimas, para melhor expressão da variabilidade genotípica, como será visto mais adiante. Ainda, tais resultados significam uma acentuada economia de recursos humanos, financeiros e de tempo que seriam dispendidos caso a avaliação dos genótipos se estendesse por vários anos.

Tabela 10 - Estimativas de divergência genética entre pares de genitores expressas pela distância generalizada de Mahalanobis (D^2) referentes a cinco características, avaliadas no período 1986-90 ($D^2_{CONJ.}$) e comparadas por correlação de Spearman aos D^2 's obtidos, no mesmo período, para cada um dos cinco anos (D^2_1 , D^2_2 , D^2_3 , D^2_4 e D^2_5). CEPEC, Itabuna, Ba.

Anos	1986/90	1986	1987	1988	1989	1990
Genitores ^a	$D^2_{CONJ.}$	D^2_1	D^2_2	D^2_3	D^2_4	D^2_5
1 e 2	11,77	16,29	7,87	62,31	36,61	40,93
1 e 3	7,96	10,68	14,90	20,04	31,42	26,90
1 e 4	32,16	36,01	23,83	45,52	121,83	63,28
1 e 5	1,94	3,72	1,92	13,98	6,36	7,58
2 e 3	11,95	14,58	29,64	64,24	3,63	21,72
2 e 4	23,51	22,41	32,29	112,97	33,90	40,90
2 e 5	8,50	8,85	12,20	66,23	19,02	29,50
3 e 4	13,10	26,84	8,64	25,87	38,60	13,35
3 e 5	5,10	5,98	12,49	4,06	20,17	15,06
4 e 5	20,86	20,38	17,87	26,60	83,41	34,58
Correlação com $D^2_{CONJ.}$		0,94	0,65	0,57	0,70	0,66

^a 1 = CC 10; 2 = SIAL 169; 3 = CEPEC 1; 4 = ICS 1; 5 = SIC 19

A estabilidade temporal da divergência encontra respaldo nos resultados já apresentados na Tabela 1. Não se obteve significância para a interação cultivares por anos, no período 1986-90, para todos os caracteres avaliados. A ausência de interação significou que os cultivares não alteraram suas classificações e comportamentos médios relativos, ao longo dos anos. Dado que a estatística D^2 estima a divergência genética pela diferença quadrática entre médias transformadas dos genitores,

envolvendo todas as características, é evidente que a ausência da interação contribuiu para a estabilidade da estatística.

A estabilidade temporal aqui discutida inclui um aspecto de natureza estatística. Trata-se da precisão diferenciada com que foram obtidas as estimativas de D^2 . Enquanto $D^2_{CONJ.}$ foi estimado a partir de um período combinado de cinco anos, os demais D^2 's foram obtidos de um simples ano. Óbvio, então, que o erro de $D^2_{CONJ.}$ é bem menor que os erros associados a cada D^2 anual. No entanto, essa dificuldade foi contornada pelas magnitudes dos coeficientes de correlação. Correlações de magnitudes tão altas atestaram a estabilidade de D^2 .

A estabilidade parcial de D^2 tem sido relatada por SHERIFF (1992) e UPADHYAY & MURTY (1970) em milheto. Em trigo, SOMAYAJULU et alii (1970) relataram que grupos amplamente divergentes mantiveram-se distintos, em todos os ambientes experimentais utilizados, ficando a instabilidade restrita àqueles grupos pouco divergentes. Um padrão estável de agrupamento por D^2 , em diferentes ambientes (locais e anos), foi detectado por BHATT (1976) em um conjunto de 12 genótipos de trigo, envolvendo seis caracteres de qualidade para panificação. Essa estabilidade ambiental de D^2 , ainda segundo BHATT (1976), apresentou-se independente da estabilidade fenotípica dos seis caracteres individuais.

JANORIA et alii (1976) detectaram um padrão estável de agrupamento, por técnicas de taxonomia numérica, entre cultivares de arroz do tipo anão avaliados em dois ambientes de fertilidade. Utilizando-se das mesmas técnicas em feijão vigna, GHADERI et alii (1979) relataram padrões similares de agrupamento sob combinação de 18 ambientes de fertilização, densidades e épocas de plantio. Todavia, SINGH & GILL (1984) encontraram variação nos padrões de agrupamento por D^2 em algodão anual, em diferentes ambientes agrônômicos.

Outro fato emergente da observação da Tabela 10, referente a estabilidade da divergência genética, é que para um dado ano, no caso 1986, a associação entre D^2 obtido de um período de cinco anos e D^2 do referido ano foi muito estreita ($r = 0,94$). A explicação para isso esteve na condição climática favorável ocorrida em 1986 (Figura 3), não repetida nos quatro anos seguintes. A despeito do ligeiro incremento no peso de sementes úmidas por fruto, todos os demais componentes de produção exibiram quedas acentuadas a partir de 1986 (Figura 3). Condições ambientais ótimas facilitaram a completa expressão dos genótipos, com respeito aos vários caracteres. De modo que um ambiente sem estresse possibilita que D^2 discrimine melhor os genótipos. A necessidade de um ambiente ótimo para melhor expressão da

divergência genética por D^2 tem sido apontada por SHERIFF (1992); SINGH & GUPTA (1979) e UPADHYAY & MURTY (1970) em milheto; SOMAYAJULU et alii (1970) e BHATT (1976) em trigo e, VARMA & GULATI (1982) em cevada.

Numero de frutos sadios por planta Numero de frutos colhidos por planta

Peso de sementes umidas por planta

Peso de sementes umidas por fruto

Percentagem de frutos doentes/planta

Figura 3 - Gráficos de performance média de cinco genitores de cacau referentes a cinco caracteres, avaliados no período de 1986 a 1990. CEPEC, Itabuna, Ba.

4.2. Estudo II: Divergência fenética entre clones de cacauero

A avaliação de populações locais de cacau tem uma importância fundamental no programa de melhoramento da espécie. Essa importância deve-se ao fato das populações locais serem altamente adaptadas a condições ambientais específicas e serem fontes de variabilidade genética. Adicionalmente, a avaliação dessas populações locais pode proporcionar informações sobre a sua origem e evolução. A divergência aqui estimada baseou-se em dados fenotípicos obtidos em coleção de germoplasma, sem a observância dos princípios experimentais. Nesse caso, a divergência tem uma natureza fenética, não reproduzindo necessariamente, conforme SNEATH & SOKAL (1973), a diversidade genética entre os acessos.

Dentre os objetivos propostos no estudo I estava a investigação da consistência das diversas técnicas multivariadas aplicadas à pesquisa da divergência entre cultivares. Fundamentalmente, o cumprimento desse objetivo validou toda a investigação referente ao estudo II, já que tais técnicas mostraram-se satisfatórias, em maior ou menor grau e suficientemente discriminativas. Sendo assim, a divergência entre 26 clones de cacau, estabelecidos no banco de germoplasma do CEPEC, foi acessada pela distância Euclidiana média e por análise de componentes principais. O estudo II enfoca então a divergência entre clones pertencentes a um mesmo grupo racial (forasteiro amazônico), selecionados na população de cacaueros da Bahia. A geração desse conhecimento pode revigorar o programa de hibridação entre clones locais.

O conhecimento da divergência, entretanto, não deve se limitar à seleção de genitores para futuros programas de hibridação. Essa informação pode e deve ser aplicada para se avaliar, descrever e classificar os recursos genéticos visando a sua preservação. A limitação de tempo, recursos financeiros e de pessoal e área exige que os recursos genéticos promissores para melhoramento tenham prioridade na preservação. Significa dizer que, materiais pertencentes a diferentes grupos de similaridade devam ser preservados, enquanto aqueles que compõem um mesmo grupo devam ser descartados ou intercambiados com outras Instituições detentoras de germoplasma da espécie. Então, a divergência entre acessos, avaliada por estatísticas multivariadas, pode proporcionar uma descrição sintética da afinidade genética entre populações. Por último, também os caracteres mais importantes para a divergência são informados pelo emprego dessas técnicas e podem, então, orientar as futuras caracterizações do germoplasma.

Técnicas multivariadas vêm sendo aplicadas em coleções de germoplasmas. CAMUSSI et alii (1983) em populações de milho italianas; SMITH (1984; 1988), SMITH et alii (1985a; 1985b), SMITH & SMITH (1987; 1988; 1989a;

1989b) e MELCHINGER et alii (1991) em germoplasmas de milho dos E.U.A.; NARAYAN & MACEFIELD (1976) em *Cicer arietinum*; BROWN (1991) em algodão anual; SPAGNOLETTI ZEULI & QUALSET (1987) em trigo; SINGH et alii (1991) em feijão; HUSSAINI et alii (1977) em *Eleusine coracana* e WILSON et alii (1990) em *Pennisetum glaucum* são alguns dos exemplos de estudos de germoplasmas por técnicas multivariadas. Maiores progressos no melhoramento de plantas, segundo HALLAUER & MIRANDA FILHO (1981), podem ser obtidos através do conhecimento mais completo da constituição do germoplasma e do maior entendimento das relações entre os acessos.

4.2.1. A distância Euclidiana média

A distância fenética entre pares de clones de cacauzeiros foi inicialmente avaliada através da distância Euclidiana média. A estatística d_{ij} , foi então estimada de dados padronizados envolvendo os 13 caracteres em estudo. E, a matriz de dissimilaridade proporcionada por d_{ij} , (Tabela 11) permitiu destacar os clones 2 e 20 como os mais divergentes entre si ($d_{ij} = 3,05$), cabendo aos clones 3 e 14 a menor divergência ($d_{ij} = 0,33$), em relação aos demais pares de clones.

Sintomaticamente, a distância Euclidiana foi sensível bastante para discriminar o par mais divergente (2 e 20) como sendo formado por um clone SIC e um SIAL. Por outro lado, o par 3 e 14 formado por clones da série SIC foi o mais similar. Como visto em material e métodos, os 15 primeiros clones pertencentes à série SIC originaram-se de seleções feitas exclusivamente em um único município sul baiano, enquanto os 11 clones restantes da série SIAL resultaram de seleção conduzida em diversos municípios da mesma região. Logo, uma mais expressiva variabilidade devia ser esperada entre os clones SIAL, como aliás ficou demonstrado. Esses resultados também sugerem que as séries formadas pelos clones 3 e 14 podem ser reunidas em uma mesma população de melhoramento.

O conhecimento dos pares de genitores de maior divergência orienta o processo de hibridação. Assim, os pares mais divergentes devem ser utilizados em cruzamentos para a geração de híbridos mais heteróticos. Por outro lado, a informação acerca dos pares mais similares é útil nos programas envolvendo retrocruzamento. Nesses programas, o emprego de genitores similares, diferenciados basicamente pelo gene a ser transferido, permite recuperar o genitor recorrente mais rapidamente.

Tabela 11- Estimativas da divergência fenética entre pares de clones expressas pela distância Euclidiana média ($\times 10^{-2}$), com base em 13 características avaliadas em 26 clones de cacau. CEPEC, Itabuna, Ba.

C*	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	128	152	135	164	136	148	93	81	85	87	82	91	143	177	165	92	183	157	298	185	169	169	95	151	158
2		172	43	153	150	145	114	135	107	88	83	94	165	211	206	132	269	155	305	207	174	130	151	182	184
3			155	117	93	156	85	87	146	95	143	141	33	115	77	99	166	55	199	93	123	171	77	93	73
4				122	128	135	100	134	105	86	80	98	144	187	189	129	261	142	275	185	146	114	137	165	167
5					64	127	98	147	135	118	143	127	102	87	124	153	202	121	172	109	50	126	111	92	126
6						102	68	109	111	100	123	114	79	86	85	129	168	112	185	98	72	121	74	75	102
7							126	137	91	123	125	89	151	136	130	169	201	177	216	144	125	62	120	124	182
8								63	93	51	79	96	69	127	114	68	184	88	238	126	112	140	56	101	91
9									103	59	86	102	85	153	113	55	166	103	261	141	153	163	56	124	105
10										82	79	47	139	157	153	123	209	159	262	169	145	96	106	124	167
11											69	69	92	150	132	66	197	89	255	143	132	132	77	116	118
12												87	127	180	168	95	224	145	287	182	161	135	100	160	145
13													139	156	151	123	208	148	258	162	138	94	108	120	167
14														108	83	93	171	65	200	94	113	167	65	95	59
15															82	165	156	145	128	64	63	157	104	64	142
16																143	123	119	163	69	112	164	80	79	112
17																	197	98	274	150	158	183	78	139	119
18																		192	234	165	184	245	141	163	173
19																			226	123	133	186	102	120	77
20																				138	149	223	218	165	230
21																					86	164	100	81	138
22																						132	108	82	146
23																							149	139	203
24																								93	98
25																									131

* Clones 1 = SIC 2; 2 = SIC 17; 3 = SIC 18; 4 = SIC 19; 5 = SIC 23; 6 = SIC 24; 7 = SIC 250; 8 = SIC 433; 9 = SIC 628; 10 = SIC 644; 11 = SIC 662; 12 = SIC 680; 13 = SIC 747; 14 = SIC 765; 15 = SIC 823; 16 = SIAL 70; 17 = SIAL 88; 18 = SIAL 105; 19 = SIAL 169; 20 = SIAL 244; 21 = SIAL 283; 22 = SIAL 325; 23 = SIAL 357; 24 = SIAL 557; 25 = SIAL 581 e 26 = SIAL 638.

O método de Tocher aplicado à matriz de distâncias Euclidianas discriminou quatro grupos de similaridade (Tabela 12). O grupo I, composto por 16 clones ou 62% do total, foi o mais numeroso. Seguiram-se os grupos II, composto por 8 clones, e os grupos III e IV singulares, constituídos por um simples clone cada. Aqui também o clone 20 apresentou-se isolado em relação a todos os outros, estando o clone 2 separado no grupo II. De outro modo, os clones 3 e 14, acessados como os mais similares, agruparam-se no grupo I. Entretanto, essa técnica de agrupamento não foi satisfatória no sentido de relacionar a similaridade com a variabilidade das séries clonais, uma vez que um mesmo grupo deteve clones SIC e SIAL juntos, marcadamente no grupo I.

O processo seletivo aplicado a ambas as séries clonais parece ter sido semelhante e ter operado a intensidades próximas. Esse fato pode explicar a presença de clones SIC e SIAL no mesmo grupo. As séries SIC e SIAL foram obtidas através de seleção de plantas individuais na população de cacauzeiros existente nos municípios do sul da Bahia. No caso dos SIC's foram selecionados os cacauzeiros com um índice de frutos inferior a 32, após quatro anos de colheitas sucessivas, iniciadas em 1934. Em 1953 teve início a seleção dos clones SIAL, finalizada cinco anos depois. O critério adotado foi um índice de frutos inferior a 40 (SORIA, 1963). O índice de frutos corresponde ao número de frutos necessários para produção de 1 kg de amêndoas secas.

Tabela 12 - Agrupamentos de similaridade entre 26 clones (Códigos) de cacau referentes a 13 características estabelecidos pelo método de Tocher, usando a distância Euclidiana média. CEPEC, Itabuna, Ba.

Grupos	Clones			
I	SIC 18 (3)	SIC 765 (14)	SIAL 169 (19)	SIAL 638 (26)
	SIC 433 (8)	SIAL 557 (24)	SIC 628 (9)	SIC 662 (11)
	SIAL 88 (17)	SIC 24 (6)	SIAL 70 (16)	SIAL 581 (25)
	SIAL 283 (21)	SIC 23 (5)	SIAL 325 (22)	SIC 823 (15)
II	SIC 17 (2)	SIC 19 (4)	SIC 680 (12)	SIC 747 (13)
	SIC 644 (10)	SIC 2 (1)	SIC 250 (7)	SIAL 357 (23)
III	SIAL 105 (18)			
IV	SIAL 244(20)			

A figura 4 retrata o agrupamento dos clones pelo método hierárquico do vizinho mais próximo, a partir da matriz de dissimilaridade gerada pela distância Euclidiana média de dados padronizados. Para PEETERS & MARTINELLI (1989) a análise de agrupamento pode ser empregada para classificar acessos em coleções de germoplasmas e estimar o potencial de variabilidade desses acessos. A classificação subjetiva dos grupos foi similar àquela obtida com a aplicação do método de otimização de Tocher. Dos quatro grupos discriminados, dois apresentaram-se compostos por um único clone cada. A concordância entre esses dois métodos de agrupamento foi satisfatória.

Figura 4 - Dendrograma representando a divergência fenética entre 26 clones de cacau obtido pelo método do vizinho mais próximo, usando a distância Euclidiana média. Códigos dos clones estão na Tabela 12.

4.2.2. Distâncias Euclidianas intra e intergrupos

Em estudos como esse, envolvendo um grande número de clones, surge a possibilidade de considerar-se não o clone individualmente, mas sim o grupo de clones. Nesse sentido, a decisão do melhorista será a de selecionar clones parentais baseado nas distâncias intergrupos. Aliás, essa é a razão maior da análise de cluster, ou seja, avaliar a divergência entre clones através de uma medida de dissimilaridade e, a seguir, agrupar tais clones em grupos de máxima similaridade interna.

A tabela 13, então, apresenta as distâncias intra e intergrupos, com base nas distâncias Euclidianas, usando o método de Tocher. Assim, distância Euclidiana intragrupo corresponde à média das distâncias entre os pares de clones que constituem o grupo. Por sua vez, as distâncias intergrupos correspondem às médias das distâncias entre os pares de clones pertencentes aos diferentes grupos. As distâncias intragrupos III e IV são nulas, dado que tais grupos incluem um simples clone cada. Entretanto, essa distância intragrupo é semelhante em se tratando dos grupos I e II. A distância máxima de 2,66 foi registrada entre os grupos II e IV, indicando ampla diversidade entre esses dois grupos. A mínima distância intergrupos (1,41) foi obtida entre os grupos I e II (Tabela 13).

Tabela 13 - Estimativas de distâncias Euclidianas intra e intergrupos obtidas a partir de 13 características, avaliadas em 26 clones de cacau. CEPEC, Itabuna, Ba.

Grupos ^a	I	II	III	IV
I	1,0022	1,4155	1,7171	2,0021
II		1,0297	2,2513	2,6571
III			0,0000	2,3417
IV				0,0000

^a G I: SIC 18 (3); SIC 23 (5); SIC 24 (6); SIC 433 (8); SIC 628 (9); SIC 662 (11); SIC 765 (14); SIC 823 (15); SIAL 70 (16); SIAL 88 (17); SIAL 169 (19); SIAL 283 (21); SIAL 325 (22); SIAL 557 (24); SIAL 581 (25) e SIAL 638 (26).

G II: SIC 2 (1); SIC 17 (2); SIC 19 (4); SIC 250 (7); SIC 644 (10); SIC 680 (12); SIC 747 (13) e SIAL 357 (23).

G III: SIAL 105 (18).

G IV: SIAL 244 (20).

Entretanto, não é suficiente considerar a divergência entre grupos como critério único para orientar os cruzamentos, como aliás ficou demonstrado no estudo 1. A performance *per se* dos clones também deve ser considerada, sobretudo a performance relativa aos caracteres de maior relevância econômica. Nesse sentido, o

caráter peso de amêndoas secas (PAS) foi utilizado como referencial para a avaliação *per se* dos clones, por ser aquele que melhor representa a produção comercial no estudo. Observou-se na Tabela 14 que o maior PAS (1,21 g) foi apresentado pelo grupo IV, constituído exclusivamente pelo clone 20 (SIAL 244). O clone 20 deteve o maior PAS, a partir do maior peso de sementes úmidas (PSU de 1,84 g) e do maior peso total de fruto (PF de 699 g). Logo, o grupo IV, alia a máxima divergência à melhor performance média, devendo-se esperar híbridos superiores a partir de seus cruzamentos com clones do grupo II (clones 2, 4, 12, 13, 10, 1, 7 e 23). Não se deve descartar, entretanto, os cruzamentos envolvendo clones 18 e 20, ou seja, os grupos III e IV, respectivamente. Tais grupos exibem performance médias satisfatórias, estando suficientemente distantes entre si para gerarem híbridos heteróticos. Informações sobre acessos que detenham genes para resistência a uma doença de importância local também devem ser consideradas, no momento da escolha dos tipos parentais.

Tabela 14 - Valores médios intragrupos referentes a 13 características avaliadas em 26 clones de cacau. CEPEC, Itabuna, Ba.

G ^a	CF	DF	C/D	PF	PC	EG	ES	NSF	PSU	PAS	CA	LA	EA
	mm	mm		g	g	mm	mm		g	g	mm	mm	mm
I	135	86	1,56	476	357	8,9	7,0	40	1,58	0,99	23,8	12,2	7,5
II	132	83	1,60	432	328	8,9	7,2	42	1,29	0,81	22,5	11,3	7,0
III	115	90	1,28	479	396	10,5	8,2	24	1,61	1,06	24,4	12,5	7,6
IV	159	90	1,77	699	546	11,4	9,2	40	1,84	1,21	24,1	12,5	7,7

^a G I: SIC 18 (3); SIC 23 (5); SIC 24 (6); SIC 433 (8); SIC 628 (9); SIC 662 (11); SIC 765 (14); SIC 823 (15); SIAL 70 (16); SIAL 88 (17); SIAL 169 (19); SIAL 283 (21); SIAL 325 (22); SIAL 557 (24); SIAL 581 (25) e SIAL 638 (26).

G II: SIC 2 (1); SIC 17 (2); SIC 19 (4); SIC 250 (7); SIC 644 (10); SIC 680 (12); SIC 747 (13) e SIAL 357 (23).

G III: SIAL 105 (18).

G IV: SIAL 244 (20).

4.2.3. A divergência fenética por meio dos componentes principais

A importância relativa dos componentes principais está na percentagem da variância total que suas raízes características ou autovalores (λ_j) correspondentes retêm. A Tabela 15 evidenciou que os dois primeiros componentes explicaram 71,37% da variação total disponível e estiveram intimamente relacionados aos caracteres de fruto (CF, DF, C/D, PF e PC) e de sementes (PSU e PAS). Os quatro

primeiros componentes explicaram juntos 89,64% dessa mesma variação. Aos nove componentes restantes coube a explicação de apenas 10,36% do total da variação.

Tabela 15 - Estimativas dos autovalores (λ_j) associados aos componentes principais e de suas importâncias relativas e acumuladas, obtidas no estudo de 13 caracteres avaliados em 26 clones de cacau. CEPEC, Itabuna, Ba.

Componentes	Autovalores	%	% Acumulada
1	6,1171	47,05	47,05
2	3,1609	24,31	71,37
3	1,7974	13,83	85,19
4	0,5782	4,45	89,64
5	0,4604	3,54	93,18
6	0,3257	2,50	95,69
7	0,2941	2,26	97,95
8	0,1260	0,97	98,92
9	0,0768	0,59	99,51
10	0,0350	0,27	99,78
11	0,0233	0,18	99,96
12	0,0044	0,03	99,99
13	0,0005	0,01	100,00

Na literatura, o emprego de descritores nos estudos de divergência em bancos de germoplasma tem demonstrado que a variância tende a se dispersar em vários componentes. PEREIRA (1989) e CURI (1993) em mandioca e BARROS (1991) em cajú confirmaram tal tendência. A observação da Tabela 15 evidenciou que o presente estudo não seguiu a mesma tendência, muito embora tenha empregado descritores de germoplasma. A razão da discordância pode estar no tipo de descritor utilizado. Aqui os descritores têm uma natureza agrônômica, restritos apenas a caracteres de fruto e semente, representando, com certa aproximação, os componentes de produção. Nos trabalhos supra citados, os descritores foram de natureza majoritariamente botânica ou botânico-agrônômica.

A visualização da divergência foi proporcionada pelo exame do gráfico de dispersão de escores dos componentes principais (Figura 5). Dado que os dois

primeiros componentes explicaram 71,37% da variação total adotou-se um gráfico bidimensional, cujas coordenadas foram estabelecidas por esses dois primeiros componentes. Percebe-se a formação de dois grupos, constituídos exclusivamente pelos clones 20 e 18, no caso os grupos definidos como IV e III, respectivamente, pelo método de Tocher. De modo subjetivo, ainda dois outros grupos poderiam ser formados, envolvendo os clones restantes. Um grupo (II) seria formado pelos clones 1, 2, 4, 7, 10, 12, 13 e 23. E o outro (I) envolveria os clones 3, 5, 6, 8, 9, 11, 14, 15, 16, 17, 19, 21, 22, 24 e 25. O clone 20 foi o mais divergente em relação aos demais, destacando-se como o mais promissor para hibridação, sobretudo se cruzado com os clones 2 e 18. Tais clones guardam distância máxima entre si. Já os clones 3 e 14 e os clones 10 e 13 constituíram os pares mais similares. Esse agrupamento subjetivo, promovido para ser concordante com o agrupamento feito pelo método Tocher, demonstrou uma distorção pouco pronunciada na passagem das variáveis de um espaço p-dimensional para o bidimensional.

Figura 5 - Gráfico de dispersão de escores dos dois primeiros componentes principais (CP₁ e CP₂) referentes a 26 clones de cacau. Códigos dos clones estão na Tabela 14. Os círculos representam os grupos que estão identificados por algarismos romanos, correspondentes aos da Tabela 12.

A importância de um estudo dessa natureza está no fato de que os clones constituem-se nos verdadeiros genitores dos futuros híbridos. Assim, a informação sobre a divergência entre os clones pode ser obtida do próprio germoplasma, durante trabalhos rotineiros de caracterização, e ser utilizada diretamente para realização dos cruzamentos mais promissores. Para maior eficiência no processo preditivo os clones podem ser dispostos em ensaios com repetição, repetidos em vários locais e anos, sempre que possível. Por outro lado, a divergência sendo quantificada nos próprios genitores S_0 suprime a exigência de autocompatibilidade dos genitores. Assim, o estudo de divergência fenética em clones S_0 fica viabilizado, quaisquer que sejam os sistemas de incompatibilidade presentes nos referidos clones. Obviamente, que se existirem informações sobre a capacidade de combinação dos clones sob estudo, essas deverão ser incluídas no processo decisório sobre os possíveis cruzamentos. O estudo I demonstrou a importância desse tipo de informação.

O comportamento produtivo das seleções locais SIC, EEG, CEPEC e Catongo em cruzamento com clones introduzidos trinitários e alto amazônicos ICS 1, SCA 6 e 12, UF 613, IMC 67 e ICS 8 foi estudado por MARIANO & BARTLEY (1981). Lamentavelmente as séries das seleções locais não foram informadas pelos autores. Essas seleções locais, ainda segundo os autores, não se comportaram diferencialmente com relação à capacidade combinatória para a produção. A reduzida capacidade combinatória foi justificada pela origem comum das seleções locais. Entretanto, o clone 20 (SIAL 244), por exemplo, não tem confirmado esse comportamento. SIAL 244 tem se destacado pela sua performance produtiva em ensaio no CEPEC, em cruzamentos com clones trinitários. Desse modo, SIAL 244 x SPA 12 e SIAL 244 x SPA 17 produziram 3,7 kg e 3,5 kg por planta, respectivamente, em 1986, no quarto ano de campo, como informaram MONTEIRO & PEREIRA (1986).

É importante ressaltar que o clone não é empregado em plantios comerciais de cacau. DIAS (1993) analisou as causas do fracasso dos plantios clonais em cacau, na década de 50, sucedidos pelos plantios de híbridos biclonais propagados por sementes. A partir dessa análise, o autor propôs a adoção de plantios multiclonais em propriedades cacaueiras com alto nível tecnológico. Tais plantios clonais poderiam proporcionar grandes benefícios à lavoura de cacau, representados, sobretudo, pela reprodução integral das combinações híbridas superiores e pelo maior ganho genético possível. Inspirado nos modelos bem sucedidos da silvicultura clonal, o autor defendeu os plantios multiclonais, a partir de blocos monoclonais de cinco a dez hectares. Ainda que

sem dispor de um número ideal de clones para recomendação, algo como 10 a 15 clones, testados para cada região, foram sugeridos.

O efeito da interação genótipo por anos pode estar influenciando os resultados referentes à divergência entre clones, já que a caracterização se deu em apenas um ano (Tabela 16). Embora o estudo I tenha detectado que a avaliação da divergência com base em apenas um ano climaticamente favorável foi consistente é prudente realizar caracterizações durante dois ou mais anos. É que esse conjunto de clones encontra-se instalado em um banco de germoplasma, sem a observação dos rigores experimentais. No estudo I, ao contrário, os tratamentos foram ensaiados experimentalmente, controlando-se os efeitos ambientais. Por outro lado, a estimação do período mínimo de colheitas sucessivas necessário para se acessar o valor fenotípico real dos clones pode ser conduzida pelo coeficiente de repetibilidade, como aplicado por DIAS (1991) e DIAS & SOUZA (1993).

Dados bioquímicos, a exemplo das isoenzimas utilizadas por LANAUD (1986), e dados moleculares, em combinação com os dados morfológicos, tradicionalmente avaliados, podem ser submetidos também a análises multivariadas. Os dados bioquímicos e moleculares, ao contrário dos morfológicos, são comparáveis no tempo e no espaço, devido não serem substancialmente influenciados pelo ambiente. Contudo, alguma concordância entre esses vários tipos de dados tem sido verificada. Análises multivariadas aplicadas a dados morfológicos e a marcadores isoenzimáticos avaliados em 15 populações de *Pinus rigida* apresentaram padrões de variabilidade muito similares (FRYER, 1987). Todavia, considerando-se a disponibilidade apenas de dados morfológicos esses, sempre que possível, devem ser avaliados em experimentos com repetição, repetidos em vários locais e anos. O uso de testemunhas comuns também é recomendável, a fim de tornar os resultados comparáveis.

Fundamentalmente, ambas as técnicas multivariadas utilizadas demonstraram, de modo concordante, a existência de divergência fenética entre clones das populações de forasteiros amazônicos da região cacaueira baiana. Entretanto, é comum na literatura de cacau encontrarem-se relatos tratando da reduzida variabilidade presente nessas mesmas populações (BONDAR, 1938; MARIANO & BARTLEY, 1981). Esses autores, dentre outros, postulam que as populações cacaueiras baianas resultaram de introduções restritas, provenientes do baixo Amazonas e, em face disso, apresentam estreita base genética. O presente estudo contraria esse paradigma da uniformidade, demonstrando a necessidade de seu reexame. Os fatos sugerem, então, que as estratégias de seleção e de hibridação nessas populações podem ser viáveis.

Tabela 16 - Valores médios por clone, média geral e coeficientes de variação fenotípica de 26 clones de cacau, obtidos em 1981, referentes a 13 características de frutos e sementes. CEPEC, Itabuna, Ba.

C ^a	CF	DF	C/D	PF	PC	EG	ES	NSF	PSU	PAS	CA	LA	EA
	mm	mm		g	g	mm	mm		g	g	mm	mm	mm
1	115	85	1,35	392	301	8,0	6,9	38	1,31	0,79	22,8	11,1	7,3
2	136	80	1,70	395	283	7,8	5,9	44	1,24	0,79	22,8	10,9	6,7
3	132	86	1,53	475	353	8,8	6,8	42	1,75	1,03	24,4	13,0	7,3
4	138	79	1,75	433	315	8,1	6,2	45	1,34	0,84	22,9	10,9	7,0
5	149	85	1,75	520	403	9,2	7,3	41	1,49	0,90	24,0	11,4	7,7
6	141	86	1,64	452	337	10,0	7,7	40	1,50	0,98	23,7	11,7	7,6
7	140	85	1,65	465	371	10,7	8,4	39	1,28	0,89	21,7	11,8	6,9
8	132	85	1,55	421	302	8,5	6,6	42	1,44	0,95	23,5	11,8	7,5
9	120	85	1,41	408	298	8,7	6,4	40	1,52	0,93	23,1	12,5	7,2
10	126	84	1,50	427	329	9,0	8,2	44	1,23	0,80	22,6	11,3	7,0
11	129	84	1,54	434	324	8,1	6,3	42	1,39	0,88	23,4	12,0	7,0
12	125	80	1,56	395	288	8,1	6,7	42	1,33	0,82	22,1	11,7	7,3
13	130	85	1,53	457	356	9,1	7,2	42	1,24	0,74	22,8	11,4	6,8
14	133	85	1,56	474	338	8,6	6,9	42	1,72	1,01	24,0	12,8	7,6
15	142	90	1,58	553	460	9,7	8,2	39	1,60	1,03	23,5	11,9	7,8
16	134	89	1,51	499	372	10,3	7,9	37	1,66	1,09	23,6	13,0	7,4
17	120	85	1,41	422	303	7,3	5,7	43	1,55	0,99	23,2	12,0	7,3
18	115	90	1,28	479	396	10,5	8,2	24	1,61	1,06	24,4	12,5	7,6
19	136	85	1,60	464	333	7,8	5,9	41	1,63	1,03	25,1	12,8	7,2
20	159	90	1,77	699	546	11,4	9,2	40	1,84	1,21	24,1	12,5	7,7
21	145	91	1,59	544	415	9,3	7,4	39	1,80	1,11	23,2	12,3	7,4
22	147	87	1,69	552	430	9,4	7,5	39	1,51	1,03	23,7	11,0	7,6
23	144	83	1,73	491	378	10,5	8,4	45	1,32	0,85	22,0	11,1	6,7
24	126	86	1,47	453	346	8,8	7,1	38	1,57	0,99	23,0	12,1	7,5
25	138	91	1,52	521	408	9,7	7,9	43	1,47	0,97	24,1	12,0	7,4
26	133	84	1,58	420	298	8,7	6,6	39	1,65	0,97	24,9	13,3	7,9
A	134	85	1,57	471	357	9,1	7,2	40	1,49	0,95	23,4	11,9	7,3
B	7,89	3,73	7,67	14,2	17,1	11,4	12,6	9,91	12,0	12,0	3,62	5,86	4,57

^a 1 = SIC 2; 2 = SIC 17; 3 = SIC 18; 4 = SIC 19; 5 = SIC 23; 6 = SIC 24; 7 = SIC 250; 8 = SIC 433; 9 = SIC 628; 10 = SIC 644; 11 = SIC 662; 12 = SIC 680; 13 = SIC 747; 14 = SIC 765; 15 = SIC 823; 16 = SIAL 70; 17 = SIAL 88; 18 = SIAL 105; 19 = SIAL 169; 20 = SIAL 244; 21 = SIAL 283; 22 = SIAL 325; 23 = SIAL 357; 24 = SIAL 557; 25 = SIAL 581 e 26 = SIAL 638.

A = médias e B = Coeficiente de variação fenotípica

Fonte: CASTRO & BARTLEY (1983)

A existência de diversidade nas populações cacaueiras do sul da Bahia pode ser atribuída a hibridações naturais ocorridas entre os cultivares que constituem tais populações. As cultivares Comum, Pará e Maranhão formam a base das populações de cacau da Bahia (SORIA, 1963), coexistindo na mesma região. Algumas dessas populações atuais exibem grande variabilidade nos caracteres de fruto e semente,

o que conforme SORIA (1963) denota sua origem híbrida. É oportuno lembrar que mesmo cacaueiros autocompatíveis como os que constituem as populações baianas (VELLO, 1971) apresentam taxas médias de cruzamento de 43% (VELLO & GARCIA, 1967). Embora as flores do cacaueiro sejam hermafroditas, as anteras encontram-se recobertas pela cógula das pétalas e, sem a intervenção de uma minúscula mosca (*Forcipomyia sp*), o pólen viscoso não contacta espontaneamente o estigma.

VELLO et alii (1969) registraram a performance produtiva de cinco híbridos, uma progênie de polinização livre e uma progênie autofecundada envolvendo apenas clones locais (Tabela 17). Já no terceiro ano de campo as produções variaram de 1.319 kg/ha para o híbrido SIAL 325 x SIAL 407 até 2.327 kg/ha de amêndoas secas no caso de SIAL 88 x SIAL 90. Observa-se ainda que as progênies autopolinizadas e de polinização livre de SIAL 93 apresentaram produções similares, evidenciando ausência de depressão endogâmica.

Tabela 17 - Valores médios de produção de amêndoas secas, de diferentes tipos de progênies de cacau, envolvendo clones locais.

Clones		Produção de amêndoas secas (kg/ha)
SIAL 93	autopolinizado	1880
SIAL 93	polinização livre	1857
SIAL 73	x SIAL 21	1787
SIAL 325	x SIAL 407	1319
SIAL 399	x SIAL 93	1282
SIAL 13	x SIAL 246	1554
SIAL 88	x SIAL 90	2327

Fonte: VELLO et alii (1969)

A quantificação da diversidade entre clones locais só tem sentido no caso da divergência fenética refletir a divergência genética. A própria literatura de cacau traz indicações fortes de que o fato possa ocorrer. ENGELS (1983), ao comparar as relações genéticas entre 32 clones de cacau da Costa Rica com as similaridades fenéticas expressas por dendrograma, detectou que os clones com um tipo parental comum conhecido agruparam-se no mesmo cluster. Seus resultados confirmaram então que as relações fenéticas entre os clones representavam, com boa aproximação, as suas relações

genéticas. A representação gráfica dos clones por análise de fator repetiu os resultados da análise de cluster.

Com respeito aos descritores empregados nesse estudo, no geral, o controle genético pode ter superado o ambiental, reforçando a representatividade da distância fenética em termos genéticos. Em um dialelo 7 x 7 entre os clones SCA 6, P 7, Catongo, UF 29, CC 42, UF 613 e UF 676 verificaram-se as seguintes estimativas de herdabilidade no sentido amplo, para caracteres de fruto e semente: 0,89 para CF; 0,90 para DF; 0,86 para o PF; 0,85 para PC; 0,77 para EG; 0,53 para ES e 0,85 para o PSU (RAMIREZ MORA, 1987). Para semente, as seguintes estimativas foram obtidas: 0,54 para LA; 0,84 para CA; 0,64 para a ES e 0,73 para o PAS (LOPEZ BAEZ, 1984).

4.2.4. Relação entre distância geográfica e divergência fenética

Nesse estudo, a divergência genética não se mostrou relacionada à distribuição geográfica, já que clones do mesmo local foram distribuídos em diferentes grupos. O fato contraria uma premissa há muito assumida como verdadeira pelos melhoristas de cacau. Basta observar que o programa de melhoramento do cacau no Brasil, iniciado na década de 60, enfatizou a hibridação entre clones locais, basicamente das séries SIC e SIAL, com clones introduzidos, procedentes de diversos países. Entretanto, VELLO et alii (1972), relatando resultados de ensaios comparativos entre híbridos locais x locais e locais x introduzidos, concluíram que "muitas combinações entre clones locais, com aparentemente menor diversidade genética, foram mais produtivas que cruzamentos de clones locais por trinitários. Contudo, a produção em larga escala desses híbridos é dificultada pela ausência do caráter autoincompatibilidade nos clones da Bahia".

Então, os resultados obtidos por VELLO et alii (1972), retratados na Tabela 18, reforçam a tese da divergência entre clones baianos, suficiente até para gerar híbridos heteróticos, quando cruzados entre si. Os resultados das combinações híbridas obtidos por VELLO et alii (1972) são perfeitamente comparáveis, embora envolvam diferentes idades, locais e épocas de plantio, devido à utilização de uma testemunha comum a todos os ensaios (Tabela 18). Dentro desse contexto, o melhorista de cacau não deve considerar uma área geográfica particular como uma entidade uniforme em termos de divergência genética. Portanto, os resultados sugerem que a seleção de clones parentais para hibridação seja feita com base na estimativa de divergência genética entre eles. Precisa antes avaliar os genótipos com base na performance *per se* e, se possível, na capacidade geral de combinação, como ficou demonstrado no estudo I.

Tabela 18 - Valores médios de produção de amêndoas secas de híbridos, da cultivar Catongo (Testemunha comum) e da relação híbrido/testemunha de cacau para cada idade, obtidos em diversos experimentos, no período de 1966 a 1971. CEPEC, Itabuna, Ba.

Agrupamentos	3º ano de plantio			4º ano de plantio			5º ano de plantio			6º ano de plantio			7º ano de plantio		
	Hib.	Test.	H/T	Hib.	Test.	H/T	Hib.	Test.	H/T	Hib.	Test.	H/T	Hib.	Test.	H/T
Híbridos**															
Locais* x UF 613	303	254	1,2	473	495	0,9	786	1207	0,6	1477	1784	0,8	1596	1687	0,9
Locais x Locais	263	121	2,2	591	377	1,6	1421	1156	1,2	1616	1391	1,2	2102	1687	1,2
Locais x Catongo	269	105	2,6	529	352	1,5	1266	833	1,5	1967	976	2,0	2423	1687	1,4
Locais x DR 2	419	257	1,6	413	392	1,0	1239	1035	1,0	1785	1544	1,1	2858	1687	1,7
Locais x SIAL CB	100	30	3,3	842	212	4,0	1802	1891	0,9	1835	2505	0,7	-	-	-
Locais x ICS 1	448	277	1,6	728	595	1,2	1255	1051	1,2	2192	1828	1,2	-	-	-
Locais x CEPEC	757	440	1,7	321	424	0,8	718	679	1,0	895	1066	0,8	-	-	-
Locais x Sca 12	1340	283	4,7	1719	844	2,0	-	-	-	-	-	-	-	-	-
Locais x Sca 6	1250	283	4,4	1594	844	1,9	-	-	-	-	-	-	-	-	-
Locais x IMC 67	1017	438	2,3	1091	844	1,3	-	-	-	-	-	-	-	-	-
Locais x ICS 8	501	312	1,6	979	844	1,2	-	-	-	-	-	-	-	-	-

* Referente a todos os clones selecionados na Bahia, Espírito Santo e Pará, incluindo SIAL CB e CEPEC. Em sua maioria são clones das séries SIC e SIAL, utilizados em cruzamentos. Todos os demais clones são introduzidos.

** Os agrupamentos híbridos retratam a performance produtiva de 172 híbridos.

Fonte: VELLO et alii (1972)

A ausência de relação entre diversidade ecológica e geográfica com divergência genética tem sido relatada por GHADERI et alii (1979) e DAS & DAS GUPTA (1984) em feijão vigna, SINGH et alii (1981); SHERIFF & SHIVASHANKAR (1992); SHERIFF (1992); UPADHYAY & MURTY (1970) e JAIN et alii (1981) em milheto, SINGH & SINGH (1976) em pimentão, SINGH et alii (1979) em arroz, KATIYAR & SINGH (1979) em grão de bico, KALLOO & SIDHU (1982) em melão cantalupe, AMALRAJ (1982) em algodão perene, SINGH & GILL (1984) em algodão anual, JOSHI & VASHI (1992) em sorgo, SOMAYAJULU et alii (1970) em trigo, VARMA & GULATI (1982) em cevada, GUPTA et alii (1991) em mostarda indiana e CAMUSSI et alii (1983) em populações italianas de milho.

A hibridação em cacau envolvendo clones locais permanece pouco explorada. O principal motivo reside, sem dúvidas, na ausência da autoincompatibilidade genética nesses clones. Não obstante, esquemas de polinização manual, já aplicados em larga escala em cacau, podem viabilizar a produção desses híbridos. Tais híbridos do tipo local x local ofereceriam como vantagem uma maior uniformidade no tamanho das amêndoas, sendo essa uma característica importante para a indústria chocolateira. Essa uniformidade foi detectada por VELLO et alii (1972) nesses tipos de híbridos. Outra grande vantagem dos híbridos do tipo local x local é a possibilidade de se unir heterose para precocidade e rendimento à uma maior adaptação às condições da região cacaueira baiana, garantindo a estabilidade da produção (ver Tabela 18).

4.2.5. Importância e descarte de caracteres

Como ficou demonstrado no estudo I, as técnicas multivariadas são sensíveis o bastante para discriminar os caracteres de menor importância para a divergência. Tais caracteres referem-se àqueles pouco variáveis e/ou redundantes, comuns em estudos envolvendo um grande número de variáveis. Assim, a otimização do conjunto original de variáveis é processada através do descarte. Esse procedimento conduz a uma maior racionalidade na avaliação dos cultivares, na medida em que trabalha objetivamente um conjunto reduzido de variáveis, de real importância para a divergência. Então o descarte significa redução de custos operacionais, de mão de obra e de tempo, dispendidos na avaliação de acessos. O cacaueiro conta com 48 descritores botânico-agronômicos para sua caracterização (ENGELS, 1981). Esse número, aparentemente excessivo, pode ser otimizado pelo emprego das técnicas multivariadas de avaliação da divergência entre acessos. As próprias análises possibilitam selecionar descritores úteis, que reflitam relações genéticas mais que fenotípicas.

Para a identificação e descarte dos caracteres menos importante para a divergência entre clones foi adotado o critério de JOLLIFFE (1972; 1973), complementado pelo procedimento utilizado por CURI (1993). Os caracteres descartáveis foram aqueles que exibiram os maiores coeficientes de ponderação, também chamados elementos dos autovetores, nos últimos componentes principais (Tabela 19). A cada caráter descartado procedeu-se a uma nova análise do conjunto de caracteres remanescentes, com nova identificação dos maiores coeficientes de ponderação. O processo de descarte sucessivo se repetiu até o limite de definição, determinado pela impossibilidade de discriminarem-se os maiores coeficientes e/ou pela inconsistência em descartar caracteres não redundantes. No limite de definição nenhuma distorção no processo de agrupamento conduzido pelo método de Tocher foi admitida.

O primeiro descarte envolveu o caráter C/D ou seja a razão entre o comprimento e o diâmetro do fruto (Tabela 19). A redundância desse caráter foi certamente a principal causa do seu descarte. Observando a Tabela 20 depara-se com uma correlação elevada ($r = 0,88$) entre C/D e CF. O descarte seguinte recaiu sobre o caráter PF - peso total do fruto (Tabela 19). O terceiro caráter a ser descartado foi a espessura do gomo (EG). Foram observadas correlações altas entre esses dois últimos caracteres e o caráter peso da casca (PC) e também entre EG e ES (Tabela 20). O quarto descarte envolveu o peso de sementes úmidas - PSU (Tabela 19), também devido à alta correlação com o peso de amêndoas secas - PAS, como consta da Tabela 20.

Na quinta etapa de descarte o processo discriminatório torna-se inconsistente. O caráter peso da casca, PC, sendo o próximo a ser descartado é altamente correlacionado ao peso total do fruto, PF ($r = 0,97$), além de ser o mais variável dentre todos (Tabela 16). Ocorre que o peso total do fruto foi descartado na segunda etapa (Tabela 19). A inconsistência aumenta se o descarte prosseguir. Assim, na sexta etapa o peso das amêndoas secas, PAS, deveria ser descartado. Observando a Tabela 20 depara-se com uma alta correlação entre o peso de sementes úmidas (PSU), já descartado na quarta etapa, e esse último caráter ($r = 0,93$). Portanto, o limite de definição pareceu estar na quinta etapa de descarte e, assim sendo, o processo todo foi interrompido na etapa anterior com descarte dos caracteres C/D, PF, EG e PSU.

Os caracteres PF e PSU descartados têm valor agrônômico. Entretanto, tiveram reduzida importância, no conjunto dos caracteres avaliados, devido a presença do peso da casca (PC) e das amêndoas secas (PAS). As correlações fenotípicas entre PF e PC foram de 0,97, entre PF e PAS de 0,69 e de 0,93 entre o PSU e o PAS (Tabela 20). O PAS sim é um caráter de grande importância na seleção de cacaueiros,

conhecido como índice de sementes. Embora correlacionado ao peso do fruto e das sementes úmidas, o peso das amêndoas secas tem apresentado alta herdabilidade (LOPEZ BAEZ, 1984), o que viabiliza a seleção direta sobre esse caráter.

Tabela 19 - Estimativas dos coeficientes de ponderação dos autovetores associados aos últimos componentes principais, referentes a 13 características, avaliadas em 26 clones de cacau. CEPEC, Itabuna, Ba.

Caráter	Últimos Componentes Principais						
	13	12	11	10	9	8	7
CF	0,6588	0,0106	0,2915	-0,1013	-0,0320	-0,4408	-0,4389
DF	-0,3287	-0,0615	0,0242	-0,1137	0,0124	-0,3388	-0,2375
C/D	-0,6602	-	-	-	-	-	-
PF	-0,1002	-0,7191	-	-	-	-	-
PC	0,0949	0,6711	-0,2575	0,3316	-0,6041	-	-
EG	0,0219	0,0592	-0,6401	-	-	-	-
ES	-0,0206	-0,0511	0,6048	-0,2063	0,4186	0,3620	0,5835
NSF	-0,0043	0,0919	-0,2003	0,0731	0,0991	0,2421	0,3274
PSU	-0,0172	0,0783	-0,0254	-0,7193	-	-	-
PAS	0,0355	0,0816	0,0939	0,4647	0,5019	0,5625	-
CA	0,0165	0,0293	0,0124	-0,0449	0,2335	0,2669	0,5264
LA	-0,0088	-0,0120	0,0736	0,2618	-0,3538	-0,3192	-0,1577
EA	-0,0144	-0,0145	-0,1321	0,1225	-0,1311	-0,1078	-0,0349

O traço refere-se ao descarte simulado do caráter associado ao correspondente coeficiente.

Para um conjunto originalmente formado por 13 variáveis foram descartadas quatro, ou seja, cerca de 31%. Percentual similar foi obtido por CURI (1993) estudando 20 caracteres, em 28 cultivares de mandioca. Deve ser ressaltado que a otimização do conjunto de variáveis pela técnica do descarte das menos importante pode gerar perda de informação. Entretanto, nesse estudo a perda foi inexistente. Desse modo, o método de Tocher aplicado ao conjunto formado pelas nove variáveis remanescentes (Tabela 21), a partir das estimativas de distâncias Euclidianas médias entre pares de clones, proporcionou agrupamento similar àquele proporcionado pelo Tocher aplicado ao conjunto original, formado pelas 13 variáveis (Tabela 12). O resultado atestou a eficiência da metodologia de descarte utilizada.

Tabela 20 - Coeficientes de correlação fenotípica entre 13 caracteres avaliados em 26 clones de cacau. CEPEC, Itabuna, Ba.

Caráter	DF	C/D	PF	PC	EG	ES	NSF	PSU	PAS	CA	LA	EA
CFa	0,22	0,88**	0,75**	0,68**	0,50**	0,44*	0,29	0,27	0,37	0,13	-0,09	0,17
DF		-0,25	0,68**	0,72**	0,59**	0,57**	-0,52**	0,63**	0,74**	0,43*	0,44*	0,54**
C/D			0,42*	0,33	0,23	0,17	0,52**	-0,04	0,01	-0,08	-0,31	-0,09
PF				0,97**	0,69**	0,68**	-0,16	0,58**	0,69**	0,31	0,18	0,41*
PC					0,76**	0,76**	-0,26	0,48*	0,62**	0,23	0,09	0,38
EG						0,93**	-0,39*	0,25	0,43*	-0,01	0,13	0,20
ES							-0,31	0,13	0,32	-0,11	-0,01	0,19
NSF								-0,35	-0,42*	-0,31	-0,35	-0,44*
PSU									0,93**	0,69**	0,78**	0,69**
PAS										0,61**	0,66**	0,67**
CA											0,64**	0,64**
LA												0,48*

a CF, DF, C/D, PF e PC referem-se ao comprimento, diâmetro, razão comprimento/diâmetro, peso do fruto e peso da casca do fruto. EG, ES, NSF e PSU representam as espessuras do gomo e do sulco, respectivamente, seguidos pelo número de sementes e peso úmido dessas sementes por fruto. PAS, CA, LA e EA referem-se ao peso de amêndoas secas e ao comprimento, largura e espessura da amêndoa.

* ** significativo aos níveis de 5% e de 1% de significância, respectivamente.

Caso a metodologia de descarte adotada fosse aquela preconizada por JOLLIFFE (1972; 1973) *in totum* seriam descartados 10 dos 13 caracteres, ou seja, cerca de 76%. Convém lembrar que essa metodologia prevê o descarte de tantas variáveis quantos forem os componentes principais com autovalores inferiores a 0,7. No caso, observando a Tabela 15, depara-se com 10 componentes que atendem a essa condição. Sem dúvida que esse método, por ser muito drástico, pode causar uma distorção acentuada no processo de agrupamento dos genótipos.

Tabela 21 - Agrupamentos de similaridade entre 26 clones (Códigos) de cacau referentes a nove características, estabelecidos pelo método de Tocher, com base nas distâncias Euclidianas médias. CEPEC, Itabuna, Ba.

Grupos	Clones			
I	SIC 18 (3)	SIC 765 (14)	SIAL 169 (19)	SIAL 638 (26)
	SIC 433 (8)	SIAL 557 (24)	SIC 24 (6)	SIC 628 (9)
	SIC 662 (11)	SIAL 88 (17)	SIAL 70 (16)	SIAL 581 (25)
	SIC 23 (5)	SIAL 283 (21)	SIAL 325 (22)	SIC 823 (15)
II	SIC 17 (2)	SIC 19 (4)	SIC 680 (12)	SIC 747 (13)
	SIC 644 (10)	SIC 2 (1)	SIAL 357 (23)	SIC 250 (7)
III	SIAL 105 (18)			
IV	SIAL 244 (20)			

4.3. Considerações finais

A viabilidade da predição da performance média e heterótica de híbridos de cacau, a partir da estimação da divergência multivariada entre seus genitores na geração S_1 , foi demonstrada no estudo I. Obviamente, que essa metodologia deve ser testada, posteriormente, num dialelo envolvendo um conjunto mais amplo de cultivares. Se possível, esse dialelo deve ser conduzido em mais de um local e por vários anos. É fundamental, ainda, conhecer o comportamento dessa metodologia de predição quando aplicada nos períodos iniciais de produção do dialelo, ou seja, a partir do quinto ou sexto ano de plantio, quando a competição intergenotípica ainda não se estabilizou.

A restrição imposta à predição da performance do híbrido pela divergência entre genitores S_1 autocompatíveis foi contornada, empregando-se genitores clonais S_0 no estudo II. A vantagem dessa estratégia é que a divergência passa a ser

estimada, diretamente, nos verdadeiros genitores dos futuros híbridos, quaisquer que sejam os sistemas de incompatibilidade presentes nos clones. Uma vez consolidada a metodologia de predição, os ensaios dos genitores clonais passam a ser de fundamental importância. Também aqui deve-se atentar para ensaios clonais repetidos em diferentes locais e anos, com utilização de testemunhas comuns.

Sem a preocupação de apresentar a divergência entre os S_1 's, como sendo aquela que seria estimada nos respectivos genitores S_0 , é possível abordar a questão, com base nos conhecimentos da cultura e nos dados apresentados nesse trabalho. A questão central é a presença e a magnitude relativa da depressão endogâmica nos S_1 's. A abordagem, então, recai sobre a direção e a magnitude dos desvios de dominância (d), para os locos controlando os componentes de produção avaliados. Locos sem dominância ($d = 0$) não causam depressão nem heterose. Pois bem, a produção média de amêndoas úmidas (PSUP) dos cultivares foi de 2,72 kg/planta/ano (Tabela 7), equivalente a uma produção estimada de 1208,80 kg/ha/ano, para um fator de conversão de 40% de cacau úmido para cacau seco, como recomendado por CARLETTO et alii (1983). A produção média estimada dos correspondentes 20 híbridos, por sua vez, foi de 1879,81 kg/ha/ano. Ou seja, os cultivares S_1 produziram apenas 35,7% menos que seus híbridos e 61% mais que a média regional. Esses números sugerem um valor baixo para d , o que justifica a heterose média pouco expressiva dos híbridos e a reduzida depressão dos seus genitores S_1 . Outra forma de atestar a depressão pouco expressiva dos S_1 's é pela análise do estande. Em 1990, último ano do estudo, o estande médio variou de 13,2 a 16,0 plantas por parcela, com relação aos cultivares ICS 1 e CEPEC 1, respectivamente.

Outra abordagem referente a depressão endogâmica é a magnitude da carga genética, ou seja, o acúmulo de genes recessivos deletérios, em homozigose nos S_1 's. As pesquisas em cacau, no entanto, sugerem que a espécie tem uma reduzida carga genética. Assim, a despeito de muitos estudos classificarem o cacau como uma espécie alógama (SORIA, 1961; 1964; VELLO, 1971; VELLO & NASCIMENTO, 1971), em função de suas peculiaridades florais e da presença de alelos de autoincompatibilidade, outros, a exemplo de Cope, citado por ASCENSO (1962), consideram-na parcialmente autogâmica, com taxas de cruzamentos variando de 18% a 100%. Taxas de cruzamento de 30% têm sido verificadas em cacaueiros de Trinidad (HARLAND & FRECHVILLE, 1927) e de 43% em Ilhéus, Brasil (VELLO & GARCIA, 1967) e Tafo, Gana (POSNETTE, 1950). Convém lembrar que cacaueiros do oeste Africano tiveram origem nas populações de cacaueiros do Brasil. A partir desses dados, pode-se inferir que cacau

é uma espécie de sistema reprodutivo misto, que pratica alogamia ou autogamia, em função da população considerada. Ora, espécie de sistema misto apresenta carga genética intermediária em relação às alógamas e autógamias, justificando também a reduzida depressão endogâmica dos S_1 's. A literatura de cacau tem tratado da reduzida depressão de alguns clones de cacau. Em especial, a tolerância à endogamia dos clones ICS 1, clones da série SIAL e clones de Catongo, usados no estudo I, tem sido relatada por BARTLEY (1969; 1971); CARLETTO et alii (1977) e VELLO et alii (1969); SORIA (1963) e SORIA & ESQUIVEL (1969), respectivamente. SORIA (1963), por exemplo, relatou a manutenção do vigor e da produtividade do cultivar Catongo até a terceira geração de endogamia.

Em aplicações futuras, para teste da metodologia de predição de híbridos, não deve existir preocupação quanto à estimação da divergência entre cultivares S_1 . Em princípio, não existe razão para que a divergência manifestada por cultivares S_1 seja distinta daquela que seria obtida entre os correspondentes S_0 . Mesmo ocorrendo depressão entre os S_1 's, não haverá prejuízos para a divergência estimada, desde que tal depressão seja de magnitude relativa semelhante. Ademais, essa situação de manipular S_1 's somente ocorrerá nos testes da metodologia. Na maioria das situações, a divergência será estimada entre clones S_0 . Por último, a metodologia multivariada de quantificação da diversidade genética, aqui proposta, poderá auxiliar, no futuro, a recomendação de clones para plantios comerciais. De modo que, para cada região sejam recomendados apenas os clones de alta performance e de grande divergência entre si.

5. CONCLUSÕES

Para o estudo I as seguintes conclusões foram pertinentes:

- a. A divergência genética entre cinco cultivares de cacau, avaliados por cinco componentes de produção, foi quantificada pela distância de Mahalanobis (D^2), distância Euclidiana média padronizada (d_e) e distância Euclidiana estimada dos dois primeiros componentes principais (d_{cp}) e das duas primeiras variáveis canônicas (d_{vc}). Essas duas primeiras variáveis acumularam 92,14% da variação total e permitiram visualizar a divergência genética em um gráfico bidimensional. O método de Tocher discriminou dois grupos divergentes entre si. O grupo I formado pelos cultivares CC 41, SIAL 169, CEPEC 1 e SIC 19 e o grupo II formado pelo cultivar ICS 1;
- b. Uma estreita associação foi observada entre as estatísticas que consideram a matriz de dispersão residual, no caso D^2 e d_{vc} . O mesmo foi observado entre as estatísticas d_e e d_{cp} . Em ambos os casos, correlações altas ($r > 0,95$) foram obtidas. Porém, entre D^2 e d_e e entre d_{cp} e d_{vc} uma fraca concordância foi verificada ($r = 0,50$ e $0,66$, respectivamente). Desse modo, o estudo demonstrou que as estatísticas que consideram a matriz de dispersão residual entre caracteres devam ser preferidas, sempre que possível sua estimação;
- c. Os caracteres de menor contribuição para a divergência genética, identificados por componentes principais, foram o número de frutos colhidos e o peso de sementes úmidas por planta, NFCP e PSUP, respectivamente. Por variáveis canônicas, foram identificados o número de frutos sadios (NFSP) e PSUP, como os de menor importância;
- d. A eficiência da predição da performance média de híbridos pela estimativa da divergência entre genitores S_1 , expressa por D^2 , foi elevada para os componentes mais íntimos da produção comercial de cacau. Correlações moderadas e altas foram verificadas entre D^2 e PSUP ($r = 0,68$) e entre D^2 e PSUF ($r = 0,76$). A performance heterótica dos híbridos também foi satisfatoriamente predita pela divergência entre os genitores. Uma correlação de $0,65$ foi encontrada entre D^2 e PSUP. Em termos de PSUP e PSUF, o cultivar ICS 1, o mais divergente dentre todos, gerou os híbridos de melhor performance média. O segundo par de cultivares mais divergente entre si, constituído por SIAL 169 e ICS 1, gerou também o híbrido F_1 de melhor performance média para os mesmos

caracteres. Finalmente, o híbrido de pior performance média CEPEC 1 x SIC 19 foi gerado do cruzamento envolvendo o segundo par de cultivares mais similar. Os resultados sugeriram ainda uma associação entre distância parental e efeitos de capacidade específica nos híbridos;

e. A distância geográfica esteve relacionada à distância genética, com base no agrupamento promovido pelo método de Tocher. Todavia, a concordância entre essas duas distâncias não foi total. O cultivar trinitário CC 41 foi agrupado junto às seleções baianas, representadas pelos clones amazônicos SIAL 169, CEPEC 1 e SIC 19. A própria origem híbrida do cultivar CC 41, resultado do cruzamento entre crioulos e amazônicos, justificou o agrupamento;

f. A divergência entre os cultivares, estimada por D^2 , apresentou um padrão estável ao longo dos cinco anos de estudos. Em consequência, a avaliação da divergência genética pode ser conduzida com base em um simples ano, especialmente se o ano apresenta condições climáticas ótimas.

Para o estudo II as seguintes conclusões foram verificadas:

a. Quatro grupos de similaridade fenética foram detectados pelo método de Tocher e, subjetivamente evidenciados pelo dendrograma, obtido pelo método do vizinho mais próximo e pelo gráfico bidimensional formado a partir dos escores dos primeiros componentes principais. Os dois primeiros componentes absorveram 71,37% da variação total. O estudo demonstrou a existência de divergência entre os clones locais, sugerindo a possibilidade de síntese de híbridos heteróticos;

b. A divergência entre os clones das séries SIC e SIAL não se mostrou relacionada à distribuição ecogeográfica. O fato, além de contrariar o paradigma da uniformidade das populações cacaueiras baianas, sugeriu a viabilidade das estratégias de seleção e hibridação nessas populações;

c. Os caracteres razão entre comprimento e diâmetro do fruto (C/D), peso do fruto (PF), espessura do gomo da casca (EG) e peso das sementes úmidas (PSU) foram identificados como os de menor contribuição para a divergência fenética entre os 26 clones estudados. Os quatro caracteres acima, representando 31% do total avaliado, foram descartados por serem redundantes ou pouco variáveis. A metodologia de descarte adotada foi satisfatória. Finalmente, os métodos multivariados provaram ser úteis na manipulação de coleções de germoplasma, otimizando o conjunto de descritores e identificando acessos em duplicata.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ADAMS, M.W. An estimation of homogeneity in crop plants, with special reference to genetic vulnerability in the dry bean, *Phaseolus vulgaris* L. **Euphytica**, 26(3): 665-79, 1977
- AMALRAJ, S.F.A. Genetic divergence in *Gossypium barbadense* L.. **Genetica Agraria**, 36(1/2): 23-30, 1982
- ANAND, I.J. & RAWAT, D.S. Genetic diversity, combining ability and heterosis in brown mustard. **The Indian Journal of Genetics & Plant Breeding**, 44(2): 226-34, 1984
- ANDERSON, T.W. **An introduction to multivariate statistical analysis**. New York, John Wiley, 1958. 374p.
- ARUNACHALAM, V. Genetic distance in plant breeding. **The Indian Journal of Genetics & Plant Breeding**, 41(2): 226-36, 1981
- ARUNACHALAM, V. & BANDYOPADHYAY, A. Limits to genetic divergence for occurrence of heterosis - experimental evidence from crop plants. **The Indian Journal of Genetics & Plant Breeding**, 44(3): 548-54, 1984
- ARUNACHALAM, V.; BANDYOPADHYAY, A.; NIGAM, S.N.; GIBBONS, R.W. Heterosis in relation to genetic divergence and specific combining ability in groundnut (*Arachis hipogaea* L.). **Euphytica**, 33(1): 33-9, 1984
- ARUNACHALAM, V. & RAM, J. Geographic diversity in relation to genetic divergence in cultivated sorghum. **The Indian Journal of Genetics & Plant Breeding**, 27(3): 369-80, 1967
- ASCENSO, J.C. O seminário sobre seleção e melhoramento do cacau da VIII Conferência Interamericana do Cacau. **Estudos Agronomicos**, 3(3): 135-41, 1962
- ATANDA, A.O. Heterotic pod production of double over single crosses in *Theobroma cacao* L. In: INTERNATIONAL COCOA RESEARCH CONFERENCE, 4., Saint Augustine, 1972. **Proceedings**. Port-of-Spain, Government of Trinidad and Tobago, 1972. p.82-9

- ATANDA, A.O. Heterosis in crosses of *Theobroma cacao*. **Experimental Agriculture**, 9(1): 23-9, 1973
- ATANDA, A.O. & TOXOPEUS, H. A proved case of heterosis in *Theobroma cacao* L. In: INTERNATIONAL COCOA RESEARCH CONFERENCE, 3., Accra, 1969. **Proceedings**. Tafo, Cocoa Research Institute of Nigeria, 1971. p.545-51
- BAGCHI, S.K. A preliminary study on the genetic divergence of *Acacia nilotica* through seed parameters. **Indian Forester**, 118(6): 416-24, 1992
- BARROS, L.M. Caracterização morfológica e isoenzimática do cajueiro. Piracicaba, 1991. 256p. (Doutorado - Escola Superior de Agricultura "Luiz de Queiroz")
- BARTLEY, B.G.D. Twenty years of cacao breeding at the Imperial College of Tropical Agriculture, Trinidad. In: CONFERÊNCIA INTERNACIONAL DE PESQUISAS EM CACAU, 2., Salvador/Itabuna, 1967. **Memórias**. São Paulo, CEPLAC, 1969. p.29-34
- BARTLEY, B.G.D. Yield variation in the early productive years in trials with cacao (*Theobroma cacao* L.). **Euphytica**, 19(2): 199-206, 1970
- BARTLEY, B.G.D. First generation inbreds as parents in hybrids of *Theobroma cacao* L. **Tropical Agriculture**, 48(1): 79-84, 1971
- BHATT, G.M. Comparison of various methods of selecting parents for hybridization in common bread wheat (*Triticum aestivum* L.). **Australian Journal of Agricultural Research**, 24(4): 457-64, 1973
- BHATT, G.M. An application of multivariate analysis to selection for quality characters in wheat. **Australian Journal of Agricultural Research**, 27(1): 11-8, 1976
- BONDAR, G. A cultura do cacau na Bahia. Salvador, Instituto de Cacau da Bahia, 1938. 205p. (Boletim Técnico, 1)
- BROWN, J.S. Principal component and cluster analyses of cotton cultivar variability across the U.S. cotton belt. **Crop Science**, 31(4): 915-22, 1991
- BURLEY, J. & BURROWS, P.M. Multivariate analysis of variation in needles among provenances of *Pinus kesiya* Royle ex Gordon (syn. *P. khasya* Royle; *P. insularis* Endlicher). **Silvae Genetica**, 21(3/4): 69-77, 1972
- BURTON, A.J.; RAMM, C.W.; PREGITZER, K.S.; REED, D.D. Use of multivariate methods in forest research site selection. **Canadian Journal of Forest Research**, 21(11): 1573-80, 1991
- CALAMASSI, R.; PUGLISI, S.R.; VENDRAMIN, G.G. Genetic variation in morphological and anatomical needle characteristics in *Pinus brutia* Ten. **Silvae Genetica**, 37(5/6): 199-206, 1988

- CAMUSSI, A.; OTTAVIANO, E.; CALINSKI, T.; KACZMAREK, Z. Genetic distances based on quantitative traits. **Genetics**, **111**(4): 945-62, 1985
- CAMUSSI, A.; SPAGNOLETTI ZEULI, P.L.; MELCHIORRE, P. Numerical taxonomy of italian maize populations: genetic distances on the basis of heterotic effects. **Maydica**, **28**(4): 411-24, 1983
- CARLETTI, G.A.; GARCIA, J.R.; MAGALHÃES, W.S. Evaluación de híbridos y líneas endocriadas de cacao en Bahia. In: INTERNATIONAL COCOA RESEARCH CONFERENCE, 5., Ibadan, 1975. **Proceedings**. Ibadan, CRIN, 1977. p. 106-9
- CARLETTI, G.A.; MONTEIRO, W.R.; BARTLEY, B.G.D. Critérios para seleção de híbridos com cacaueiros. **Revista Theobroma**, **13**(4): 315-20, 1983
- CASTRO, G.C.T. & BARTLEY, B.G.D. Caracterização dos recursos genéticos do cacaueiro. 1. Folha, fruto e semente de seleções da Bahia das séries SIC e SIAL. **Revista Theobroma**, **13**(3): 263-73, 1983
- CHAUDHARY, B.D. & SINGH, V.P. Genetic divergence in some indian and exotic barley varieties and their hybrids. **The Indian Journal of Genetics & Plant Breeding**, **35**(3): 409-13, 1975
- CHEESMAN, E.E. Notes on the nomenclature, classification and possible relationships of cocoa populations. **Tropical Agriculture**, **21**(8): 144-59, 1944
- COOLEY, W.W. & LOHNES, P.R. **Multivariate data analysis**. Malabar, Robert E. Krieger, 1986. 364p. (Central)
- CRESS, C.E. Heterosis of the hybrid related to gene frequency differences between two populations. **Genetics**, **53**(2): 269-74, 1966
- CRUZ, C.D. Aplicação de algumas técnicas multivariadas no melhoramento de plantas. Piracicaba, 1990. 188p. (Doutorado - Escola Superior de Agricultura "Luiz de Queiroz")
- CUATRECASAS, J. Cacao and its allies. A taxonomic revision of the genus *Theobroma*. **Bulletin of the United States National Museum**, **35**(6): 379-614, 1964
- CURI, R. Dinâmica evolutiva e caracterização de germoplasma de mandioca (*Manihot esculenta*, Crantz) na agricultura autóctone do sul do Estado de São Paulo. Piracicaba, 1993. 103p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz")

- DAS, P.K. & DAS GUPTA, T. Multivariate analysis in black gram [*Vigna mungo* (L) HEPPER}. **The Indian Journal of Genetics & Plant Breeding**, 44(2): 243-7, 1984
- DIAS, L.A.S. Repetibilidade e capacidade de combinação em cacau (*Theobroma cacao* L.) nas condições do sul da Bahia. Piracicaba, 1991. 111p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz")
- DIAS, L.A.S. Propagação vegetativa vs reprodução seminal em cacau. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA PARA O PROGRESSO DA CIÊNCIA, 45., Recife, 1993. **Anais**. São Paulo, SBPC, 1993. p. 19
- DIAS, L.A.S. & SOUZA, C.A.S. Aplicação do coeficiente de repetibilidade na seleção de cacaueiros em plantação comercial. In: CONGRESSO NACIONAL DE GENÉTICA, 39., Caxambú, 1993. **Resumos**. Ribeirão Preto, SBG, 1993. p. 364
- DRAKE, D.W. Contrasting success of natural hybridization in two *Eucalyptus* species pairs. **Australian Journal of Botany**, 28(2): 167-91, 1980
- ENGELS, J.M.M. **Genetic resources of cacao: A catalogue of the CATIE collection**. Turrialba, CATIE, 1981. 196p. (Technical Bulletin, 7)
- ENGELS, J.M.M. A systematic description of cacao clones. III. Relationships between characteristics and some consequences for the cacao breeding. **Euphytica**, 32(3): 719-33, 1983
- ENGELS, J.M.M.; BARTLEY, B.G.D.; ENRIQUEZ, G.A. Cacao descriptors, their states and *modus operandi*. **Turrialba**, 30(2): 209-18, 1979
- ENRIQUEZ, G.A. & SORIA, J. **Catalogo de cultivares de cacao**. Turrialba, IICA, 1967. 551p.
- ESQUIVEL, O. & SORIA, J. Algunos datos sobre la variabilidad de algunos componentes del rendimiento en poblaciones de híbridos interclonales de cacao. **Cacao**, 12(4): 1-8, 1967
- FALCONER, D.S. **Introduction to quantitative genetics**. 3. ed. London, Longman, 1989. 438p.
- FALKENHAGEN, E.R. & NASH, W.Sr. Multivariate classification in provenance research. **Silvae Genetica**, 27(1): 14-23, 1978
- FERREIRA, D.F. Métodos de avaliação da divergência genética em milho e suas relações com os cruzamentos dialélicos. Lavras, 1993. 68p. (Mestrado - Escola Superior de Agricultura de Lavras)
- FISHER, R.A. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, 7: 179-88, 1936

- FRYER, J.H. Agreement between patterns of morphological variability and isozyme band phenotypes in pitch pine. *Silvae Genetica*, **36**(5/6): 199-206, 1987
- GHADERI, A.; ADAMS, M.W.; NASSIB, A.M. Relationship between genetic distance and heterosis for yield and morphological traits in dry edible bean and faba bean. *Crop Science*, **24**(1): 37-42, 1984
- GHADERI, A.; SHISHEGAR, M.; REZAI, A.; EHDAIE, B. Multivariate analysis of genetic diversity for yield and its components in mung bean. *Journal of American Society Horticultural Science*, **104**(6): 728-31, 1979
- GLENDINNING, D.R. & VERNON, A.J. Inter-varietal competition in cocoa trials. *Journal Horticultural Science*, **40**(4): 317-9, 1965
- GODOI, C.R.M. *Análise estatística multidimensional*. Piracicaba, ESALQ, Departamento de Matemática. 1985. 187p.
- GOODMAN, M.M. The races of maize: I. The use of Mahalanobis' generalized distances to measure morphological similarity. *Fitotecnica Latinoamericana*, **4**(1): 1-22, 1967
- GRAFIUS, J.E. A geometry for plant breeding. *Crop Science*, **4**(3): 241-6, 1964
- GUPTA, V.P.; SEKHON, M.S.; SATIJA, D.R. Studies on genetic diversity, heterosis and combining ability in indian mustard [*Brassica juncea* L. (CZERN & COSS)]. *The Indian Journal of Genetics & Plant Breeding*, **51**(4): 448-53, 1991
- HALLAUER, A.R. & MIRANDA FILHO, J.B. *Quantitative genetics in maize breeding*. Ames, Iowa State University Press, 1981. 468p.
- HARLAND, S.C. & FRECHVILLE, G.E. Natural crossing and the genetics of axil spot in cacao. *Genetica*, **9**(1): 279-88, 1927
- HUSSAINI, S.H.; GOODMAN, M.M.; TIMOTHY, D.H. Multivariate analysis and the geographical distribution of the world collection of finger millet. *Crop Science*, **17**(2): 257-63, 1977
- JAGADEV, P.N. & SAMAL, K.M. Multivariate analysis in niger (*Guizotia abyssinica* CASS). *The Indian Journal of Genetics & Plant Breeding*, **51**(4): 392-4, 1991
- JAGADEV, P.N.; SAMAL, K.M.; LENKA, D. Genetic divergence in rape mustard. *The Indian Journal of Genetics & Plant Breeding*, **51**(4): 465-7, 1991
- JAIN, A.K.; DHAGAT, N.K.; TIWARI, A.S. Genetic divergence in finger millet. *The Indian Journal of Genetics & Plant Breeding*, **41**(3): 346-8, 1981
- JANORIA, M.P.; RHODES, A.M.; SHRIVASTAVA, M.N. Phenetic similarity among dwarf rice cultivars, grown in two environments, as estimated by numerical taxonomic methods. *Annals of Botany*, **40**(2): 845-9, 1976

- JESWANI, L.M.; MURTY, B.R.; MEHRA, R.B. Divergence in relation to geographical origin in a world collection of linseed. **The Indian Journal of Genetics & Plant Breeding**, 30(1): 11-25, 1970
- JOLLIFFE, I.T. Discarding variables in a principal component analysis. I. Artificial data. **Applied Statistics**, 21(2): 160-73, 1972
- JOLLIFFE, I.T. Discarding variables in a principal component analysis. II. Real data. **Applied Statistics**, 22(1): 21-31, 1973
- JOSHI, P. & VASHI, P.S. Mahalanobis generalized distance and genetic diversity in sorghum. **The Indian Journal of Genetics & Plant Breeding**, 52(1): 85-93, 1992
- KALLOO, J.D. & SIDHU, A.S. Genetic divergence in muskmelon (*Cucumis melo* L.). **Genetica Agraria**, 36(1): 1-8, 1982
- KATIHAR, R.P. & SINGH, S.P. Genetic divergence in chickpea. **The Indian Journal of Genetics & Plant Breeding**, 39(2): 354-8, 1979
- KENDALL, M.G. **A course in multivariate analysis**. London, Charles Griffin, 1963. 185p.
- KHARE, D. & SINGH, C.B. Divergence analysis in *Vicia faba* L. for nutritional and antinutritional attributes. **The Indian Journal of Genetics & Plant Breeding**, 52(1): 58-62, 1992
- LANAUD, C. Utilisation des marqueurs enzymatiques pour l'étude génétique du cacaoyer: *Theobroma cacao* L. II. Etude du polymorphisme de six systèmes enzymatiques. **Café Cacao Thé**, 30(4): 271-80, 1986
- LOPEZ BAEZ, O. Herencia de ciertos caracteres de la semilla del cacao (*Theobroma cacao* L). Turrialba, 1984. 93p. (Mestrado - Universidad de Costa Rica)
- MALUF, W.R. & FERREIRA, P.E. Análise multivariada da divergência genética em feijão-vagem (*Phaseolus vulgaris* L.). **Horticultura Brasileira**, 1(2): 31-4, 1983
- MALUF, W.R.; FERREIRA, P.E.; MIRANDA, J.E.C. Genetic divergence in tomatoes and its relationship with heterosis for yield in F₁ hybrids. **Revista Brasileira de Genética**, 6(3): 453-60, 1983
- MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. **Multivariate analysis**. London, Academic Press. 1979. 521p.
- MARIANO, A.H. Avaliação de cultivares híbridas de cacauero: Ensaios 34 e 35. In: CEPEC. **Informe de Pesquisas**. Ilhéus, CEPLAC, 1986. p.13-5
- MARIANO, A.H. & BARTLEY, B.G.D. Comportamento das seleções baianas na produção de híbridos de cacaueiros. In: CONFÉRENCE INTERNATIONALE

- SUR LA RECHERCHE CACAOYÈRE, 7., Douala, 1979. **Actes**. London, J. de Lafforest and Transla-Inter, 1981. p.527-33
- MARIANO, A.H. & YAMADA, M.M. Avaliação de cultivares híbridas de cacauero: Ensaio 23. In: CEPEC. **Informe de Pesquisas**. Ilhéus, CEPLAC, 1986. p.10
- MAZE, J.; PARKER, W.H.; BRADFIELD, G.E. Generation-dependent patterns of variation and population differentiation in *Abies amabilis* and *A. lasiocarpa* (Pinaceae) from north-coastal British Columbia. **Canadian Journal of Botany**, 59(3): 275-82, 1981
- MELCHINGER, A.E.; MESSMER, M.M.; LEE, M.; WOODMAN, W.L.; LAMKEY, K.R. Diversity and relationships among U.S. maize inbreds revealed by restriction fragment length polymorphisms. **Crop Science**, 31(3): 669-78, 1991
- MERGEN, F.; LESTER, D.T.; FURNIVAL, G.M.; BURLEY, J. Discriminant analysis of *Eucalyptus cinerea* x *Eucalyptus maculosa* hybrids. **Silvae Genetica**, 15(5/6): 148-54, 1966
- MIRANDA, J.E.C.; CRUZ, C.D.; COSTA, C.P. Predição do comportamento de híbridos de pimentão (*Capsicum annuum* L.) pela divergência genética dos progenitores. **Revista Brasileira de Genética**, 11(4): 929-37, 1988
- MOLL, R.H.; LONNQUIST, J.H.; VÉLEZ FORTUNO, J.; JOHNSON, E.C. The relationship of heterosis and genetic divergence in maize. **Genetics**, 52(1): 139-44, 1965
- MOLL, R.H.; SALHUANA, W.S.; ROBINSON, H.F. Heterosis and genetic diversity in variety crosses of maize. **Crop Science**, 2(3): 197-8, 1962
- MONTALVAN DEL AGUILA, R. Determinação de divergência genética em germoplasma de arroz (*Oryza sativa* L.) através de análises eletroforéticas de proteínas de grão. 1990. 83p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz")
- MONTEIRO, W.R. & PEREIRA, M.G. Avaliação de cultivares híbridas de cacauero: Ensiao 37. In: CEPEC. **Informe de Pesquisas**. Ilhéus, CEPLAC, 1986. p.17
- MORRISON, D.F. **Multivariate statistical methods**. New York, McGraw-Hill, 1976. 415p.
- MURTY, B.R. & ANAND, I.J. Combining ability and genetic diversity in some varieties of *Linum usitatissimum*. **The Indian Journal of Genetics & Plant Breeding**, 26(1): 21-36, 1965

- MURTY, B.R. & ARUNACHALAM, V. The nature of divergence in relation to breeding system in some crop plants. **The Indian Journal of Genetics & Plant Breeding**, 26(2): 188-98, 1966
- NARAYAN, R.K.J. & MACEFIELD, A.J. Adaptive responses and genetic divergence in a world germplasm collection of chick pea (*Cicer arietinum* L.). **Theoretical and Applied Genetics**, 47(4): 179-87, 1976
- OPEKE, L.K. & JACOB, V.J. Preliminary observations on 106 crosses from a diallel crossing programme in *Theobroma cacao* L. In: CONFERÊNCIA INTERNACIONAL DE PESQUISAS DE CACAU, 2., Salvador/Itabuna, 1967. **Memórias**, São Paulo, CEPLAC, 1969. p.94-9
- PEETERS, J.P. & MARTINELLI, J.A. Hierarchical cluster analysis as a tool to manage variation in germplasm collection. **Theoretical and Applied Genetics**, 78(1): 42-8, 1989
- PEREIRA, A.V. Utilização de análise multivariada na caracterização de germoplasma de mandioca (*Manihot esculenta*, Crantz). Piracicaba, 1989. 180p. (Doutorado - Escola Superior de Agricultura "Luiz de Queiroz")
- PEREIRA, M.G.; CARLETTO, G.A.; DIAS, L.A.S. **Avaliação de híbridos de cacauzeiros nas condições de Linhares - ES.** Ilhéus, CEPLAC, 1987. 40p. (Boletim Técnico, 150)
- PIMENTEL GOMES, F. **Curso de estatística experimental.** 13. ed. Piracicaba, Nobel, 1990. 468p.
- PINTO, G.C.P.; MAGALHÃES, W.S.; MARQUES, E.S.; RODRIGUES, E.M. Seleção de matrizes superiores nos cacauais da Bahia. In: CONFERÊNCIA INTERNACIONAL DE PESQUISAS EM CACAU, 2., Salvador/Itabuna, 1967. **Memórias**. São Paulo, CEPLAC, 1969. p.105-13
- PIRES, C.E.L.S. Diversidade genética de variedades de cana-de-açúcar (*Saccharum* spp.) cultivadas no Brasil. Piracicaba, 1993. 120p. (Doutorado - Escola Superior de Agricultura "Luiz de Queiroz")
- POSNETTE, A.F. The pollination of cacao in the gold coast. **Tropical Agriculture**, 25(3): 155-63, 1950
- POTTS, B.M. & REID, J.B. Hybridization between *Eucalyptus obliqua* L'Hérit and *E. pulchella* Desf. **Australian Journal of Botany**, 31(3): 211-29, 1983
- POTTS, B.M. & REID, J.B. Analysis of a hybrid swarm between *Eucalyptus risdonii* Hook. f. and *E. amygdalina* Labill. **Australian Journal of Botany**, 33(5): 543-62, 1985a

- POTTS, B.M. & REID, J.B. Variation in the *Eucalyptus gunnii-archeri* complex. II. The origin of variation. **Australian Journal of Botany**, 33(5): 519-41, 1985b
- POUND, F.J. The genetic constitution of the cacao crop. In: ANNUAL REPORT ON CACAO RESEARCH, 2., Port-of-Spain, 1932. Port-of-Spain, IMPERIAL COLLEGE OF TROPICAL AGRICULTURE, 1933. p.10-24
- PULCINELLI, C.E.; COUTO, H.T.Z.; JESUS, R.M.; GARCIA, A. Divergência genética entre populações de eucalipto. In: CONGRESSO NACIONAL DE GENÉTICA, 39., Caxambú, 1993. **Resumos**. Ribeirão Preto, SBG, 1993. p.355
- RABOIN, L.M.; PAULIN, D.; CILAS, C.; ESKES, A.B. Analyse génétique de quelques caractères quantitatifs des fleurs du cacaoyer (*Theobroma cacao* L.). Leur intérêt pour l'évaluation de la diversité de l'espèce. **Café Cacao Thé**, 37(4): 273-82, 1993
- RAM, J. & PANWAR, D.V.S. Intraspecific divergence in rice. **The Indian Journal of Genetics & Plant Breeding**, 30(1): 1-10, 1970
- RAMANUJAM, S.; TIWARI, A.S.; MEHRA, R.B. Genetic divergence and hybrid performance in mung bean. **Theoretical and Applied Genetics**, 45(5): 211-4, 1974
- RAMIREZ MORA, L.G. Herencia de ciertos caracteres de la mazorca y del arbol de cacao (*Theobroma cacao* L.). Turrialba, 1987. 82p. (Mestrado - Universidad de Costa rica)
- RAO, C.R. **Advanced statistical methods in biometric research**. New York, John Wiley, 1952. 390p.
- REYES, E.H. & PEREZ, A. Resultado de un ensayo comparativo de progenies híbridas entre cacao Amazónico por Criollo, Amazónico por Trinitario y Criollo por Trinitario. In: INTERNATIONAL COCOA RESEARCH CONFERENCE, 3., Accra, 1969. **Proceedings**. Tafo, Cocoa Research Institute of Nigeria, 1971. p.590-9
- RESENDE, M.D.V. & HIGA, A.R. Aplicación de técnicas de análisis multivariada en el estudio de la divergência genética entre procedências de *Eucalyptus viminalis*. In: JORNADAS SOBRE EUCALIPTOS DE ALTA PRODUCTIVIDAD, Buenos Aires, 1991. **Actas**. Buenos Aires, CIEF, 1991. p.139-54
- RESENDE, R.M.S. Aplicação de técnicas de análise multivariada e eletroforese de isoenzimas em estudos de relações fenéticas no gênero *Laelia* secção *Parviflorae*. Piracicaba, 1991. 130p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz")

- RICHERS, R.H. & JEFFERS, J.N.R. Multivariate analysis of the elms of northern France. II. Pooled analysis of the elm populations of northern France and England. **Silvae Genetica**, **27**(3/4): 85-95, 1978
- RUSSEL, T.A. The vigour of some cacao hybrids. **Tropical Agriculture**, **29**(4/6): 102-6, 1952
- SHAMSUDDIN, A.K.M. Genetic diversity in relation to heterosis and combining ability in spring wheat. **Theoretical and Applied Genetics**, **70**(3): 306-8, 1985
- SHERIFF, R.A. Divergence analysis in finger millet (*Eleusine coracana* GAERTN). **The Indian Journal of Genetics & Plant Breeding**, **52**(1): 72-4, 1992
- SHERIFF, R.A. & SHIVASHANKAR, G. Genetic divergence in foxtail millet (*setaria italica* B.). **The Indian Journal of Genetics & Plant Breeding**, **52**(1): 29-32, 1992
- SINGH, D. The relative importance of characters affecting genetic divergence. **The Indian Journal of Genetics & Plant Breeding**, **41**(2): 237-45, 1981
- SINGH, T.H. & GILL, S.S. Genetic diversity in upland cotton under different environments. **The Indian Journal of Genetics & Plant Breeding**, **44**(3): 506-13, 1984
- SINGH, S. & GUPTA, P.K. Genetic divergence in pearl millet. **The Indian Journal of Genetics & Plant Breeding**, **39**(2): 210-5, 1979
- SINGH, S.P.; GUTIÉRREZ, J.A.; MOLINA, A.; URREA, C.; GEPTS, P. Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. **Crop Science**, **31**(1): 23-9, 1991
- SINGH, Y.P.; KUMAR, A.; CHAUHAN, B.P.S. Genetic divergence in pearl millet. **The Indian Journal of Genetics & Plant Breeding**, **41**(2): 186-90, 1981
- SINGH, A. & SINGH, H.N. Genetic divergence in chilli. **The Indian Journal of Genetics & Plant Breeding**, **36**(3): 425-30, 1976
- SINGH, V.P.; SWAMINATHAN, M.S.; MEHRA, R.B.; SIDDIQ, E.A. Divergence among dwarfs of cultivated rice. **The Indian Journal of Genetics & Plant Breeding**, **39**(2): 315-22, 1979
- SMALL, E. A numerical analysis of morpho-geographic groups of cultivars of *Humulus lupulus* based on samples of cones. **Canadian Journal of Botany**, **59**(3): 311-24, 1981
- SMITH, J.S.C. Genetic variability within U.S. hybrid maize: Multivariate analysis of isozyme data. **Crop Science**, **24**(6): 1041-6, 1984

- SMITH, J.S.C. Diversity of United States hybrid maize germplasm: Isozymic and chromatographic evidence. **Crop Science**, **28**(1): 63-69, 1988
- SMITH, D.F. & SMITH, J.S.C. Prediction of heterosis using pedigree relationship, biochemical and morphological data. In: ANNUAL ILLINOIS CORN BREEDERS SCHOLL, 23., 1987. p.1-21
- SMITH, J.S.C. & SMITH, O.S. Associations among inbred lines of maize using electrophoretic, chromatographic, and pedigree data. 1. Multivariate and cluster analysis of data from 'Lancaster Sure Crop' derived lines. **Theoretical and Applied Genetics**, **73**(5): 654-64, 1987
- SMITH, J.S.C. & SMITH, O.S. Associations among inbred lines of maize using electrophoretic, chromatographic, and pedigree data. 2. Multivariate and cluster analysis of data from Iowa Stiff Stalk Synthetic derived lines. **Theoretical and Applied Genetics**, **76**(1): 39-44, 1988
- SMITH, J.S.C. & SMITH, O.S. The description and assessment of distance between inbred lines of maize: I. The use of morphological traits as descriptors. **Maydica**, **34**(2): 141-50, 1989a.
- SMITH, J.S.C. & SMITH, O.S. The description and assessment of distances between inbred lines of maize: II. The utility of morphological, biochemical, and genetic descriptors and a scheme for the testing of distinctiveness between inbred lines. **Maydica**, **34**(2): 151-61, 1989b.
- SMITH, J.S.C.; GOODMAN, M.M.; STUBER, C.W. Genetic variability within U. S. maize germplasm: I. Historically importance lines. **Crop Science**, **25**(3): 550-5, 1985a
- SMITH, J.S.C.; GOODMAN, M.M.; STUBER, C.W. Genetic variability within U. S. maize germplasm: II. Widely used inbred lines 1970 to 1979. **Crop Science**, **25**(4): 681-5, 1985b
- SMITH, J.S.C.; SMITH, O.S.; BOWEN, S.L.; TENBORG, R.A.; WALL, S.L. Similarities among a group of elite maize inbreds as measured by pedigree, F₁ grain yield, grain yield, heterosis, and RFLPs. **Theoretical and Applied Genetics**, **80**(6): 833-40, 1990
- SNEATH, P.H.A. & SOKAL, R.R. **Numerical taxonomy**. San Francisco, W.R. Freeman, 1973. 573p.
- SOMAYAJULU, P.L.N.; JOSHI, A.B.; MURTY, B.R. Genetic divergence in wheat. **The Indian Journal of Genetics & Plant Breeding**, **30**(1): 47-58, 1970

- SORIA, J. El mejoramiento del cacao. In: HARDY, F., ed. **Manual de cacao**. San Jose, IICA. 1961. p.358-80
- SORIA, J. Observaciones sobre las variedades y cultivares de cacao en Bahía, Brasil. **Cacao**, 8(1): 1-6, 1963
- SORIA, J. El vigor híbrido y su uso en el mejoramiento genético de cacao. **Fitotecnia Latinoamericana**, 1(1): 59-78, 1964
- SORIA, J. Notes on the native cacaos in the vicinity of Iquitos (Peru) and of Alto Beni (Bolivia). **Cacao**, 10(4): 14-6, 1965
- SORIA, J. The latest cocoa expeditions to the Amazon basin. **Cacao**, 15(1): 5-15, 1970
- SORIA, J. & ESQUIVEL, O. Algunos resultados del programa de mejoramiento genético de cacao en el IICA - Turrialba. In: CONFERÊNCIA INTERNACIONAL DE PESQUISAS EM CACAU, 2., Salvador/Itabuna, 1967. **Memórias**. São Paulo, CEPLAC, 1969. p.35-42
- SOUZA, A.L.; HOSOKAWA, R.T.; KIRCHNER, F.F.; MACHADO, S.A. Análises multivariadas para manejo de floresta natural na reserva florestal de Linhares, Espírito Santo: Análises de agrupamento e discriminante. **Revista Árvore**, 14(2): 85-101, 1990
- SPAGNOLETTI ZEULI, P.L. & QUALSET, C.O. Geographical diversity for quantitative spike characters in a world collection of durum wheat. **Crop Science**, 27(2): 235-41, 1987
- SRIVASTAVA, P.S.L. & ARUNACHALAM, V. Heterosis as a function of genetic divergence in triticales. **Zeitschrift für Pflanzenzüchtung**, 78(2): 269-75, 1977
- TIBBITS, W.N. Germination and morphology of progeny from controlled pollinations of *Eucalyptus nitens* (Deane & Maiden) Maiden. **Australian Journal of Botany**, 36(6): 677-91, 1988
- TIMOTHY, D.H. Genetic diversity, heterosis and use of exotic stocks in maize in Columbia. **Statistical Genetics and Plant Breeding**, 982: 581-93, 1963
- TOXOPEUS, H. Botany, types and populations. In: WOOD, G.A.R. & LASS, R.A., ed. **Cocoa**. 4 ed.. New York, Longman, 1985a. cap. 2., p.11-37. (Tropical Agriculture Series).
- TOXOPEUS, H. Planting material. In: WOOD, G.A.R. & LASS, R.A., ed. **Cocoa**. 4 ed.. New York, Longman, 1985b. cap. 4., p.80-92. (Tropical Agriculture Series).
- UPADHYAY, M.K. & MURTY, B.R. Genetic divergence in relation to geographical distribution in pearl millet. **The Indian Journal of Genetics & Plant Breeding**, 30(3): 704-15, 1970

- VAN LAAR, A. Multivariate analysis: A way to better understanding of complexity. **South African Forestry Journal**, 141: 34-41, 1987
- VAN LAAR, A. **Forest biometry**. Stellenbosch, Sappi Forests, 1991. 590p.
- VARMA, N.S. & GULATI, S.C. Genetic divergence in 2-rowed and 6-rowed barley. **The Indian Journal of Genetics & Plant Breeding**, 42(3): 314-8, 1982
- VELLO, F. Influência da origem do grão de pólen sobre alguns caracteres fenotípicos do fruto de cacau (*Theobroma cacao* L.). **Revista Theobroma**, 1(3): 3-10, 1971
- VELLO, F. & GARCIA, J.R. Estudos sobre polinização do cacau: I. Determinação da porcentagem de polinização cruzada em condições de campo. In: CEPEC. **Informe Técnico de Pesquisas**. Ilhéus, CEPLAC, 1967. p.10-12
- VELLO, F.; GARCIA, J.R.; MAGALHÃES, W.S. Produção e seleção de cacaueiros híbridos na Bahia. **Revista Theobroma**, 2(3): 15-35, 1972
- VELLO, F.; MARIANO, A.H.; GARCIA, J.R.; NASCIMENTO, T.F.; MAGALHÃES, W.S. O programa de melhoramento genético do cacau na Bahia. In: CONFERÊNCIA INTERNACIONAL DE PESQUISAS EM CACAU, 2., Salvador/Itabuna, 1967. **Memórias**. São Paulo, CEPLAC, 1969. p.43-56
- VELLO, F. & NASCIMENTO, I.F. Influência da origem do pólen na produção do cacau. **Revista Theobroma**, 1(1): 7-14, 1971
- VELLO, N.A. & PIRES, C.E.L.S. Estratégias para a seleção de parentais. **Revista Brasileira de Genética**, 15(1): 45-9, 1992 (Suplemento 1)
- VENCOVSKY, R. & BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto, R.B.G., 1992. 496p
- WILCHES, O.M. Evaluacion de treinta y cuatro variedades de mani mediante tecnicas multivariadas. **Revista ICA**, 18(1): 67-76, 1983
- WILSON, J.P.; BURTON, G.W.; ZONGO, J.D.; DICKO, I.O. Diversity among pearl millet landraces collected in central Burkina Faso. **Crop Science**, 30(1): 40-3, 1990
- WILKS, S.S. Certain generalizations in the analysis of variance. **Biometrika**, 24: 471-4, 1932