

# Construção de uma coleção nuclear de arroz para o Brasil

Tabare Abadie<sup>(1)</sup>, Celia Maria Torres Cordeiro<sup>(2)</sup>, Jaime Roberto Fonseca<sup>(3)</sup>, Rosa de Belém das Neves Alves<sup>(2)</sup>, Marília Lobo Burle<sup>(2)</sup>, Cláudio Brondani<sup>(3)</sup>, Paulo Hideo Nakano Rangel<sup>(3)</sup>, Emílio da Maia de Castro<sup>(3)</sup>, Heloísa Torres da Silva<sup>(3)</sup>, Marlene Silva Freire<sup>(4)</sup>, Francisco José Pfeilsticker Zimmermann<sup>(3)</sup> e José Ronaldo Magalhães<sup>(5)</sup>

<sup>(1)</sup>Universidad de la República, Facultad de Agronomía, Garzón 780, Montevideo, Uruguay. E-mail: tabadie@inia.org.uy <sup>(2)</sup>Embrapa Recursos Genéticos e Biotecnologia, PqEB - Parque Estação Biológica, W5 Norte Final, CEP 70770-900 Brasília, DF. E-mail: celia@cenargen.embrapa.br, rbelem@fca.unesp.br, miburle@ucdavis.edu <sup>(3)</sup>Embrapa Arroz e Feijão, Rod. Goiânia a Nova Veneza, Km 12, Fazenda Capivara, CEP 75375-000 Santo Antônio de Goiás, GO. E-mail: jfonseca@cnpaf.embrapa.br, brondani@cnpaf.embrapa.br, heloisa@cnpaf.embrapa.br, fjpz@cnpaf.embrapa.br, phrangel@cnpaf.embrapa.br, emilio@cnpaf.embrapa.br <sup>(4)</sup>Caixa Postal 565, CEP 37270-000 Campo Belo, MG. <sup>(5)</sup>Embrapa Gado de Leite, Rua Eugênio Nascimento, Dom Bosco, CEP 36038-330 Juiz de Fora, MG. E-mail: josemag@cnpgl.embrapa.br

**Resumo** – A coleção de germoplasma de arroz da Embrapa consiste aproximadamente de 10.000 acessos. O objetivo desse trabalho foi estabelecer a Coleção Nuclear (CN) dessa coleção utilizando as informações e dados disponíveis sobre seus acessos. A estratégia CN foi introduzida no manejo de recursos genéticos vegetais com o principal objetivo de ampliar e sistematizar o uso desses recursos. Uma CN deve ser selecionada procurando reter a variabilidade genética existente na coleção inteira (CI) com um mínimo de redundância. Os acessos da coleção de arroz foram classificados em três estratos: a) variedades tradicionais do Brasil (VT); b) linhagens/cultivares melhoradas do Brasil (LCM); e c) linhagens/cultivares introduzidas (LCI). As variedades tradicionais foram ainda classificadas segundo o sistema de cultivo (terras altas, várzeas e facultativo). Os três estratos foram representados na Coleção Nuclear, mas ênfase maior foi dada às variedades tradicionais, que constituíram 308 acessos. Os acessos foram alocados para cada sistema de cultivo, proporcionalmente ao produto do logaritmo do número de variedades tradicionais pelo índice de Shannon (medida de diversidade) de cada um deles. A seleção dos acessos foi feita com o auxílio do Sistema de Informação Geográfica (SIG). A CN brasileira de arroz está formada por 550 acessos

**Termos para indexação:** *Oryza sativa*, SIG, diversidade genética, recursos genéticos, terras altas, várzeas.

## Constructing a rice core collection for Brazil

**Abstract** – The Rice Germplasm Collection of Embrapa consists of approximately 10,000 accessions. This study aimed to establish a core collection using the currently available information data for those accessions. The strategy Core Collection (CN) was introduced in the management of plant genetic resources with the main purpose of improving the use of these resources. CN should be selected in order to preserve the genetic variability of the whole collection (CI), with minimum redundancy. The accessions within the rice collection were classified into three strata: a) landraces from Brazil (VT); b) breeding materials from Brazil (LCM); and c) introductions (LCI). The landraces were further classified according to crop system (uplands, lowlands and facultative). These three strata were represented in the Core Collection, but more emphasis was considered in representing the landraces, which are represented by 308 accessions. The accessions were allocated, for each crop system, proportionally to the product of the logarithm of the number of landraces by the Shannon Diversity Index (a measure of genetic diversity) within each crop system. Curators and breeders, supported by a Geographical Information System (GIS) made the selection of the accessions. The final Brazilian rice Core Collection consists of 550 accesses.

**Index terms:** *Oryza sativa*, GIS, genetic diversity, genetic resources, uplands, lowlands.

## Introdução

O arroz (*Oryza sativa* L.) é cultivado em todos os continentes, sendo um alimento nutritivo na dieta de mais da metade da população mundial. Foi domesticado na Ásia, onde evoluiu sob forte pressão de seleção natural

e artificial, em vários grupos ecogeográficos, entre eles os grupos *indica* e *japonica*, que mostram a sincronia em sua diferenciação morfológica e molecular (Li et al., 2000).

A cultura espalhou-se para o Oriente Médio por volta do século VIII e daí para a Europa nos séculos IX e X.

A introdução no Brasil foi feita pelos portugueses no século XVI (Pereira, 2002), nas regiões costeiras da Bahia, e no século seguinte no Maranhão. Espalhou-se com sucesso por todo o país permitindo que, nos séculos XVIII e XIX, o Brasil exportasse arroz.

Tradicionalmente o arroz tem sido plantado no Brasil em dois ambientes: várzeas e terras altas. Em várzeas, o arroz é plantado em áreas naturalmente inundadas (arroz de várzea), em várzeas com irrigação controlada (arroz irrigado) ou em várzeas sem irrigação (várzea úmida). Em terras altas o arroz é cultivado em terras firmes sem irrigação (arroz de sequeiro). No Brasil, as diferenças entre os dois ambientes de cultivo são bastante claras, demandando, inclusive, que diferentes programas de melhoramento com objetivos distintos e base genética diferenciada sejam estabelecidos para cada um deles. Devido à história das introduções de arroz durante o período colonial, o germoplasma usado pelos pequenos agricultores em várzeas e terras altas pode também ser confundido com os dois diferentes 'pools' gênicos, dos grupos *indica* e *japonica*, respectivamente (Guimarães & Sant'Ana, 1999).

Chang & Bardenas (1965) consideram que não existe uma nítida diferença morfológica ou fisiológica entre os genótipos dos dois sistemas de cultivo. No entanto, é bem conhecido que a pressão de seleção natural e artificial exercida no ecossistema de terras altas induziu adaptações específicas, especialmente no hábito de crescimento radicular (Pinheiro, 1999).

O melhoramento genético foi um dos principais responsáveis pelo aumento verificado na produtividade do arroz no Brasil no início da década de 80. Após esse grande avanço, os ganhos genéticos, quando obtidos, foram de pequena magnitude, configurando uma situação de patamar de produtividade. Isto tem sido explicado em parte pela reduzida base genética das populações utilizadas nos programas de melhoramento. Rangel et al. (1996), analisando as genealogias de 42 cultivares de arroz de terras baixas recomendadas para cultivo no Brasil no período 1980–1992, verificaram que 68% do 'pool' gênico destas cultivares eram provenientes de apenas dez ancestrais. A base genética do arroz de terras altas é também estreita: 81% do 'pool' gênico das cultivares liberadas no período 1971–1993 foram oriundos de apenas onze ancestrais (Montalban et al., 1998).

A ampliação da base genética das cultivares de arroz do Brasil tem sido proposta como estratégia para se romper, a médio e longo prazos, o atual patamar de produtivi-

dade e evitar a vulnerabilidade genética. Uma das alternativas para isso é usar a diversidade presente na coleção de germoplasma, especialmente entre as variedades tradicionais (VT), que têm sido pouco usadas no melhoramento genético. No entanto, os recursos disponíveis para caracterização e avaliação de germoplasma que vão permitir o uso da variabilidade genética armazenada na coleção de germoplasma são cada vez mais limitados. Daí a necessidade de se estabelecer uma Coleção Nuclear (CN) de arroz para o Brasil.

Esta estratégia permite priorizar e concentrar a aplicação de recursos de modo a formar uma base de informação mais completa sobre este conjunto de acessos levando efetivamente a uma ampliação do uso do germoplasma. Uma CN consiste de um limitado conjunto de acessos que procura representar a diversidade genética da coleção inteira (CI) com um mínimo de redundância (Van Hinthum, 1999). Procedimentos adequados para o seu desenvolvimento são disponíveis em Brown & Spillane (1999) e Van Hinthum (1999).

A amostragem de uma coleção de germoplasma para estabelecer uma CN requer um esforço integrado envolvendo curadores, melhoristas e geneticistas. Definição sobre o tamanho da CN, de modo a facilitar o seu uso, e a escolha de critérios de classificação para os quais os acessos da CI tenham informação, são essenciais. A definição de estratégias de alocação do esforço amostral para procurar maximizar a diversidade genética capturada na CN também é um ponto importante.

O desenvolvimento de CN de arroz tem sido discutido por vários autores (Vaughan, 1991; Vaughan & Jackson, 1995; Jackson et al., 1999). No caso de grandes coleções como a do International Rice Research Institute (IRRI), nas Filipinas, composta de mais de 100.000 acessos, o desenvolvimento de uma CN representa um desafio, já que mesmo uma amostra relativamente pequena alcança tamanho de difícil manuseio na maioria das situações (Jackson et al., 1999). Apesar disso, Jackson et al. (1999) sugerem que a importância de CN para fins de conservação e pesquisa genômica justifica o estabelecimento de CN de diferentes tamanhos para atender diferentes demandas. Zeng et al. (2001) estudaram a diversidade ecológica e genética de acessos de arroz da província de Yunnan, China, com o objetivo de estabelecer uma CN.

O objetivo deste trabalho foi desenvolver uma CN para o germoplasma de arroz do Brasil, conservado na Embrapa, utilizando as informações e dados disponíveis atualmente sobre os seus acessos.

## Material e Métodos

A coleção de germoplasma de arroz até o ano de 2002 estava constituída por 9.890 acessos, mantidos pela Embrapa Arroz e Feijão, em Santo Antônio de Goiás, GO, e Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, em condições de conservação de médio e longo prazo.

Um tamanho entre 500 e 600 acessos para a CN foi considerado aceitável para um manejo adequado na caracterização morfológica, molecular e avaliação em campo. Os acessos da coleção foram classificados inicialmente em três estratos: a) variedades tradicionais do Brasil (VT), coletadas durante as quatro últimas décadas em plantios de pequenos agricultores que praticam agricultura de subsistência – neste grupo também se incluíram as variedades comerciais mantidas pelo agricultor por mais de 30 anos (2.402 acessos) (Figura 1); b) linhagens/cultivares melhoradas (LCM) oriundas de programas de melhoramento do Brasil (3.448 acessos); c) linhagens/cultivares introduzidas (LCI) no Brasil oriundas de programas de melhoramento de outros países (4.040 acessos).

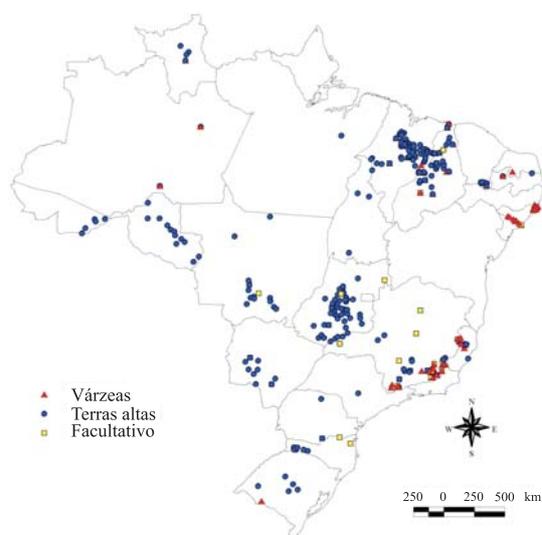
O ambiente (várzea ou terras altas) foi tomado como um segundo critério para estratificação do grupo VT. Esta estratificação foi avaliada usando um modelo de regressão logística para quantificar a associação entre os ambientes e os seguintes descritores: tipo de grão (TG), pubescência da glumela (PG), cor da glumela

(CG), cor do ápulo (CA), pubescência da folha (PF), arista (A), ângulo da folha bandeira (AFB), altura da planta (AP), tipo de panícula (TP) e exerceção da panícula (EP). Os sete primeiros descritores foram considerados na análise em uma escala nominal e os três últimos em escala ordinal. Um grupo de 1.041 acessos (43% do grupo VT) foi usado nesta análise, sendo 158 acessos de várzea e 883 acessos de terras altas. Para a construção do modelo foi adotada a técnica de ‘stepwise’, com um nível de significância de 0,05 para introduzir ou excluir descritores no modelo.

Inicialmente, a origem geográfica não foi sugerida pelos melhoristas e curadores como um critério de classificação para o grupo VT. No entanto, a extensão do país e a diversidade geográfica das regiões de coletas surgiram a importância deste critério. Um conjunto de 68 VT de arroz de várzea coletado nos Estados do Maranhão e Minas Gerais, pertencente à coleção da Embrapa, e caracterizado quanto a dez descritores agrônômicos (altura da planta, área foliar, comprimento da panícula, número de espiguetas por panícula, porcentagem de grãos cheios por panícula, massa de 100 grãos em uma amostra de 20 panículas, número de perfilhos e de panículas por m<sup>2</sup>, dias para o florescimento e produção em grãos por parcela) (Rangel et al., 1991), foi submetido a uma análise canônica discriminante para testar o efeito da origem geográfica. Embora esses dados sejam limitados – referem-se apenas a VT de arroz de várzea em duas regiões de cultivo –, o estudo pode oferecer subsídio sobre o uso do critério geográfico, pois as regiões representadas têm alta densidade de coleta e são importantes regiões de cultivo de VT.

Na constituição da CN procurou-se guardar aproximadamente a proporção de 50% de VT, 25% de LCM e 25% de LCI. O estrato VT foi constituído de uma proporção mais alta porque ele representa a variabilidade genética do arroz adaptada às diversas condições de cultivo no Brasil, sendo por isso um recurso genético único e de grande valor para uso e estudos que demandem uma representação dessa variabilidade.

Para delinear a alocação dos 300 acessos que representarão as VT na CN, utilizaram-se os dados de caracterização de 1.078 acessos de VT quanto a dez descritores morfológicos. Este conjunto de dados foi constituído pelos 1.041 acessos previamente referidos



**Figura 1.** Locais de coleta das variedades tradicionais (VT) de arroz da coleção de germoplasma da Embrapa.

para uso na avaliação das diferenças entre os sistemas de cultivo em terras altas e várzeas, mais um grupo de 37 acessos classificados como facultativos quanto ao sistema de cultivo. Este último grupo inclui os acessos que foram descritos nos dados de passaporte como sendo plantados em ambos os sistemas de cultivo. A diversidade em cada sistema de cultivo foi estimada usando o Índice de Shannon (IS), segundo a fórmula:

$$IS = - \sum_{i=1}^{n_j} \sum_{j=1}^{10} p_{ij} \times \log(p_{ij})$$

sendo  $p_{ij}$  a frequência fenotípica do  $i$ -ésimo estado do  $j$ -ésimo descritor e  $n_j$  o número de estados do  $j$ -ésimo descritor. Este índice é comumente usado em estudos de avaliação de diversidade a partir de descritores discretos (Galwey, 1995). Mantendo-se o número de estados constante, IS é maior quando os estados dos descritores são igualmente comuns do que quando alguns estados são comuns e outros raros. Este índice é particularmente adequado para delinear a amostragem para CN porque é também sensível à riqueza de alelos, no caso presente ao número de estados do descritor.

Ao se admitir que as VT da coleção da Embrapa possam ser tomadas como uma amostra aleatória das VT das regiões do Brasil onde ocorreram as expedições de coleta, pode-se usar o método 'bootstrap' para estimar o erro padrão, intervalo de confiança (IC) e obter estimativas do IS corrigidas quanto à tendenciosidade. Os IC aproximados foram estimados pelo método BC<sub>A</sub> (correção para o viés e aceleração), usando uma macro SAS (SAS Institute, 2002).

Quatro alternativas de alocação de amostra, entre aquelas adotadas para seleção de uma CN, foram avaliadas para o estrato VT: i) aleatória simples; ii) estratificada proporcional ao número de acessos de cada sistema de cultivo na coleção; iii) estratificada proporcional ao logaritmo do número de acessos de cada sistema de cultivo na coleção; e iv) estratificada proporcional ao produto do IS pelo logaritmo do número de acessos de cada sistema de cultivo na coleção.

Estas alternativas foram avaliadas gerando-se uma distribuição empírica do IS calculado em cada uma de 3.000 amostras. Nas alternativas estratificadas, cada uma das amostras resultou de uma combinação de três subamostras com tamanhos ditados pelo tipo de alocação avaliado, constantes na Tabela 1, e retiradas com repo-

sição da amostra original de cada tipo de cultivo. Na alternativa aleatória simples cada amostra de 300 acessos foi retirada com reposição da amostra original total. A estratégia escolhida foi aquela em que o 95<sup>o</sup> e 5<sup>o</sup> percentís da distribuição amostral empírica do IS foram mais elevados, indicando que 90% das amostras geradas apresentaram uma capacidade de retenção de variabilidade genética entre estes valores de IS. Assim, a amostragem foi delineada não para representar o padrão de diversidade genética, medida pelo IS, das variedades tradicionais das regiões coletadas, (representadas na coleção de germoplasma), mas sim, para escolher um tipo de amostragem que, entre os três clássicos propostos para CN, incluindo-se adicionalmente a amostragem aleatória simples, gere amostras que freqüentemente apresentem uma representação mais ampla dessa diversidade genética, como medida pelo IS.

As coordenadas geográficas dos locais de coleta dos acessos de VT identificadas por sistemas de cultivo foram introduzidas no SIG Arcview 3.2 e os pontos lançados sobre o mapa político do Brasil ( Figura 1).

## Resultados e Discussão

O sistema de cultivo mostrou ser um bom critério para classificar os acessos de arroz da amostra. O modelo logístico ajustado foi significativamente diferente do modelo nulo (apenas intercepto):  $\chi^2$  para o teste da razão de verossimilhança = 303,89; gl = 12;  $p < 0,0001$ ;  $R^2_{ajustado} = 0,44$ . Todos os oito descritores incluídos no modelo foram significativos em  $P < 0,01$ , com exceção de TP, que foi significativo em  $P = 0,026$ . Os descritores não incluídos no modelo (A e AP) nunca atingiram o nível de significância requerido para um descritor entrar no modelo ( $P = 0,05$ ). O descritor PF foi o primeiro a ser incluído no modelo. Além disso, a inclusão de PF

**Tabela 1.** Alocação da amostra de 300 acessos de variedades tradicionais (VT) de arroz em três diferentes sistemas de cultivo<sup>(1)</sup>.

Amostra estratificada	Várzea	Terras altas	Facultativo
Proporcional a $N_h$	39	246	15
Proporcional ao $\log(N_h)$	95	126	79
Proporcional ao $\log(N_h)IS_h$	86	128	86

<sup>(1)</sup> $N_h$  é o número de acessos da coleção inteira no  $h$ -ésimo estrato, conforme inventário de janeiro de 2002.

melhorou os níveis de significância dos descritores TP ( $P = 0,09$  para  $P = 0,032$ ) e EP ( $P = 0,54$  para  $P < 0,0001$ ), sugerindo que a pubescência da folha é um descritor cujos estados têm um desempenho importante em estruturar a associação destes descritores com a razão de chances (várzeas/terras altas).

É importante ressaltar que embora apenas caracteres morfológicos tenham sido usados, eles são relevantes na caracterização do arroz. Alguns deles são incluídos no Índice de Cheng, citado por Li et al. (2000). Dos seis descritores usados neste índice, três – pubescência da folha, comprimento/largura do grão e cor da casca na emissão da panícula – são relacionados àqueles usados nesta modelagem. O índice de Cheng é reconhecido como um bom indicador para diferenciação entre os grupos *indica* e *japonica* e tem se mostrado altamente correlacionado com o diagnóstico desta diferenciação, feito com marcadores RFLP (Li et al., 2000). A adaptação a determinado sistema de cultivo é uma aptidão muito complexa, provavelmente determinada por um grande número de genes. Assim, a adoção desta aptidão para estratificar a CI indiretamente significa organizar a sua diversidade quanto a um grande número de outros caracteres.

Quanto ao critério geográfico, o teste de ‘Wilks’ produziu  $F_{(10,57)} = 3,20$ ,  $P = 0,0025$ , indicando que a variável canônica discriminante que separa o centróide dos dois grupos é significativa. O coeficiente de correlação canônica que mede a associação entre os grupos e os escores sobre variável canônica é igual a 0,60. Isto significa que 0,36 da variação total entre os escores sobre a variável canônica está associada à origem geográfica das VT.

O resultado indica que a classificação por este critério ajudaria a obter uma adequada representação da variabilidade genética associada à diversidade geográfica. A ação combinada da seleção natural e artificial pretende produzir específica adaptação das VT às diferentes condições ecogeográficas. Entre outros, um componente de destaque desta adaptação geográfica pode ser atribuído à resposta ao fotoperíodo. As divergências no comprimento do fotoperíodo são importantes nas diferentes regiões de coleta. Em adição, existe variabilidade genética para a sensibilidade ao fotoperíodo como tem sido mostrado pela adaptação diferenciada das cultivares produzidas pelo melhoramento no Brasil (Steinmetz & Meirelles, 1999). No entanto, devido à ausência de um conjunto de dados de avalia-

ção agronômica que permitam uma quantificação do impacto da estratificação por região geográfica, dentro de cada sistema de cultivo, na recuperação desse tipo de variabilidade dentro da CN, este critério não foi usado para definir a estratégia de amostragem, porém a diversidade geográfica foi levada em conta na seleção dos acessos.

Existe muita discussão sobre quais são os descritores adequados para serem usados no desenvolvimento de uma CN – morfológicos, genéticos, moleculares, ecogeográficos. Uma CN poderia ainda ser definida a partir de dados de passaporte. Na verdade, bons dados de passaporte (origem geográfica, descritores ecológicos) provavelmente não têm rival no fornecimento de informação concisa e não dispendiosa sobre a estrutura da variabilidade genética de uma coleção (Galwey, 1995). Resultados relatados por Basigalup et al. (1995) e Malosetti & Abadie (2001) indicam que CN desenvolvidas usando este tipo de informação foram tão ou mesmo mais diversas que CN desenvolvidas usando outros tipos de informação. Um levantamento conduzido pelo IPGRI nos bancos de germoplasma visando avaliar vários aspectos envolvidos na implementação de CN demonstrou que geralmente mais de um critério é usado para estabelecer uma CN. Entre eles, os mais usados foram a origem geográfica dos acessos (95%), caracteres morfológicos (77%), grupos taxonômicos específicos e intra-específicos (63%), regiões ecogeográficas (34%) e marcadores genéticos (10%) (Brown & Spillane, 1999).

No presente caso os dois critérios – geográfico e sistema de cultivo – propostos e avaliados para estratificar a CN de arroz estão entre aqueles mais usados para estabelecer CN (Brown & Spillane, 1999), tendo em vista que o sistema de cultivo está associado a significativa diferenciação entre caracteres morfológicos.

#### Alocação da amostra

As estimativas do IS e seus IC a 0,90 foram 6,58 (6,27–6,93), 7,27 (7,14–7,41) e 8,09 (7,69–8,31) para os sistemas de cultivo em várzeas, terras altas e facultativo, respectivamente. Os intervalos de confiança não apresentam sobreposição, indicando que as pequenas diferenças observadas entre as estimativas do IS são significativas e coincidem com o que é comumente conhecido sobre a diversidade do arroz no Brasil, isto é, que as VT de arroz de terras altas têm uma diversidade genética mais elevada que aquelas de várzeas.

Um interessante aspecto do grupo das VT coletadas no Brasil é a existência de um terceiro grupo, o facul-

tativo, que apresenta um IS mais elevado que os outros dois. Este grupo merece estudos adicionais que esclareçam a sua origem e a sua relação com os outros grupos. Ele pode ser o produto de uma classificação errônea ou um verdadeiro grupo de transição entre os dois sistemas de cultivo principais. Se prevalecer esta última situação, o grupo pode ter evoluído localmente por meio de introgressões ou recombinações entre os dois grupos principais, constituindo uma importante fonte de variação para os programas de melhoramento. A mais elevada diversidade capturada pelo IS é decorrente de uma distribuição mais uniforme das freqüências dos estados dos descritores, não se observando uma predominância de freqüências extremas, no intervalo 0–1.

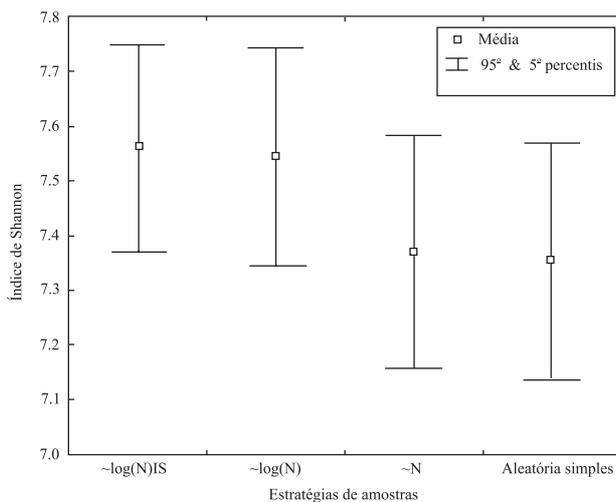
Entre as estratégias de amostragens avaliadas, aquelas com alocação proporcional ao logaritmo do número de acessos de cada estrato, ponderada ou não pelo IS, foram as que melhor atenderam o critério proposto (Figura 2). A partir deste resultado, a alocação da amostra de 300 acessos foi definida com a seguinte composição: 86 acessos de várzeas, 128 acessos de terras altas e 86 de acessos facultativos.

### Seleção dos acessos para compor a CN

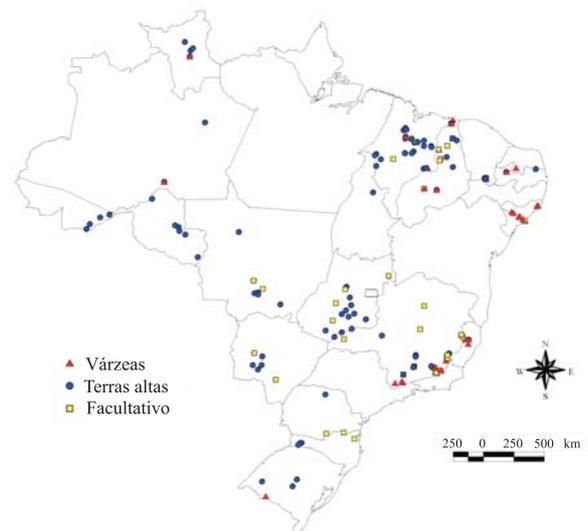
A importância do critério geográfico ofereceu subsídios para conduzir a seleção das VT levando em conta também a diversidade geográfica. Assim, procurou-se assegurar uma distribuição geográfica proporcional ao

número de VT coletadas em cada uma das unidades da federação, respeitado o número de acessos estabelecido para cada sistema de cultivo (Figura 3). As VT da CN têm uma distribuição espacial semelhante àquela da CI (Figura 1), mas com a representação dos sistemas de cultivo mais equilibrada, como consequência do próprio procedimento amostral adotado. Procurou-se representar também todos os tipos de grãos existentes nos acessos coletados em cada uma das unidades. Em adição, a experiência dos curadores sobre as características e uso das VT a partir de seus nomes comuns, além de aspectos ecogeográficos específicos das regiões, foram elementos usados para selecionar as VT típicas de cada região e que melhor representavam a diversidade deste estrato. Em algumas regiões onde a diferenciação ecológica tem sido descrita como relevante para os programas de melhoramento da cultura, este critério foi também usado (Burle et al., 2001). Os curadores selecionaram onze acessos das VT, para os quais não se dispunha das coordenadas geográficas dos locais de coleta. Neste grupo predominaram as antigas variedades comerciais, cuja denominação indicava algum aspecto importante para a cultura.

No caso das LCM, a seleção dos acessos para CN foi feita também para os sistemas de cultivo várzeas e terras altas, que correspondem a diferentes programas de melhoramento. A seleção dos acessos do estrato das



**Figura 2.** Estatísticas da distribuição empírica do IS para diferentes estratégias de amostragem das variedades tradicionais, a partir de dez descritores.



**Figura 3.** Locais de coleta das variedades tradicionais (VT) da coleção nuclear (CN) de arroz da Embrapa.

LCI foi conduzida procurando representar os países de origem e a relevância do material para os programas.

Uma CN de 550 acessos, 5,6% da CI, foi assim selecionada com a composição final apresentada na Tabela 2. Essa composição é diferente daquela escolhida inicialmente. Pequenas alterações foram feitas para atender as adaptações sugeridas pelos curadores.

É comumente estabelecido que uma amostragem de 10% dos acessos fornece um tamanho básico para uma CN. Esta generalização é proposta por Brown (1989), que estimou uma amostragem de 10% para reter no mínimo 70% dos alelos presentes na CI, com frequência superior a  $10^{-4}$ . Este resultado foi obtido com base no modelo de alelos neutros e sua teoria amostral aplicada aos alelos que Brown (1989) classificou como raros e amplamente dispersos, em contraponto aos outros três grupos de alelos: raros x localizados; comuns x dispersos; e comuns x localizados. No entanto, como o próprio Brown (1989, 1995) e Allard (1992) enfatizam, esse grupo de alelos não é necessariamente o que mais se destaca em uma coleção de germoplasma. Na verdade, alelos não-neutros que são selecionados natural ou artificialmente para adaptação a específicas condições ecológicas ou agrícolas (grupo dos alelos comuns e localizados) são potencialmente mais relevantes para os programas de melhoramento (Allard, 1992). A recomendação de Brown (1989) é que a representação desses alelos na CN seja maximizada, respeitando-se o número de acessos considerado aceitável para a CN e adotando-se uma adequada classificação da coleção. Deste modo, o tamanho final da CN de arroz do Brasil foi uma solução de compromisso entre essas considerações práticas e aqueles princípios teóricos.

**Tabela 2.** Composição dos 550 acessos da Coleção Nuclear (CN) de arroz no Brasil.

Estratos da coleção nuclear	Número	(%) <sup>(1)</sup>
Variedades tradicionais (VT)		
Várzeas	77	3,2
Terras altas	148	6,2
Facultativo	83	3,4
Linhagens/cultivares melhoradas (LCM)	94	2,7
Linhagens/cultivares introduzidas (LCI)	148	3,7

<sup>(1)</sup>Porcentuais em relação ao número de acessos em cada estrato na coleção inteira.

## Conclusão

A Coleção Nuclear de Arroz do Brasil está formada por 550 acessos (5,6% da coleção inteira), sendo 308 variedades tradicionais, 94 linhagens/cultivares melhoradas e 148 cultivares/linhagens introduzidas.

## Referências

- ALLARD, R.W. Predictive methods for germplasm identification. In: STALKER, H.T.; MURPHY, J.P. (Ed.). **Plant breeding in the 1990's**. Oxon: Wallingford, 1992. p.119-146.
- BASIGALUP, D.H.; BARNES, D.K.; STUCKER, R.E. Development of a core collection for perennial *Medicago* plant introductions. **Crop Science**, v.35, p.1163-1168, 1995.
- BROWN, A.H.D. The case for core collection. In: BROWN, A.H.D.; FRANKEL, O.H.; MARSHALL, D.R.; WILLIAMS, J.T. (Ed.). **The use of plant genetic resources**. Cambridge: University Press, 1989. p.136-156.
- BROWN, A.H.D. The core collection at the crossroads. In: HODKING, T.; BROWN, A.H.D.; VAN HINTUM, T.J.L.; MORALES, E.A.V. (Ed.). **Core collections of plant genetic resources**. New York: John Wiley & Sons, 1995. p.3-9.
- BROWN, A.H.D.; SPILLANE, C. Implementing core collections-principles procedures, progress, problems and promise. In: JOHNSON, R.C.; HODGKIN, T. (Ed.). **Core Collections for today and tomorrow**. Rome: IPGRI, 1999. p.1-9.
- BURLE, M.L.; FONSECA, J.R.; ALVES, R.B.N.; CORDEIRO, C.M.T.; FREIRE, M.S.; MELO, L.A.M.P.; RANGEL, P.H.N.; SILVA, H.T. **Caracterização de germoplasma de arroz de acordo com o ambiente de origem**: mapeamento em SIG. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2001. 51p. (Documentos, 63).
- CHANG, T.T.; BARDENAS, E.A. **The morphology and varietal characteristics of the rice plant**. Los Baños: IRRI, 1965. 40p. (Technical Bulletin, n.4).
- GALWEY, N.W. Verifying and validating the representativeness of a core collection. In: HODKING, T.; BROWN, A.H.D.; VAN HINTUM, T.J.L.; MORALES, E.A.V. (Ed.). **Core collections of plant genetic resources**. New York: John Wiley & Sons, 1995. p.87-198.
- GUIMARÃES, E.P.; SANT'ANA, E.P. Sistemas de cultivo. In: VIEIRA, N.R.A.; SANTOS, A.B.; SANT'ANA, E.P. (Ed.). **A cultura do arroz no Brasil**. Goiânia: Embrapa Arroz e Feijão, 1999. p.17-35.
- JACKSON, M.T.; PHAM, J.L.; NEWBURY, H.J.; FORD-LLOYD, B.V.; VIRK, P.S. A core collection for rice-needs opportunities and constraints. In: JOHNSON, R.C.; HODGKIN, T. (Ed.). **Core collections for today and tomorrow**. Rome: IPGRI, 1999. p.18-27.
- LI, R.; JIANG, T.B.; XU, C.G.; LI, X.H.; WANG, X.K. Relationship between morphological and genetic differentiation in rice (*Oryza sativa* L.). **Euphytica**, v.114, p.1-8, 2000.
- MALOSETTI, M.; ABADIE, T. Sampling strategy to develop a core collection of Uruguayan maize landraces based on morphologi-

- cal traits. **Genetic Resources and Crop Evolution**, v.48, p.381-390, 2001.
- MONTALBAN, R.; DESTRO, D.; SILVA, E.F.; MONTAÑO, J.C. Genetic base of Brazilian upland rice cultivars. **Journal of Genetics & Breeding**, v.59, p.203-209, 1998.
- PEREIRA, J.A. **Cultura do Arroz no Brasil**: subsídios para a sua história. Teresina: Embrapa Meio-Norte, 2002. 226p.
- PINHEIRO, B.S. Características morfofisiológicas da planta relacionadas a produtividade. In: VIEIRA, N.R.A.; SANTOS, A.B.; SANT'ANA, E.P. (Ed.). **A cultura do arroz no Brasil**. Goiânia: Embrapa Arroz e Feijão, 1999. p.117-148.
- RANGEL, P.H.N.; CRUZ, C.D.; VENCOSKY, R.; FERREIRA, R.P. Selection of local lowland rice cultivars based on multivariate genetic divergence. **Revista Brasileira de Genética**, v.14, p.431-453, 1991.
- RANGEL, P.H.N.; GUIMARÃES, E.P.; NEVES, P.C.F. Base genética das cultivares de arroz (*Oryza sativa* L.) irrigado do Brasil. **Pesquisa Agropecuária Brasileira**, v.31, p.349-357, 1996.
- SAS INSTITUTE (Cary, Estados Unidos). **Jackknife and bootstrap analyses**. Disponível em: <<http://ftp.sas.com/techsup/download/stat/jackboot.html>> Acesso em: jun. 2002.
- STEINMETZ, S.; MEIRELLES, E.J.L. Clima. In: VIEIRA, N.R.A.; SANTOS, A.B.; SANT'ANA, E.P. (Ed.). **A cultura do arroz no Brasil**. Goiânia: Embrapa Arroz e Feijão, 1999. p.59-88.
- VAN HINTHUM, T. The general methodology for creating a core collection. In: JOHNSON, R.C.; HODGKIN, T. (Ed.). **Core Collections for today and tomorrow**. Rome: IPGRI, 1999. p.10-17.
- VAUGHAN, D.A. Choosing rice germplasm for evaluation. **Euphytica**, v.54, p.147-154, 1991.
- VAUGHAN, D.A.; JACKSON, M.T. The core collection as a guide to the whole collection. In: HODKING, T.; BROWN, A.H.D.; VAN HINTUM, T.J.L.; MORALES, E.A.V. (Ed.). **Core collections of plant genetic resources**. New York: John Wiley & Sons, 1995. p.229-239.
- ZENG, Y.; LI, Z.; YANG, Z.; WANG, X.; SHEN, S.; ZHANG, H. Ecological and genetic diversity of rice germplasm in Yunnan, China. **Plant Genetic Resources Newsletter**, v.125, p.24-28, 2001.

---

Recebido em 19 de dezembro de 2003 e aprovado em 22 de setembro de 2004