

MOLECULAR ANALYSIS OF GENE BANKS FOR SUSTAINABLE CONSERVATION AND INCREASED USE OF CROP GENETIC RESOURCES

M. E. Ferreira

ferreira@cenargen.embrapa.br
Catholic University of Brasilia - UCB
EMBRAPA Genetic Resources and Biotechnology (CENARGEN)
Parque Estação Biológica 70770-900
Caixa Postal 02372 - Brasília – DF - BRAZIL

Summary

The effective use of crop genetic resources stored in genebanks by breeding programs is limited. The number of accessions deposited in genebanks, however, is continuously growing. Slow germplasm characterization has been pointed out as a major cause of this discrepancy. Molecular marker technology offers opportunities to increase the use of crop genetic resources deposited in genebanks. High throughput genotyping of germplasm accessions allows for the examination of genetic relationships and sampling of core collections representative of the allelic richness of the genebank. Core collections can be used for intensive phenotypic evaluation of traits of agronomic importance and re-sequencing of candidate genes associated with their control. Single nucleotide polymorphism variation between the accessions of the collection can be associated with phenotypic variation. The integration of genomic technology and characterization of germplasm banks will play an important role in the sustainable conservation and increased use of crop genetic resources.

Keywords

Core collection, microsatellite markers, genetic diversity, candidate genes, SNPs

The conservation, management and use of germplasm maintained in genebanks poses a number of challenges to the researchers dedicated to the investigation of plant genetic resources. Common problems include, for example, the development of strategies for sampling representative individuals in natural populations, the improvement of tools and technology for long term conservation or high throughput genetic analysis of an increasingly high number of stored accessions. Central to sustainable conservation is the knowledge of the genetic diversity present in a genebank. This is also key to the potential exploitation of genebanks by breeding programs. Therefore, the characterization of the accessions maintained in the collection and the examination of the genetic relationship between them is important for the sustainable conservation and increased use of crop genetic resources. Germplasm characterization of plant accessions deposited in genebanks has been limited and it is probably a major cause for the limited use of accessions in breeding programs.

Germplasm characterization refers to the observation, measurement and documentation of heritable plant traits in a collection. The resulting data allows the identification and classification of accessions, building a catalog of descriptors with embedded biological information that are essential to collection management or to direct use in agriculture. The characterization of plant germplasm, therefore, aims at the description and understanding of the genetic diversity of the organisms under study. Today, germplasm characterization has

been developed based mostly on morphological descriptors, agronomic descriptors (traits) and molecular marker technology.

The current characterization of plant germplasm collections relies strongly on morphological descriptors. Morphological descriptors are reliable, easy to study and relatively low cost to evaluate. However, the use of morphological descriptors present some limitations, such as (a) limited polymorphism, lowering the potential success of an extended classification approach, which would require a high number of descriptors in order to compensate the small number of morphotypes; (b) potential environmental influence on the phenotype, making even more complex the process of evaluation and information exchange. Here one should be careful with false positives when the environment affects specific morphotypes; (c) impact of a morphological descriptor in the viability of the individual.

Germplasm characterization based on agronomic traits, on the other hand, is particularly useful in crops of economic importance. The amount of data related to agronomic traits that is available by crop germplasm evaluation is limited. Due to reasons that include relatively high costs and difficulties of large scale experimental trials, the use of agronomic evaluation to characterize germplasm collections is far from the actual necessity of uncovering the phenotypes of agronomic interest in accessions of a collection. More should definitely be done in this area in order to stimulate a higher use of stored germplasm in breeding programs. A complete agronomic trait evaluation of crop germplasm in the next few years, though, seems to be practically impossible to achieve.

Recently, germplasm characterization based on molecular methods has experienced a great development. The methods that reveal sequence polymorphism in the DNA structure, known as molecular markers, fomented a revolution in the speed and quality of large scale plant germplasm characterization. The routines of germplasm characterization have relied increasingly on the use of such methods. Among all different classes of molecular markers available for evaluating genetic diversity, microsatellites or simple sequence repeats (SSRs) [1;2] are well known for their potentially high information content and versatility as molecular tools [3]. Hundreds of microsatellite markers have been developed for different species, having their chromosomal location and polymorphism levels determined. The use of fluorescently labeled microsatellite marker panels greatly increased the capacity of semi-automated genotyping of a large number of accessions, allowing for a faster and highly informative characterization of accessions deposited in genebanks.

The objective of this paper is to discuss how molecular analysis of genebanks can impact on sustainable conservation and increased use of crop genetic resources.

The Brazilian Upland Rice Genebank: a case study

Classical attempts to directly use accessions deposited in germplasm banks in breeding programs have been somehow limited to the identification of sources of genes of interest, such as resistance to plant pathogens or pests, and the transfer of those genes to elite material of the program. Linkage drag has usually restrained the breeder from the initiative of using accessions from the germplasm bank, since the advanced material of the breeding program is far more attractive than the inherent risk of using germplasm of unknown pedigree, performance or phenotypic adaptation. When the risk is taken, the accessions are usually submitted to a screening condition that reveals the presence of a gene of interest and typically a backcross program is initiated to transfer the gene to an elite line or cultivar. This procedure, however, is usually limited to traits of monogenic control. Complex traits require more elaborate methods, such as inbred backcross line [4] or AB-QTL mapping, which integrates the classical backcross method with linkage information based on molecular markers [5]. This procedure has been effective to demonstrate how to transfer quantitative trait loci (QTL)

identified in wild relatives deposited in genebanks to elite lines or cultivars of a cultivated species [6; 7].

More recently, genome sequencing opened the possibility of finding candidate genes for complex traits (candidate QTLs) in the genome of a crop species and, using the germplasm bank, identify genes of agronomic importance. Marker technology allied with detailed phenotypic characterization of germplasm banks can, therefore, be potentially useful in gene discovery. The use of this approach is based upon the integration of knowledge of genetic diversity of the germplasm bank, genomic information available for the chosen crop and intensive phenotyping of the trait of choice. A case study on how molecular analysis of genebanks can have an impact on sustainable conservation and increased use of crop genetic resources is discussed below using rice as a model.

Rice has one of the largest *ex situ* germplasm collections in the world, comprised of accessions of cultivated (*Oryza sativa* L. and *O. glaberrima* Steud.) and wild species. Because of its widespread use in the planet, *O. sativa* is considered the most important cultivated species of rice. Rice is a daily staple food in the diet of billions of people in the world, including millions of Brazilians. The production and consumption of rice in Brazil is comparable to figures observed in some Asian countries. About 60% of the rice production in the country is irrigated (*O. sativa* spp. *indica*), while upland rice (*O. sativa* spp. *japonica*) accounts for 40% of the total. A national genebank of *japonica* rice is maintained by Embrapa. This genebank contains approximately 4,000 accessions, including landraces collected in villages and isolated rural areas of the country, where rice has been cultivated since its introduction in Brazil centuries ago. This germplasm is a source of genes that control traits of economic importance, such as drought tolerance and resistance to plant pathogens. Most of these rice accessions have not yet been studied in depth. The phenotypical and genotypical characterization of the collection is just beginning.

The traits of economic interest in plants almost always involve a complex genetic control, generally determined by various genes and a strong interaction with the environment. The genetic control of drought tolerance in plants, a trait of great importance for the genetic improvement of rice and other grass species, for example, is typically quantitative. Several QTLs showing different effects on the phenotypic variation are involved in the control of drought tolerance in rice. Some QTLs might have a strong or major effect on the phenotype, others might have a small but still significant effect on the trait. The genetic and physical mapping of the genome, based on the use of molecular markers and cloned segments of bacterial artificial chromosomes (BACs), has allowed for great advances in the understanding of the genetic control of quantitative traits, at times making it possible to isolate the genes associated with quantitative trait control by the positional cloning approach [8]. This approach overcame the classic problem of lack of information regarding the gene product of loci of economic interest (usually a quantitative trait), so useful when reverse genetics is possible. In other words, the lack of knowledge of the protein related to a trait could be circumvented in gene isolation studies using detailed genetic mapping of the region around the locus of interest, followed by sequencing selected clones of large genomic insert libraries and identification of the desired gene by complementation studies. Genetic mapping allows for the break down and examination of the genetic control of a complex trait, making it possible to identify the number of genes that are involved in the expression of a certain trait, locate these genes or gene regions on the chromosomes, measure their impact on the phenotypic variation and understand gene interaction in the manifestation of the trait.

An alternative to positional cloning for the isolation of genes controlling quantitative traits has been advocated with the availability of massive DNA sequence data obtained for some species by whole genome sequencing projects. The approach relies on the power of

association tests to verify the correlation of sequence variation at candidate genes with phenotypic variation for the trait of interest [9]. Recent developments in the area of human genetics, particularly in the identification of factors that cause genetic diseases, have been based on genetic association tests. These tests have been used for many years in genetic studies, especially in classical medical genetics, but the genomic information gathered recently affords opportunity for the development of new strategies of analysis. The association tests seek to identify nucleotide variation observed at a genomic region with phenotypic variation for a trait, leading to the identification of DNA regions which control the phenotypes of interest. These tests do not rely on segregating populations, but rather on the evaluation of allelic variation observed in the prospective loci in relation to the phenotype of interest in natural populations. In view of the developments of genomic technology, allelic variations are currently detected by DNA sequencing, characterized as mutations and small insertions and/or deletions (indels) known as SNPs (“Single Nucleotide Polymorphisms”) [10].

The rice genome was sequenced in its entirety in 2002 [11; 12], with a great part of the sequences deposited in public databases over the last few months. It is logical to assume that the use of bioinformatics and the re-sequencing of a great number of genes located on some rice chromosomes can be an effective way to correlate the allelic variability detected at these loci with the phenotypic variability for resistance to abiotic stresses. Therefore, re-sequencing of candidate genes in a set of germplasm accessions and phenotypic evaluation of the same set for drought tolerance could be an interesting approach to discover genes involved in the control of traits of economic importance in grass species. Candidate genes found in regions in the vicinity of QTLs for response to the abiotic stresses are immediate targets for drought tolerance gene isolation in this species. Genotypic and phenotypic analysis can be carried out with accessions sampled in germplasm collections, especially landraces of upland rice showing genetic variability for drought tolerance. For this purpose, the following steps should be considered:

(a) genetic maps making ample use of molecular markers (especially microsatellite or STRs) are constructed in order to locate genomic regions (QTLs) associated with the control of drought tolerance in grasses. Several populations of grass species (rice, maize and sorghum) segregating for alleles associated with the control of drought tolerance are then examined. Microsatellite markers are used to cover the entire rice genome by means of large-scale genotyping and mapping. Genotyping is performed using high throughput equipment such as automatic DNA sequencers. By multiplexing many PCR products into a single lane on a polyacrylamide gel, a large increase in genotyping throughput is achieved. Multiplexed PCRs greatly increase the amount of collected information in a single reaction. The segregating populations are then submitted to extensive phenotypic evaluation through different bioassays and replicated experiments in the field. The genotypic and phenotypic information is used to identify the genomic regions which control the traits of interest (Quantitative Trait Loci - QTL). The QTL x environment interactions are estimated using the data obtained in different locations. In addition, anchor-markers used for mapping purposes on different grass species are employed in the construction of genetic maps of rice and other grass species, allowing the information obtained in the rice genome to be transferred to these other species via syntenic map analysis, and vice-versa [13]. Synteny analysis, therefore, provides a comparative look at the genetic control of drought tolerance in the genome of different species.

(b) The genomic regions delimited by QTL analysis are studied in detail using the rice genome sequence database. The microsatellite markers that flank a specific drought tolerance QTL in a chromosome are used to delimit in the physical map the BAC clones that potentially harbor the gene(s) associated with the trait. All ORFs (open reading frames) and controlling

segments in the region are re-annotated and candidate genes are selected for detailed sequence analysis. Genotypic variation (SNPs – single nucleotide polymorphisms) are detected after sequence alignment.

(c) The candidate genes selected in the vicinity of a QTL are re-sequenced in a sample of rice variety accessions selected from the genebank. This sample represents a core collection [14] of the germplasm studied, i.e., it should represent the maximum possible level of the genetic diversity existing in a complete germplasm bank.

(d) In depth phenotypic evaluation of a core collection is undertaken to access diversity of agronomic trait (drought tolerance) present in the upland rice genebank. The measurement of drought tolerance is based on indirect observations of morphological, physiological, biochemical and agronomical responses of the plant variety to abiotic stress. Root morphology, for example, is usually dissected in many ways to provide clues about the ability of the plant to survive and compete under drought conditions. In this case, the root system architecture, as measured by the number and arrangement of secondary roots, dry matter weight, root abundance in different layers of soil, root penetration ability, among others, are quantified as an indirect way to access drought tolerance. These measurements can be statistically treated as components of drought tolerance. It is also very common to use grain yield under drought conditions to compare genotypic responses of different plant varieties to water deficit stresses. Enzyme activity and metabolic bioassays can also offer biochemical means of studying drought tolerance in laboratory conditions.

It is expected that significant associations between sequence polymorphism and phenotypic variation will indicate and allow the isolation of genes that control drought tolerance in rice. This information could be readily transferred to other grass species such as maize and sorghum. The breeding programs will certainly benefit from the data based on the molecular analysis of genebanks, stimulating an increase in use of stored genetic resources. The integration of genomic technology and characterization of germplasm banks will play an important role in the sustainable conservation of genebanks.

THE ROLE OF BIOTECHNOLOGY
Villa Gualino, Turin, Italy – 5-7 March, 2005

REFERENCE LIST

- [1] Tautz, D (1989) Hypervariability of simple sequences of a general source for polymorphic DNA markers. *Nucleic Acids Research*, 17:6463-6471
- [2] Weber, RD and May, PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* 44:388-396
- [3] Ferreira, ME and Grattapaglia, D (1996) Introdução ao uso de marcadores moleculares em análise genética. 220pp. Embrapa -SPI
- [4] Wehrhahn, C and Allard, RW (1965) The detection and measurement of the effects of individual genes involved in the inheritance of a quantitative character in wheat. *Genetics* 51:109-119
- [5] Tanksley, SD and Nelson, JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor Appl Genet* 92:191-203
- [6] Xiao, J; Grandillo, S; Ahn, SN; McCouch, SR; Tanksley, SD; Li, J and Yuan, L (1996) Genes from wild rice improve yield. *Nature* 384:223-224
- [7] Brondani, C; Rangel, PHN and Ferreira, ME (2002) QTL mapping and introgression of yield related traits from *Oryza glumaepatula* to *O. sativa* using microsatellite markers. *Theoretical and Applied Genetics* 104:1192-1203
- [8] Tanksley SD; Ganai, MW and Martin, GB (1995) Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. *Trends in Genetic* 11(2):63-68
- [9] Risch, NJ (2000) Searching for genetic determinants in the new millennium. *Nature* 405 15:847-856
- [10] Collins, FS; Guyer, MS and Chakravarti, A (1997) Variations on a theme: cataloging human DNA sequence variation. *Science* 278, 1580-1581
- [11] Goff, SA et al. (2002) A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*) *Science* 296 5 92-100
- [12] Yu, J. et al. (2002) A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *indica*). *Science* 296 5 79-92
- [13] Moore, G; Devos, KM; Wang, ZM and Gale, MD (1995) Grasses, line up and form a circle. *Curr. Biol.* 5, 737-739
- [14] Frankel, OH and Brown AHD (1984) Current plant genetic resources – a critical appraisal. In.: *Genetics: new frontiers*, Vol. 4, Oxford & IBH Publishing Co., New Delhi, p. 1-11