S. Kresovich · A. K. Szewc-McFadden · S. M. Bliek · J. R. McFerson

**Abundance and characterization of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (rapeseed)**

Received: 20 August 1994 / Accepted: 3 February 1995

**Abstract** A size-fractionated library of *Brassica napus* L. (rapeseed), composed of 15,000 clones, was screened for the presence of GA-, CA-, and GATA-simple-sequence repeats (SSRs). GA-SSRs were four- and five-fold more abundant than CA- and GATA-SSRs, respectively, and present at a frequency of approximately one SSR for every 100 kb of DNA. Following the sequencing of 124 positive clones, primer pairs were designed and evaluated for seven selected SSRs. Products were amplified in an array of individuals of *B. napus*, *B. oleracea* and *B. rapa*, demonstrating that the seven SSRs were conserved among species. Two SSRs were polymorphic. Among 11 accessions, the dinucleotide (GA)-repeat, B.n.9A, yielded 12 fragments, while the tetranucleotide-repeat (GATA), B.n.6A2, revealed two fragments. Automated, fluorescence-based detection of polyacrylamide gels has been employed to simultaneously increase throughput, reduce unit cost, improve analytical resolution, and expedite data acquisition of SSR analysis. Though initial financial investment and technical capabilities may prevent some from directly employing our documented approach, SSR analysis warrants further investigation as a tool in genetic studies for enhancing both the conservation and utilization of genetic resources.

**Key words** Genetic analysis · Fluorescence-based detection · STR · Microsatellite DNA · Multiplex PCR

---

**Introduction**

Effective conservation of plant genetic resources requires a comprehensive understanding of genetics. This understanding attempts to integrate organizational levels (molecules, chromosomes, cells, individuals, populations, species, and ecosystems) as well as to firmly establish the relationship of genotype to phenotype. With increasing demand by scientists for ready access to useful and representative genetic resources, the present is a challenging yet opportune time to enhance our knowledge of the genetics and genome organization of crop species for improving conservation planning and practice. In particular, curators of plant genetic resources require a more accurate measurement of the following genetic parameters: (1) identity: the determination that an accession in the genebank is catalogued correctly and true to type; (2) relationship: the degree of relatedness among genotypes in an accession or among accessions in a collection; (3) structure: the amount of genetic variation present, and how it is partitioned among individuals, accessions, and collections; and (4) location: the presence of a desired gene/gene complex in a specific accession and of a the mapped site of a desired DNA sequence on a particular chromosome in an individual (Kresovich et al. 1992).

The development and application of molecular genetic markers provide the opportunity to reveal DNA sequence polymorphisms useful to discriminate genetic variation among individuals and within populations. As with applications in plant breeding (Ragot and Hoisington 1993), not all molecular markers are suitable for use in all applications of plant genetic resources conservation. Both restriction fragment length polymorphisms (RFLPs) and random amplified polymorphic DNA polymorphisms (RAPDs) have been used extensively to characterize plant genetic resources ( Tanksley et al. 1989; Rafalski et al. 1991). Though useful in selected applications, both RFLPs and RAPDs exhibit limitations and can be misused, misrepresented, or misinterpreted. For example, in certain situations it may be difficult to infer molecular genotype from...
phenotype, to detect the state (heterozygotic or homozygotic) at a locus, or to identify allelic relationships. Also, the feasibility of employing either technique to resolve questions that require a large number of samples (>1000) to be profiled is questionable based on technical (discriminatory ability, sensitivity, reproducibility, and need for further genetic analysis and diagnostics development) as well as operational (protocol characteristics, time, and cost) considerations.

To examine the use of molecular genetic markers for improved conservation of plant genetic resources, we have focused our research efforts on marker development, integration of advanced technology, and improved analysis of generated data. This effort has been founded on: (1) the application of polymerase chain reaction (PCR)-based DNA amplification, (2) the utilization of repetitive sequence DNA as markers, and (3) fluorescence-based automated DNA detection (Ziegel et al. 1992). Among the classes of repetitive DNA sequences that have proven amenable for PCR amplification is the simple-sequence repeat (SSR) (Weber and May 1989), a variety of di-, tri-, tetra-, and penta-nucleotide tandem repeats that are dispersed throughout the genome. Based on studies in humans, the advantages of an SSR locus include: (1) its great abundance and distribution, (2) its ability to be “tagged” in the genome, (3) its high level of polymorphism (Fregeau and Fourney 1993; Kimpton et al. 1993), (4) its clearly defined genetic definition (Queller et al. 1993), and (5) its ease of detection via automated systems (Morgante and Olivieri 1993; Rafalski and Tingey 1993). However, technical drawbacks that currently restrict the use of this type of marker on a large scale are also apparent. Cregan (1992) highlighted technical concerns about the development and application of SSR markers, including the costly and time-consuming nature of the identification of polymorphic loci and the possible detection problems incurred when alleles differ only slightly in length.

At present, information on SSR markers in plant species is limited (Condit and Hubbell 1991; Akkaya et al. 1992; Lagercrantz et al. 1993; Senior and Heun 1993; Thomas and Scott 1993; Thomas et al. 1993; Saghai-Maarooof et al. 1994; Wang et al. 1994; Rongwen et al. 1995). Moreover, primary applications of SSR markers have focused on the use of these landmarks for genome mapping. Like other investigators, we have been concerned about the potentially great investment needed for the generation of 1000–1000s of SSR markers needed for mapping. However, the utilization of SSR markers to resolve curatorial needs may be far less ambitious in both scope and size, e.g., a goal of 20–30 highly polymorphic loci (with polymorphism information content values of 0.5) per crop species. The objective of this initial phase of research was (1) to establish measures of abundance, and (2) to characterize selected di- and tetra-nucleotide SSRs, isolated from a size-fractionated genomic library of Brassica napus L., for ultimate use in resolving genetic identity, relatedness, and structure.

### Materials and methods

#### Construction of small insert genomic libraries

Genomic DNA was isolated from leaf tissue of B. napus L. ‘Jet Neuf’ using a modified CTAB technique with subsequent purification employing a cesium chloride gradient (Sambrook et al. 1989). Following isolation, DNA was digested with TaqI and electrophoresed on a low-melting-temperature agarose gel (Nusieve GTG, FMC). The 100–500-bp fragments were excised and isolated from the gel via agarase digestion (Dumais et al. 1987). Selected fragments were ligated to CiaI-digested pGEM-7Zf(+) (Promega) that had been dephosphorylated with calf intestinal alkaline phosphatase (Sambrook et al. 1989). Recombinants were transformed in E. coli JM109 cells (Promega) using a standard protocol provided by the supplier.

Screening of libraries for clones containing selected di- and tetra-nucleotide repeats

Colonies were transferred onto Duralose UV filters (Stratagene) following protocols provided by the supplier. Two dimeric repeats, (GA)$_{10}$ and (CA)$_{10}$, and a single tetrameric repeat, (GATA)$_{4}$, oligonucleotide (supplied by Applied Biosystems, Inc.) were labeled with $^{32}$P by a 5' Terminus Labeling System (Gibco, BRL). The selection of a core repeat sequence was based on surveys of databases for the abundance and informativeness of loci in human and plant species, particularly B. napus. The oligonucleotides were used as probes to screen 15,000 clones each from the B. napus genomic library, according to the protocol supplied by Stratagene. Hybridizations and subsequent washes were conducted at 45°C under high stringency. Each filter was screened twice and only positives colonies detected in both screens were sequenced.

#### Sequencing of positive clones

Plasmid DNA of GA$_{4}$, CA$_{4}$, and GATA-positive clones was isolated using a modified phenol-chloroform protocol supplied by Applied Biosystems, Inc. Clones were sequenced from M13 sequencing primer sites of the pGEM 7Zf(+) using either the Taq DyeDeoxy Terminator Cycle Sequencing Kit or the Prism Sequenase Terminator Double-Stranded DNA Sequencing Kit on an upgraded ABI model 370 DNA Sequencing System (Applied Biosystems, Inc.).

#### Primer design

Primers complementary to the flanking regions of the repeats (Table 1) were designed using Designer PCR (Research Genetics). Two key criteria in primer pair formulation included a primer $T_m$ of 65–66°C and a primer pair $T_m$ variance of 1.0°C. These stringent criteria were employed to preclude problems with spurious banding patterns generated during amplification. Because one of our operational goals is to perform multiplex PCRs with SSRs, a uniform annealing temperature across all SSRs was essential. Also, primer pairs were generated to produce amplified DNA fragments between approximately 96 and 247 bp in length. Primers were synthesized by Operon Technologies, Inc. or on an ABI model 392 nucleic-acid synthesizer and were purified by either HPLC or by OPC purification cartridges (Applied Biosystems, Inc.), respectively.

#### PCR profiling

A test array of 11 Brassica accessions was selected to represent a range of diversity within the U.S. National Plant Germplasm System Brassica collections. Three species, B. napus, B. oleracea, and B. rapa, were represented in the array (Table 2). Ten plants of each accession were grown in the greenhouse for DNA isolation. Tissue