A study of genetic diversity among *Brassica napus* and *Brassica juncea* germplasm collections using Simple Sequence Repeat (SSR) molecular markers

J Batley\(^A\), A Vecchies\(^A\), R Mogg\(^B\), J Bond\(^B\), N Cogan\(^A\), C Hopkins\(^A\), N Gororo\(^C\), S Marcroft\(^C\), J Forster\(^A\), G Spangenberg\(^A\) and D Edwards\(^A\)

\(^A\) Plant Biotechnology Centre, Department of Primary Industries, La Trobe University, Bundoora, 3084, Australia
\(^B\) Centre for Ecology and Hydrology, CEH Dorset, Winfrith Technology Centre, Dorchester, DT2 8DH, UK
\(^C\) Agriculture Victoria, Victorian Institute for Dryland Agriculture, Horsham, Victoria 3401, Australia.

**Abstract**

*Brassica* species represent a broad range of crops. This reflects the high degree of genetic diversity and related phenotypic plasticity. An understanding of the genetic basis of *Brassica* diversity aids both breeding and the discovery of rare alleles and traits. Single Sequence Repeat (SSR) molecular markers are popular tools for assessing genetic diversity due to their high degree of polymorphism and their transferability across related species. We have applied public SSR markers to genotype international germplasm collections of *B. napus* and *B. juncea* as well as commercial *Brassica* lines grown in Australia. Results illustrate the genetic relationship between these lines and also provide a measure of the level of genetic uniformity within lines.

**Keywords:** microsatellites, perfect alleles, candidate genes, genotyping

**Introduction**

Estimation of genetic diversity in a crop species can assist in the evaluation of germplasm collections as potential gene pools to improve the performance of cultivars. Genotyping accessions and assessing the level of genetic diversity within a germplasm collection can be of use to plant breeders in a number of ways. Molecular markers can assist in the choice of parents for mapping populations, marker-assisted selection and back-crossing schemes. When selecting parents for a linkage mapping population, it is desirable to choose parents which are as diverse as possible. This will increase the probability of polymorphic alleles being detected at as many loci as possible, which are essential when determining linkage distance between loci. Furthermore, genotyping individual and cultivars within germplasm collections allows breeders to narrow the search for new alleles at a locus of interest. This leads to the identification of 'perfect' alleles from candidate genes in individuals or cultivars, which can be used to introduce desirable traits into new varieties. By detecting correlations between desired phenotypic traits and markers in mapping populations and searching for the allele associated with the desired trait in the germplasm collection, perfect alleles may be detected in a line which can then be used as a source of material for new and improved cultivars. Genotypic information from germplasm collections can also be used to assign lines and populations to heterotic groups, monitor changes in allele frequencies in cultivars or populations and therefore assess divergence of identical lines in different countries, to study the evolutionary history of wild relatives, to verify pedigrees and fill in the gaps in incomplete pedigree or selection history (Warburton and Hoisington 2001). Therefore, to improve the characterisation and conservation of crop genetic resources, the assessment of genetic identity, relatedness and structure among cultivars, varieties and populations becomes a priority (Phippen et al. 1997). The use of molecular markers can then facilitate *Brassica* breeding by means of marker assisted selection to improve agronomically important traits.

**Simple sequence repeats**

Molecular markers, such as Simple Sequence Repeats (SSRs) are a powerful tool with which to generate information contained within germplasm collections and are becoming increasingly important in plant variety testing (Donini et al. 2000). SSRs, also known as microsatellites, are short stretches of DNA sequence occurring as tandem repeats of mono-, di-, tri-, tetra-, penta- and hexa-nucleotides. Previous studies of SSRs in crop species, including SSRs in *Brassicas* (Charters et al. 1996; Saal et al. 2001;
SzeweMcFadden et al. 1996) showed that they are a valuable tool for characterising germplasm. This is due to their properties of genetic co-dominance, abundance, dispersal throughout the genome, multi-allelic variation, high reproducibility and amenability to high throughput automated allele detection and sizing (Schlötterer 1998). For example, the utility of SSRs for distinguishing between B. napus varieties has recently been demonstrated by Tommasini et al. (2003), in which ten commercial European varieties were discriminated using 15 SSR markers.

The ability to conserve genetic resources and utilise genetic diversity within the Brassica breeding programs requires knowledge of genetic diversity both within and between Brassica lines. The objectives of the present study were to characterise an international germplasm collection, consisting of over 100 B. napus and B. juncea lines (currently maintained by Agriculture Victoria), using genomic and EST (expressed sequence tags) derived SSR markers and to develop a diversity panel for selection of markers to assist Australian canola varietal production. The markers developed from expressed sequences can not only be used for diversity studies, molecular mapping, cultivar identification and marker-assisted selection, but also for identifying gene-trait relations in Brassicas. We describe here the development of novel EST-SSR markers and their application, along with 15 genomic SSRs, to assess the genetic diversity within and between the Brassica lines in this germplasm collection. Further genetic diversity analysis has also been performed on related germplasm from 17 lines, maintained in the UK along with the 30 B. napus cultivars grown commercially in Australia.

**EST SSR discovery and characterisation**

Using an automated script developed in-house, primer pairs have been designed to more than 1000 unique SSR loci, discovered from more than 20000 Expressed Sequence Tags (ESTs), from cDNA libraries constructed from Brassica species (B. nigra, B. juncea and B. napus) (Edwards et al. 2003). Two hundred and seventy of these SSR markers, representing a variety of SSR repeat motifs and lengths, have been characterised in 8 samples of B. napus (A and C genomes), B. juncea (A and B genomes) and B. carinata (B and C genomes). Of these 270 markers, 218 (80.2%) are polymorphic, amplifying between 2 and 11 alleles (mean = 3.6 alleles per locus). Forty (14.7%) of the primer pairs amplify multiple loci, either from multiple genomes of Brassica or from different members of a gene family.

**Assessment of genetic diversity of Brassica germplasm collections**

Forty-three B. juncea (Indian mustard) and 86 B. napus (oilseed rape and turnip) accessions were obtained from all canola and mustard-producing countries of the world to form the germplasm collection. Between two and twelve plants from each accession were genotyped with fifteen SSRs from BrassicaDB (http://jic-bioinfo.bbsrc.ac.uk/BrassicaDB/) along with 10 novel EST markers, for an initial assessment of relationships between accessions, and diversity within and between lines. DNA was extracted from all the remaining plants in the accessions (up to 20) for further diversity analysis. Considerable variation was observed within accessions. Less than 5% B. juncea and 10% B. napus showed identical genotypes between all individuals within a line, reflecting the outbreeding nature of the species and the country of origin. Thirty commercial Australian B. napus cultivars also showed significant variation within some of the lines. While the majority of varieties grouped together, diversity within other lines reflected the complexity of the background pedigree.

Six of the genomic SSR markers were genome specific (Figure 20) and five amplified all three genomes, representing a conserved region within all three genomes. Polymorphism levels were similar between the two species, with an average of 11.1 alleles/locus (4–31) detected in B. napus, compared to a mean of 10.6 alleles/locus (2–24) in B. juncea. EST-SSRs showed less polymorphism than genomic SSRs with an average of 6.5 alleles/locus, compared to 12.6 alleles/locus with genomic SSRs. Genetic diversity was apparent between geographic regions as well as between individuals within the same lines. B. juncea accessions largely grouped according to geographical origin with most regions forming distinct groups. The Australian lines did not group together. This may reflect the recent initiation of the Australian B. juncea canola breeding program and the sources of material for this program. B. napus accessions also grouped according to geographic origin and breeding history.

No new alleles were detected in feral B. napus samples. Phenetic analysis indicates likely origins for some of the feral lines. For example one feral plant clustered with the cultivar Falcon, indicating this variety is the likely source, while two other feral plants grouped with the French varieties, again indicating the likely source. Identical accessions maintained in germplasm collections in Australia and the United Kingdom showed varying genotypes, with differences at three genomic SSR loci and one EST-SSR locus, reflecting the differences in maintenance of the collections in each country.
Figure 20: EST Microsatellite locus demonstrating amplification specificity in the B and C genomes of *Brassica*, (a) amplification in *B. napus* (B genome), (b) amplification in *B. carinata* (B and C genomes) and (c) amplification in *B. juncea* (B genome). No amplification was observed from the A genome.

Conclusions

The outcomes from this project are the selection of a diversity panel for the assessment of further *Brassica* SSR and SNP markers to be used in current and future canola breeding programs, along with information on the genome specificity of SSR markers within the amphidiploid *Brassica* species. This information will greatly assist the application of molecular markers to Australian canola variety production and identify patterns of variation which can be used to design further breeding programs and conserve *Brassica* genetic resources.

References


