Genetic and morphological diversity in the *Brassicas* and wild relatives

Robert Redden[1], Megan Vardy[2], David Edwards[2], Harsh Raman[3] and Jacqueline Batley[4]
[1]Department of Primary Industries, Victorian Institute for Dryland Agriculture, Private Bag 260, Horsham Vic 3401, Australia
[2]School of Land, Crop and Food Sciences, University of Queensland, Brisbane, QLD 4072, Australia
[3]EH Graham Centre for Agricultural Innovation (an alliance between NSW Department of Primary Industries and Charles Sturt University), Wagga Wagga Agricultural Institute, PMB, Wagga Wagga, NSW 2650, Australia
[4]ARC Centre of Excellence for Integrative Legume Research and School of Land, Crop and Food Sciences, University of Queensland, Brisbane, Australia j.batley@uq.edu.au

ABSTRACT

Germplasm collections are valuable gene pools providing diverse genetic material that may be applied for the improvement of cultivars and advanced agronomic productivity. An assessment of genetic and genomic diversity within these collections can be used to assign lines and populations to diverse groups; to study the evolutionary history of wild relatives; to verify pedigrees and fill in the gaps in incomplete pedigree or selection history, to monitor changes in allele frequencies in cultivars or populations and to help narrow the search for new alleles at loci of interest. Towards the application of genetic diversity analysis to *Brassica* germplasm improvement, we are characterising genetic diversity between 28 species and wild relatives of *Brassica* from an international germplasm collection, using 60 Simple Sequence Repeat (SSR) molecular markers. Results on the SSR polymorphism and phylogenetic relationship among species and wild relatives of *Brassica* will be presented.

Key words: Genetic diversity – Simple Sequence Repeats (SSRs) – wild *Brassicas* – germplasm collection

INTRODUCTION

The *Brassica* genus contains many agronomically important crop species with a range of adaptation for cultivation under varied agroclimatic conditions. Of these, oil seed rape (*Brassica napus* L. ssp *oleifera*), is an important oilseed crop grown in moist and cool climates throughout the world. The genomic relationship of the six cultivated *Brassica* species of economic importance has been determined (U, 1935). The genomes have been denoted as the A, B and C genomes, with three monogenomic diploid species, namely *B. rapa* syn. *campestris* (AA, 2n=20; chinese cabbage and turnip), *B. nigra* (BB, 2n=16; black mustard) and *B. oleracea* (CC, 2n=18; cabbage, brussel sprouts, cauliflower and broccoli). The remaining three cultivated species, *B. napus* (AACC, 2n=38; canola, swede), *B. carinata* (BBCC, 2n=34; Ethiopian mustard), and *B. juncea* (AABB, 2n=36; Indian mustard) are amphidiploid hybrid taxa, evolving through hybridisation between the monogenomic diploid species.

Crop domestication frequently results in limited genetic diversity. Wide genetic diversity for improved crop productivity is largely unexplored in the wild relatives of *Brassica* crops and there is not a clear association between genetic distance and taxonomic classification within the Brassicaceae (Gomez-Campo and Prakash, 1999; Warwick et al., 2000). Even though there has been extensive morphological phenotyping and the Brassicaceae include the model system *Arabidopsis thaliana*, relatively little is known of the evolutionary history of the Brassicaceae (Beilstein et al., 2006). Genotyping individuals and cultivars within germplasm collections can be used to provide insight on the evolutionary history of the family, as well as to help breeders narrow the search for new alleles at loci of interest and assist in the identification of 'perfect' marker alleles from candidate genes that can then be introduced into new varieties along with their associated desirable traits.

The present study includes the use of SSRs, also known as microsatellites, to understand the genetic relationship among wild and cultivated Brassicaceae accessions obtained from an international germplasm collection, representing 28 species. It has been reported previously that SSRs are a valuable tool for characterising germplasm in various crop species, including *Brassica* species (Saal et al. 2001; Lowe et al. 2004; SzewcMcFadden et al. 1996). This is due to their properties of high reproducibility, co-dominance, abundance, wide dispersal throughout the genome and multi-allelic variation (Powell et al. 2006). Improvements in molecular marker technology offer a potential improvement in the efficiency and affordability of variety testing. Furthermore, due to the economic importance of cultivated *Brassica* species, large investments have been made in the development of *Brassica* SSRs, many of which are available to the scientific community, (http://www.brassica.info/ssr/SSRinfo.htm). In addition, the SSRs developed from expressed sequences, produced during gene discovery projects are gene specific molecular markers (Hopkins et al., 2007). The objectives are to determine the extent of genetic diversity between *Brassica* species and their wild relatives and to estimate phylogeny in the Brassicaceae, correlate this with morphological and inter-specific hybridisation data and suggest strategies to broaden the germplasm base among cultivated *Brassica* breeding programs.

MATERIALS AND METHODS

Selection of Plant material

Fifty lines from the Brassicaceae, comprised of 28 species, from the Australian Temperate Field Crop Collection (ATFCC) held at DPI-Horsham, Australia, were selected. These were comprised of the cultivated species and a selection of their primary, secondary and tertiary progenitors from both the rapa/oleracea and nigra lineages (Table 1). Plants were grown in a glasshouse, employing recommended cultural practices. Three week old seedling tissue was harvested and stored at –20°C until needed.

Table 1. Description of samples studied, including taxon, genebank source country, whether samples are wild or cultivated and information on their relationship within the Brassicaceae

Taxon	Genebank	Sample type	Gene Pool
B. barrelieri	Spain	wild	secondary[2]
B. carinata	Ethiopia	landrace	allotetraploid
B. carinata	Pakistan	landrace	allotetraploid
B. deflexa	Turkey	wild	secondary[2]
B. deflexa ssp. leptocarpa	Iran	wild	secondary[2]
B. incana	Italy	wild	primary[2]
B. incana	Italy	wild	primary[2]
B. juncea	China	advanced cultivar	allotetraploid
B. juncea	Pakistan	Breeder's Line	allotetraploid
B. maurorum	Algeria	weed	primary[1]
B. montana	Italy	wild	primary[2]
B. montana	France	wild	primary[2]
B. nigra	India	landrace	diploid progenitor
B. nigra	Ethiopia	landrace	diploid progenitor
B. nigra var. abyssinica	Ethiopia	traditional cultivar/landrace	diploid progenitor
B. oleracea var. gongylodes	Austria	unknown	diploid progenitor
B. oxyrrhina	Australia	wild	secondary[1]
B. rapa ssp. campestris	Algeria	landrace	diploid progenitor
B. rapa ssp. chinensis	China	traditional cultivar/landrace	diploid progenitor

Table 1 (Cont). Description of samples studied, including taxon, genebank source country, whether samples are wild or cultivated and information on their relationship within the Brassicaceae

Taxon	Genebank	Sample type	Gene Pool
B. rapa ssp. nipposinica	Japan	traditional cultivar/landrace	diploid progenitor
B. ruvo	Italy	advanced cultivar oleraceae	primary [2]
B. souliei ssp. amplexicaulis	Australia	unknown	secondary[1]
B. spinescens	Algeria	wild	primary [1]
B. tournefortii	Australia	wild	secondary[1/2]
B. tournefortii	France	wild	secondary[1/2]
B. tournefortii	Spain	wild	secondary[1/2]
B. tournefortii	India	wild	secondary[1/2]
Capsella bursa-pastoris	Australia	wild	tertiary
Crambe abyssinica	Former Soviet Union	unknown	primary or secondary [1]
Crambe abyssinica	Sweden	unknown	primary or secondary [1]
Crambe abyssinica	Poland	unknown	primary or secondary [1]
Crambe abyssinica	Canada	unknown	primary or secondary [1]
Crambe abyssinica	Ethiopia	unknown	primary or secondary [1]
Diplotaxis tenuifolia	Australia	wild	secondary[1/2]
Eruca sativa	Pakistan	unknown	secondary[2]
Eruca sativa	Turkey	unknown	secondary[2]
Guizotia abyssinica	India	wild	tertiary
Guizotia abyssinica	India	wild	tertiary
Guizotia abyssinica	South Africa	wild	tertiary
Hirschfeldia incana	Australia	wild	secondary[1]
Hirschfeldia incana	Australia	wild	secondary[1]
Lepidium campestre	Australia	wild	tertiary
Matthiola incana	Spain	wild	tertiary
Matthiola longipetala subsp. bicornis	Australia	wild	tertiary
Sinapis alba	UK	selection/cultivar	secondary[1]
Sinapis alba	Sweden	selection/cultivar	secondary[1]
Sinapis alba	Israel	unknown	secondary[1]
Sinapis alba	Australia	unknown	secondary[1]
Sisymbrium erysimoides	Australia	wild	tertiary
Sisymbrium sp.	Australia	wild	tertiary

[1]nigra lineage
[2]rapa/oleracea lineage

DNA isolation

Genomic DNA was extracted individually from one individual from each line, to determine the degree of diversity between each variety, using the DNeasy® Miniprep Kit (QIAGEN) according

to the manufacturers' instructions. DNA was resuspended in 250 μl of buffer AB and stored at -20°C. The final DNA concentration ranged from 5 to 35 ng/μl.

SSR Analysis

Sixty genomic and EST derived SSRs will be used in the present study. The SSRs were selected, from a total of 200 screened in *B. napus*, for representative coverage of the genome and polymorphism information content (PIC) (Lowe et al., 2004; Tommasini et al., 2003). SSR primer pairs will be synthesised with the forward primer of each pair 5' end-labelled with either a 6-FAM or HEX (Sigma-Genosys) or NED or PET (Applied Biosystems) fluorescent tag. PCR amplifications and capillary electrophoresis will be performed by the Australian genome Research Facility (AGRF). Allele sizes will be determined using GeneMapper® 3.7 (Applied Biosystems).

Data analysis

The SSR loci will be scored as dominant, due to the amplification of greater than two loci per SSR primer combination in the amphidiploid species. SSR bands will be recorded as present (1) or absent (0) for construction of a binary matrix. A similarity matrix will be calculated using the Jaccard coefficient (Sneath and Sokal 1973), pooling together alleles across all loci, using NTSYS-pc software (Rohlf, 1998). Cluster analysis will be performed with NTSYS-pc to generate a dendrogram using the unweighted pair group with arithmetic average clustering (UPGMA; Sneath and Sokal 1973). For each SSR the PIC (Polymorphism Information Content) was calculated, according to the formula described by Smith et al. (1997), defined as PIC = $1-\Sigma_{Pi}^{2}$ where *Pi* is the allele frequency for the *i*-th allele.

RESULTS

In this study, a total of 60 SSR markers will be used for the analysis of genetic relationships of fifty Brassicaceae accessions. Summary details of the SSR markers, including their amplification in the different species, total number of polymorphic alleles, size range of polymorphic alleles, number of alleles produced per primer pair, number of accession specific alleles and the PIC for each SSR, will be presented.

In order to understand the genetic relationships and phylogeny among the species studied, cluster analysis will be performed using NTSYS software version 2.1 based on the combined profiles of all the loci produced by the SSR markers. The Jaccard genetic similarities between pairs of accessions will be used to generate a dendrogram based on UPGMA analysis, and assess pair-wise similarities between the accessions. The results will also be used to assign species of unconfirmed lineage to the correct origin.

DISCUSSION

The results obtained in this work will demonstrate that SSR markers can be effectively used for the genetic analysis of a germplasm collection. The results will assist in understanding the phylogeny and evolutionary history of the Brassicaceae and how this correlates with the existing morphological and inter-specific hybridisation classifications. We can determine the relationships between the different lineages. The information can be utilised to develop further studies to investigate the wide genetic diversity of important agronomic traits such as disease resistance and abiotic stresses which will be present in the wild germplasm. The rapid advances in next-generation sequencing, with the production of thousands of millions of nucleotide bases at an ever reducing cost provides the opportunity to apply this data for genetic diversity analysis. In the future, we can identify which Brassicaceae will be useful to sequence and use next generation sequencing data, molecular markers and morphological variation to study diversity across *Brassica* species and wild relatives.

ACKNOWLEDGEMENTS

The authors would like to acknowledge funding support from the Grains Research and Development Corporation (Project DAN00117), the University of Queensland (Project UQECR038 2008003553) and the Australian Research Council (Projects LP0883462 and

DP0985953). Support from the Australian Genome Research Facility (AGRF) and the ARC Centre of Excellence for Integrative Legume Research (CILR) is gratefully acknowledged.

REFERENCES

Beilstein, M.A., I.A. Al-Shehbaz, and E.A. Kellogg, 2006: Brassicaceae phylogeny and trichome evolution. Am. J. Bot. 96, 607-609.

Gomez-campo, C. and S. Prakash, 1999: Origin and domestication. In: C. Gomez-Campo (ed), Biology of Brassica coenospecies, pp. 33-52. Elsevier, Amsterdam.

Hopkins, C.J., N.O.I Cogan, M. Hand, E. Jewell, J. Kaur, X. Li, G.A.C. Lim, A.E. Ling, C. Love, H. Mountford, M. Todorovic, M. Vardy, G.C. Spangenberg, D. Edwards, and J. Batley, 2007: Sixteen new simple sequence repeat markers from *Brassica juncea* expressed sequences and their cross-species amplification. Mol. Ecol. Notes 7, 697-700

Lowe, A.J., C. Moule, M. Trick, and K.J. Edwards, 2004: Efficient large-scale development of microsatellites for marker and mapping applications in Brassica crop species. Theor. Appl. Genet. 108**,** 1103-1112.

Powell, W., G.C. Machray, and J. Provan, 1996: Polymorphism revealed by simple sequence repeats. Trends in Plant Science 1, 215-222.

Rohlf, F.J., 1998:. NTSYS-pc. Numerical Taxonomy and Multivariate Analysis System. Version 2.02e. Exeter Software, Setauket, New York.

Saal, B., J. Plieske, J. Hu, C.F. Quiros, and D. Struss, 2001: Microsatellite markers for genome analysis in Brassica. II. Assignment of rapeseed microsatellites to the A and C genomes and genetic mapping in *Brassica oleracea* L. Theor. Appl. Genet. 102, 695-699.

Sneath, P.H.A., and R.R. Sokal, 1973: Numerical taxonomy. WH Freeman & Co., San Francisco

Smith, J.S.C., E.C.L. Chin, L. Shu, O.S. Smith, S.J. Wall, M.L. Senior, S.E. Mitchell, S. Kresovich, and J. Ziegle, 1997: An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. Theor. Appl. Genet. 95, 163-173

Szewc-McFadden, A.K., S. Kresovich, S.M. Bliek, S.E. Mitchell, and J.R. McFerson, 1996: Identification of polymorphic, conserved simple sequence repeats (SSRs) in cultivated Brassica species. Theor. Appl. Genet. 93, 534-538.

Tommasini, L., J. Batley, G. Arnold, R.G. Cooke, P. Donini, D. Lee, J. Law, C. Moule, M. Trick, and K.J. Edwards, 2003: Application of fluorescence-based automated sizing of Multiplex Simple Sequence Repeats (SSR) marker sets for Testing Distinctness and Uniformity of Brassica napus L. Varieties. Theor. Appl. Genet 106, 1091-1101

U, N., 1935: Genome analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilisation. Jpn. J. Bot. 7, 389-452.

Warwick, S.I., A. Francis, and J. La Fleche, 2000: Guide to wild germplasm of brassica and allied crops (tribe *brassicaceae, brassicaceae*). 2nd edition, Eastern cereal and Oilseed Research Centre, Agriculture and Agri-Food Canada, Ottawa, Ont. [http://res2.agr.ca/ecorc-creco/cwmt/tech.htm].