

FINAL REPORT

DAV00058

Australian participation in the Multinational Brassica Genome Project (MBGP)

PROJECT DETAILS

PROJECT CODE: DAV00058

PROJECT TITLE: AUSTRALIAN PARTICIPATION IN THE MULTINATIONAL BRASSICA GENOME PROJECT (MBGP)

START DATE: 31.12.2003

END DATE: 01.01.2008

SUPERVISOR: GERMAN SPANGENBERG (EXECUTIVE DIRECTOR)

ORGANISATION: VICTORIAN DEPARTMENT OF ENVIRONMENT AND PRIMARY INDUSTRIES

CONTACT NAME: GERMAN SPANGENBERG

Summary

The project was initiated to develop genomic resources in *Brassica*, namely bacterial artificial chromosome (BAC) libraries, maps, expressed sequence tag (EST) libraries and databases based on the sequencing of the *Brassica rapa* genome.

The Multinational Brassica Genome Project (MBGP) is an international collaboration with sequencing, database development and management tasks shared among a wide international consortium.

This project report specifically relates to the Australian contribution to the MBGP which was to sequence approximately 1/10th of the *B. rapa* genome through the sequencing of BAC clones and associated activities in database development and management.

This sequence information will be made available to Australian canola breeders.

Report Disclaimer

This document has been prepared in good faith on the basis of information available at the date of publication without any independent verification. Grains Research & Development Corporation (GRDC) does not guarantee or warrant the accuracy, reliability, completeness or currency of the information in this publication nor its usefulness in achieving any purpose.

Readers are responsible for assessing the relevance and accuracy of the content of this publication. GRDC will not be liable for any loss, damage, cost or expense incurred or arising by reason of any person using or relying on information in this publication. Products may be identified by proprietary or trade names to help readers identify particular types of products but this is not, and is not intended to be, an endorsement or recommendation of any product or manufacturer referred to. Other products may perform as well or better than those specifically referred to. Check www.apvma.gov.au and select product registrations listed in PUBCRIS for current information relating to product registration.

Copyright

Grains Research and Development Corporation. This publication is copyright. Apart from any use as permitted under the Copyright Act 1968, no part may be reproduced in any form without written permission from the GRDC.

Old or Archival Reports (Projects that concluded in 2007 or earlier)

The information contained in these older reports is now several years old, and may have been wholly or partially superseded or built upon in subsequent work funded by GRDC or others. Readers should be aware that more recent research may be more useful for their needs. Findings related to agricultural chemical use are also potentially out of date and are not to be taken as a recommendation for their use.

Conclusions

Through the successful completion of this project, Australia has significantly contributed to the amount of sequence information that is available internationally for *Brassica*.

Sequencing activities in seven of the 10 chromosomes of *B. rapa* and development of genomic tools, such as molecular markers, sequence and bioinformatics databases have greatly enhanced the genomic resources available to *Brassica* scientists and breeders.

In particular, several resources have been developed that have direct applicability to canola (*Brassica napus*) molecular breeding.

The dissection of the A and C genomes of canola is critical in the development of market ready markers. Despite the incomplete sequence data, the comparison of data from *B. napus* and *Brassica oleracea* facilitates this and will thereby allow breeders to identify genome specific sources of improvement for traits such as disease resistance.

Recommendations

Since the Australian contribution to the MBGP has been achieved, it is recommended that:

- Australian canola breeders be advised that the Australian genome sequence information is available for their use. A suitable vehicle for this communication may be a presentation to the breeders and the National Brassica Germplasm Enhancement Program.
- GRDC approves submission of the sequence information derived in this project to the international database according to the protocols developed by the MBGP community.

Outcomes

The availability of the complete *Brassica* genome sequence would lead to faster incorporation of valuable traits into commercial germplasm, more robust methods of variety improvement and a broader range of valuable traits identified and incorporated at a greatly reduced cost.

These technologies would revolutionise the *Brassica* crops currently grown in Australia and lead to the rapid introduction of designer varieties.

Australia's contribution of approximately 1/10th of the genome sequence information for *B. rapa* is a significant international contribution towards the development of sequence based tools for *Brassica* breeding.

While the complete genome sequence of *B. rapa* has not yet been completed, progress has been made in generating BAC sequence data for seven of the 10 chromosomes. These sequences can be mined for molecular markers to improve map coverage, and thus aid directed plant breeding. The sequences can also be mined for genes of agronomic importance for further characterisation and use in transgenic research. Transgenic crops are proving highly effective at increasing yields with a reduced environmental footprint. The genomic sequence data will enable the production of transgenic *Brassica* crops.

Achievements/Benefits

Brassica species play an important role in global agriculture and horticulture, as well as being the closest crop relatives to *Arabidopsis*. The species are characterised by a wide range of adaptations that have been domesticated into crops including oilseed rape/canola and swede (*B. napus*); cabbage, cauliflower, broccoli, Brussels sprout (*B. oleracea*); Chinese cabbage, pak choi, turnip and oil (*B. rapa*) and mustards (*B. nigra*, *B. juncea*, *B. carinata*). *Brassica* crops contribute both to the economies and health of populations around the world. Crop improvement is a key route to ensuring continued benefits arise from these foods and plant products. A wide range of genetic and genomic resources is available from *Brassica*, as well as easy access to information derived from *Arabidopsis*. Although considerable progress has been made in the genetic analysis of agronomic and related plant traits in *Brassica*, there are currently limitations in identifying and characterising the behaviour of the underlying genes, genomic regulatory networks and associated metabolism. A range of major challenges focuses on harnessing genetic diversity through an information-led approach to crop improvement. There is a wide range of valuable traits for which genetic variation exists but where understanding is required. These include improving harvest index and yields in the context of changing climate and the need to reduce inputs, optimising quality fit for purpose, and identifying scope for added value through nutritional, prophylactic health or non-food use. Given the unique position of *Brassic*as with their diverse crop types and metabolites, the vision for maximising the benefits from recent advances in genomics can be summed up in terms of "global information, local implementation". Genomic information is the key to exploiting knowledge gained at the level of gene expression, biochemistry, metabolism and physiology. In particular, the *Brassica* research communities need to be able to locate genes in their relevant genomic context, in order to understand regulation and pressures on evolution. For effective crop-based studies, it is also essential to be able to navigate between trait and gene, and thus integrate information from agronomy, breeding, genetics and genomics. This will provide a functional understanding of relevant systems to underpin future crop and product development within a sustainable framework.

The Multinational Brassica Genome Project is an international collaboration with sequencing, database development and management tasks shared among a wide international consortium.

The Steering Committee for the MBGP selected *B. rapa* as the first species to be sequenced, as it has the smallest genome (ca. 550 Mb), the lowest frequencies of repetitive sequences, and communal BAC libraries and mapping populations are available. Consequently, the Brassica rapa Genome Sequencing Project (BrGSP) was launched, based on the Chinese cabbage line Chiifu-401, and associated BAC library resources developed in Korea. The 'A' genome of *B. rapa* is also present in major oil crops of *B. napus* and *B. juncea*.

Substantial progress was made in this project towards sequencing seven of the 10 chromosomes of the A genome of *B. rapa* by the international consortium. This international effort has greatly enhanced the publicly available genomic information for *Brassica*. A complete summary of the international program may be found on the website www.brassica.info

This Australian project has delivered

- characterisation of *B. rapa* BACs by end sequencing.
- *in silico* discovery of 2,857 SSR markers within the BAC end sequences.
- primer pairs for these SSR markers designed for use by the MBGP community.
- 110,000 BAC end sequences deposited in a public database.
- a *B. rapa* sequence database established with new sequence data being added as it is produced.
- approx.42 Mbp of novel BAC sequence data from 424 BAC clones, approximating to 1/10 of the genome of *B. rapa*.

The Australian contribution to the MBGP has been achieved and the sequence information is available for contribution to the international *Brassica* sequence database upon approval from GRDC.

Intellectual property summary

The intellectual property (IP) developed in this project was firstly made available to the participants in MGBP. A process is underway to make much of this data more broadly available and utilised.

Additional information

Love CG, Batley J, Lim G, Robinson AJ, Savage D, Singh D, Spangenberg GC, Edwards D (2004) New computational tools for Brassica genome research. *Comparative and Functional Genomics*. 5:276-280

Savage D, Batley J, Erwin T, Logan E, Love CG, Lim G, Mongin E, Barker G, Spangenberg G, Edwards D (2005) SNPServer: A real-time SNP discovery tool. *Nucleic Acids Research* 33:W493-495.

Love C, Robinson A, Lim G, Hopkins C, Batley J, Barker G, Spangenberg G, Edwards D (2005) ASTRA: An integrated database for Brassica genomic research. *Nucleic Acid Research* 33, Database issue: 656-659.

Love C, Logan E, Erwin T, Spangenberg G, Edwards D (2006) Analysis of the Brassica A and C genomes and comparison with the genome of *Arabidopsis thaliana*. *Acta Horticulturae* 706, 99-104.

Love C, Logan E, Erwin T, Hopkins C, Batley J, Spangenberg G, Edwards D (2006) Integrating and interrogating diverse Brassica data within an Ensembl structured database. *Acta Horticulturae* 706, 77-82.

Hopkins C, Mogg R, Gororo N, Salisbury P, Burton W, Love C, Spangenberg G, Edwards D, Batley J (2006) An assessment of genetic diversity within and between Brassica napus and Brassica juncea lines from international germplasm collections. *Acta Horticulturae* 706, 115-119.

Sexton AC, Cozijnsen AJ, Keniry A, Jewell E, Love CG, Batley J, Edwards D, Howlett BJ (2006). Transcription profile of genes from three developmental stages of the plant pathogen *Sclerotinia sclerotiorum*. *FEMS Microbiology Letters* 258, 150-160.

Burgess B, Mountford H, Hopkins CJ, Love C, Ling AE, Spangenberg GC, Edwards D, Batley J (2006). Identification and characterisation of simple sequence repeat (SSR) markers derived in silico from Brassica oleracea genome shotgun sequencing. *Molecular Ecology Notes* 6, 1191-1194.

Jewell E, Robinson A, Savage D, Erwin T, Love CG, Lim GAC, Li X, Batley J, Spangenberg GC, Edwards D (2006). SSRPrimer and SSR Taxonomy Tree: Biome SSR discovery. *Nucleic Acids Research* 34, 656-659.

Erwin TA, Jewell EG, Love CG, Lim GAC, Li X, Chapman R, Batley J, Stajich JE, Mongin E, Stupka E, Ross B, Spangenberg G, Edwards D (2007). BASC: an integrated bioinformatics system for Brassica research. *Nucleic Acids Research* 35, 870-873.

Since completion of the program, the full genome sequence of *B. rapa* has been determined and may be accessed through the Brassica dB (BRAD) URL: <http://brassicadb.org/brad>.

The results are described in the following publications:

The Plant Cell, Vol. 25: 1541-1554, May 2013, www.plantcell.org

Deciphering the Diploid Ancestral Genome of the Mesohexaploid Brassica rapa C W Feng Cheng,^a Terezie Mandáková,^b Jian Wu,^a Qi Xie, c Martin A. Lysak,^{b,1} and Xiaowu Wang^{a,1,2}

- a. Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China
- b. Plant Cytogenomics, Central European Institute of Technology (CEITEC) and Faculty of Science, Masaryk University, CZ-602 00 Brno, Czech Republic
- c. State Key Laboratory of Plant Genomics, National Center for Plant Gene Research (Beijing), Institute of Genetics and Developmental

Biology, Chinese Academy of Sciences, Beijing 100101, China

Nature Genetics 43, 1035-1039 (2011), doi:10.1038/ng.919

The genome of the mesopolyploid crop species *Brassica rapa*