

FINAL REPORT

CMB00013

Australian Winter Cereals Molecular Marker Program - wheat genome sequencing consortium

PROJECT DETAILS

PROJECT CODE: CMB00013

PROJECT TITLE: AUSTRALIAN WINTER CEREALS MOLECULAR MARKER PROGRAM - WHEAT GENOME SEQUENCING CONSORTIUM

START DATE: 01.10.2005

END DATE: 30.06.2006

SUPERVISOR: PROFESSOR RUDI APPELS

ORGANISATION: MOLECULAR PLANT BREEDING PTY LTD

CONTACT NAME: RUDI APPELS

Summary

Chromosome 3B of wheat is an early target for the International Wheat Genome Sequencing Consortium (IWGSC). The pilot project for sequencing 50 bacterial artificial chromosome (BAC) clones in a 12cM region with the Sr2, *Stagonospora nodorum*, leaf rust (Lr) and fusarium head blight (FHB) resistance loci has been successful in positioning Australia within the IWGSC. A flour allergen locus was found. The project is the first to globally target 4Mb of the wheat genome. The work was carried out in collaboration with the Chinese Academy for Agricultural Sciences (CAAS) group in Beijing (Prof Jia Jizeng and Dr Xuiying Kong), who provided the actual sequencing (6 x coverage was found to be sufficient). The data were immediately transferred to Murdoch University for assembly of sequences.

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

A significant long term impact of this one year investment by the GRDC has been the positioning of Australia in the IWGSC. Prof Rudi Appels has remained as a co-chair of IWGSC and in January 2007 is organising the IWGSC workshop at the annual Plant and Animal Genomics meeting in San Diego. Furthermore, two key individuals in the IWGSC, Dr Catherine Feuillet and Kellye Eversole, travelled to Canberra in August to meet with the leadership of the Australian Research Council (ARC) and GRDC, and with Dr WJ Peacock in his capacity as Chief Scientist. As a direct result, the Chief Scientist requested Rudi Appels to prepare a document for an investment by Australia in wheat and barley genomics. The document was prepared with Prof Peter Langridge after an international level discussion at the International Triticeae Mapping Initiative (ITMI) meeting in South Australia (SA) in late August. Discussions have followed involving the Department of Agriculture, Fisheries and Forestry (DAFF) (Dr Peacock's recommendation), as well as ARC (by Prof A Robertson, University of Western Australia (UWA)) and GRDC (this report).

The focus on the group 7 chromosomes for Australia to take leadership in is an exciting prospect and is already being accepted internationally as colleagues are planning work on the group 7 chromosomes in consultation with Rudi Appels (recent examples include a commitment in Czechoslovakia to build a BAC library for 7D and the listing of Rudi Appels as a collaborator in a successful National Science Foundation (NSF) proposal by Dr K Devos, USA). The bioinformatics being carried out at the Centre for Comparative Genomics (CCG), (Prof. M Bellgard, Murdoch University) is now recognised internationally as a result of the present proposal and this means the top groups in the field around the world are engaged.

The project has demonstrated the technical achievement that wheat genome sequencing is feasible despite the presence of the much discussed content of repetitive DNA sequences and has contributed to pioneering the annotation of the wheat genomic sequence.

Recommendations

The revolution in genomics provides direct access to the genetic code underpinning the attributes society values in crops such as wheat and barley. The availability of physical map and sequence data for a species has an enormous impact on the types of research that can be conducted and can be rapidly translated to practical outcomes through support of gene discovery, analysis of molecular diversity and the information it provides on genome structure and behaviour.

New breeding strategies, both conventional and molecular, result from these outputs. These, in turn, provide new options for breeders leading to increased flexibility and sophistication and increased rates of genetic gain. This is expected to be particularly important for intensively bred species, such as wheat and barley, where the opportunities for gain are becoming increasingly restricted, yet the need for improved adaptation is becoming more urgent as there is a steady decline in the

quality of the agricultural environment. The genomics revolution in wheat and barley will provide access to the gene complexes that control the way these plants respond to the environment. Manipulation of these genes through genetically modified (GM) or non-GM technologies will provide plant types that can deal with climate change while maintaining market demands for grain quality and health attributes.

The GRDC currently supports research programs by many groups in the genetic and molecular analysis of wheat and barley. This covers research into disease resistance, tolerance to abiotic stresses and processing quality. This work has underpinned highly successful breeding programs that have ensured regular advances in crop yield and quality. In many respects, the current viability of the cropping industry in Australia has been dependent upon the ability to rapidly identify, access and use technological advancements. For this reason, it is important that Australia expands its role in the current international proposals to develop physical maps and ultimately to sequence the wheat and barley genomes. This role will ensure effective leveraging of overseas investments and ensure early access to the data and information in order to help maintain the vitality and competitiveness of research and delivery programs. A detailed proposal for Australia's participation has been documented for the Chief Scientist of Australia (Dr WJ Peacock).

Outcomes

Social outcome

A significant long term impact of this one year investment by the GRDC has been the positioning of Australia in the IWGSC. Prof Rudi Appels has remained as a co-chair of IWGSC and in January 2007 is organising the IWGSC workshop at the annual Plant and Animal Genomics meeting in San Diego.

Environmental outcome

Based on the sequencing work in this report, two regions of interest are of particular interest. One is the gene for Sr2 which was one of the targets of this genomic analysis, and is now confined to a 200kb region. The molecular markers identified from this region are now providing much better markers for Sr2 than those previously available and will contribute to the management of stem rust disease problems in wheat crops. The other gene called expansin EXPB11 was an unexpected discovery but needs further investigation because it is an endosperm specific protein and contributes to allergic reactions of humans to wheat flour. Other genes have also been annotated (present at a low density) and still need basic functional studies. New retrotransposable elements have been identified and are now part of a PhD thesis by James Breen (at Murdoch University).

Economic outcome

The long term investment in wheat genome sequencing in Australia, as initiated by this project, will have significant economic outcomes in contributing to producing wheat varieties better able to cope with climate change and biotic and abiotic stresses. This will be achieved through providing access to the genomic details of the gene networks controlling plant performance.

Achievements/Benefits

Chromosome 3B of wheat is an early target for the IWGSC. The pilot project for sequencing 50 BAC clones in a 12cM region which carries the Sr2, *S. nodorum* (Sn), Lr and FHB resistance loci, has been successful in positioning Australia within the IWGSC. A flour allergen locus was also found. The project is widely recognised as the first one globally to target 4Mb of the wheat genome. The arrangement for determining the DNA sequence worked extremely well - the actual sequencing of the wheat BAC clones was carried out in CAAS, Beijing (6 x coverage was found to be sufficient) and the data immediately transferred to Murdoch University for assembly of sequences. The compilation and analysis of the linear DNA sequences was carried out by Dr David Dunn.

The immediate targets within the gwm389-gwm493/20Mb region of chromosome 3BS are two regions, the Sr2-Sn and Lr-FHB regions. The INRA group led by Catherine Feuillet has, to date, identified six large BAC contigs ranging in size from 250kb to 1.6Mb for a total length of 4.1Mb. There remains a significant challenge to obtain a physical contig covering the 20Mb region, but the INRA group now has all the BAC end sequences determined and this will assist in anchoring other contigs in the region. A group in the USA (involving Bikram Gill) is sequencing BACs (5-10) in the immediate FHB region and another group (Jeff Bennetzen) has seed funding from NSF for sequencing 250 BACs selected at random. In Australia, the (near) phase 2 level assembly of three contigs with lengths of 1.8Mb, 1.1Mb and 1.0Mb has been completed.

A major challenge has been the anchoring of the BAC clones to a genetic map and collaboration with the Canberra group (CSIRO-PI, Evans Lagudah and Wolfgang Spielmeier) has been extensive. Initially, the group of BACs 113-H12, 115-G14, 043-0105 were thought to cover the proximal end of the Sr2 locus. This was subsequently shown not to be the case and these clones have been put to the side of the major contig, pending localisation within the genetic map. The sequences for this small contig are now providing probes to allow mapping in Cranbrook x Halberd. The new BACs, 010-D05 and 042-J02, have been confirmed to cover the proximal end of the Sr2 locus. This means the Sr2 locus can now be confined to the overlapping BACs, 028-F08, 070-N04, 058-L24, 042-G11, 043-E17. This particular array has the DNA sequence well validated with very few gaps to be filled in order to complete a finished sequence. The latter is being carried out in collaboration with the group in Canberra. Significantly, some new probes designed from the BAC sequences of the neighbouring BACs 122-F21, 038-E11, as well as 028-F08, are polymorphic in Australian germplasm and are being trialled as new markers (essentially perfect markers since they are physically so close to the Sr2 gene) in breeding lines.

It is clear that the detailed analysis of the remainder of the BACs in the Sr2 contig, as well as the gwm533 and STS104 contigs, will contribute significantly to delineating the entire region of interest which carries a resistance gene for *S. nodorum*, *Septoria tritici*, FHB resistance and possibly Yr27. All of these (except perhaps FHB resistance) are disease resistance genes of significance in Australian wheat breeding programs. In addition to the impact on the national wheat breeding objective, this project has had international impact (evident from the presentation at the ITMI meeting in SA, August 2006) in the following aspects:

- Gap closure works well with overlapping BACs in wheat genome sequence assemblies.
- High levels of repetitive DNA can be dealt with, but greater redundancy in overlapping BAC sequences is probably required relative to that of a smaller genome.
- Many targets of simple sequence repeats (SSRs) are evident for designing new Sr2/Sn region markers, for use in tracking this small chromosome segment in a breeding program.
- BAC sequencing reveals some ambiguity in the DNA fingerprinting-based assembly of BAC, particularly in complex regions such as that represented by the SSR gwm533.

It is evident that new probes designed from the BAC sequences of the neighbouring BACs 122-F21, 038-E11, as well as 028-F08, are polymorphic in Australian germplasm. These are now being trialled as new markers (essentially perfect markers since they are physically so close to the Sr2 gene) in breeding lines. Figure 2 shows the analysis of an SSR from 028-F08 in the Cranbrook x Halberd, where Cranbrook is the source of the Sr2 locus from Hope and is quite distinct from the simple polymerase chain reaction (PCR) product in Halberd. The Canberra group has confirmed that the probe behaves the same way in assaying the Hope Sr2 locus in their cross, and is completely linked.

To date, the only other regions sequenced to the level achieved in this proposal is a small contig of 250kb at the FHB locus. A large 3Mb region is currently being sequenced at the FHB-Rph7 locus by the group in INRA (France).

Specific technical advances

(1) A deliberate decision was made at the outset of this project that a 6 x coverage of BACs with respect to amount of sequencing that was carried was sufficient for the assembly of the sequence of the minimum tiling path (MTP). This decision was based on observations of the sequence assembly of two BAC clones that were virtually the same (due to the selection of BAC's process). The decision allowed 30-40% more BACs to be sequenced since the contract price was set at a per BAC basis and the degree of coverage of the BAC with respect to the number of random clones sequenced. It is significant to note, therefore, that this decision was vindicated by work carried out in ET5 (Meredith Carter) where one of the clones sequenced and processed by the normal process (carried out in this report) was also sequenced, on contract, to completion by the Beijing Genomics Institute (funded by the CCG). The comparison of the two assemblies indicated the process used in this report achieved better than 98% coverage.

(2) An inconsistency between the physical assembly in the Sr2 region of 3BS and the genetic map of Arrino x Forno (used in Switzerland) by one of the collaborating groups was examined in detail. The sequence assembly of the group in Perth was confirmed by checking alternative assemblies and this issue is currently being investigated further with the group in Switzerland by looking at the actual genetic map using the curation procedures established in GA17. The occurrence of a genuine, small inversion between varieties is important because it signals a degree of caution may be required in declaring the 'correct' genome sequence has been achieved.

(3) A key aspect of the current project was to provide evidence for the accuracy of the DNA fingerprinting used to initially

compile the MTPs. In the 'gwm533' contig, it has been shown by the sequence assembly carried out in Perth that the DNA fingerprinting process was 'fooled' by the complexity of repetitive sequences in this region and included two BAC clones that should not in fact be there. This is currently under further investigation because it is important in the detailed analysis of the wheat genome. In addition, it provided further evidence of the complexities of using gwm533 as a molecular marker for Sr2 in breeding programs because this was a direct example of the gwm533 ambiguity at the level of high resolution DNA fingerprinting.

(4) The discovery of a near perfect match to Ta EXPB11, a published cDNA that was derived from a cDNA library produced in Canberra from cv. Wyuna, has provided the first genomic sequence of a potentially important human allergen in wheat flour. This is a significant technical achievement that is being published after some follow up functional studies and provides an insight into the new discoveries that will come our way besides the targeted gene regions.

The outputs from the project have been made available to selected colleagues in the IWGSC and will be made available more generally in March 2007, by which time it is expected that some ambiguities in the genome sequence assembly will be resolved. The sequence information has been offered to colleagues in the GRDC-Australian Winter Cereals Molecular Marker Program (AWCMMP) and collaborators within Australia.

Other research

Determining the detailed structure and function of the wheat and barley genomes is now widely recognised internationally as a vital activity for these cereal crops if they are to be developed to take their place in an agriculture system facing the challenges of climate change, as well as increased demands for quality attributes targeted to health and specific end-products.

Additional information

The assembly of wheat chromosome 3B for structure-function studies - a paper being prepared for submission to Science or equivalent by the IWGSC (first author Dr Catherine Feuillet).

The expansin B11 locus on chromosome 3B of wheat - a paper being prepared for submission to Plant Physiology or equivalent by the IWGSC (first author Dr David Dunn from Murdoch University).

Oral presentations on the project have been provided at the IWGSC/ITMI meeting in Victor Harbor (SA, August 2006) and the IWGSC meeting at the PAG in San Diego (California USA January 2007).