

FINAL REPORT

UQ00026

AWCMMP - Component ET8: Pedigree-based genome mapping for marker-assisted selection and recurrent parent recovery in wheat and barley

PROJECT DETAILS

PROJECT CODE: UQ00026

PROJECT TITLE: AWCMMMP - COMPONENT ET8: PEDIGREE-BASED GENOME MAPPING FOR MARKER-ASSISTED SELECTION AND RECURRENT PARENT RECOVERY IN WHEAT AND BARLEY

START DATE: 01.07.2002

END DATE: 31.12.2004

SUPERVISOR: DR EMMA MACE

ORGANISATION: DEPARTMENT OF PRIMARY INDUSTRIES AND FISHERIES

CONTACT NAME: EMMA MACE

Summary

This project has investigated and developed systems for implementing marker-assisted selection (MAS) in order to improve the efficiency of selection and increase the rate of genetic gain in breeding programs. Protocols for implementation of pedigree-based whole genome marker analysis in northern region winter cereal breeding programs have been tested in order to provide a) useful information about the use of molecular markers for routine genomic analysis, b) the integration of genotypic, phenotypic and pedigree information for targeted wheat and barley lines, c) the genomic impacts of strong selection pressure in case study pedigrees, and d) directions for future pedigree-based marker development and analysis.

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

This project has investigated the protocols and systems necessary to implement the pedigree-based whole genome (PBWG) approach using the Queensland Department of Primary Industries and Fisheries (QDPI&F) and Enterprise Grains Australia (EGA) northern region focussed wheat and barley breeding programs as case studies.

A core set of 126 wheat varieties and breeding lines from pedigrees relevant to the northern region with a focus on Condor and Cook related material has been catalogued, and a set of 93 barley genotypes were selected with lineage from Triumph and/or Koru, together with the other parents of the chosen advanced breeding lines, which are representative of a significant proportion of the Northern Barley Improvement Program (NBIP) gene pool.

One hundred and eighteen SSR markers were selected for assessment in wheat spread at approximately 10cM intervals on chromosomes 2B, 2D, 3B, 3D, 4A, 4B, 4D and 7A. These chromosomes were chosen on the basis of location of genes and putative quantitative trait loci (QTL) associated with height, milling yield, flour colour, black point (BP) resistance and rust resistances. In addition to the simple sequence repeat (SSR) markers, 63 diversity arrays technology (DART) markers were applied to 93 genotypes by Triticarte P/L. 56 SSR markers linked to traits of importance to the barley industry have been applied to DNA fingerprinting of the 66 barley genotypes. In addition to the SSR markers, 297 polymorphic DART markers were applied to the same genotypes by Triticarte P/L.

Pedigree data for the 126 wheat lines and 93 barley lines have been obtained, cleaned, validated and entered into the Pedigree-Based Marker-Assisted Selection System (PBMAS). PBMAS is the software developed for this project. Modules developed include: import, export and storage of pedigree data; printable graphical display of pedigree data including functions to calculate the inbreeding coefficient (COI) and coefficient of performance (COP); import, export and storage of molecular marker data; and printable display of graphical genotypes colour coded for parental origin (using identity-by-descent) for selected combinations of genotypes and chromosomes.

Marker data have been analysed in conjunction with both existing and new phenotypic data. The analysis of the PBWG mapping approach will allow for an enhanced understanding of the processes of selection within a breeding program, including identification of regions under selection, and will act as a basis of validation of markers and QTL for the more effective implementation of markers, through the confirmation of QTL location, the identification of markers which will work in other, distantly related populations, and in the identification of QTL which are effective in other populations. The outcomes from this project will encourage more sophisticated use of markers in targeted breeding populations as selection is applied to conserve and combine critical genomic regions for both environmental adaptation and grain and end-use quality.

Recommendations

When making selection decisions cereal breeders need to integrate information from a range of sources such as yield trials, quality laboratories and pathology trials in a relatively short timeframe. To date, it can be argued that the use of this information has been inefficient, and additionally, the recent deployment of marker technologies has greatly complicated the situation by greatly increasing the amount of information that can potentially be used by breeders. The development of high throughput genotyping technologies such as DART will further intensify this trend. Increasingly, the major limitation to the efficient use of this information is the need to take data from a range of sources and generate appropriate summaries in a timely fashion.

In this project, it has been determined that PBWG mapping offers significant potential to increase the genetic gain made in plant breeding programs by identifying:

- 1) Regions of the genome currently under selection in the breeding program
- 2) The frequency and distribution of particular QTL alleles within a breeding program
- 3) The ancestral origin of regions under selection and/or regions containing QTL of economic importance
- 4) QTL and estimation allelic effects of particular QTL in breeding populations.

It is recommended that these strategies identified for the PBWG approach are implemented in the northern barley and wheat breeding programs. The interaction between breeders, molecular biologists, statisticians and programmers using real data sets is critical to the continued development and utilisation of appropriate tools.

It is recommended that there is more focus on building core sets of resources which are essential pre-requisites for implementing PBWG, including identifying core panel sets of molecular markers flanking QTL. DART markers may offer the

necessary low-cost, high throughput genotyping approach to be used in conjunction with trait-linked microsatellite markers.

It is further recommended that the qualitative outcomes identified – such as information about the frequency, distribution and ancestral origin of known QTL, and information about historical selection for particular regions – are further developed in order to identify a) parents for crossing, b) breeding populations for validation of MAS to help confirm QTL, c) useful or deleterious linkage between genes, and d) provide new targets for marker development.

Outcomes

PBWG mapping offers significant potential to increase the genetic gain made in plant breeding programs through the development of systems for MAS. With in-breeding plants, genomic regions associated with traits of importance are traditionally identified using populations derived from single crosses of inbred lines. This strategy however is frequently sub-optimal for direct application in breeding programs for numerous reasons. QTL trait association strategies frequently use obscure or out-dated parents. This strategy competes directly with breeding programs for significant resources for phenotypic evaluation. It is also limited in coverage because it addresses only the genes segregating in the populations under study. Additionally, this strategy provides no information about the frequency or value of the QTL in the breeding populations and therefore requires a further step of validating the usefulness of markers for direct application in the breeding program. Due to the wide-spread adoption of this QTL analysis approach with specific doubled haploid (DH) populations, marker application is being impeded through the lack of information about the polymorphism and marker-trait association in genetic material relevant to Australian northern region cereal improvement.

PBWG provides a vehicle for incorporating marker technologies into applied breeding programs by bridging the gap between development and implementation. The PBWG marker concept uses pedigree information to identify markers linked to traits based on identity by descent in breeding populations. The approach makes efficient use of pedigree, phenotypic and genotypic information collected in the normal breeding program, thereby exploiting this valuable resource. The PBWG approach integrates very closely with marker-assisted recurrent parent recovery (MARPR), a technology already applied to many crops including barley, maize and rice, that can speed variety development in back-crossing strategies.

Effective use of a PBWG-MARPR marker approach requires relevant pedigree and phenotypic data, low-cost, high throughput genotyping and a data management and analysis system that combines pedigree information, and genotypic and phenotypic data. Implementation of these components allows linkage to advanced molecular marker and related research, which directly addresses constraints and opportunities arising within the context of the breeding program. The current project has piloted the design and development of a system that integrates these components of PBWG and MARPR marker application for use in applied breeding programs. The beneficial outcomes of such research will impact on many levels within the grains industry, from increased on-farm productivity to the accelerated release of elite wheat and barley varieties. These benefits will impact nationally and regionally and will contribute to a healthy regional grains industry.

Achievements/Benefits

This project aimed to develop the protocols and systems necessary to implement the PBWG approach using the QDPI&F and EGA northern region focussed wheat and barley breeding programs as case studies. Within each case study, the overall strategy was to:

- Identify appropriate interrelated sets of germplasm
- Identify a suite of robust and polymorphic molecular markers
- Fingerprint the germplasm and verify the pedigrees
- Collate phenotypic, genotypic and pedigree information
- Verify and repeat phenotypic information as necessary
- Trace markers through the pedigree and detect evidence of selection for particular alleles, and
- Compare regions under selection with the location of QTL for traits of importance known to be under selection in the breeding programs, as a means of QTL validation.

Marker profiles were made of 126 wheat lines and 93 barley lines. Wheat lines were related to the wheat variety Condor. These included breeding lines reaching S4 stage, intermediate lines, parents and donor lines. Barley lines were those derived from Triumph and/or Koru. A suite of 118 SSR markers were identified spaced at around 10cM across eight wheat chromosomes – 2B, 2D, 3B, 3D, 4A, 4B, 4D and 7A. The selection of these chromosomes was based on knowledge available at the time on locations of putative QTLs for traits of interest in populations related to Cook or Condor. These traits included milling yield,

flour colour, BP resistance, pre-harvest sprouting resistance, rust resistance genes, cereal cyst nematode (CCN) resistance (Cre 3), crown rot (CR) resistance, height genes, granule-bound starch synthase (GBSS) and the RLN resistance gene Rln1. Putative milling yield and colour b QTL locations were supplied by Anke Lehmensiek from her data mining project and by Adele Schmidt from her study of the Kukri¹ x Lang¹ DH population. The aim was also for this marker set to be complementary to the set used in the association mapping project based in Adelaide. This work was out-sourced to Aggenomics Pty Ltd. Genetic location of markers was based primarily on the composite map of Professor Rudi Appels (wheat.pw.usda.gov/cgi-bin/graingenes), but additional information was used from other sources including Roder et al (1998) Genetics 149: 2007-2023; Pestrova et al (2000) Genome 43: 689-697; Somers et al (2004) Theor Appl Genet 109: 1105-1114; Chalmers et al (2001) Aust. J. Agric Res. 52: 1089-1119, Schmidt et al (pers com), Lehmensiek et al (pers com), Collard et al (pers com) and Zwart et al (pers com).

Fifty six markers linked to traits of importance to the barley industry in the northern region have been identified, optimised and catalogued. Markers from each chromosome were selected to ensure adequate genome coverage. The traits included disease resistances (net form of net blotch (NFNB), spot form of net blotch (SFNB), leaf rust, stem rust, powdery mildew, barley yellow dwarf virus (BYDV), leaf scald), malting quality traits (hot water extract, diastatic power, beta-amylase, protein, kernel discolouration, plump grain, pre-harvest sprouting) and agronomic and environmental tolerance traits (aluminium and manganese tolerance, flowering, dwarfing). These markers were applied to a set of 93 interrelated genotypes derived from Triumph and/or Koru, representative of a significant proportion of the Northern Barley Improvement Program (NBIP) gene pool. The initial polymerase chain reaction (PCR) assays were outsourced to the New South Wales (NSW) DPI (Wagga Wagga), and following the appointment of Emma Mace in September 2003, the PCR assays were completed at the Hermitage Research Station (HRS), Warwick.

During the course of this project DArT has become available through Triticarte Pty Ltd. 121 wheat DNA samples were submitted to Triticarte for DArT analysis in October 2003 for analysis using the DArT 1.0 chip. First results were received in January 2004. The remainder was received in September 2004 for a total of 114 lines for 148 markers. The resultant dendrogram was largely consistent with known pedigree information, but these markers were not mapped. These samples, and some new samples, have been submitted in February 2005 for reanalysis on the DArT 2.0 chip, a subset of markers from which have reported to be mapped. The 93 barley genotypes were also submitted to Triticarte for DArT analysis in two separate batches. The first batch, with results delivered in October 2003, consisted of 297 polymorphic DArT markers across 58 genotypes. The second batch, with results delivered in January 2005, consisted of 409 polymorphic DArT markers across 35 genotypes. 120 DArT markers were found to be common between the two data sets. It is anticipated that between 50% to 75% of these markers will be mapped in the near future and the mapping information will be utilised in the second phase of this project.

Genotypic information for the full suite of SSR markers and DArT obtained for targeted wheat and barley lines has been stored with corresponding phenotypic data in the database associated with PBMASS. Accurate pedigree information for wheat and barley lines included in this study has been accumulated, verified and entered into the PBMASS database. Because much of the pedigree information still existed in hand-written hard copy, significant effort was required to clean and verify the available records. The process was confounded by the discovery of significant capture errors in the International Maize and Wheat Improvement Centre (CIMMYT) databases, which hold pedigree information on Australian wheats.

Available phenotypic data for wheat and barley lines from targeted pedigrees, including foliar disease response, yield, agronomy and quality data, from 1986 onwards, has been accessed, collated and contributed to the PBMASS database. Specifically, barley quality and agronomics data has been collated from a maximum of nine trial sites annually from 1994 to 2003, in addition to merging with the historical data sets available for barley quality from 1976 to 2003. Wheat quality data has been captured for all material in S4 trials from 1986 to 2003, together with pathology data for foliar and root diseases.

Since the aim of this project was largely about developing systems which enable the application of genomic analysis and MAS in breeding programs, a strong part of its focus was the integration of genotypic, phenotypic and pedigree data. The key to successfully integrating the phenotypic, pedigree and genotypic data sets was the standardisation of the naming convention for genotypes. Significant effort was spent in collating all the aliases in use in the wheat and barley breeding programs and selecting standardised preferred names for each genotype. The standardised genotype identifier therefore acts as the backbone onto which the other pieces of information (genotypic, phenotypic, pedigree) are attached, ordered and accessed.

Development of a specification for database requirements to manage the multiple aliases for each genotype and to achieve

integration of the phenotypic, pedigree and genotypic data types has been undertaken, and consideration and assessment of associated software for presentation and analysis is well advanced. The PBMAS software developed as an initiative of the combined QDPI&F plant breeding programs provides a user-friendly desktop tool to assist plant breeders in utilising pedigree, molecular marker and phenotypic data when selecting breeding lines. The software, designed to run on Microsoft Windows 2000/XP provides a C++ graphical user interface (GUI) to a Microsoft Access database. Although the database structure is suitable for storing large volumes of data PBMAS has been designed as a desktop tool and is not intended as a mass storage mechanism. Data input and output (excluding graphical representations) is via MS Excel spreadsheets or comma separated (csv) files.

Current development provides concurrent visualisation of pedigree, graphical genotype and phenotypic data, and future extensions will incorporate the ability to select lines based on combinations of these three data types. Pedigree information can be input in either 'female parent'/'male parent' format or as a Purdy style string which greatly reduces the number of individual entries required in input files. For Purdy style input the software parses the pedigree string and builds the pedigree from the gleaned information. Purdy style pedigree strings are also generated dynamically for output. Extensive checking of pedigrees is carried out during the import process including checking of aliases/synonyms and selection histories. Graphical pedigree representations are provided to display either ancestors, descendants or both, for a selected line for a specified number of generations. The graphical genotype display draws a chromosome map for combinations of genotypes and chromosomes, colour-coded for identity by descent (IBD). IBD calculations recurse (to a specified number of generations) through the pedigree tracing each allele to determine the ancestral origin of each region. PBMAS utilises existing marker data for ancestors and descendants to infer missing values, for an inferred value to be accepted – the probability of it being correct must satisfy a predefined criteria. The inference of missing data is essential for the IBD calculations to work to their potential, and current development only incorporates basic inference mechanisms. There is large potential for further development in this area. The allele report in PBMAS calculates proportions of alleles derived from a specified target genotype for a selected list of genotypes and loci both by IBD and identity by state (IBS). Phenotypic data can be represented graphically as a chart for a selected group of trials, genotypes and traits.

Regarding the barley data analysis, a total of 258 alleles were identified across 93 genotypes from 56 loci, with an average of 4.6 alleles per locus. Table 1 (see Attachment 2) lists the location, allele sizes, gene diversity and standard errors for each locus. The gene diversity values were calculated for each SSR locus using Nei's unbiased statistic, $H = n(1 - \sum p_i^2) / (n - 1)$, where 'n' is the number of individuals analysed and 'p_i' is the frequency of the ith allele at one locus. The genotypic data set was divided between the ancestral genotypes and their descendants in order to observe specific allele substitution events and the effects of selection over time through the changes in allele frequencies. Figure 1 (see Attachment 1) details an overview of diversity changes across the genome between ancestors and descendants (e.g. chromosome 6H can be shown to have been heavily under selection through the decrease in gene diversity values from 0.73 in the ancestral lines to 0.5 in the descendants). This figure illustrates that as well as a shift in the frequency of alleles, there are also differences in the number of alleles between the two groups. Figure 2 gives a more detailed breakdown of a few key loci under selection within the NBIP barley gene pool, and Table 2 details the marker-trait associations for each locus.

The shifts in allelic profiles combined with information of the pedigrees of the genotypes involved, offers the opportunity to monitor the flow of alleles through the ancestral lineage and therefore identify regions of the genome that have been preferentially selected for or against. Constant selection for the same alleles in multiple pedigrees highlights associations between the markers and the superior alleles that influence the specific trait in question. The IBD graphical genotype approach, coded for within the PBMAS software, also offers a powerful methodology to trace markers through a pedigree and detect evidence of selection for particular alleles. A graphical genotype of chromosome 1H detailing the haplotypes of a subset of the NBIP pedigree, as calculated through IBD, is given in Figure 3. Figure 4 details the pedigree structure of the subset of genotypes selected. The whole genome pedigree mapping approach also provides a mechanism for QTL validation.

QTL validation has also been explored in a different approach combining single marker analysis to detect potential associations between marker (genotypic) classes and their respective phenotypic values with a locus-by-locus analysis of molecular variance (AMOVA) to obtain an estimate of how each locus contributes to the differentiation between phenotypic classes. This was performed on data sets combining genotypic and disease response data for all 93 genotypes, specifically for disease ratings for two pathotypes of NFNB, and three components of SFNB resistance – overall disease score, lesion size and extent of chlorosis. Table 3 details the loci identified with significant association with the disease scoring data for both pathotypes for NFNB. Figure 5 illustrates the genomic locations of these loci and details the studies that have also previously identified QTL for NFNB resistance in the same genomic location. The correspondence between the loci significantly

associated with both NFNB and SFNB resistance, as identified through the use of the barley pedigree set and those previously identified through conventional QTL mapping studies utilising segregating populations, illustrates the potential for extracting maximum value from whole genome mapping of genotypes in a known pedigree structure for QTL validation purposes.

The outcomes from this project will encourage more sophisticated use of markers in targeted breeding populations as selection is applied to conserve and combine critical genomic regions for both environmental adaptation and grain and end-use quality. The use of a focussed set of genotypes means that the marker technology will be quickly available for use with a large proportion of the gene pools in the respective case study breeding programs.

Other research

Further research and development (R&D) opportunities were identified during the course of this project to develop and implement practical selection tools for wheat and barley breeding programs that provide summaries integrating information from a variety of sources, including high throughput, whole-genome genotyping pedigree analysis based on IBD, phenotypic evaluation, and earlier QTL studies. Opportunities for building on the qualitative outcomes achieved in this project have been identified which included a) information about the frequency, distribution and ancestral origin of known QTL, and b) information about historical selection for particular regions. This will involve expanding both the number of markers used for genotyping and the number of lines genotyped in both wheat and barley.

Regarding the wheat component, further R&D opportunities have emerged regarding the generation of genotypic information using SSRs on the other 14 chromosomes of wheat, to complement the SSR and DArT data collected along eight of the 21 wheat chromosomes in this project. Some important traits for which QTL have been located since the planning stages of this project include resistance to CR and root lesion nematode (RLN). Other Australian Winter Cereals Molecular Marker Program (AWCMMP) projects (GA1) are underway to identify genetic regions associated with water absorption, extensibility, and a range of other important quality attributes. It is anticipated that by 2006 some of the genetic regions identified in these studies will be available for validation using this approach. Additionally, opportunities for studying a further pedigree based on the variety Hartog have emerged, using a combination of SSRs and DArT technology. The comparison of regions under selection in these two pedigrees will be very informative and may also provide insight into some of the epistatic reactions which may come in to play when genetic material from these two pedigrees is combined.

Regarding the barley component, further opportunities for exploring the potential of DArT markers for direct application in a molecular breeding strategy within the NBIP have emerged. The DArT markers could be used to validate QTL being developed in other AWCMMP projects (GA1) for feed quality and disease resistances and, through the dense marker coverage provided, to enhance our understanding of the genomic regions under selection within the breeding program.

In summary, the technology applications identified during the course of the project should be implemented and further built on in the wheat and barley breeding programs in order to enhance breeder's capacity for informed decision-making on new crosses and appropriate selection strategies, to produce relevant information for use in interpreting associations between traits and genomic regions under selection through the pedigrees targeted in each breeding program and for trait selection, parent characterisation and MARPR. The implementation of the PBWG mapping approach will allow for an enhanced understanding of the processes of selection within a breeding program, including identification of regions under selection, and will act as a basis of validation of markers and QTL for the more effective implementation of markers, through the confirmation of QTL location, the identification of markers which will work in other, distantly related populations, and in the identification of QTL which are effective in other populations.

Intellectual property summary

The PBMASS software will be freely available in its current form and with no guarantee of upgrades, as an executable file via CD following email request, to AWCMMP participants and GRDC affiliated breeding/research programs, with a licence agreement containing relevant disclaimers.

Additional information

Eisemann B, Banks P, Butler D, Christopher M, DeLacy I, Jordan D, Mace E, McGowan P, McIntyre L, Poulsen D, Rodgers D, Sheppard J, 2004. Pedigree-based genome mapping for marker-assisted selection and recurrent parent recovery in wheat and barley. A poster-paper presented at the 4th International Crop Science Congress, Brisbane 26 Sept-1 Oct 2004 and at the XIII Plant & Animal Genome Conference, San Diego, Jan 15-19 2005.

McIntyre CL, Christopher M, Rodgers D, Delacy I, Eisemann BR, 2004. Pedigree-based genome mapping for marker-assisted selection in wheat. A poster-paper presented at the Linkage Disequilibrium Workshop, Barossa Valley Apr 4-7 2004 and at the XII Plant & Animal Genome Conference, San Diego, Jan 10-14 2004.