Selection to accommodate genotype-by-environment interactions in the Northern Wheat Improvement Program

Summary
Wheat breeders must make decisions about advancing potential varieties through three stages of yield evaluation (Stage 1 is crossing and nursery observation) given an industry expectation of continual identification of superior varieties and therefore a time constraint on the time for testing. In Australian dryland environments, trials are subject to substantial yield variation due to inter-annual variability in climate, particularly rainfall and frost, which also interacts with regional variation in soil type and depth. These conditions generate substantial genotype by environment (GxE) interactions between genotype and environment that make it difficult to identify varieties that are superior in all locations and years. Given restricted resources in terms of the number of years and locations that can be tested, breeders wish to optimise the flow of potential varieties through the program.

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
1. This project has demonstrated that it is practical to build a simulation module of an operating wheat breeding program and to use simulation to evaluate alternative scenarios to optimise the breeding program. Although only limited information was available when constructing genetic models for the simulation experiments, the genetic models adequately represented the current state of knowledge regarding the Leslie Research Centre (LRC) wheat production environment, and they were suitable for providing guidance on the appropriateness of alternative multi-environment trial (MET) strategies in the face of GxE interactions.
2. The advantages of weighted selection of the Germplasm Enhancement Program (GEP) component of the National Wheat Improvement Program (NWIP) were confirmed by simulation (1999 report and Podlich et al., 1998). These simulations demonstrate that where the effects of GxE interactions can be differentiated in different environment types (based on location or other classifications), then selection can be improved by ensuring that trial results are weighted to balance the sampling of environments (locations and years) against actual occurrence of environments in the real world.
3. Given the effect of environment characterisation on the balance of variance components for G and GxE interaction, in the probe trials it is useful to apply environment characterisation (EC) in determining genotype best linear unbiased prediction (BLUP) values within the breeding program. Specific conclusions on how to assign weightings in different parts of the breeding program were not able to be made as there was insufficient information from breeding trials (compared with the GEP trials) to parameterise the simulation of environment types.

Recommendations
1. Simulation research, using the QU GENE application module ‘LRC’, showed that increasing the number of trials at each MET stage resulted in improved genetic gain when GxE interaction was important in the target population environment (TPE) (Jensen, 2003). However, diminishing returns were apparent as MET size increased, such that an increase in MET size from two to three resulted in a greater improvement than was seen as MET size increased from three to four. Importantly, for the LRC wheat breeding program, substantial benefits were shown to arise when preliminary yield evaluation (PYE) MET size was increased from two trials per year to three per year. Thus, it is important for the program to consistently realise its current target of three PYE trials per year. The loss of a single trial results in a substantial genetic penalty for the program. Wegener et al. (1999) suggested that it would be worthwhile to build extra trials into the program to mitigate the potential loss of a trial or trials, and the simulation studies provide extra support for this conclusion in relation to all of the MET stages. For the genetic models considered in the simulation studies, the optimum number of METs for each stage were six to eight for the PYE stage; seven to eight for the strain stage; and twenty for the
regional stage. These results are dependent on heritability, and the genetic models used in the study. Clearly, it is unrealistic for the LRC wheat breeding program to implement these large numbers of trials within their current resources, and nor should they until further information is available regarding the breeding program’s TPE and the genetic basis of the GxE interactions.

2. In line with previous experience (e.g. project UQ29), it can be further recommended that, combined with the use of improved experimental design and analysis, no more than two replications need be utilised at any stage. Given that more than two replicates have often been used in strain and regional trials, resource savings from reducing these might allow planting of an additional PYE and strain trial each year.

3. Germplasm is tested in two years for both the PYE and strain MET stages and three years for the regional MET stage. The simulation studies indicated that all yield data from a given MET stage should be used when making selection decisions in the second year (and third year for regional METs) of the stage (Jensen, 2003). Failing to use the previous information results in a reduction in gain from selection. Data from the PYE MET stage should only be considered when making selection decisions in later stages if it is weighted with the appropriate genetic covariance for additive genetic effects. The lines tested in the strain stage are derivatives of those selected from the PYE stage, and weighting the PYE data based on genetic covariance would appropriately account for the genetic relationship between the entities tested in the different MET stages. Using PYE MET data in later stages, without this weighting, can result in reduced genetic performance for the selected lines.

4. With respect to environment characterisation, the previous recommendation of Cooper et al. (1996) that testing programs try to include at least one ‘well managed’ (i.e. irrigated) site is supported by the analysis of the probe trials. This ensures that genotypes with substantial yield potential are not ‘overlooked’ should the sample of trials be particularly unrepresentative.

Outcomes

Expected Outcome (benefits)

Economic Outcomes

Improved varieties remain perhaps the most cost-effective methods of delivering research to growers. To the degree that recommendations are implemented by breeding programs, this research to increase the efficiency of wheat breeding programs allows realisation of an incremental permanent increase in return to the Australian grains industry from investment by growers, GRDC and taxpayers.

Achievements/Benefits

Overview of Project Achievements

The aims of this project were:

1. To optimise multi-environment testing in the three stages of testing in the NWIP.
2. To determine the influence of drought stress on GxE interactions of wheat yield in the NWIP, and
3. To evaluate and implement in the NWIP a weighted selection strategy that accommodates the effects of GxE interactions and increases the chances of making good selection decisions.

To enable optimisation, this research project developed a computer simulation model of the breeding program and assembled and analysed large numbers of historical trials to determine input parameters for the simulation. Given that the data assembly was delayed by changes in personnel, the simulations were run concurrently using a range of variance component values based on analyses from project UQ29. To further accommodate GxE interaction, simulation was used to classify the degree of drought experienced to try to explain differences in genotype performance.

Database development:

1. Update of the Queensland (QLD) Regional Variety Trial (RVT - Stage 4) database developed in UQ29. The database now contains all genotype mean Best Linear Unbiased Estimates (BLUEs) and error estimates from trials conducted between 1972 and 2001. Genotype names were standardised on QT names where appropriate.
2. Establishment of a database (1984-2001) of 191 strain (Stage 3) trials containing raw means and fitted (BLUEs) comprising...
97,566 observations. Raw data from these and PYE trials pre-1990 or so were gathered from computer file archives or re-punched from paper records. Fitted values were estimated following extensive re-analysis by fitting spatial and column-row statistical models to each trial. Genotype names were standardised on QT names so they could be linked to the pedigree database.

3. Establishment of a database (1984-2001) of 142 PYE (Stage 2) trials containing raw means and fitted BLUEs comprising 78,131 observations. Fitted values were obtained similarly to those in strain trials. Pedigree names were cleaned but not completely resolved for these trials. Extensive manual pedigree tracking has been begun by the QLD Primary Industries (QDPI) independently and also with assistance in the new project on association mapping (UQ00026).

4. Establishment of a database of BLUEs for 16-20 probe genotypes from 105 trials undertaken between 1988 and 2000 in 23 locations by Mark Cooper, Alan Peake, Nicole Jensen with UQ staff and QDPI breeders.

These databases provide the base information to evaluate alternative methods of analysing individual trials and optimisation of genotype selections when combining data within and across stages of selection. The probe database is complemented by a database of information (climate, soil, management data) that enables the trials to be simulated using the Agricultural Production Systems sIMulator (APSIM). The databases are a substantial resource for future study of wheat breeding programs in the northern region. The simulations that were undertaken by Jensen can be repeated to look specifically at weighted selection and allocation of METs using the new variance component values determined here.

Estimation of genetic parameters:

5. Software code written in S-Plus and utilising SAMM (the S-Plus implementation of ASREML). This code enabled automated computation of genetic parameters (variances, heritabilities and correlations) from datasets of regional, strain and PYE trials. It also created databases of these parameters as well as BLUE and BLUP estimates of genotype yields that can be used in further investigations. Databases were constructed of the fitted models, derived genetic parameters (including G and GxE variance, etc.) for all stages of selection.

These parameters can be used as inputs to the analyses of optimal trial numbers (using standard formulae) and to initialise simulations of the breeding program. They also allow, for the first time, direct comparison of genetic parameters from three stages of an Australian wheat breeding program over long time periods.

Software development:

6. Simulation of the breeding program

a. Details of all steps in the LRC pedigree breeding program were documented (see Chapters 3 and 4 of Jensen, 2003)
b. Early during the project, a QU-GENE module (LRC) was written to simulate the LRC wheat improvement breeding program in version 1 of QU-GENE (Jensen 2003).
c. Following the outbreak of yellow spot in 1998, another module (LOWRISK) was written to determine the feasibility of combining multiple, simply inherited traits into a single variety and was used to prepare a confidential report to GRDC.
d. The results of studies of the optimisation of the LRC program using the QU-GENE LRC module are given in Chapter 6 of Jensen (2003).
e. Given developments in UQ123, a new generic breeding module (QU-Line) was designed for version 2 of QU-GENE and has been parameterised to simulate the NWIP. The QU-Line module allows evaluation of alternative breeding scenarios, as well as weighted and unweighted selection strategies.

7. Crop simulation and EC tools

a. Database developed by QDPI (REMS), and partly funded by UQ113, to store weather, soil and management (sowing, fertiliser, irrigation, etc) data related to any trial.
b. Automated procedure to simulate a set of trials (using APSIM) and do EC using the statistical language.
c. Interpretation of breeding trials using environment characterisations.

Automated procedure using public R software to take EC results, combine results with observed yield data from the same trials and analyse the effects of weighted and unweighted selection.

When the above procedure was applied to the probe trial database undertaken over a period of 10 years, it was found that
stress patterns were not consistently separating central and southern locations, i.e. there is substantial overlap between the regions in the types of drought stress patterns being observed. The ratio of genotypic variance to that for GxE interaction was greater within stress types than within regions i.e. EC explains part of the GxE interaction that is not captured by location and year analyses.

Other research

In this study, the new variance components from early stages of selection were not able to be determined prior to the undertaking of the simulation study. Obviously, it would be advisable to repeat the simulation study using these new variance estimates as guides in parameterising the simulation.

An obvious extension of this research is to utilise historical climate and soil information from the different breeding trials to attempt to explain GxE interaction within these stages. For example, weather data and soil characterisations can now be obtained relatively simply as long as the location and planting dates of trials are known. Again once environmental characterisation data are available to define the TPE for the LRC wheat breeding program, the simulation experiments examining appropriate numbers of trials at each MET stage could be re-run.

Another key remaining question is how much information should be weighted from the high-input trials that seem to be quite useful predictors of performance. Both of these questions need to be addressed using the actual trial information and in combination with the simulations capability that is available for the LRC pedigree breeding program.

The databases developed in this project have already assisted in the development of ideas for the new GRDC project (UQ00026 Pedigree-based genome mapping for MAS and recurrent parent recovery in wheat and barley) that was begun by University of Queensland (UQ), QDPI and CSIRO last year. Large amounts of yield data (that was re-punched or organised here) are available to that project.

Intellectual property summary

At the outset of the project, it was indicated that the QU-Gene software and developments of it in the project would remain the property of UQ. QU-Gene and QU-Gene modules are available under licence from UQ. Licences vary in their specification and payment for usage (if applicable) depending on the prior intellectual property (IP) invested by licensees and developing partners. The new breeding module will be available for licence to Australian wheat breeding programs. A confidential report was made to GRDC on the simulation of a LOWRISK selection program.

Additional information

Publications


Confidential report to GRDC on simulation of LOWRISK breeding strategies for combining simply inherited traits in the Northern Wheat Improvement Program (submitted).