View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Sydney eScholarship

Marker/trait associations identified in spring wheat using 25 years of CIMMYT international trials

Arief V[1], DeLacy IH[1,2], Dieters MJ[1], Crossa J[4], Godwin I[1], Batley J[1], Davenport G[3], Dreisiqacker S[3], Duveiller E[3], Edwards D[2], Huttner E[5,6], Lambrides C[1], Manes Y[3], Payne T[3], Singh RP[3], Warburton M[3], Wenzl P[5,6], Kilian A[5,6], McLaren G[3,4], Braun H-J[3], Crouch J[3], Ortiz R[3], Basford KE[1,2]

[1]School of Land Crop and Food Sciences, The University of Queensland, [2]Australian Centre for Plant Functional Genomics (ACPFG), [3]International Maize and Wheat Improvement Center (CIMMYT), [4]International Rice Research Institute (IRRI), [5]Diversity Arrays Technology P/L Canberra, [6]Triticarte P/L Canberra.

ABSTRACT

This study identified marker/trait associations (MTA) by the joint analysis of phenotypic data from a comprehensive set of field trials and dense DArT genome scans (1,447 polymorphic markers). MTA were identified for 21 traits (three rusts, grain yield, five agronomic characters, two quality traits, and 10 other foliar diseases) using data collected from the first 25 years of CIMMYT's Elite Spring Wheat Yield Trials (ESWYTs). Genotypic data were generated for a set of 645 partially duplicated lines. Three structures were used: (1) family structure based on pedigree information, (2) temporal structure based on ESWYT cycle, and (3) spatial structure based on the CIMMYT mega-environment classification. MTA were identified for each trait using a t-test with a p-value less than or equal to 0.001 declared significant. This approach identified numerous associations for each trait. The DArT genome scans were consistent across duplicated lines and enabled the identification of introgressed segments based on haplotypes. The results of this study improved our understanding of the germplasm used in the CIMMYT breeding programs, including a better characterization of parental lines. The results will also assist the selection of new crosses and provide a path towards 'haplotype' breeding.

INTRODUCTION

Association analysis is commonly used to identify marker/trait associations (MTAs) for the identification of quantitative trait loci (QTL), i.e. genes controlling the expression of complex traits. In plants QTL research has mostly focused on the analyses of designed experimental populations (usually from a bi-parental cross) evaluated over a limited number of trials in one or a few years. However, this approach suffers from several disadvantages[1, 2] which result in the information often not being directly applicable to breeding programs[3]. The detection of MTAs using plant breeding populations is expected to overcome these limitations.

MTAs can be due to any factors that cause gametic phase disequilibrium, including linkage, selection, migration or drift[4]. These factors occur in non-random mating populations and are referred to as being structured[5]. A Plant Breeding Population (PBP) is typically a highly structured population; it consists of lines tested in an historical set of field trials which are the result of intense selection among and within a large number of genealogically connected families. It is suggested that integrating population structure[5] into the analysis is one strategy for increasing the proportion of identified MTA that are due to linkage. However, for a PBP, it is important to document every MTA regardless of what caused it.

Population structure can be known and may be based on: the family ancestry[6]; a temporal subset[7]; the spatial origin of test[6] of the lines; or a combination of the above. If the population structure is not known, then population structure can be inferred using phenotypic or marker data. However, it is preferable to use the known structure if it is available. In historical PBP data, the three known causes of structure are all available. In this study, we investigate MTA in a PBP using family ancestry (pedigree data), temporal subset (ESWYT entry), and spatial origin (agro-ecological zone of test) of the lines.

MATERIALS AND METHODS

The phenotypic data were obtained from the first 25 cycles of the CIMMYT's Elite Wheat Yield Trials (ESWYTs) for 21 traits (three rusts, grain yield, five agronomic characters, two quality traits, and 10 other foliar diseases) on 685 lines. The data were analysed using the restricted maximum likelihood[8], as implemented in the ASREML software[9] and with a separate model for each structure. The BLUPs (Best Linear Unbiased Predictions) derived from this analysis were then used in the association analysis.

Marker data for this analysis consisted of 1447 polymorphic DArT markers[10] (supplied by Triticarte Pty. Ltd.) for 599 of 685 lines tested in the ESWYTs, with several lines duplicated. Five hundred DArT markers had previously been assigned to wheat chromosomes of which 300 are presented in at least one of nine linkage maps supplied by Triticarte (www.triticarte.com.au/content/wheat_diversity_ analysis.html). The other markers were assigned to chromosomes based on the classification of the markers across lines. Each marker was allocated to the

chromosome of the first mapped marker with which it grouped. Classification of the markers across entries was conducted using the Czekanowski metric and modified Ward's method as the clustering strategy. Markers were then ordered using the proximity matrix for each chromosome.

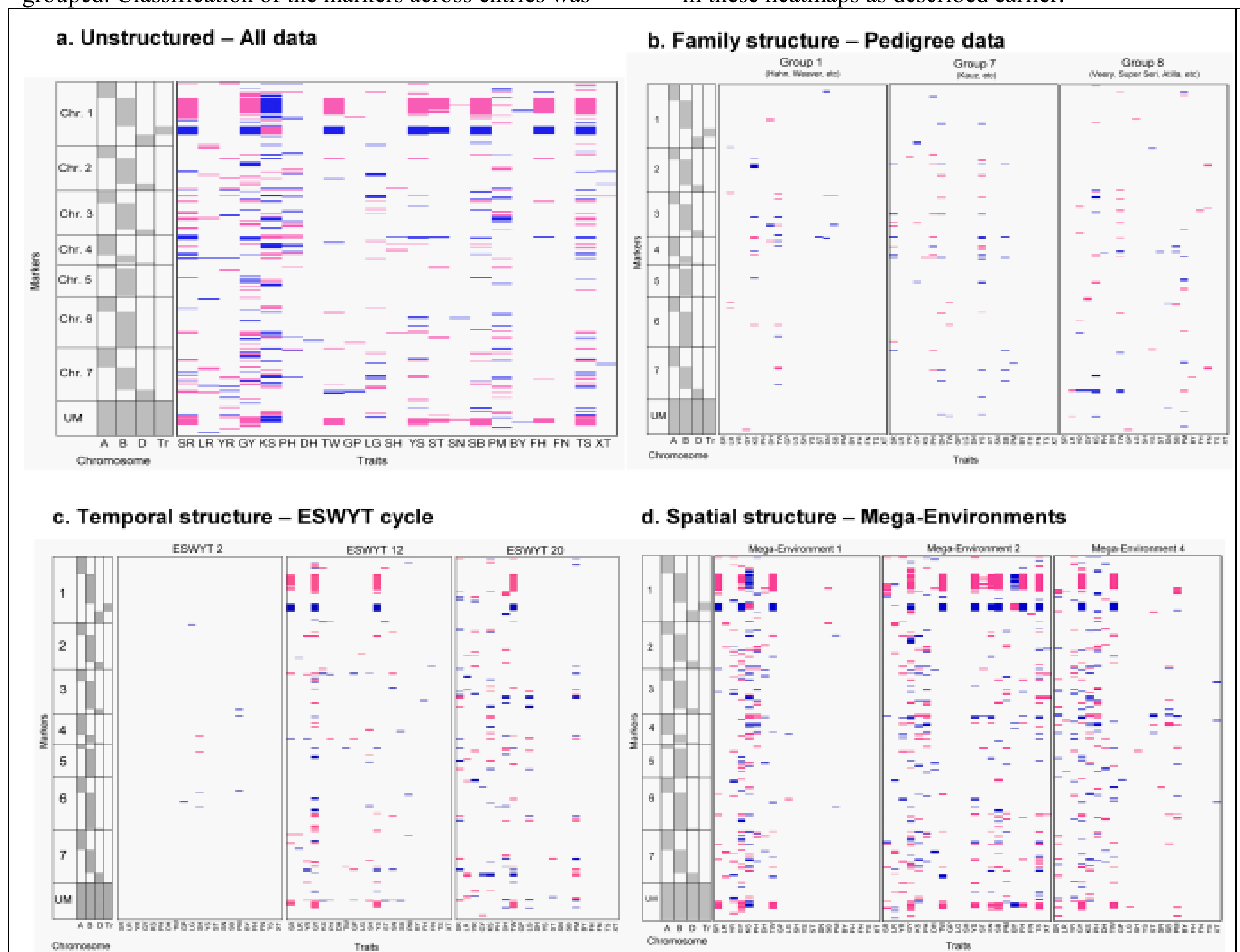

Fig. 1: Heatmap for marker-trait associations (MTA) results using (a) all data, (b) family structure, (c) temporal structure, and (d) spatial structure. Pink indicates a negative association and blue indicates a positive association. SR=stem rust; LR=leaf rust; YR=stripe rust; GY=grain yield; KS=kernel size; PH=plant height; DH=days to heading; TW=test weight; GP=grain protein; LG=lodging; SH=spike shattering; YS=stripe rust on the spike; ST=septoria trtitici blotch; SN=septoria nodorum blotch; SB=spot blotch; PM=powdery mildew; BY=barley yellow dwarf; FH=fusarium head blight; FN=fusarium leaf blotch; TS=tan spot; and XT=xanthomonas

Family structures were obtained using pattern analysis on the coefficient of parentage (COP) matrix among the 599 lines. The COP matrix was calculated based on pedigree information using the BROWSE module in IWIS3[11]. A modification of Ward's method was used as the clustering strategy, and the 11-group-level was chosen for investigation of structure. Spatial structures were based on the CIMMYT mega-environment (ME) classifications[12], and temporal structures were based on each cycle of ESWYTs.

Association analyses were conducted using a simple t-test. A threshold value of $p \leq 0.001$ was used to declare a MTA significant. The results were presented as heatmaps which display the positive and negative associations for the 21 traits. The markers were ordered in these heatmaps as described earlier.

RESULTS AND DISCUSSIONS

A large number of MTA were identified for most of the 21 traits investigated (Fig. 1) and association profiles differed for each structure used in the analysis. For diseases, the profiles also were dependent on the presence of the pathogen in any year or location. Haplotype analysis using duplicate lines showed that DArT markers were consistent within lines.

In all data sets, most of MTA identified occurred as aggregates of two or more markers. Similar results were also observed when a heatmap was generated using the Synthetic-Opata map (data not presented here). For this paper we considered and aggregated MTAs as indicative of a single QTL if it consisted of three or more markers.

Based on MTA using unstructured data (all data), we identified markers that were suspected to indicate the 1B1R translocation. These results were confirmed by comparing haplotypes of the lines known to be carrying or not carrying the translocation. These markers were assigned to the 1B1R translocation and used in the heatmaps. The same approach was used to identify markers for Lr19 translocation in chromosome 7D. Several markers were identified as potentially indicating this translocation. Since none of the DArT markers were derived from germplasm containing the Lr19 translocation, we used the absence of those markers to indicate the presence of Lr19 translocation. We also found that the presence of those markers were negatively associated with leaf rust.

MTA for family structure based on pedigree information showed different profiles for the three groups, of the 11 groups identified, presented here (Figure 1b). MTA for the 1B1R translocation disappeared, because within each group there were either no or very few polymorphisms for the 1B1R translocation, i.e. most of the group members were uniform for the 1B1R translocation.

Different MTA profiles were observed in each of the first 25 ESWYTs, as illustrated for ESWYTs 2, 12 and 20 in Figure 1c – due to both the subset of lines tested as well as line by year interaction. The lines tested in the ESWYT changed from ESWYT to ESWYT with most lines tested in only one ESWYT, a few repeated but none repeated in all 25 ESWYTs examined. Although MTA profiles were different across ESWYT, there was no obvious pattern observed from year to year in the number of MTA detected. In some ESWYTs very few MTA were identified, such as in ESWYT2, while in others, such as ESWYT 12 and 20, many were identified (Figure 1c). MTA identified for the 1B1R translocation also changed depending on the lines tested in each ESWYT. The numbers of lines with the 1B1R translocation first increased and then started to decline in the later ESWYTs.

MTA in three mega-environments (Figure 1d) also showed different profiles, especially for diseases. ME1, ME2 and ME4 are all low latitude (less than 40°). ME1 consists of low-rainfall and irrigated, ME2 of high-rainfall, and ME4 of water stressed environments[12]. The diseases present in these MEs also differ. ME1 has rusts and karnal bunt; ME2 has rusts, Septoria, scab, barley yellow dwarf, *Xanthomonas*, powdery mildew, bunts, tan spot and root diseases; ME4 has rusts, Septoria, scab, bunts, smuts, root and crown rots, and nematodes[12]. The differences in disease occurrence were reflected in the MTA results (Figure 1d). A great number of MTA for disease were observed in ME2 compared to the other MEs (Figure 1d). In the three MEs presented in this paper, the same set of lines was used, and hence the differences observed in the MTA were not due to differences in the sets of lines tested.

The approach reported here provides a deeper understanding of what is happening in a breeding program, and will assist in the selection of new crosses and provide a path towards haplotype breeding. The results provided here were obtained without complete genetic map information for the markers but will be substantially enhanced when such a map becomes available.

ACKNOWLEDGEMENT

- University of Queensland and CIMMYT for strategic initiative funds
- Endeavour International Postgraduate Research Scholarship

REFERENCES

1 Eisemann B. *et al.*, paper presented at the 4th International Crop Science Congress: New directions for a diverse planet, Brisbane Australia 2004.
2 Parisseaux B., Bernardo R. (2004) *Theoretical and Applied Genetics,* **109:** 508.
3 Beavis W. D., in *Molecular dissection of complex traits* Paterson A. H., Ed. (CRC Press, Boca raton, 1998) pp. 145-162.
4 Jannink J.-L., Walsh B., in *Quantitative genetics, genomics and plant breeding* Kang M. S., Ed. (CABI Pub, Wallingford, U.K., 2002) pp. 59-68.
5 Pritchard J. K., Rosenberg N. A. (1999) *Am. J. Hum. Genet.,* **65:** 220.
6 Jakobsson M. *et al.* (2008) *Nature,* **451:** 998.
7 Beer S. C. *et al.* (1997) *Journal of Agricultural Genomics,* **3 [online]:** URL: http://wheat.pw.usda.gov/jag/papers97/paper197/indexp197.html.
8 Peterson R. F., Campbell A. B., Hannah A. E. (1948) *Canadian Journal of Research Section C,* **26:** 496.
9 Gilmour A. R., Gogel B. J., Cullis B. R., Welham S. J., Thompson R. (2006) *ASReml User Guide Release 2.0* (VSN International Ltd., Hemel Hempstead, HP1 1ES, UK), pp. 267.
10 Jaccoud D., Peng K., Feinstein D., Kilian A. (2001) *Nucleic Acids Research,* **29:** e25.
11 McLaren C. G. (2007) *IRRI, Philippines. http://cropwiki.irri.org/icis/index.php/TDM_GMS_Browse. Cited 4 August 2007*.
12 Rajaram S., van Ginkel M., Fischer R. A., in *Proceedings of the 8th International Wheat Genetic Symposium* Li Z. S., Xin Z. Y., Eds. (Beijing, China, 1995) pp. 1101-1106.