Genome-wide association mapping in barley

Yahiaoui S., Casas A.M., Moralejo M.A., Medina B., Gracia M.P., Ciudad F.J., Codesal P., Molina-Cano J.L., Lasa J.M., Igartua E.

in

Molina-Cano J.L. (ed.), Christou P. (ed.), Graner A. (ed.), Hammer K. (ed.), Jouve N. (ed.), Keller B. (ed.), Lasa J.M. (ed.), Powell W. (ed.), Royo C. (ed.), Shewry P. (ed.), Stanca A.M. (ed.).
Cereal science and technology for feeding ten billion people: genomics era and beyond

Zaragoza : CIHEAM / IRTA
Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 81

2008
pages 123-126

Genome-wide association mapping in barley

S. Yahiaoui*, A.M. Casas, M.A. Moralejo**, B. Medina*, M.P. Gracia*, F.J. Ciudad***, P. Codesal***, J.L. Molina-Cano**, J.M. Lasa* and E. Igartua***

*Department of Genetics and Plant Production, Aula Dei Experimental Station, CSIC, P.O. Box 202, E-50080 Zaragoza, Spain

**Centre UdL-IRTA, Av. Rovira Roure 191, E-25198 Lérida, Spain

***ITA, Instituto de Tecnología Agraria, Junta de Castilla y León, P.O. Box 172, E-47071 Valladolid, Spain

Introduction

Conventional methods for mapping QTL in plants involve the use of traditional biparental populations, and linkage mapping (LM). Association mapping (AM), based on linkage disequilibrium (LD), offers an alternative method for QTL mapping (Remington *et al*., 2001; Flint-Garcia *et al*., 2003, 2005). It utilizes ancestral recombination events to make marker-phenotype associations. Association analysis has been employed only recently in plants, due in part to the confounding effects of population structure (Thornsberry *et al*., 2001; Kraakman *et al*., 2004; Aranzana *et al*., 2005; Breseghello and Sorrells, 2006). We tested the feasibility of carrying out association analysis between several morphologic and agronomic traits in 156 landraces of the Spanish Barley Core Collection (SBCC), for which polymorphic data were available.

In a previous study, 225 barley accessions were genotyped with 73 markers, distributed along the barley genome (Fig. 1). Population structure was evaluated with the software STRUCTURE (Falush *et al*., 2003), using data on 64 SSR. For association analysis, only markers with band frequencies in between 5% and 95% were used. To assess the association of each maker locus with traits, we fitted multiple regression models of the trait response on either the set of allele indicator variables (models W, P4), or on that set and another one including four variables describing the probability of membership of each genotype to each of the four genetic groups (model P), as reported by Kraakman *et al*. (2006). Multiple testing was addressed using the Bonferroni correction.

Results and discussion

The 225 genotypes were assigned to four genetically distinct populations, clearly separating the Spanish landraces from other European materials (Table 1). Association analysis was carried out only for 156 SBCC landraces and population structure was taken into account in the analysis (Table 2).

Table 1. Distribution of genotypes in four Model-based populations

Germplasm groups	Number of entries	Populations inferred with Structure			
		Pop. 1	Pop. 2	Pop. 3	Pop. 4
Spanish 6-row	152	18	0	49	85
Spanish 2-row	11	0	9	1	1
Reference 6-row	33	31	2	0	0
Reference 2-row	29	6	23	0	0

If population structure was not considered, a high number of marker-trait associations were detected (311, model W). Many of these associations could be false positives and, indeed, that number was reduced to 65 when population structure was included in the model (model P, Table 2). Similarly, by controlling population structure, Thornsberry *et al*. (2001) decreased the number of false-

positive associations by almost fivefold for some traits in an association study of flowering time in maize. On the other hand, Flint-Garcia *et al*. (2005) pointed out that structured association analysis of traits highly correlated with population structure will result in false negatives (lack of power), as could be the case for traits such as thousand kernel weight or length of spike in our data. All of them are related to spike morphology, which is highly correlated with population structure, since one of the populations is almost exclusively made of two-row entries (Pop. 2). We also evaluated the associations only within population 4 (model P4), since it showed a low LD level (data not shown). Some of the detected associations in this group were not found in the larger sample.

Table 2. Number of loci that revealed positive associations with the phenotypic variables considered, according to the analysis model used (without population structure: W; with population structure: P, or within population 4: P4)

Variables	W	P	P4	Variables	W	P	P4
Yield (YLD)	21	9	2	Malt extract (MET)	1	0	0
Heading (HED)	17	12	16	Stem pigmentation (SPG)	3	3	0
Lodging (LDG)	14	2	1	Auricle pigmentation (APG)	2	0	4
Plant height (PHT)	8	0	0	Hairy sheath (HSH)	13	4	0
Test weight (TWG)	34	0	1	Glume and glume awn (GGA)	9	4	0
Thousand kernel weight (TKW)	19	2	2	Lemma type (LMT)	18	3	0
Spikes m^{-2} (SPM)	3	0	0	Lemma awn/ hood (LMA)	0	0	0
Kernel m^{-2} (KNM)	0	0	1	Lemma awn barbs (LMB)	0	0	0
Kernel per spike (KPS)	6	0	0	Glume colour (GLC)	0	0	0
Leaf rust (LRS)	21	4	4	Awn colour (AWC)	0	0	0
Net Blotch (NBL)	2	0	0	Length of rachilla hairs (LRH)	11	5	3
Powdery mildew (PML)	14	0	0	Length of spike (LSP)	32	5	3
Screening (SCR)	9	2	5	Spike density (DSP)	3	2	1
Plumpness (PPS)	9	0	0	Spikelets per spike (SPS)	1	0	1
Protein (PRT)	9	0	1	Row number (RWN)	32	8	0

The positions of markers that showed high association with trait values, were compared with those of loci or QTL known to govern the trait in other studies, based on a consensus bin map (Fig. 1), and on the revision of the results found in a number of studies corresponding to 60 biparental populations (not shown). In spite of a high rate of false positives, due to population structure, apparent associations were still identified, for most of the phenotypic traits tested: grain yield, days to heading, lodging, thousand grain weight, leaf rust resistance, row number, stem pigmentation, spike density or spike length, when population structure was taken into account in the analysis. In several cases, the associations detected agreed with the positions previously identified in classical mapping populations.

For the trait days to heading, 12 associations were detected and 8 of them were in the same position as QTL reported in the literature. Conversely, the 2 loci more significantly associated with resistance to leaf rust under natural infection (Bmac399 and MWG699) did not correspond to any known gene. The marker Bmag223, that is closely linked to the gene srh for rachilla hair length (Costa et al., 2001), showed a very high association with length of rachilla hairs. Kraakman et al. (2006) also identified an association between this marker and the same trait in a study of modern spring barley cultivars. The loss in statistical power and the uncertainty about the significance of associations due to multiple testing in AM seem counterbalanced by the increase in the scope of genetic diversity analyzed, and the permanent nature of the populations and data analyzed. In view of these results, the authors are optimistic about the feasibility of AM in barley and about the value of the SBCC as an appropriate resource for gene and allele discovery.

Some of the associations revealed with this landrace collection occurred in regions containing QTLs detected in biparental mapping populations, but we also found potentially new marker-trait associations that seem promising and could be studied in the future, in specific biparental crosses segregating for those regions. LM and AM can be used as complementary methods.

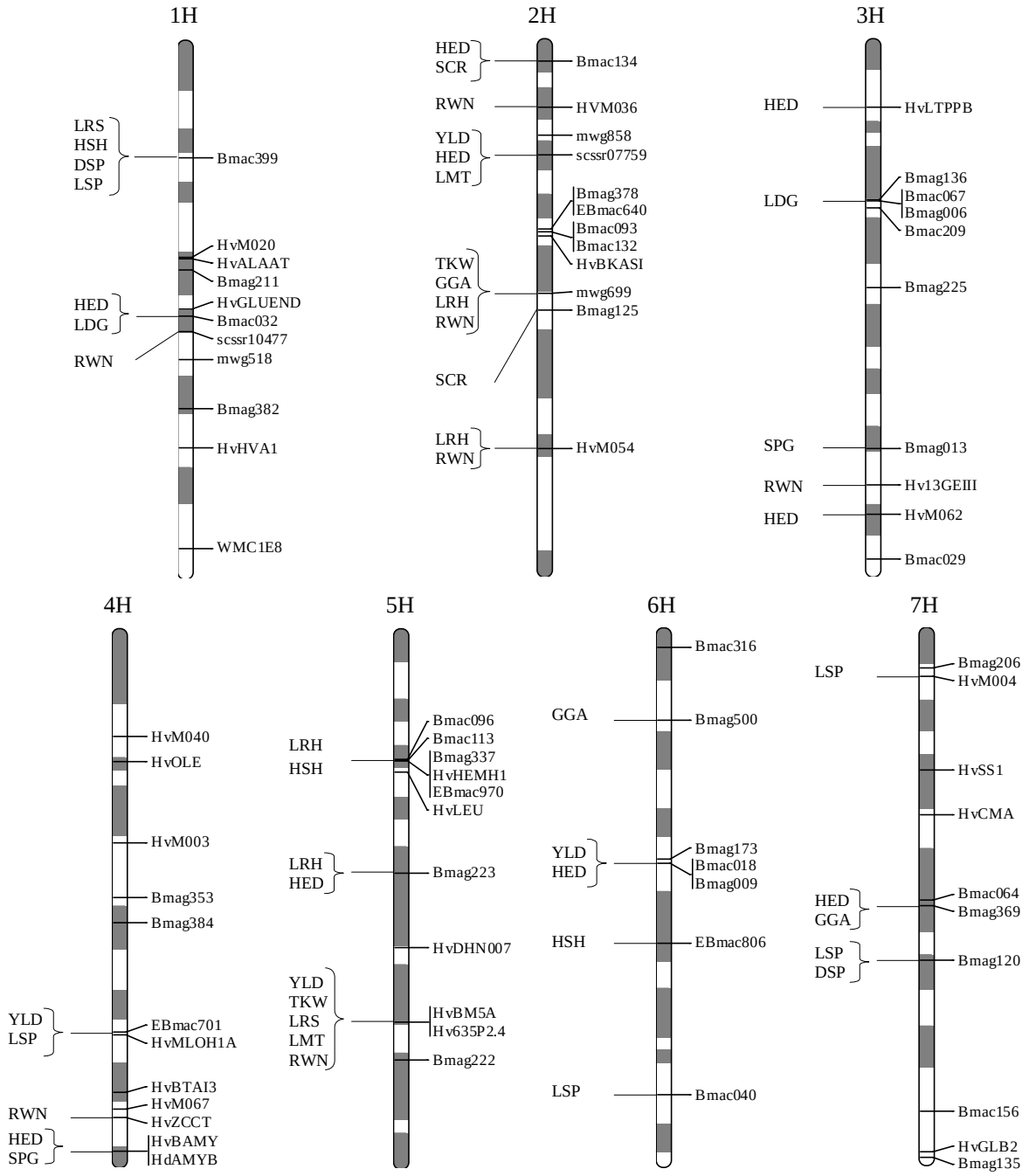

Fig. 1. Overview of a consensus map showing the position of the markers tested, barley chromosome bins, and the associations found with different traits in the model P (taking into account population structure).

This work was supported by INIA (project RTA01-088-C3) and by the European Regional Development Fund (ERDF).

References

Aranzana, M.J., Kim, S., Zhao, K., Bakker, E., Horton, M., Jacob, K., Lister, C., Molitor, J., Shindo, C., Tang, C., Toomajian, C., Traw, B., Zheng, H., Bergelson, J., Dean, C., Marjoram, P., Nordborg, M. (2005). *Plos Genetics,* 1: 531-539.

Breseghello, F., Sorrells, M.E. (2006). *Genetics,* 172: 1165-1177.
Costa J.M., Corey A., Hayes, P.M., Jobet, C., Kleinhofs, A., Kopisch-Obusch, A., Kramer, S.F., Kudrna, D., Li, M., Riera-Lizarazu, O., Sato, K., Szűcs, P., Toojinda, T., Vales, M.I., Wolfe, R.I. (2001). *Theor Appl Genet.,* 103: 415-424.
Falush, D., Stephens, M., Pritchard, J.K. (2003). *Genetics,* 164: 1567-1587.
Flint-Garcia, S.A., Thornsberry, J.M., Buckler, E.S. (2003). *Ann Rev Plant Biol.,* 54: 357-374.
Flint-Garcia, S.A., Thuillet, A.C., Yu, J., Pressoir, G., Romero, S.M., Mitchell, S.E., Doebley, J. Kresovich, S., Goodman, M.M., Buckler, E.S. (2005). *Plant J.,* 44: 1054-1064.
Kraakman, A.T.W., Niks, R.E., van den Berg, P.M.M.M., Stam, P., Van Eeuwijk, F.A. (2004). *Genetics,* 168: 435-446.
Kraakman, A.T.W., Martínez, F., Mussiraliev, B., van Eeuwijk, F.A., Niks, R.E. (2006). *Mol. Breeding,* 17: 41-58.
Remington, D.L., Thorsnsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M.M., Buckler, IV E.S. (2001). *Proc Natl Acad Sci.*, USA, 98: 11479-11484.
Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D., Buckler, IV E.S. (2001). *Nat Genet.,* 28: 286-289.