Genome-Wide Associations for Water-Soluble Carbohydrate Concentration and Relative Maturity in Wheat Using SNP and DArT Marker Arrays

Ben Ovenden,*1 Andrew Milgate,* Len J. Wade,1,2 Greg J. Rebetzke,§ and James B. Holland**††

*New South Wales Department of Primary Industries, Yanco Agricultural Institute, Yanco, New South Wales 2703, Australia, 1New South Wales Department of Primary Industries, Wagga Wagga Agricultural Institute, Wagga Wagga, New South Wales 2650, Australia, 2Charles Sturt University, Graham Centre, Wagga Wagga, New South Wales 2678, Australia, §Commonwealth Scientific and Industrial Research Organisation, Agriculture and Food, Canberra, Australian Capital Territory 2601, Australia, **Plant Science Research Unit, United States Department of Agriculture-Agricultural Research Service, and ††Department of Crop and Soil Sciences, North Carolina State University, Raleigh, North Carolina 27695-7620

ORCID ID: 0000-0003-2015-1650 (B.O.)

ABSTRACT Improving water-use efficiency by incorporating drought avoidance traits into new wheat varieties is an important objective for wheat breeding in water-limited environments. This study uses genome-wide association studies (GWAS) to identify candidate loci for water-soluble carbohydrate accumulation—an important drought-avoidance characteristic in wheat. Phenotypes from a multi-environment trial with experiments differing in water availability and separate single nucleotide polymorphism (SNP) and diversity arrays technology (DArT) marker sets were used to perform the analyses. Significant associations for water-soluble carbohydrate accumulation were identified on chromosomes 1A, 1B, 1D, 2D, and 4A. Notably, these loci did not collocate with the major loci identified for relative maturity. Loci on chromosome 1D collocated with markers previously associated with the high molecular weight glutenin Glu-D1 locus. Genetic × environmental interactions impacted the results strongly, with significant associations for carbohydrate accumulation identified only in the water-deficit experiments. The markers associated with carbohydrate accumulation may be useful for marker-assisted selection of drought tolerance in wheat.

KEYWORDS water-soluble carbohydrates, nonstructural carbohydrates, association analysis, genotype-by-environment interaction, molecular marker

Reduction in grain yield and quality due to drought decrease the sustainability of farming systems, and threatens global food security (Ray et al. 2012; Reynolds et al. 2016). Incorporating traits that improve water-use efficiency (WUE) in water-limited environments into elite breeding germplasm is an important aim for wheat genetic improvement (Rebetzke et al. 2009; Reynolds et al. 2015). Water soluble carbohydrate (WSC) accumulation and remobilization are promising traits that could contribute to improved grain-filling under water-limited conditions, and, consequently, improved WUE (Bidinger et al. 1977; Pheloung and Siddique 1991; Gebbing and Schnyder 1999; Foulkes et al. 2007; Paskowski et al. 2016). Carbohydrate accumulation occurs when the crop synthesizes assimilate at a rate greater than sink requirement. In wheat, most of the carbohydrate is stored in the form of fructans, with a minor component of sucrose and hexose (Schnyder 1993; Wardlaw and Willenbrink 1994). Both the accumulation and remobilization of WSC is modified by environmental conditions that alter the balance between sources and sinks of assimilate. In particular, the availability of source carbon (as sucrose) affects accumulation (Xue et al. 2013). The WSC can be remobilized for use in growth or respiration (Kiniry 1993). However, the main sink for remobilization is the developing grain (Schnyder 1993; van Herwaarden et al. 1998; Takahashi et al. 2001), with remobilized WSC contributing as much as 30–50% of grain yield
under terminal drought conditions, and 10–20% under well-watered conditions (Bidinger et al. 1977; Pheloung and Siddique 1991; Schnyder 1993; Gebbing and Schnyder 1999; Piaskowski et al. 2016).

Flowering time is a key trait associated with WSC accumulation owing to the nature of WSC accumulation across crop growth stages (Passioura 1996; Rebetzke et al. 2008). Accumulation of WSC increases from before anthesis to a peak at 7–20 d after anthesis (Gebbing 2003; Ehdiae et al. 2008; Zhu et al. 2010) where WSC concentration (WSCC) can reach as much as 40% of total stem weight (Schnyder 1993). After anthesis, WSC levels decline due to remobilization to other sinks. Under water deficit conditions, this peak can sometimes occur before anthesis (Goggin and Setter 2004), and remobilization is earlier and proportionally greater (Bidinger et al. 1977; Virgona and Barlow 1991).

A number of studies have reported on genetic control and quantitative trait loci (QTL) for WSC accumulation and related characters in wheat (Snape et al. 2007; Yang et al. 2007; Rebetzke et al. 2008; McIntyre et al. 2010; Pinto et al. 2010; Bennett et al. 2012). Mapping populations typically varied for the major developmental genes for photoperiod sensitivity and reduced plant height, which can indirectly cause much of the observed phenotypic variability for grain yield and other traits (Rattey et al. 2009; Bennett et al. 2012; Edae et al. 2014). The biparental populations assessed in Rebetzke et al. (2008) were segregating for the photoperiod sensitivity loci Ppd-D1 and the semi-dwarfing loci Rht-B1 and Rht-D1, and these loci collocated with QTL for WSCC, WSC total amount per square meter, and WSC per tiller.

Genomic strategies show significant promise for the improvement and understanding of drought tolerance traits (Langridge and Reynolds 2015). The primary objectives of this study were to identify markers associated with WSCC by genome-wide association studies (GWAS) and characterize the dependency of marker associations on environments. We conducted GWAS separately for two molecular marker sets (SNP and DArT markers) and for each experiment to assess the variability of marker-trait associations due to genotype × environment (G × E) interaction (Oldmeadow et al. 2011; Zila et al. 2013). We also conducted GWAS for relative maturity to ascertain if loci with significant associations with WSCC were due to the indirect effects of relative maturity on WSCC.

**MATERIALS AND METHODS**

**Genotypes used in this study**

The genotypes for the GWAS analyses were selected from evaluation trials conducted in multiple environments in 2009 and 2010. Each field trial contained 990 genotypes. Some genotypes were not repeated at every experiment, with a total of 1314 genotypes tested. Thus, for relative maturity GWAS, all 990 genotypes were used in 2009 experiments, and 972 genotypes were used in 2010 experiments. The accumulation of WSC varies with plant development (Ehdiae et al. 2008), so the subset used for WSCC GWAS consisted of 312 breeding lines from the 2009 experiments constrained to a 3–5 d difference in anthesis date as well as 46 commercially grown varieties. For the second year of this study in 2010, except for 11 breeding lines excluded from the overall experiments, the same breeding lines and varieties were evaluated for WSCC.

**Experimental design and site locations**

Experiments with contrasting irrigation and rainfed treatments were grown at Yanco Agricultural Institute (Yanco, Australia) and Coleambally Community Experimental Demonstration Farm (Coleambally, Australia) in 2009 and 2010. A split-plot design was used, in which the main-plot factor was irrigation treatment (irrigated or rainfed), and the 990 genotype entries (including the subset of genotypes for WSC measurement) were the subplot factor. There were two replicates of each treatment at each location. Genotype placement was optimized with the spatial design package DiGGer (Coombes 2002). For the laboratory phase measuring WSC using near-infrared spectroscopy (NIRS), an experimental design structured by day of measurement and NIRS instrument carousel and well was implemented to account for extraneous variation originating from laboratory processes. Samples from both field locations were pooled into one laboratory phase experimental design for each year, and the placement of genotypes within the laboratory experimental phase was also optimized with DiGGer (Coombes 2002), with partial replication of 20% of experiment field plots sampled (i.e., a replication level of 1.2), following the methods in Cullis et al. (2006) and Smith et al. (2006).

All experiments were sown on a full soil profile of moisture, achieved by flood irrigating each site 4–6 wk before sowing, so that the focus on water deficit conditions would be in the later stages of crop growth. Sowing dates were targeted for the first 2 wk of May, and sowing rates were 115 kg ha⁻¹ in the irrigated and 70 kg ha⁻¹ in the rainfed treatments, respectively. Presowing nitrogen was targeted to be 120 kg N ha⁻¹ from the combination of deep soil nitrogen (following soil testing—data not shown) and fertilizer applied at sowing. Irrigated experiments were fertilized supplementally through the growing season to a total of ~300 kg N ha⁻¹ consistent with predicted N demand by the crop. Experiments were subject to a strict weed, pest, and disease control regime to maximize yield potential. Soil moisture at each experiment was monitored using gypsum block AM400 soil moisture data loggers (Hansen, Wenatchee, WA). Onsite weather stations (Davis Instruments, Hayward, CA) were used to record rainfall and air temperature. Irrigation treatments were flood irrigated when soil water potential fell below ~75 kPa. Both sites had below average rainfall and above average temperatures in 2009, while conditions at both sites in 2010 had higher rainfall and lower temperatures than average.

**Measurements and observations**

Relative maturity at a common date around the median flowering date for all entries within each experiment was determined using the Zadoks decimal score for plant development (Zadoks et al. 1974). Scores for each field experiment were taken when most lines were in the range Z50–Z69 (head emergence to completion of anthesis).

Lines selected for WSCC analysis were sampled from a 50-cm long section of row (0.09 m²) when the irrigated treatments at each site were ~180° d postanthesis, following the sampling method of Rebetzke et al. (2008). For WSC analysis, ~5–10 stalks (including leaves, leaf sheaths, and heads, but not senesced plant material) were subsampled from each biomass sample, and ground to pass through a 2 mm-sized screen. Ground biomass samples were homogenized, desiccated, and scanned by NIRS with a Bruker Multi-purpose Analyzer (Bruker Optik GmbH, Ettlingen, Germany) and OPUS software (version 5.1). Scanned spectra were transformed using the first derivative and multiplicative scatter correction. Calibrations to obtain predicted WSCC values from spectra measurements were constructed using the “Quant 2 Method” component of the OPUS software with a randomly selected 10% subset of samples. WSCC for the calibration samples was determined using the alkaline ferricyanide method (Piltz and Law 2007).
Statistical methods for phenotype values

A multiplicative mixed linear model was used to analyze the multi-experiment phenotype data for both traits following Gilnour et al. (1997) and Beeck et al. (2010). The linear mixed model is given by

\[ y = X\tau + Z_g g + Z_u u + \eta \]

where \( y \) is the \((n \times 1)\) data vector of the response variable across \( p \) experiments with \( N \) plots per experiment; \( \tau \) is a \((t \times 1)\) vector of fixed effects (including linear trends across range and row) with associated design matrix \( X \). The term \( u \) is a random component with associated design matrix \( Z_u \) and contains experimental blocking structures used to capture extraneous variation (including field range and row for both traits, and laboratory day of measurement, NIRS carousel and well for WSCC only).

The residual error is \( \eta = (\eta_1, \ldots, \eta_p) \), which, at the \( j \)th experiment, was assumed to have distribution \( \eta_j \sim N(0, \sigma^2_{\eta_j} I_N) \), where \( \sigma^2_{\eta_j} \) is the residual variance for the \( j \)th experiment and \( R_j \) is a matrix that contains a parameterization for a separable autoregressive AR1 @AR1 process to model potential spatial correlation of the observations for the relative maturity analysis. For WSCC analysis, unique residual variances for each year were modeled.

The term \( g \) is a random component with associated design matrix \( Z_g \) used to model the genotype within experiment effects, which combine the genotype and \( G \times E \) interaction effects. Organizing the genotype within environment effects as a matrix of rows corresponding to genotypes and columns corresponding to environments facilitates modeling \( g \) as a multiplicative \( k \)-factor analytic (FA) model (Smith et al. 2001):

\[ g = (A \otimes I_m) f + \delta \]

where \( A \) is a matrix with \( j \)th column containing the \( j \)th factor loadings for the \( p \) experiments, \( f \) is a vector of genotype scores across the \( p \) experiments, and \( \delta = (\delta_1, \ldots, \delta_p) \) is a residual genetic term, where, at the \( j \)th experiment, \( \delta_j \sim N(0, \sigma^2_{\delta_j} I_m) \), and \( \sigma^2_{\delta_j} \) is the residual genetic variance for the \( j \)th experiment. The term \( I_m \) represents an \( m \times m \) identity matrix.

The variance model for the combined genotype and \( G \times E \) effects is given by

\[ \text{var}(g) = \left( AA' + \psi \right) \otimes I_m \]

where \( \psi \) is a diagonal matrix of the \( p \) environment specific variances.

For each analysis, the most parsimonious FA model was identified using the Akaike Information Criterion (AIC) (Akaike 1974). The nongenetic random effects were maintained in the model if they were significant according to log likelihood ratio tests relative to the full model with all nongenetic random effects (Stram and Lee 1994). Fixed effects were tested for significance using Wald F-statistics (Kenward and Roger 1997).

Empirical best linear unbiased predictors (E-BLUPs) for phenotypic values were obtained from the FA models for each individual experiment (Kelly et al. 2007; Cullis et al. 2010). For both relative maturity and WSCC, experiments were clustered using the matrix of genetic correlations between experiments (Cullis et al. 2010). All data were analyzed using the software package ASReml-R (Butler et al. 2009), in the R statistical software environment (R Development Core Team 2012).

Genotyping methods

Two separate marker sets were used: 985 lines from the overall experiment were genotyped using the Illumina 9k Infinum iSelect beadchip array (Cavanagh et al. 2013), resulting in 4883 polymorphic SNPs across the population. Similarly, 955 lines were genotyped with Diversity Arrays technology (DArT) (Akbari et al. 2006) resulting in 2013 polymorphic markers across the population. Genotyping included all 358 lines phenotyped for WSCC. Genotype information for SNP and DArT marker datasets were prepared separately for analysis using the R software package Symbreed (Wimmer et al. 2012). Imputation of missing values (3.5% for SNPs and 15% for DArTs) was performed using the software package Beagle (Browning and Browning 2009). Each marker dataset was filtered for duplicated and nonmonomorphic markers, as well as markers with minor allele frequency of <5%. The resulting 4162 SNP markers and 1773 DArT markers were used to compute a separate scaled identity by descent relationship matrix (K) after Endelman and Jannink (2012) for each marker dataset.

Consensus maps were used for marker physical positions. For the DArT dataset this study used the Wheat Interpolated Maps (version 6) as a reference to locate the positions of DArT markers (Dr Andrzee Kilian, Diversity Arrays Pty Limited, personal communication), and for the SNP dataset the 9K Consensus Map (version 4) was used (Dr Matthew Hayden, DEPI Victoria, personal communication.).

Linkage disequilibrium analysis

Patterns of linkage disequilibrium (LD) in the SNP and DArT marker sets were estimated using the methods of Breseghello and Sorrells (2006). Pairwise LD estimates \( (r^2) \) were calculated with the software package PLINK (Purcell et al. 2007) for unlinked loci pairs and for syntenic loci separately. Syntenic \( r^2 \) was plotted against pairwise genetic distance from the consensus maps for all chromosomes on each genome with a second degree locally weighted polynomial regression (LOESS) curve fitted to each scatter plot (Cleveland 1979). All of the unlinked \( r^2 \) estimates were square-root-transformed to approximate a normal distribution, and the 5% quantile of that distribution was determined following Breseghello and Sorrells (2006). The intersection of the LOESS curve and the 5% quantile for unlinked marker pairs was taken as an estimate of the extent of LD decay within each genome following Laido et al. (2014).

GWAS methods

Separate association analyses for each trait at each experiment were performed using the phenotype E-BLUPs described above. Associations using SNP and DArT marker sets were performed separately. The compressed mixed linear model approach (Zhang et al. 2010) was implemented in the R software package Genome Association and Prediction Integrated Tool (GAPIT) (Lipka et al. 2012) as follows:

\[ y = X\beta + Z_u u + \eta \]

where \( y \) is the vector of E-BLUPs for one trait measured in one experiment, \( \beta \) is a vector of fixed effects for the corresponding design matrix \( X \), including a molecular marker. The vector of overall genetic line effects \( u \) with associated design matrix \( Z_u \) is modeled as \( Var(u) = K\sigma^2_u \); where \( K \) is the relationship matrix and \( \sigma^2_u \) is the estimated additive genetic variance. \( \eta \) is the vector of random residuals. False discovery rates (FDR) were estimated separately for each experiment following Benjamini and Hochberg (1995) with a nominal threshold of 10% to declare significant associations.

Data availability

Supplemental Material, File S1 contains a detailed description of all Supplemental files. File S2 contains phenotype information for WSCC.
RESULTS

Genotype \times environment interactions

Consistent with experimental weather conditions, genetic correlations for WSCC between experiments showed two distinct environment groups, with the Yanco and Coleambally 2009 rainfed experiments, which experienced terminal water deficit, forming one cluster, and the other experiments collectively representing a well-watered cluster. Within these clusters, the genetic correlations were maintained at $r_G = 0.87$ between the two rainfed experiments that make up the water deficit environment cluster, and ranged from $r_G = 0.74–0.98$ in the well-watered environment cluster. Between the two clusters, genetic correlations ranged from $r_G = 0.02–0.35$. For relative maturity, no environmental clustering was evident, and genetic correlations between all experiments were very high, ranging from $r_G = 0.92$ to $r_G = 0.99$.

LD and minor allele frequency

We evaluated the distribution of LD within chromosomes separately for each marker set and for each of the three wheat genomes. LD was more extensive with respect to linkage distances within the D genome for the DArT marker set, as the average DArT marker LD did not decrease below the 5% quantile for unlinked marker pairs ($r^2 = 0.0456$) until the distance between markers was $\geq 25$ cM (Figure 1). By comparison, the average LD was $< 0.0456$ at distances of 16–18 cM for the A and B genomes respectively. In contrast, LD decreased below the 5% quantile ($r^2 = 0.0470$) for the SNP marker set at 21 cM for both the D and B genomes, while LD for the A genome was higher at 24 cM (Figure 2).

The minor allele frequency distribution for SNP markers was similar to that for the DArT markers, although the DArT markers had a lower proportion of the rarest allele class (MAF = 5–7.25%; Figure 3). SNPs

Figure 1 Pairwise LD estimates ($r^2$) plotted against Euclidian pairwise marker distances for markers on the same consensus chromosome for the DArT marker set.

Figure 2 Pairwise LD estimates ($r^2$) plotted against Euclidian pairwise marker distances for markers on the same consensus chromosome for the SNP marker set.

File S3 contains phenotype information for relative maturity at flowering time. File S4 contains SNP genotypes for each individual. File S5 contains DArT genotypes for each individual.
with MAF below 5% were not included in the analysis because of their reduced power for GWAS.

GWAS

Both marker sets displayed a low degree of population structure (Figure 4) with no obvious patterns among genotypes. For genomic relationship matrices computed from either SNP or DArT markers separately, the first two eigenvectors collectively explained only 15% of the variation in genomic relationships, indicating a lack of strong subpopulation structure in the association population. Only markers identified as statistically significant (with \( P \)-values below the 10% FDR threshold) in more than one experiment in the association studies were considered reliable associations (Table 1 and Table 2).

The GWAS for relative maturity using the DArT markers identified only one marker (on chromosome 2D), which was detected as significant in three experiments, all of them in the well-watered environment cluster (10COLE_IRR, 10COLE_RFD, and 10YANA_IRR; Table 2). This marker is located >30 cM from the marker associated with WSCC on the same chromosome. In contrast, GWAS for relative maturity using the SNP marker set identified 17 markers significant in more than one experiment (Table 2). SNP associations with relative maturity within specific experiments did not follow the pattern of environmental clustering observed for WSCC. Rather, five of 17 SNP associations with relative maturity were observed in all experiments. Some environmental-specificity was observed for relative maturity associations, but this did not reflect differences between well-watered and water-limited conditions. For example, four of 17 SNP associations were detected at both treatments within the same year-location combination (Table 2), suggesting G \( \times \) E patterns for relative maturity due to local weather patterns rather than water availability.

Among the SNPs associated with relative maturity, 11 markers were located within 3 cM of each other on the consensus map on chromosome 5A, and three markers collocated on chromosome 5B. Additionally, one marker was identified on each of chromosomes 2D, 4B, and 5D. Four of the significant markers on 5A and the marker on 2D were detected in all eight experiments.

The range in MAF of trait-associated loci ranged from 0.107 to 0.491 for the DArT marker set, and at least one relatively rare SNP allele was detected for relative maturity (MAF = 0.058 on chromosome 4B). The highest MAF for associated loci in the SNP marker set was 0.400.

DISCUSSION

Comparison of analysis at individual experiments

GWAS results for both relative maturity and WSCC show that significant associations can be experiment-specific, and relatable to the overall G \( \times \) E relationships between experiments for each trait. Genetic correlations between all experiments were very high for relative maturity (indicating limited G \( \times \) E interaction), and significant loci were detected in all experiments. In contrast, G \( \times \) E was strong for WSCC, with factor analysis revealing two distinct environment types, corresponding to well-watered and water-deficit environments. Reflecting these differences in overall G \( \times \) E patterns between relative maturity and WSCC, several markers were associated with relative maturity...
GWAS detected associations between markers on chromosomes 1A, 1B, 1D, 2D, and 4A with WSCC measured in water-limited conditions. No markers were associated with both WSCC and relative maturity, in contrast to Rebetzke et al. (2008), where flowering time loci explained large proportions of variation for WSCC. The results herein may reflect the sampling methods used for phenotyping WSCC, or because the association population lines for WSCC were selected to be constrained for development.

Among the markers significantly associated with WSCC in the two water deficit environment experiments, wPt-7359 on 1B has not been previously reported in trait associations, but wPt-800147 on 4A was associated with grain yield under water deficit conditions (Masoudi et al. 2015). Marker wPt-3743 on 1D was associated with a range of other traits, including grain yield and resistance to yellow rust, powdery mildew, and leaf rust (Crossa et al. 2007), grain yield and spike length under salt stress conditions (Azadi et al. 2015), and spike number (tiller number) per plant (Cai et al. 2014). Marker wPt-9592 on 1A was previously associated with grain yield under water deficit conditions (in particular, terminal drought; Crossa et al. 2007), heading date after vernalization (Le Gouis et al. 2012), and seed dormancy (Singh et al. 2010).

Marker wPt-3743 on chromosome 1D was previously reported to be located near the high molecular weight glutenin Glu-D1 locus and the storage protein activator gene locus SPA-D (Plessis et al. 2013; Deng et al. 2015; Jin et al. 2015). Marker wPt-733835 is also in this region (Jin et al. 2015). The Glu-D1 locus is important for selection as, along with the Glu-A1 and Glu-B1 loci, it is responsible for a large percentage of the phenotypic variation for dough quality. The combination of glutenin alleles present at the Glu-1D locus will largely determine the end use and grain quality class of wheat varieties (Payne 1987; Whiting 2004). Glutenin protein complexes play an important role in conferring elasticity and strength in wheat dough (Plessis et al. 2013), and the Glu loci have been shown to collocate with QTL for nitrogen and dry matter accumulation in grain (Charmet et al. 2020).

Rebetzke et al. (2008) identified QTL for WSCC per tiller that collocated with the glutenin loci Glu-A1, and Glu-B1. The Glu-D1 and SPA-D loci contribute to phenotypic variation for grain yield and grain number through the plant response to nitrogen (Bordes et al. 2013). Potentially the Glu loci could be involved with the inheritance of WSCC through an interaction between nitrogen use, tiller number, and grain weight. WSCC is influenced by nitrogen content, as higher nitrogen availability in the plant drives sink demand for assimilate (van Herwaarden et al. 2015). The

| Marker     | Chromosome | Distance (cM) | P.Value | Minor Allele Frequency | FDR Value | Range | Nearby Genes |
|------------|------------|---------------|---------|------------------------|------------|-------|--------------|
| wPt-9592   | 1A         | 62.7          | 0.000520 | 0.000789               | 0.491      | 0.0725 – 0.0999     | wPt-9592 |
| wPt-665784 | 1A         | 69.3          | 0.000520 | 0.000789               | 0.491      | 0.0725 – 0.0999     | wPt-665784 |
| wPt-7359   | 1A         | 83.3          | 0.000520 | 0.000789               | 0.491      | 0.0725 – 0.0999     | wPt-7359 |
| wPt-800147 | 4A         | 1.56 x 10^-5  | 0.000520 | 0.000789               | 0.491      | 0.0725 – 0.0999     | wPt-800147 |

**Table 1** Markers significant for water-soluble carbohydrate concentration at >1 experiment for the DArT marker set.
| Experiments | Marker | Chromosome | Distance (cM) | P-Value Range | Minor Allele Frequency | FDR Value Range | Nearby Genes |
|-------------|--------|------------|--------------|---------------|------------------------|----------------|--------------|
| 10COLE_IRR, 10COLE_RFD, 10YANA_IRR, 10YANA_RFD | wPt-730744 | 2D | 73.0 | 4.85 $\times 10^{-5}$ – 7.91 $\times 10^{-5}$ | 0.161 | 0.0482–0.0787 | Ppd-D1 |
| 09COLE_IRR, 09COLE_RFD, 09YANA_IRR, 09YANA_RFD, 10YANA_IRR, 10YANA_RFD | wsnp_CAP12_c812_428290 (IWA989) | 2D | 57.9 | 3.91 $\times 10^{-16}$ – 1.414 $\times 10^{-14}$ | 0.134 | 1.50 $\times 10^{-12}$ – 5.41 $\times 10^{-11}$ | Ppd-D1 |
| 09YANA_IRR, 09YANA_RFD | wsnp_BE422566B_Ta_1_2 (IWA76) | 4B | 43.3 | 3.74 $\times 10^{-1}$ – 0.000171 | 0.0582 | 0.0498–0.507 |
| 09COLE_IRR, 09COLE_RFD, 09YANA_IRR, 09YANA_RFD, 10YANA_IRR, 10YANA_RFD | wsnp_AJ612027A_Ta_2_1 (IWA1) | 5A | 66.2 | 6.30 $\times 10^{-9}$ – 3.74 $\times 10^{-6}$ | 0.136 | 0.0258–0.0446 | Vrn-A1 |
| 09COLE_IRR, 09COLE_RFD, 09YANA_IRR, 09YANA_RFD, 10YANA_IRR, 10YANA_RFD | wsnp_C508_1008029 (IWA4087) | 5D | 61.0 | 3.90 $\times 10^{-5}$ – 0.000332 | 0.0202 | 0.0258–0.0978 | Vrn-A1 |

Table 2 Markers significant for relative maturity at >1 experiment for the DArT and SNP marker datasets

Experiment is given as year-site-irrigation treatment. Chromosome and position are from the consensus maps. Nearby genes are from CMAP GrainGene database [http://wheat.pw.usda.gov/cmap/](http://wheat.pw.usda.gov/cmap/) searches within the LD blocks estimated for each genome.
Loci associated with relative maturity

Significant associations were identified in the vicinity of a number of the known major flowering time loci, including the main photoperiod and vernalization loci under selection in wheat breeding germplasm pool globally (Yan et al. 2004; Eagles et al. 2009, 2010; Cane et al. 2013; Slafier et al. 2015). Both DArT and SNP markers were identified close to the photoperiod-sensitivity locus Ppd-D1 on chromosome 2D on the consensus map (Table 1). Given the importance of the Ppd-D1 locus to selection of growth duration and adaptability (Kamran et al. 2014; Slafier et al. 2015), the SNP and DArT markers identified here may prove useful to supplement other markers for this locus, such as those outlined in Cane et al. (2013).

The analyses were able to detect significant associations near the Vrn-A1, Vrn-B1, and Vrn-D1 loci across multiple experiments, although only SNP markers were identified as statistically significant, including 11 markers near Vrn-A1, three markers near Vrn-B1, and one marker was detected near Vrn-D1 (Table 1). For the genotypes in this study, variation in at these loci would be expected to include alleles for both spring and winter alleles, as well as winter alleles that confer different vernalization requirements (Eagles et al. 2010, 2014; Harris et al. 2017). One marker identified (wsnp_Af6120277_6_2) was also reported to be associated with the Vrn-A1 locus by Lopes et al. (2015). Three markers identified on 5B are within 10 cm of the Vrn-B1 locus reported by Guedira et al. (2014). The single marker on chromosome 5D associated with relative maturity (wsnp_Ex_c508_1008029) at both 10YANA_IRR and 10YANA_experiments corresponds to the vicinity of the Vrn-D1 locus (Eagles et al. 2009).

An additional marker associated with relative maturity at both the 09YANA_IRR and 09YANA_RF experiment was located on chromosome 4B (wsnp_BE4225666_6_1). This marker was significant in both the water deficit environments as well as the well-watered environments in this study. QTL for heading date on this chromosome have been previously reported by Hanoqc et al. (2007), Griffiths et al. (2009), and Le Gous et al. (2012), and photoperiod sensitivity QTL on this chromosome have been reported by Shindo et al. (2003) and Sourdille et al. (2003).

Comparison of DArT and SNP molecular marker sets

DArT markers were developed using methylation-sensitive restriction enzymes, and, as such, can represent methylation polymorphisms that provide both genetic and epigenetic information (Wenzl et al. 2004; Akbari et al. 2006). It is also possible for DArT markers to be located in insertion/deletion sites (indels), although ~80% are SNPs (Kilian et al. 2005). The minor allele frequency distribution for both marker sets was close to uniform across most of the allele frequency range, in contrast to the expected inflation of rare alleles expected for markers under drift–mutation equilibrium (Hamblin et al. 2011). This frequency spectrum in the SNP marker set was noted previously by Cavanagh et al. (2013), who concluded “The observed MAF is the consequence of intentional bias in SNP selection, where common alleles were favored by choosing more broadly distributed SNPs.”

Our genotype set had more than twice as many polymorphic SNP markers than DArT markers, and the DArT marker set had a higher proportion of missing data. Across the A and B genomes, DArT markers exhibit a more rapid breakdown of LD than SNP markers, but the pattern was reversed in the D genome. The longer linkage blocks in the D genome for both SNP and DArT markers is consistent with previous reports (Wang et al. 2014). The differences in marker density, distribution and LD structure between the two marker sets are the most likely causes of the observed differences in association analyses.

Conclusions

This study highlights the need to characterize G × E interactions in multi-environment datasets, and to define target populations of environments for marker-trait associations. These populations of environments define the scope of inference for interpreting GWAS results. In this study, we identified two clusters of experiments based on their genotypic correlations for the expression of WSCC. Marker associations for WSCC were identified only in the water deficit experiments, which represented a minority of the experiments; these associations would have been missed if the trait values were averaged across all experiments.

The loci identified for WSCC have both previously been associated with performance under water limited conditions, but did not reflect linkage to major effect relative maturity loci. The marker on 1D colocalizes with the Glu-D1 locus, which may have some pleiotropic effect on WSCC. These reported associations may be useful for marker-assisted selection of WSCC in water-limited environments, independent of relative maturity.

ACKNOWLEDGMENTS

We thank Kerry Schirmer and Aaron Hutchison for their expert technical contributions and data collection. We also thank the referees for their improvements to the manuscript. The breeding lines used in this study were kindly provided by Australian Grains Technologies, InterGrain, HRZWheats (now Dow Seeds), Sydney University, and LongReach Plant Breeders. The authors gratefully acknowledge the Grains Research and Development Corporation (GRDC) of Australia funding for this project (ICF00007), and a Grains Industry Research Scholarship for B.O. The American Australian Association is gratefully acknowledged for the ConocoPhillips Education Fellowship awarded to B.O.

LITERATURE CITED

Akaike, H., 1974 A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19(6): 716–723.

Akbari, M., P. Wenzl, V. Caig, J. Carling, L. Xia et al., 2006 Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor. Appl. Genet. 113(8): 1409–1420.

Azadi, A., M. Mardi, E. Hervan, S. Mohammadi, F. Moradi et al., 2015 QTL mapping of yield and yield components under normal and salt-stress conditions in bread wheat (Triticum aestivum L.). Plant Mol. Biol. Rep. 33(1): 102–120.

Beec, C. P., W. A. Cowling, A. B. Smith, and B. R. Cullis, 2010 Analysis of yield and oil from a series of canola breeding trials. Part I. Fitting factor analytic mixed models with pedigree information. Genome 53(11): 992–1001.

Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57: 289–300.

Bennett, D., A. Izanloo, M. Reynolds, H. Kuchel, P. Langridge et al., 2012 Genetic dissection of grain yield and physical grain quality in bread wheat (Triticum aestivum L.) under water-limited environments. Theor. Appl. Genet. 125(2): 255–271.

Bidinger, F., R. B. Musgrave, and R. A. Fischer, 1977 Contribution of stored pre-anthesis assimilate to grain yield in wheat and barley. Nature 270(5636): 431–433.

Bordes, J., C. Ravel, J. P. Jaubertie, B. Duperrier, O. Gardet et al., 2013 Genomic regions associated with the nitrogen limitation response revealed in a global wheat core collection. Theor. Appl. Genet. 126(3): 805–822.
Breshegghlo, F., and M. E. Sorrells, 2006  Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics 172: 1165–1177.

Brown, W. L., and S. R. Browning, 2009  A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84(2): 210–223.

Butler, D., B. R. Cullis, A. R. Gilmour, and B. J. Gogel, 2009  ASReml-R Reference Manual, Queensland Department of Primary Industries and Fisheries, Brisbane, QLD.

Cane, K., H. A. Eagles, D. A. Laurie, B. Trevaskis, N. Vallance et al., 2013  Ppd-B1 and Ppd-D1 and their effects in southern Australian wheat. Crop Pasture Sci. 64: 100–114.

Cavanagh, C. R., S. Chao, S. Wang, B. E. Huang, S. Stephen et al., 2013  Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc. Natl. Acad. Sci. USA 110(20): 8057–8062.

Charmet, G., N. Robert, G. Brunaud, L. Linossier, P. Martre et al., 2005  Genetic analysis of dry matter and nitrogen accumulation and protein composition in wheat kernels. Theor. Appl. Genet. 111(3): 450–500.

Cleveland, W. S., 1979  Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 74(368): 829–836.

Coombes, N. E., 2002  The reactive tabu search for efficient correlated experimental designs. Ph.D. Thesis, Liverpool John Moores University, Liverpool, UK.

Crossa, J., C. Riveros, E. G. Holliday, R. Scott, P. Moscato et al., 2015  Genetic dissection on wheat flour quality traits in two related populations. Euphytica 203(1): 221–235.

Cui, F., C. Zhao, A. Ding, J. Li, L. Wang et al., 2014  Construction of an integrative linkage map and QTL mapping of grain yield-related traits using three related wheat RIL populations. Theor. Appl. Genet. 127(3): 659–675.

Cullis, B. R., A. B. Smith, and N. E. Coombes, 2006  On the design of early generational variety trials with correlated data. J. Agric. Biol. Environ. Stat. 11(4): 381–410.

Curling, T., D. A. Laurie, B. Trevaskis, N. Vallance et al., 2013  Ppd-B1 and Ppd-D1 and their effects in southern Australian wheat. Crop Pasture Sci. 64: 100–114.

Cavanagh, C. R., S. Chao, S. Wang, B. E. Huang, S. Stephen et al., 2013  Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc. Natl. Acad. Sci. USA 110(20): 8057–8062.

Charmet, G., N. Robert, G. Brunaud, L. Linossier, P. Martre et al., 2005  Genetic analysis of dry matter and nitrogen accumulation and protein composition in wheat kernels. Theor. Appl. Genet. 111(3): 450–500.

Cleveland, W. S., 1979  Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 74(368): 829–836.

Coombes, N. E., 2002  The reactive tabu search for efficient correlated experimental designs. Ph.D. Thesis, Liverpool John Moores University, Liverpool, UK.

Crossa, J., C. Riveros, E. G. Holliday, R. Scott, P. Moscato et al., 2015  Genetic dissection on wheat flour quality traits in two related populations. Euphytica 203(1): 221–235.

Cui, F., C. Zhao, A. Ding, J. Li, L. Wang et al., 2014  Construction of an integrative linkage map and QTL mapping of grain yield-related traits using three related wheat RIL populations. Theor. Appl. Genet. 127(3): 659–675.

Cullis, B. R., A. B. Smith, and N. E. Coombes, 2006  On the design of early generational variety trials with correlated data. J. Agric. Biol. Environ. Stat. 11(4): 381–410.

Curling, T., D. A. Laurie, B. Trevaskis, N. Vallance et al., 2013  Ppd-B1 and Ppd-D1 and their effects in southern Australian wheat. Crop Pasture Sci. 64: 100–114.

Cavanagh, C. R., S. Chao, S. Wang, B. E. Huang, S. Stephen et al., 2013  Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc. Natl. Acad. Sci. USA 110(20): 8057–8062.

Charmet, G., N. Robert, G. Brunaud, L. Linossier, P. Martre et al., 2005  Genetic analysis of dry matter and nitrogen accumulation and protein composition in wheat kernels. Theor. Appl. Genet. 111(3): 450–500.
Passioura, J. B., 1996 Drought and drought tolerance. Plant Growth Regul. 20(2): 79–83.

Payne, R. L., 1987 Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. Annu. Rev. Plant Physiol. 38(1): 141–153.

Pheloung, P., and K. Siddique, 1991 Contribution of stem dry matter to grain yield in wheat cultivars. Funct. Plant Biol. 18(1): 53–64.

Piaskowski, J. L., D. Brown, and K. Garland Campbell, 2016 Near-infrared calibration of soluble stem carbohydrates for predicting drought tolerance in spring wheat. Agron. J. 108(1): 285–293.

Pitzl, J., and D. Law, 2007 AFIA-Laboratory Methods Manual. Australian Fodder Industry Association Inc, Balwyn, VIC.

Pinto, R. S., M. Reynolds, K. Mathews, C. L. McIntyre, J.-I. Olivares-Villeges et al., 2010 Heat and drought adaptive QTL in a wheat population designed to minimize confounding agronomic effects. Theor. Appl. Genet. 121(6): 1001–1021.

Plessis, A., C. Ravel, J. Bordes, F. Balfourier, and P. Martre, 2013 Association study of wheat grain protein composition reveals that gliadin and glutenin composition are trans-regulated by different chromosome regions. J. Exp. Bot. 64(12): 3627–3644.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira et al., 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81(3): 559–575.

R Development Core Team, 2012 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. Available at: http://www.R-project.org/

Rathey, A., R. Shorter, S. Chapman, F. Drecrer, and A. van Herwaarden, 2009 Variation for and relationships among biomass and grain yield component traits conferring improved yield and grain weight in an elite wheat population grown in variable yield environments. Crop Pasture Sci. 60(8): 717–729.

Ray, D. K., N. Ramankutty, N. D. Mueller, P. C. West, and J. A. Foley, 2012 Recent patterns of crop yield growth and stagnation. Nat. Commun. 3: 1293.

Rebetzke, G. J., A. F. van Herwaarden, C. Jenkins, M. Weiss, D. Lewis et al., 2008 Quantitative trait loci for water-soluble carbohydrates and associations with agronomic traits in wheat. Aust. J. Agric. Res. 59(10): 891–905.

Rebetzke, G. J., S. C. Chapman, C. L. McIntyre, R. A. Richards, A. G. Condon et al., 2009 Grain yield improvement in water-limited environments, pp. 215–249 in Wheat: Science and Trade, edited by Carver, B. F.. Wiley-Blackwell, Ames, IA.

Reynolds, M., M. Tattariss, C. M. Cossani, M. Ellis, K. Yamaguchi-Shinozaki et al., 2015 Exploring genetic resources to increase adaptation of wheat to climate change, pp. 355–368 in Advances in Wheat Genetics: From Genome to Field: Proceedings of the 12th International Wheat Genetics Symposium, edited by Ogihara, Y., S. Takumi, and H. Handa. Springer, Tokyo, Japan.

Reynolds, M. P., P. Quilligan, P. K. Aggarwal, K. C. Bansal, A. J. Cavaleri et al., 2016 An integrated approach to maintaining cereal productivity under climate change, Glob. Food Secur. 8: 9–18.

Rusuksa, S., D. Lewis, G. Kennedy, R. Furbank, C. Jenkins et al., 2008 Large scale transcriptome analysis of the effects of nitrogen nutrition on accumulation of stem carbohydrate reserves in reproductive stage wheat. Plant Mol. Biol. 66(1): 15–32.

Schnyder, H., 1993 The role of carbohydrate storage and redistribution in the source-sink relations of wheat and barley during grain filling—a review. New Phytol. 123(2): 233–245.

Shindo, C., H. Tsuji moto, and T. Sasaki ma, 2003 Segregation analysis of heading traits in hexaploid wheat utilizing recombinant inbred lines. Hereditity 90(1): 56–63.

Singh, R., M. Matus-Cá diz, M. Baga, P. Hucl, and R. Chibbar, 2010 Identification of genomic regions associated with seed dormancy in white-grained wheat. Euphytica 174(3): 391–408.

Slaf er, G. A., A. G. Kantolic, M. L. Appendino, G. Tranquilli, D. J. Miralles et al., 2015 Genetic and environmental effects on crop development determining adaptation and yield, pp. 285–319 in Crop Physiology, Ed. 2, chap. 12, edited by Sadras, V. O., and D. F. Calderini. Academic Press, San Diego.

Smith, A., B. Cullis, and R. Thompson, 2001 Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57(4): 1138–1147.

Smith, A., P. Lim, and B. R. Cullis, 2006 The design and analysis of multi-phase plant breeding experiments. J. Agric. Sci. 144(5): 393.

Snape, J. W., M. J. Foulkes, J. Simmonds, M. Leverington, L. J. Fish et al., 2007 Dissecting gene × environmental effects on wheat yields via QTL and physiological analysis. Euphytica 154(3): 401–408.

Sourdille, P., T. Cadalen, H. Guyomarč, J. Snape, M. Perretant et al., 2003 An update of the Courtort × Chinese Spring intervarietal molecular marker linkage map for the QTL detection of agronomic traits in wheat. Theor. Appl. Genet. 106(3): 530–538.

Stram, D. O., and J. W. Lee, 1994 Variance components testing in the longitudinal mixed effects model. Biometrics 50: 1171–1177.

Takahashi, T., P. M. Chevalier, and R. A. Rupp, 2001 Storage and remobilization of soluble carbohydrates after heading in different plant parts of a winter wheat cultivar. Plant Prod. Sci. 4(3): 160–165.

van Herwaarden, A. F., J. F. Angus, R. A. Richards, and G. D. Farquhar, 1998 ‘Hanging-off’, the negative grain yield response of dryland wheat to nitrogen fertiliser II. Carbohydrate and protein dynamics. Aust. J. Agric. Res. 49(7): 1083–1094.

Virgonza, J. M., and E. W. R. Barlow, 1991 Drought stress induces changes in the non-structural carbohydrate composition of wheat stems. Funct. Plant Biol. 18(3): 239–247.

Wang, S. D., Wong, K. Forrest, A. Allen, S. Chao et al., 2014 Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol. J. 12(6): 787–796.

Wardlaw, I. F., and J. Willenbrink, 1994 Carbohydrate storage and mobilisation by the culm of wheat between heading and grain maturity: the relation to sucrose synthase and sucrose-phosphate synthase. Funct. Plant Biol. 21(3): 255–271.

Wenzl, P., J. Carling, D. Kudrna, D. Jaccoud, E. Huttner et al., 2004 Diversity arrays technology (DArT) for whole-genome profiling of barley. Proc. Natl. Acad. Sci. USA 101(26): 9915–9920.

Whiting, D., 2004 Wheat Varieties in Australia, 1968–2001. D. Whiting, Snowtown, SA.

Wimmer, V., T. Albrecht, H.-J. Auinger, and C.-C. Schön, 2012 Symbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics 28(15): 2086–2087.

Xue, G.-P., J. Drenth, D. Glassop, M. Koosker, and C. L. McIntyre, 2013 Dissecting the molecular basis of the contribution of source strength to high fructan accumulation in wheat. Plant Mol. Biol. 81(1–2): 71–92.

Yan, L., M. Helguera, K. Kato, S. Fukuyama, J. Sherman et al., 2004 Allelic variation at the VRN-1 promoter region in polyploid wheat. Theor. Appl. Genet. 109(8): 1677–1686.

Yang, D.-L., R.-L. Jing, X.-P. Chang, and W. Li, 2007 Identification of quantitative trait loci and environmental interactions for accumulation and remobilization of water-soluble carbohydrates in wheat (Triticum aestivum L.) stems. Genetics 176: 571–584.

Yu, M., S.-L. Mao, G.-Y. Chen, Z.-E. Pu, Y.-M. Wei et al., 2014 QTLs for uppermost internode and spike length in two wheat RIL populations and their affect upon plant height at an individual QTL level. Euphytica 2001(1): 95–108.

Zadoks, J. C., T. T. Chang, and C. F. Konzak, 1974 A decimal code for the growth stages of cereals. Weed Res. 14(6): 415–421.

Zhang, Z., E. Ersöz, C.-Q. Lai, R. J. Todhunter, H. K. Tiwari et al., 2010 Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42(4): 355–360.

Zhu, L., S. Li, Z. Liang, Z. Zhang, and X. Xu, 2010 Relationship between yield, carbon isotope discrimination and stem carbohydrate concentration in spring wheat grown in Ningxia Irrigation Region (North-west China). Crop Pasture Sci. 61(9): 731–742.

Zila, C., L. F. Samayoa, R. Santiago, A. Butrón, and J. B. Holland, 2013 A genome-wide association study reveals genes associated with fusarium ear rot resistance in a maize core diversity panel. G3 3: 2095–2104.

Communicating editor: E. Akhunov