Plant Biotechnology Journal

Plant Biotechnology Journal (2019), pp. 1–13 doi: 10.1111/pbi.13052

Elucidating the genetic basis of biomass accumulation and radiation use efficiency in spring wheat and its role in yield potential

Gemma Molero[1,†], Ryan Joynson[2,†], Francisco J. Pinera-Chavez[1], Laura-Jayne Gardiner[2], Carolina Rivera-Amado[1], Anthony Hall[2,*] and Matthew P. Reynolds[1,*]

[1]*Global Wheat Program, International Maize and Wheat Improvement Centre (CIMMYT), Texcoco, Mexico*
[2]*The Earlham Institute, Norwich, UK*

Received 13 August 2018;
revised 20 November 2018;
accepted 25 November 2018.
**Correspondence* (Tel +52 (55) 5804 2004 ext. 2227; fax 765-983-1560; email: m.reynolds@cgiar.org (MPR) and Tel +44 7812 238578; email: anthony.hall@earlham.ac.uk (AH))
[†]These authors contributed equally to this work.

Keywords: biomass, radiation use efficiency, genome-wide association studies, exotic material, wheat, yield potential.

Summary

One of the major challenges for plant scientists is increasing wheat (*Triticum aestivum*) yield potential (YP). A significant bottleneck for increasing YP is achieving increased biomass through optimization of radiation use efficiency (RUE) along the crop cycle. Exotic material such as landraces and synthetic wheat has been incorporated into breeding programmes in an attempt to alleviate this; however, their contribution to YP is still unclear. To understand the genetic basis of biomass accumulation and RUE, we applied genome-wide association study (GWAS) to a panel of 150 elite spring wheat genotypes including many landrace and synthetically derived lines. The panel was evaluated for 31 traits over 2 years under optimal growing conditions and genotyped using the 35K wheat breeders array. Marker-trait association identified 94 SNPs significantly associated with yield, agronomic and phenology-related traits along with RUE and final biomass (BM_PM) at various growth stages that explained 7%–17% of phenotypic variation. Common SNP markers were identified for grain yield, BM_PM and RUE on chromosomes 5A and 7A. Additionally, landrace and synthetic derivative lines showed higher thousand grain weight (TGW), BM_PM and RUE but lower grain number (GM2) and harvest index (HI). Our work demonstrates the use of exotic material as a valuable resource to increase YP. It also provides markers for use in marker-assisted breeding to systematically increase BM_PM, RUE and TGW and avoid the TGW/GM2 and BM_PM/HI trade-off. Thus, achieving greater genetic gains in elite germplasm while also highlighting genomic regions and candidate genes for further study.

Introduction

Bread wheat (*Triticum aestivum*) is one of the most globally important crops with 750 million tonnes produced each year (FAO, 2016) across more than 220 million hectares of land (Singh *et al.*, 2016). Due to the rapid rate of worldwide population increase and diet shifts, genetic gains in wheat would have to increase at a rate of 2.4% per year (Hawkesford *et al.*, 2013; Ray *et al.*, 2012, 2013), leading to the consensus that overall wheat yield must be doubled by 2050 if we are to keep up with demand. However, after significant increases in wheat yield in the latter half of the 20th century, improvement has slowed in recent decades (Slafer *et al.*, 2014), with a predicted increase of only 38% by 2050 at current rates (Ray *et al.*, 2013) causing a yield deficit of at least 12%. The largest proportion of wheat yield increases since *The Green Revolution* have been attributed to both changes in agronomic practice and improvements in the ratio of grain yield to total biomass (harvest index, HI) (Fischer *et al.*, 1998). However, the effect of both of these aspects is approaching plateau with very little progress being made since the 1980s in the case of HI of spring wheat (Reynolds *et al.*, 2009a,b). Yield progress has been associated with source-related traits such as photosynthetic rate and increased stomatal conductance in bread and durum wheats (Fischer, 2007; Fischer *et al.*, 1998). Recent studies reported that additional photosynthesis-related traits such as stem water-soluble carbohydrate content, aboveground biomass, crop growth rate and radiation use efficiency (RUE) have been improved in the last five decades within the semi-dwarf bread wheats (Aisawi, 2011; Aisawi *et al.*, 2015; Shearman *et al.*, 2005). In wheat, evidence for increased yield in response to CO_2 enrichment (Ainsworth and Long, 2005) highlight the importance of photosynthesis for which significant improvement in RUE is still possible (Long *et al.*, 2006; Zhu *et al.*, 2010). As such, the smallest of increases in net rate of photosynthesis and RUE could have a large impact on biomass and in turn yield if HI is maintained at current levels.

In the last decades, exotic parents have been used in breeding programmes with the aim of introducing greater diversity into elite gene pools (Singh *et al.*, 2018). The exotic parents that are most frequently used are those from the primary gene pool represented by germplasm that share a common genome but that have become isolated from mainstream gene pools such as

Please cite this article as: Molero, G., Joynson, R., Pinera-Chavez, F. J., Gardiner, L.-J., Rivera-Amado, C., Hall, A. and Reynolds, M. P. (2019) Elucidating the genetic basis of biomass accumulation and radiation use efficiency in spring wheat and its role in yield potential. *Plant Biotechnol. J.*, https://doi.org/10.1111/pbi.13052

landraces (Reynolds *et al.*, 2009a,b), which have been shown to be not only genetically, but epigenetically diverse (Gardiner *et al.*, 2018). The secondary gene pool that has also been used is represented by closely related genomes that can be utilised through inter-specific hybridisation, and would include the development of so-called 'synthetic' or 're-synthesised' wheat, where a tetraploid durum wheat has been hybridised with *Aegilops tauschii*, the ancestral donor of the D genome, to recreate hexaploid bread wheat (Mujeeb-Kazi *et al.*, 1996). This approach has been successful in introducing disease resistance as well as drought and heat adaptive traits (Cossani and Reynolds, 2015; Lopes and Reynolds, 2011; Lopes *et al.*, 2015; Reynolds *et al.*, 2007; Trethowan and Mujeeb-Kazi, 2008). Despite the range of genetic resources available, the vast majority of genetic resources remain unused in breeding (Reynolds *et al.*, 2009a,b) because of uncertainties associated with the use of undomesticated or unimproved genetic backgrounds. Landrace and synthetic material has been identified with superior biomass in comparison with elite lines under drought and heat conditions (Cossani and Reynolds, 2015; Lopes and Reynolds, 2011) and elite lines that include landrace or synthetic material in their background have been developed in recent years for heat and yield potential conditions (Reynolds *et al.*, 2017). These new elite landrace and synthetic lines are derived from parents selected for expressing higher biomass and/or RUE (Reynolds *et al.*, 2017).

The genetic basis of biomass accumulation and RUE are still unclear and as a result, the potential yield increases associated with these traits are still relatively untapped. In this study, yield traits along with biomass and RUE were measured at key growth stages to establish the phenotypic variation present in a panel formed after screening a range of elite International Maize and Wheat Improvement Center (CIMMYT) spring wheat germplasm. We also combine this data with genotypic data through genome-wide association studies (GWAS) to identify marker-trait associations (MTAs) allowing the identification of genomic regions of interest that will help to elucidate the genetic basis of biomass accumulation and RUE in wheat.

Results

The 150 elite lines were evaluated during two consecutive years under similar and optimal growth conditions with a shorter cycle during Y17 probably associated with higher temperatures registered (Table S1). The population was carefully selected to obtain a reduced range in phenology and height to avoid confounding effects. Most of the lines expressed a phenological range of 10 days for anthesis date and physiological maturity (92% and 93% of the lines respectively) and 15 cm variation in height (90%) Figures S1 and S2.

Variations in phenotypic traits in elite germplasm

In total, data for 31 agronomic and physiological traits were collected during field trials (Table 1). Days to anthesis (DTA) was used as a covariate (COV) when it was significant in the analysis but only for independent variables, therefore, phenology (DTInB and DTM), phenological patterns (RSGP and PGF) and RUE were not adjusted with this COV. Thus, only Plants per m^2, Stems per m^2 7 days after anthesis (A7), grains per square metre (GM2), Infertile SPKL per SP, SpikeL, BM_A7, LI_InB and LI_A7 were adjusted using DTA as COV. The results from the analysis of variance (ANOVA) for most traits indicated significant variations among genotypes, environments (years) and genotype × environment interactions where year was the least significant factor (Table 1).

Average grain yield during the two seasons ranged from 4846 to 7052 kg/ha (Table 1) with mean values of 6319 and 5593 kg/ha for Y16 and Y17 respectively. Days to initiation of booting (DTInB) in Y16 (66 days) was 10 days longer than in Y17 (56 days), as well as DTA and days to physiological maturity (DTM) (that were 5 and 6 days longer respectively; Table S1). Highly significant genetic variation was observed for all sink traits (i.e. those associated with grain, their number, size and partitioning to them). Genotype was significant for most of the source traits (i.e. related to carbon assimilation) with the exception of RUE_GF, LI_InB and LI_A7 (Table 1). In general, phenology and sink traits presented higher heritability than agronomic and source-related traits (Table 1). The lowest heritability was for LI_InB and LI_A7 (0.00 and 0.06 or 0.15 and 0.09 when the COV was not used).

Association between traits under yield potential conditions

General phenotypic and genetic correlations among yield and all the physiological traits evaluated are shown in Tables S2 and S3. Multiple regression analysis (stepwise) was conducted to determine subsets of variables that best explain yield and other key agronomic traits. All traits presented in Table 1 were used as yield-predicting variables in the stepwise analysis. Other traits such as HI, BM_PM, GM2, thousand grain weight (TGW) and RUET were also used as dependent variables where yield (all cases) was excluded and other traits (when indicated) were excluded based on multi-collinearity test or if they were not completely independent from the dependent variable (Table 2). For grain yield, RUET explained 38.2% of its variability whereas the combination of RUET and HI or HI and BM_PM explained 65.7% and 86.1% respectively (Table 2). For HI, DTM explained 19% of its variability showing a negative effect and the combination with height and grain weight per spike (GWSP) explained 35.6% and 49.7% respectively (Table 2). Height also presented a negative effect on HI. In the case of BM_PM, height, DTM, TGW and GM2 had a positive effect and together explained 58.2% of its variability. For TGW, Plants per m^2, together with height, SKLSP and spikes per square metre (SM2) explained 56.3% of its variability (Table 2). In the case of GM2, height, SM2, SKLSP and HI explained 48.9% of total variability (Table 2). For RUET, the independent variable that was chosen first by the model was TGW that explained 20.1% of the variation followed by the combination of TGW and GM2 that explained 47.8% highlighting the importance of sink traits determining RUET. The model combining TGW, GM2, and DTM explained 80.3% of RUET variability (Table 2).

Genotyping and anchoring to physical map

Of the 35 143 SNP markers in the 35k array, 35 110 (99.9%) could be anchored to the Refseq-v1.0 reference genome, including 509 to unassigned contigs. The chromosomal and sub-genome distributions of the total array SNPs and of those that were polymorphic in our panel can be seen in Table S4 and Table S5. 9267 SNP polymorphic loci were identified after filtering for a MAF of 5%, the distribution of these can be seen in Table S6. Chromosome 1B had the highest number of markers (910) and Chromosome 4D the lowest (57). The B genome had the highest number of markers (4551) followed by the A (3498) and D genomes (1218). The overall marker density was one

Table 1 Descriptive statistics, broad sense heritability (H^2) and ANOVA for agronomical and physiological traits of HiBAP grown for 2 years (Y15-16 and Y16-17) in northeast Mexico under full irrigated conditions

							ANOVA[§]		
Trait[†]	Mean	Min.	Max.	LSD	CV	H^2	G	Y	G×Y
Agronomic									
Grain Yield (kg/ha)	5956	4846	7052	705	6.7	0.60	***	ns	***
Height (cm)	99	85	114	6	3.0	0.80	***	ns	***
Plants per m^2	179	108	255	47	20.3	0.37	***	ns	***
Stems per m^2 E40	665	480	972	172	13.6	0.59	***	ns	***
Stems per m^2 InB	544	419	755	116	12.9	0.67	***	*	*
Stems per m^2 A7	422	280	630	95	14.3	0.62	***	*	ns
Phenology and phenological patterns									
DTInB	61	52	68	3.8	2.8	0.83	***	***	***
DTA	76	68	85	3.2	1.5	0.87	***	***	***
DTM	115	105	124	3.4	1.7	0.85	***	***	***
RSGP (%)	13.7	10.4	17.7	2.7	12.5	0.46	***	***	***
PGF (%)	33.4	29.6	39.8	2.3	4.0	0.71	***	ns	***
Sink									
HI	0.47	0.40	0.52	0.04	4.7	0.73	***	ns	*
TGW	43.9	30.0	53.8	2.9	3.4	0.94	***	***	***
GM2	13664	10382	16669	1569	6.9	0.84	***	ns	***
SM2	303	234	412	46	10.1	0.77	***	**	ns
GWSP	2.1	1.3	2.8	0.3	9.0	0.83	***	ns	ns
GSP	48.4	38.1	68.6	7.1	9.1	0.77	***	*	ns
SPKL per SP	19.9	17.2	24.2	1.7	4.4	0.77	***	*	***
Infertile SPKL per SP	1.1	0.1	1.9	0.7	31.1	0.49	***	ns	***
SpikeL (cm)	12.0	7.7	14.3	1.2	5.6	0.85	***	ns	**
Source									
BM_E40 (g/m^2)	147	99	193	36.7	13.8	0.24	*	ns	*
BM_InB (g/m^2)	425	313	514	92.0	11.8	0.34	**	*	***
BM_A7 (g/m^2)	861	701	1005	129.2	9.1	0.50	***	*	*
BM_PM (g/m^2)	1355	1104	1645	209	7.7	0.41	***	*	***
RUE_E40InB (g/MJ)	2.08	1.47	3.43	0.79	19.3	0.34	**	**	***
RUE_InBA7 (g/MJ)	2.39	1.37	3.21	0.80	21.0	0.28	**	ns	ns
RUE_GF (g/MJ)	2.02	0.96	2.96	1.05	26.2	0.11	ns	*	***
RUET (g/MJ)	2.09	1.61	2.48	0.37	8.7	0.42	***	ns	***
LI_E40 (%)[‡]	80.1	61.8	92.5	12.7	7.5	0.51	***	-	-
LI_InB (%)	95.4	85.8	99.9	6.5	3.3	0.00	ns	ns	***
LI_A7 (%)	93.5	85.7	98.8	7.4	4.7	0.06	ns	ns	ns

[†]E40: 40 days after emergence, InB: Initiation of Booting. A7: 7 days after anthesis, DTInB: days to initiation of booting, DTA: days to anthesis, DTM: days to physiological maturity, RSGP: rapid spike growth phase, PGF: percentage of grain filling duration, HI: Harvest Index, TGW: thousand grain weight, GM2: grains per square metre, SM2: spikes per square metre, GSP: number of grains per spike, GWSP: grain weight per spike, SPKL per SP: number of spikelets per spike, Infertile SPKL per SP: number of infertile spikelets per spike, BM_E40: biomass measured 40 days after emergence, BM_InB: BM at initiation of booting, BM_A7: BM measured 7 days after anthesis, BM_PM: biomass at physiological maturity, RUE_E40InB: Radiation use efficiency from canopy closure to initiation of booting, RUE_InBA7: from initiation of booting to 7 days after anthesis, RUE_GF: RUE from 7 days after anthesis until physiological maturity, RUET: radiation use efficiency from canopy closure to physiological maturity, LI_E40: light interception 40 days after emergence, LI_InB: initiation of booting, LI_A7: 7 days after anthesis.

[‡]Only 1 year data (Y15-16).

[§]*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ and not significant (ns).

marker per 1.4 Mbp. Marker density was highest in the B genome followed by the A and D genomes with one marker per 1.1, 1.4 and 3.2 Mbp respectively. The physical distribution of polymorphic loci can be seen in Table S5 and in Figure 3.

Population structure analysis and linkage disequilibrium

The extent of LD and the average trend of LD decay rate were estimated based on pairwise LD squared correlation coefficients (R^2) for all intrachromosomal SNP loci for each chromosome

Table 2 Stepwise analysis with Yield, HI, BM_PM, TGW, GM2 and RUET as dependent variables for the whole set of 150 wheat genotypes

Trait	Variable chosen	Adjusted R^2	*P*-value	Sig. *F* change	Traits excluded[†]
Yield	RUET	0.382	<0.001	<0.001	
	RUET, HI	0.657	<0.001	<0.001	
	RUET, HI, BM_PM	0.861	<0.001	<0.001	
	HI, BM_PM	0.861	<0.001	0.487	
	HI, BM_PM, GM2	0.866	<0.001	0.012	
HI	DTM(−)	0.190	<0.001	<0.001	BM_PM
	DTM(−), Height(−)	0.365	<0.001	<0.001	
	DTM(−), Height(−), GWSP	0.497	<0.001	<0.001	
	DTM(−), Height(−), GWSP, RUET(−)	0.595	<0.001	<0.001	
	DTM(−), Height(−), GWSP, RUET(−), SM2	0.861	<0.001	<0.001	
BM_PM	Height	0.127	<0.001	<0.001	RUET*, RUEGF* and HI
	Height, DTM	0.189	<0.001	0.001	
	Height, DTM, TGW	0.258	<0.001	<0.001	
	Height, DTM, TGW, GM2	0.582	<0.001	<0.001	
	Height, DTM, TGW, GM2, SM2	0.657	<0.001	<0.001	
TGW	Plants per m^2(−)	0.284	<0.001	<0.001	GM2*, GWSP
	Plants per m^2(−), Height	0.443	<0.001	<0.001	
	Plants per m^2(−), Height, SKLSP(−)	0.498	<0.001	<0.001	
	Plants per m^2(−), Height, SKLSP(−), SM2(−)	0.563	<0.001	<0.001	
	Plants per m^2(−), Height, SKLSP(−), SM2(−), GSP(−)	0.740	<0.001	<0.001	
GM2	Height(−)	0.241	<0.001	<0.001	TGW*, GSP
	Height(−), SM2	0.327	<0.001	<0.001	
	Height(−), SM2, SKLSP	0.406	<0.001	<0.001	
	Height(−), SM2, SKLSP, HI	0.489	<0.001	<0.001	
	Height(−), SM2, SKLSP, HI, DTInB(−)	0.522	<0.001	0.001	
RUET	TGW	0.207	<0.001	<0.001	BM_PM* and RUEGF*
	TGW, GM2	0.478	<0.001	<0.001	
	TGW, GM2, HI(−)	0.700	<0.001	<0.001	
	TGW, GM2, HI(−), DTM(−)	0.803	<0.001	<0.001	
	TGW, GM2, HI(−), DTM(−), BME40(−)	0.832	<0.001	<0.001	

Independent variables chosen in either of the analyses contributed significantly to the models. Up to five variables were selected.

[†]Based on multi-collinearity test, referred traits were excluded when $r > |0.700|$ to avoid collinearity (indicated with *). Additional traits were excluded from the model as they were not considered independent from the dependent variable. Yield was excluded as independent variable from the analysis. (−) indicates negative effect on the model of the trait selected.

(Figure S3). Across the whole genome, the average R^2 value was 0.105 with 33.3% of the pairwise comparisons being statistically significant at $P < 0.01$ and 1.9% of marker pair combinations in full linkage ($R^2 = 1.0$). The A genome showed an average R^2 of 0.097, 31.7% of pairwise comparisons were significant and 1.17% of marker pairs were in complete linkage. The B genome showed an average R^2 of 0.107, 34.2% of pairwise comparisons were significant and 2.17% of marker pairs were in complete linkage. The D genome showed an average R^2 of 0.132, 34.3% of pairwise comparisons were significant and 3.45% of marker pairs were in complete linkage. Chromosome 1 had the highest average R^2 (0.134) and highest proportion of pairs in complete LD (4.9%) with chromosome 7 showing the lowest for these metric with 0.081 and 0.6% respectively. A chromosomal breakdown of LD can be found in Table S6.

In order to determine the resolution of any MTA identified in this study, LD decay for the population was evaluated at the genome (Figure S3a) and sub-genome level (Figure S3b). The critical LD value of 0.301 for the population was determined by taking the 95th percentile of the square root normalised distribution of unlinked R^2 values. LD decayed below 0.301 at 8 Mbp for the whole genome and at 7, 8.6 and 12.4 Mbp for the A, B and D sub-genomes respectively. Whole genome initial LD was 0.62 and LD decayed to 50% of this value at 7.2 Mbp. Initial LD values decayed to 50% at 8.6, 10.4 and 11.1 Mbp for the A, B and D sub-genomes.

The most likely number of genetic lineages in the HiBAP panel was deduced using a combination of STRUCTURE model-based Bayesian clustering and hierarchical Ward clustering, revealing the presence of two main genetic lineages. Lineages 1 and 2 could also be further subdivided into four and two sub-lineages respectively (Figure 1). Population structure analysis facilitated the identification of likely number of clusters *(k)* as described, with most evidence suggesting *k*2 and some evidence for *k*3. Each accession was assigned to lineage 1 (gold) or lineage 2 (purple) based on their membership coefficient (Figure 1c). STRUCTURE membership coefficients also demonstrated some degree of admixture in a small number of accessions. Subpopulations 1 and 2 were composed of 97 and 51 accessions respectively. Hierarchical clustering analysis identified two main subpopulations as suggested by the Bayesian modelling approach (Figure 1a) with a high level of correlation between the two methods (96.7%). Exotic pedigree history was identified as the cause of this split where subpopulation 2 comprised almost exclusively of lines with

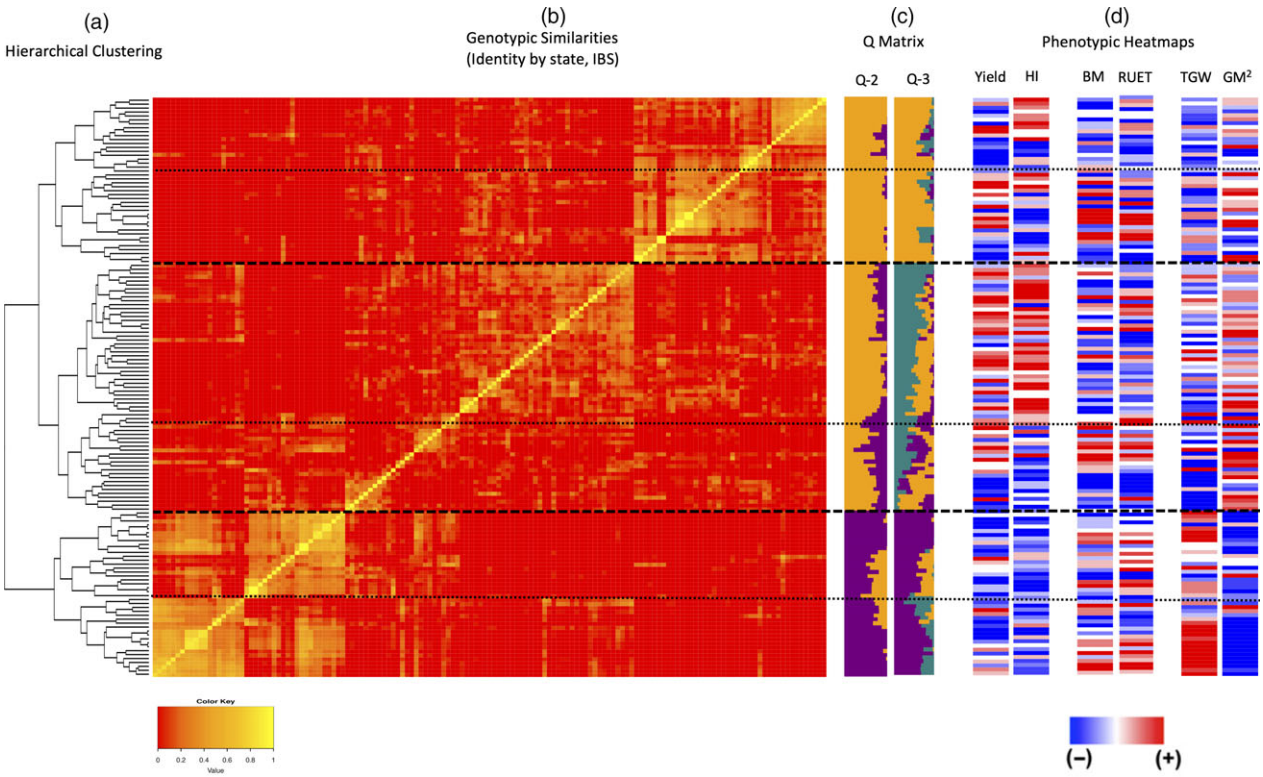

Figure 1 Population structure of 148 accessions of the HiBAP panel. (a) Showing the population structure of the HiBAP panel using hierarchical Ward clustering. (b) A heatmap depicting an identity by state (IBS) Kinship matrix of the HiBAP panel, where horizontal dashed lines depict possible subpopulations based on hierarchical clustering in 2A. (c): Bar plots based on model-based Bayesian clustering analysis using STRUCTURE v2.3.4 ordered in to match the kinship matrix heatmap. (d) Kinship matrix ordered heatmaps for multiple measured traits. Heatmaps for Yield and Harvest Index (HI) demonstrate clustering at the highest genetic level while TGW and GM2 show the inherent trade-off between grain size and grain number in this population.

landrace, synthetic wheat or both in their pedigree history (Figure 2a,b).

Familial relatedness in the panel was assessed using kinship to facilitate estimation of additive genetic variance. The kinship matrix is visualised in Figure 1b as a heatmap showing localised similarity in sub-clusters and a larger degree of kinship between members within lineage 2 than those within lineage 1.

Exotic background on elite lines

In the set of 150 lines, 99 lines were considered elite lines (i.e. progeny of crosses between elite lines), while the rest were products of pre-breeding, namely crosses between elite lines and landraces (11), synthetic hexaploids (26) or product of recent crosses involving landraces, synthetic hexaploids and elite lines (14). Similarly, a phenotypic distinction was observed, where most of the synthetic derivative lines and landraces derivatives appeared distributed in the area of BM_PM, RUET, TGW and Height (Figure 2c). This demonstrates a similar level of separation between elite, exotic derived panel members between the phenotypic and genetic data. To further evaluate these observations, a t test was conducted to compare if significant differences were observed based on the pedigree background (Table 4). No differences in grain yield were observed among the four groups while differences between elite and exotic background were observed in the case of GM2, TGW, HI, Height and biomass at different growth stages. The results indicate that BM_PM and TGW are influenced more by lines that have synthetic/landrace background while also demonstrating that elite backgrounds contribute more to higher GM2 and HI. These results could be used to indicate the successful integration of high biomass phenotypes from synthetic/landrace diversity without affecting grain yield. However, this background is also contributing to slightly taller genotypes, in which despite positive contributions to BM_PM, TGW and GWSP, there is a negative impact on GM2 and a slightly lower expression of HI (Tables S2 and S3; Figure 2c).

Genome-wide association analysis

Marker-trait association analyses were performed using BLUE means (Best Linear Unbiased Estimators) from 2 or 4 repetitions for each measured trait over two growing seasons. In total 94 MTAs were identified (Tables 3 and S7). The largest number of MTAs were detected for SM2 (9), SPKSP (7), BM_PM (6), LIE40 (6) with five MTAs identified for RUET, RUE GF, GM2, DTA and DTInB (Table 3). The A and B sub-genomes contained the highest number of identified MTAs (39) with the least identified on the D sub-genome (16). Chromosome 2B contained the highest individual number of MTAs (Table S7, Figure 3). Multi-trait MTAs were identified on 2B (GWSP/SM2), 2B and 3A(DTInB/DTA), 2D (Stm2_InB/TGW), 3B (RUE_E40InB/LIE40), 3D (DTA/PGF and DTInB/DTA/SPKLSP), 5A (BM_PM/RUET/RUE_GF/YLD), 6B (GWSP/SM2) and 7A (BM_PM/RUET/YLD) as shown in Figure S5. Identified MTAs explained between 7% and 17% of phenotypic variation (Table S7). All Manhattan plots of the GWAS results are shown in Figure S4.

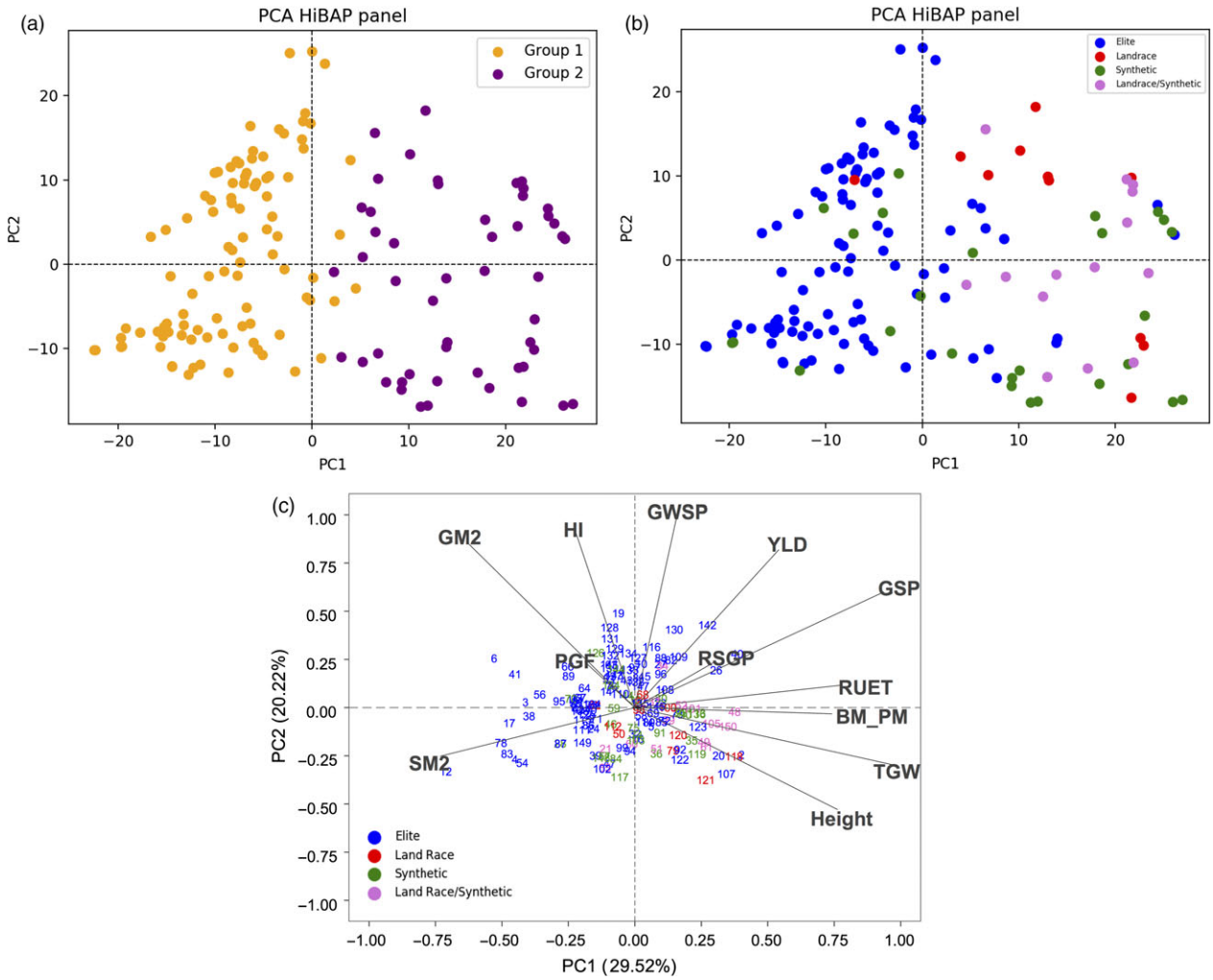

Figure 2 Principal Component Analysis (PCA) using genetic data with samples coloured by cluster determined by STRUCTURE (a) and by pedigree history (b) and phenotypic data (c).

In total, 927 unique genes within a 1 Mbp region (±1 Mb of each locus) of the 94 MTAs were identified. From those, 38 promising candidates were identified for further validation and are presented in Table S8. For example, for phenological parameters, genes previously associated with days to maturity, transition from vegetative stage, seed maturation and pollen development in wheat or other organisms were close to the markers identified. For biomass traits, genes related to sugar transport; for RUE traits, many genes related to response to light stimulus, chloroplast stroma, photosynthesis, electron transport, chlorophyll, light harvest and photosystem II; for SM2, genes previously related to tiller or culm number; for SPKLSP, one gene related to spikelet fertility and for TGW, genes related to grain weight were close to the markers identified in the chromosomes indicated in Table S8.

Discussion

Genetic and phenotypic variation in the high biomass association panel

The value of incorporating exotic germplasm into elite backgrounds has been demonstrated previously for disease resistance as well as drought and heat adaptive traits (Cossani and Reynolds, 2015; Lopes and Reynolds, 2011; Lopes *et al.*, 2015; Reynolds *et al.*, 2007; Trethowan and Mujeeb-Kazi, 2008) and recently for yield potential (Reynolds *et al.*, 2017), where it was suggested that the genetics gains were associated with a better source and sink balance. In this study, the phenotypic and genetic characterization for yield-related traits indicated that elite landraces and synthetic derivative lines presented higher values for biomass and TGW under yield potential conditions (Table 4) indicating that exotic material could be good donors for these traits. As an example, the five lines with the highest TGW expression all contained a Mexican landrace in their pedigree, OAX93.24.35, MEX94.27.1.20 and/or PUB94.15.1.12 and one of the lines contained the synthetic AE.SQUARROSA (205) background. Among the elite lines expressing the highest biomass, some of them contained the 1BL.1RS translocation that in previous studies has been associated with improved harvest biomass (Carver and Rayburn, 1994; Foulkes *et al.*, 2007; Villareal *et al.*, 1995, 1998) and with improved RUE (Foulkes *et al.*, 2007; Shearman *et al.*, 2005). The enhancement in TGW and biomass is not translated to higher yield due to a trade-off observed with GM2 and HI respectively (Tables S2 and S3). The negative relationship

Table 3 Summary of Marker-Trait Associations (MTAs) with different physiological traits and the chromosomes where they were identified

Trait	Number of MTAs	Chromosomes
Agronomic		
Grain Yield (kg/ha)	3	5A, 6A, 7A
Plants per m^2	4	1A, 2B, 3B, 5A
Stems per m^2 E40	2	2B, 6B
Stems per m^2 InB	4	1A, 2D, 3A, 6B
Phenology and phenological patterns		
DTInB	5	2B, 3A, 3D, 5B, 6B
DTA	5	2B(2), 3A, 3D(2)
RSGF (%)	4	1A, 2B(2), 4D
PGF (%)	4	3A(2), 3D, 5B
Sink		
HI	2	2B, 6A
TGW	2	2D, 6D
GM2	5	2B, 3B, 5A, 6D, 7B
SM2	9	1A(3), 2B, 3B, 5B, 6B(2), 7B
GWSP	4	1A, 1B, 2B, 6B
SPKL per SP	7	1A, 2B(2), 3D(2), 4B, 7A
Spike (cm)	3	5A, 5B, 7A
Source		
BM_E40 (g/m^2)	2	1B, 3B
BM_InB (g/m^2)	3	2A, 4B, 7A
BM_PM (g/m^2)	6	5A, 6A, 7A(2), 7B, 7D
RUE_E40InB (g/MJ)	4	2A, 2D, 3B. 6A
RUE_GF (g/MJ)	5	1A, 1D, 2A, 5A, 6A
RUET (g/MJ)	5	3D, 5A(2), 6A, 7A
LI_E40 (%)‡	6	1B, 3B(3), 5A, 6D

‡Only one-year data (Y15-16).

between TGW and GM2 has been widely reported in wheat (Acreche and Slafer, 2006; Bustos *et al.*, 2013; García *et al.*, 2013; Miralles and Slafer, 1995; Quintero *et al.*, 2018; Sadras, 2007) while the trade-off between HI and biomass has been recently reported for modern cultivars (Aisawi *et al.*, 2015). Furthermore, this panel was not chosen for yield potential *per se*, nor necessarily high final biomass, but for different sources of expression for yield potential traits including RUE at different growth stages, traits which when combined strategically in physiological pre-breeding crosses, would be expected to be complementary in terms of yield and final biomass (Reynolds and Langridge, 2016). Proof of concept already exists in that the best new lines developed using such approaches were those involving exotic parents (Reynolds *et al.*, 2017).

LD within the HiBAP population was found to be comparable to that seen in other spring wheat populations (Edae *et al.*, 2013) and we also identify the greatest degree of LD in the D genome which has been reported previously (Edae *et al.*, 2014; Sukumaran *et al.*, 2015). However, this may be an artefact of the relatively low number of polymorphic sites in this panel in the 35K wheat breeders array, something which could be investigated using *de novo* SNP discovery methods such as exome capture.

Population structure analysis identified two main subpopulations whose members were dictated by presence/absence of exotic material in the recent pedigree history of panel members and the resultant genetic variation those crosses have inferred. Of the members of subpopulation 2, ~80% had recently incorporated exotic material in their pedigree compared to only ~10% in subpopulation 1. The exotic subpopulation 2 showed significantly higher biomass at physiological maturity than the elite subpopulation while maintaining an overall yield that was not significantly lower than the elite subpopulation (Table 4). This confirms success in the effort to introduce more genetic diversity in the CIMMYT spring wheat breeding programme, specifically on the introduction of higher biomass lines that are able to maintain yields comparable to common elite varieties and represents successful manipulation of the source sink balance (Reynolds *et al.*, 2017).

Marker-trait associations

To identify novel MTAs for RUE and biomass at various growth stages, GWA was carried out using phenotyping data collected over two growing seasons and >9K SNPs. We also attempted to identify candidate genes, that can be further studied, utilizing the extensive inter-organism knowledge intersection network, Knetminer (Hassani-Pak *et al.*, 2016), a tool that identifies genes or their orthologs in other species that have been previously associated with a specific trait. Together, these methods produce novel MTAs that can be incorporated into CIMMYT marker-assisted breeding programmes along with identification of novel gene targets for future academic studies. Common MTAs were identified for multiple traits such as phenological parameters and SPKLSP, one common for BM_PM, RUET, RUE_GF and yield, Stm2_InB and TGW, SM2 and GSP and RUE_E40 and LI_E40 (Figure S5). To our knowledge, this is the first time that a common MTA has been detected for yield, biomass and RUE traits. This reinforces the idea that increasing RUE under favourable conditions is the key to improving wheat yield potential (Parry *et al.*, 2011). In addition, the identification of MTAs associated with biomass and RUE at different growth stages is the key to optimizing the photosynthetic potential of the plant along the whole crop cycle if all the genomic regions of interest are presented together in a single genotype.

In this study, MTAs relating to source traits were identified at various growth stages, including biomass accumulation, RUE and light interception. MTAs for BM_PM were identified on chromosomes 6A, 7B and 7D along with multi-trait markers on 5A and 7A also associated with RUE at various stages and yield. This suggests that the same genes/gene clusters are having a pleiotropic effect on source traits and overall yield, an observation that has been seen previously between yield and biomass in winter wheat (Mason *et al.*, 2013). QTLs for biomass at physiological maturity and anthesis have been identified on chromosome 7AL previously using double haploid mapping methods in a study spanning 24 years (Quarrie *et al.*, 2006). By utilizing high-density SNP arrays, we believe we have mapped this QTL in much higher resolution. MTAs were identified for RUE at multiple growth stages on chromosomes 1A, 1D, 2A, 2D, 3B, 5A, 6A and 7A. Of these MTAs, the marker found on 5A for RUE at grain filling and total RUE appears to play an important role in the accumulation in biomass in the later stages of plant development being also associated with biomass at physiological maturity. In our candidate gene search, we identified coleoptile phototrophism 1 (CPT1) gene in close proximity to this MTA which has been shown to effect phototropism in rice, which may be having an effect on RUE (Haga, 2005). Candidate gene searches for multiple other RUE MTAs yielded identification of genes that have roles in responses to UV and light stimulus including BTF2-like transcription factor, aldehyde dehydrogenase (ALDH), CPT1, guanosine diphosphate dissociation inhibitor (GDI2), early

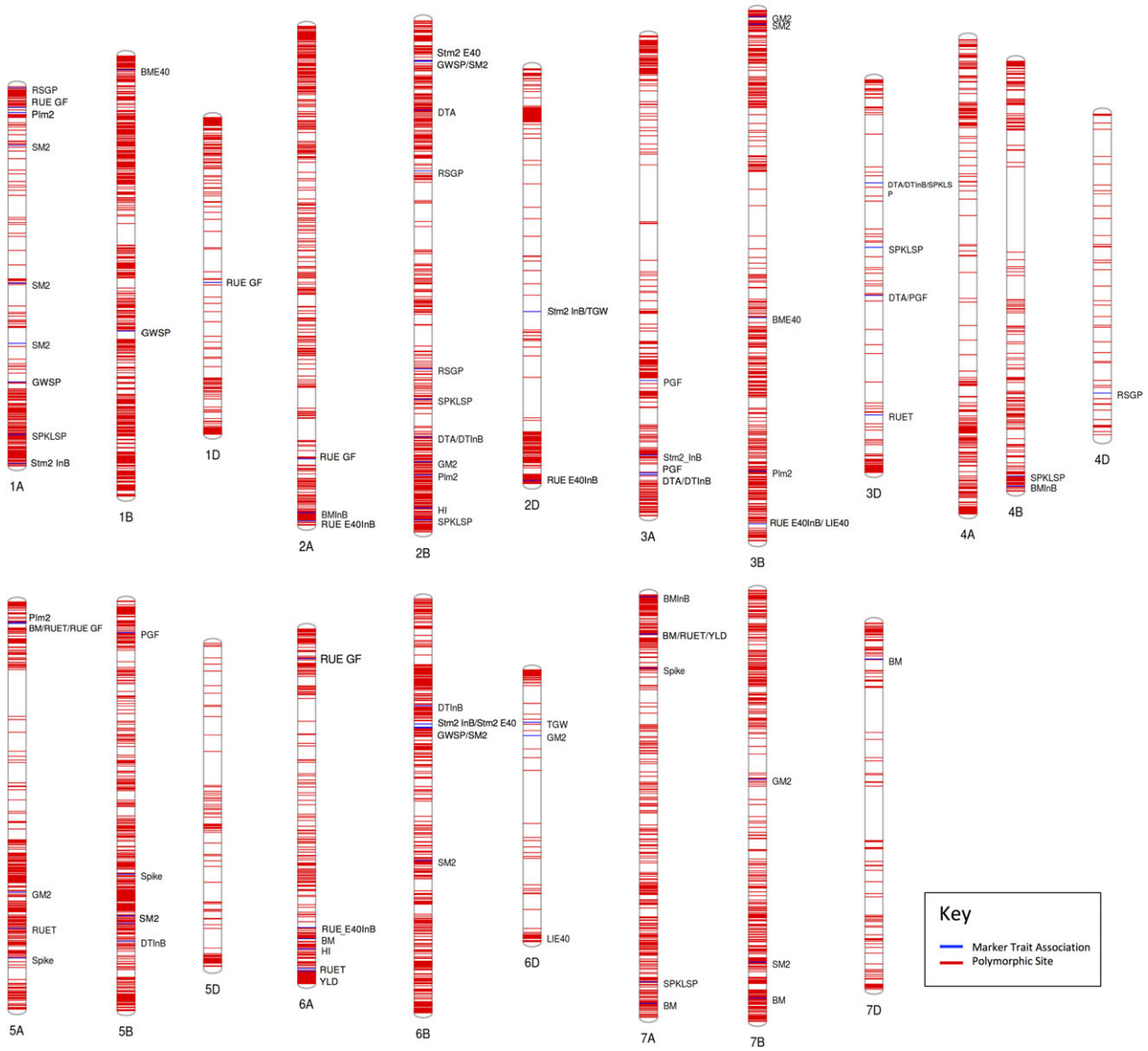

Figure 3 Chromosomal locations of MTA's identified where blue lines indicate MTA location and red lines indicate the location of a polymorphic SNP used in the GWAS.

light-inducible protein (ELIP) and glutathione-s-transferase 3 (GST3). The GST3 gene functions as an antioxidant in plants and has been shown to increase photosynthetic capacity/recovery under high light intensities (Lim *et al.*, 2005) while ALDH Arabidopsis mutants have shown reduced photosynthetic capacity and quantum yield of photosystem II (Missihoun *et al.*, 2018). Similarly, ELIP, identified close to the MTA for RUE_GF in 6A work to prevent oxidative damage to leaves by binding chlorophyll and absorbing light that exceeds photosynthetic capacity (Hutin *et al.*, 2003), thereby protecting proteins and

Table 4 Adjusted means for yield and other traits comparing elite, landrace derivatives, synthetic derivatives and lines that included landraces together with synthetic derivatives in their pedigree

Type	YLD	DTA	GM2	TGW	HI	Height	BM_E40	BM_InB	BM_A7	BM_PM	RUE_GF
Elite	597^{A}	76^{B}	$14\ 077^{A}$	42.6^{C}	0.473^{A}	98.5^{D}	145^{B}	418^{C}	856^{B}	1346^{B}	1.99^{B}
Landrace derivatives	592^{A}	79^{A}	13118^{B}	45.7^{B}	0.450^{C}	103.3^{A}	146^{AB}	452^{A}	891^{A}	1394^{A}	2.02^{AB}
Synthetic derivatives	594^{A}	76^{B}	13096^{B}	45.6^{B}	0.463^{B}	100.5^{C}	153^{A}	433^{CB}	867^{AB}	1358^{AB}	2.03^{AB}
Synthetic+Landrace derivative	593^{A}	76^{B}	12320^{C}	48.2^{A}	0.459^{B}	101.7^{B}	152^{AB}	443^{AB}	872^{AB}	1389^{A}	2.17^{A}

Means followed by the same letter are not significantly different ($P < 0.05$) according to pairwise *t* tests.

photosynthetic pigments from damage by reactive oxygen species (ROS) (Barber and Andersson, 1992; Niyogi, 1999). Identification of photoprotective genes at multiple MTAs for RUE at multiple growth indicates protection of photosynthetic machinery has a large impact on overall impact on RUE in wheat.

Multiple MTAs were identified associated with phenology and phenological patterns on different chromosomes of which the ones associated with DTA and DTInB in 2A have been previously mapped by (Bordes *et al.*, 2014) which are <1 cM (centimorgan) of our markers. However, we also demonstrate a novel pleiotropic MTA on 3D that suggests a link between phenological development and spikelet number, an occurrence that has been linked previously with allelic diversity of Ppd-1 effecting spikelet formation (Ochagavía *et al.*, 2018).

Marker-trait associations involving grain yield and sink traits were identified on multiple wheat chromosomes with the greatest presence in chromosome 2B, which is consistent with the findings of grain yield QTL studies in winter wheat (Assanga *et al.*, 2017). MTAs for TGW were identified on chromosomes 2D and 6D, of which the 6D MTA is in the same cM bin as a QTL for TGW that has been identified previously (Wang *et al.*, 2012) which we confirm here using a different population. Traits related to spike number and plant density were also measured, leading to identification of five MTAs for number of stems per m^2 and nine for number of spikes per m^2. Of which the MTAs on 2B and 7B regarding spike number have been previously mapped using traditional QTL mapping to within 2.4 and 4 cM of our markers respectively (Guan *et al.*, 2018). Candidate gene searches for these traits yielded identification of the Ran GTPase (RAN1) gene under the MTA on chromosome 7B. The wheat RAN1 gene has previously been transformed into both Arabidopsis and rice causing increased stem/tiller number in both, 3-fold higher in the case of rice (Wang *et al.*, 2006), highlighting the candidacy of this gene for further study in wheat.

Potential Implication in Wheat Breeding

The value of exotic material as donors for high expression of biomass and TGW into elite wheat backgrounds under favourable conditions highlights the importance of using these genetic resources in the breeding pipelines. The trade-offs existing between GM2/TGW and BM_PM/HI are limiting current genetic gains but the identification of molecular markers associated with all the traits could be a valuable tool for wheat improvement if molecular-assisted selection is considered.

The GWAS presented here was able to uncover associations between SNPs and yield and yield-related source and sink traits in wheat, with special emphasis on markers associated for the first time with RUE at different growth stages. Although this methodology only provides a statistical link between traits and genomic sequences, such information can be a solid starting point for functional genetic studies. SNP markers closely linked to traits identified by GWAS are being converted into KASP assays for marker-assisted selection (MAS) that will be tested in the near future.

The development of a high-density physical map with the wheat 35K array and comparative genomics provide a powerful tool in searching for potential candidate genes in wheat. Bioinformatics analysis of the mapped SNP markers in the important MTA regions for yield and yield components identified a large number of candidate genes. Many of these genes were associated with photosynthetic machinery. However, since a number of biological processes are associated with these candidate genes, more detailed experimental analyses will be needed to confirm their roles in determining yield potential-related traits.

Experimental procedures

Plant material and growth conditions

The High Biomass Association Mapping Panel (HiBAP) consists of 150 wheat spring types (149 bread wheat and one durum line used as local check) agronomically acceptable including elite high yield material, pre-breeding lines crossed and selected for high yield and biomass, synthetic derived lines and appropriate checks. The panel is the result of systematic screening under field conditions of CIMMYT genetic resources that allowed the identification of elite genotypes with favourable expression of traits of interest. These traits were biomass/RUE at different growth stages including final aboveground biomass, high biomass 7 days after anthesis, high biomass at booting stage, high biomass at canopy closure and high RUE at pre and post anthesis. In the selection of the panel, extremes in phenology and plant height were discarded. Therefore, the material has a restricted range of maturity to avoid confounding effects associated with extreme phenology, and restricted plant height to avoid confounding effects on biomass expression. The 150 lines were evaluated in two consecutive growing seasons (2015/16 and 2016/17, referred to hereafter as Y16 and Y17 respectively).

All of the field experiments were carried out at IWYP-Hub (International Wheat Yield Partnership Phenotyping Platform) situated at CIMMYT's Experimental Station, Norman E. Borlaug (CENEB) in the Yaqui Valley, near Ciudad Obregon, Sonora, Mexico (27°24′ N, 109°56′ W, 38 masl), under fully irrigated conditions. The soil type at the experimental station is a coarse sandy clay, mixed montmorillonitic typic calciorthid, low in organic matter, and slightly alkaline (pH 7.7) in nature (Sayre *et al.*, 1997). Experimental design was an α-lattice with four replications in raised beds (two beds per plot each 0.8 m wide) with four (Y16) and two (Y17) rows per bed (0.1 and 0.24 m between rows respectively) and 4 m long. The emergence dates were 7 December 2015 and 30 November 2016 for Y16 and Y17 respectively. In Y16, the experiment was sown under dry soil whereas Y17 the experiment was sown under moisture (15 days after soil irrigation). The seeding rates were 102 Kg/ha both years. Appropriate weed disease and pest control were implemented to avoid yield limitations. Plots were fertilized with 50 kg N per ha (urea) and 50 kg P per ha at soil preparation, 50 kg N per ha with the first irrigation and another 150 kg N per ha with the second irrigation. Growing conditions and main agronomical characteristics of the trial grown for 2 years are summarized in Table S1.

Agronomic and physiological measurements

Most of the variables were measured in two replicates with the exception of phenology, yield, thousand grain weight (TGW), grain number per m^2 (GM2) and the variables derived from those where four replicates were scored. Phenology of the plots was recorded along the cycle using the scale for growth stages (GS) developed by Zadoks *et al.* (1974), following the average phenology of the plot (when 50% of the shoots reached a certain developmental stage). The phenological stages recorded were initiation of booting (GS41, DTInB), anthesis (GS65, DTA)

and physiological maturity (GS87, DTM). For each plot, the duration in days from emergence to these stages was calculated.

Biomass was measured 40 (Y16) or 42 (Y17) days after emergence (BM_E40), at initiation of booting stage according to plot phenology (BM_InB), approximately 7 days after anthesis (BM_A7) and after physiological maturity (BM_PM). Samplings for BM_E40, BM_InB and BM_A7 consisted of total aboveground tissue in 0.4 m^2 from two beds, starting at least 50 cm from the end of the plot (or the previous harvest) to avoid border effects. A subsample of fresh biomass was weighted and oven-dried at 70 °C for 48 h for constant dry weight measurement. At physiological maturity, a sample of 100 (Y16) or 50 (Y17) fertile shoots was taken randomly from the harvested area to estimate yield components and HI. The sample was oven-dried, weighed and threshed to allow calculation of HI, spikes per square metre (SM2), GM2, number of grains per spike (GSP) and grain weight per spike (GWSP). Grain yield was determined in 3.2–4 m^2 using standard protocols (Pask *et al.*, 2012). To avoid border effects arising from extra solar radiation reaching border plants, 50 cm of the plot edges were discarded before harvesting. BM_PM was calculated from yield/HI. From the harvest of each plot, a subsample of grains was weighed before and after drying (oven-dried to constant weight at 70 °C for 48 h) and the ratio of dry to fresh weight was used to determine dry grain yield and TGW. Plant height and spike length (SpikeL) were measured as the length of five individual shoots or spikes per plot from the soil surface to the tip of the spike and from the spike collar to the ear tip excluding the awns in both cases. Fertile and infertile spikelets per spike (SPKL per SP) were also counted in five spikes per plot.

Percentage of rapid spike growth period (RSGP) was calculated as the difference between DTA and DTInB divided by the total length cycle (DTM). Percentage of grain filling (PGF) was calculated as the number of days between anthesis and physiological maturity divided by DTM. Radiation use efficiency was estimated as the slope of the linear regression of cumulative aboveground biomass on cumulative intercepted PAR (Monteith, 1977). Different RUE were calculated considering the different biomass sampling such as RUE_E40InB: from 40 days after emergence to initiation of booting, RUE_InBA7: from initiation of booting to 7 days after anthesis, RUE_GF: RUE from 7 days after anthesis until physiological maturity, RUET: RUE from 40 days after emergence to physiological maturity. A correction factor for RUE_GF and integrated RUET of 0.5 of PAR intercepted before canopy closure and during 25% of the grain filling period was applied (Reynolds *et al.*, 2000).

DNA extraction and genotyping

Leaf material was obtained from plants growing in the field during Y16, material from 10 individuals was taken per line and pooled for DNA extraction using the standard protocol for the DNeasy plant mini kit (Qiagen, Manchester, UK). DNA purity was assessed using a NanoDrop 2000 (Thermofisher Scientific, Loughborough, UK) and quantified fluorometrically using the Quant-iT™ assay kit (Life Technologies, Warrington, UK). The SNP markers were generated using the 35K wheat breeders array (Affymetrix, High Wycombe, UK) (Allen *et al.*, 2017) following the manufacturers protocol. Allele clustering and subsequent SNP calling were carried out using the Axiom Analysis Suite v2.0. Residual heterozygous calls were entered as missing values. Markers with a minor allele frequency of <5% were removed. Probe sequences for array loci were anchored to the IWGSC Refseq-v1.0 Chinese Spring hexaploid wheat genome assembly (Appels *et al.*, 2018) using BWA (Li and Durbin, 2009). Where sequences mapped identically to multiple chromosomes, inference was taken from the genetic map positions available for 21 708 of the array SNPs (the distribution of anchored markers can be found in Table S4). SNP markers with unknown chromosome positions were removed. After filtering, 9267 SNP markers for 148 accessions were retained, 3498 on the A genome, 4551 on the B genome and 1218 on the D genome.

Linkage disequilibrium

To estimate the level of linkage disequilibrium (LD) between markers, the square of the determination coefficient (R^2) (Hill and Robertson, 1968) was calculated for each pairwise combination of 9267 SNPs in TASSEL 5 (Bradbury *et al.*, 2007). To assess the patterns of LD decay over physical distance, pairwise R^2 values were binned by distance between SNP pairs in 50 Kbp intervals across >600 Mbp and the average R^2 value of the subsequent bins was then plotted vs physical distance. A locally weighted polynomial regression (LOESS) curve was fitted using statistical program R (R Core Team, 2017). The critical R^2 value for this population was deduced by taking the 95th percentile of the square root transformed R^2 distribution of all unlinked SNP pairwise comparisons (inter-chromosome). The physical distance at which LD fell below the population critical LD threshold was used to determine the interval size of identified molecular trait associations markers (MTAs). The extent of marker pairs in LD and the mean R^2 values were calculated for each chromosome and sub-genome.

Population structure analysis

The population structure of the panel was determined using STRUCTURE 2.3.4 (Pritchard *et al.*, 2000) using a model-based Bayesian approach and Hierarchical clustering in R. STRUCTURE was run using the admixture model with 50 000 burn-in iterations followed by 50 000 Markov Chain Monte Carlo (MCMC) iterations for assumed subpopulations (k) 1–10 with nine independent replicates for each k value. The most likely value of k was determined for each population using structureHarvester.py (Earl and vonHoldt, 2012) incorporating the delta K method of (Evanno *et al.*, 2005) where the Δk statistics deduced from the rate of change in the probability of likelihood [LnP(D)] value between each k value was used to predict the likely number of subpopulations. A consensus Q matrix was created from the independent STRUCTURE replicates using Clummpp 1.1.2 (Jakobsson and Rosenberg, 2007). Population structure plots were produced using the Pophelper R library. Genotypes were allocated to their respective subpopulations from which they showed the highest estimated membership coefficient (Q). Distance-based cluster analysis was carried out using the hclust clustering algorithm in R as implemented by the Heatmap2 package. Principal Component Analysis (PCA) was carried out using the Scikit-Lean package in python using 2386 LD-pruned SNPs. SNPs were LD-pruned by removing the member of a SNP pair with the lowest MAF where the pair had an R^2 > critical LD value (0.301) within the distance LD decayed below that value in this population (8 Mbp).

Genome-wide association analysis

Association analysis was carried out using TASSEL 5 (Bradbury *et al.*, 2007) using 9267 markers for 148 HiBAP lines (the durum wheat line was excluded). The unified mixed linear model approach was applied to the genotype/phenotype data, adjusted

using the first five eigenvectors from principal component analysis (PC1-5) or membership coefficient matrices produced by STRUCTURE (Q2-4) and kinship matrix (K) information as covariates in the regression model to reduce errors resulting from confounding population structure effects. The centred Identity by State (IBS) method of Endelman and Jannink (2012) implemented in Tassel was used to create the kinship matrix. False discovery rate (FDR) adjusted *P*-values were found to be too strict in this study, therefore a threshold of −log10(*P*-value) >3 was chosen, as used by multiple studies of this size (Liu *et al.*, 2017; Sukumaran *et al.*, 2018; Sun *et al.*, 2017; Valluru *et al.*, 2017). In order to identify potential candidate genes, genes within 1 Mbp of each MTA were submitted to KnetMiner along with keywords describing their corresponding trait (http://knetminer.rothamsted.ac.uk/) (Hassani-Pak *et al.*, 2016). If adequate evidence was available that the gene or it's orthologs in other organisms was involved in a mechanism linking to the MTA trait, genes were selected as possible candidates for further study.

Statistical analysis

Adjusted means were calculated for each trait by combining data from 2 years using DTA as covariate (fixed effect) when its effect was significant with the exception of phenology or phenological derived traits. The analysis of variance was conducted with the GLM procedure from META R 5.1 (Alvarado *et al.*, 2017), with all the effects of years, blocks within replications, replications within years, replications, genotypes and G×Y being considered as random effects. Broad sense heritability (H^2) was estimated using the MIXED procedure from META R 5.1 (Alvarado *et al.*, 2017) considering all the terms in the model (years, replications within years, genotypes and G×Y) as random effects. H^2 was estimated for each trait over the 2 years as:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{e} + \frac{\sigma^2}{re}}$$

where r = number of repetitions, e = number of environments (years), σ^2 =error variance, σ_g^2 =genotypic variance and σ_{ge}^2 = G × Y variance. Phenotypic correlations (r_p) between traits were simple Pearson correlations. Genetic correlations among traits (r_g) were calculated for cross-year means using the equation from (Cooper *et al.*, 1996) as described in detail by (Vargas *et al.*, 2013):

$$r_g = \frac{\overline{\sigma_{g(jj')}}}{\overline{\sigma_{g(j)}}\,\overline{\sigma_{g(j')}}}$$

where $\overline{\sigma_{g(jj')}}$ is the adjusted mean of all pairwise genotypic covariance between trait j and j' and $\overline{\sigma_{g(j)}}\,\overline{\sigma_{g(j')}}$ is the average of all pairwise geometric means among the genotypic variance components of the traits.

Since the traits were measured in different units, we performed the PCA based on the correlation matrix using the PRINCOMP procedure from SAS 9.1 (SAS Institute Inc., 2004), and then graphed the first two eigenvectors. Multiple linear regression analysis (stepwise) was used to analyse the relationship between the studied variables using the SPSS statistical package (SPSS Inc., Chicago, IL, USA). In this analysis, some traits were excluded when $r > |0.700|$ to avoid collinearity based on multi-collinearity test. Additional traits were removed from the model as indicated in the table when those were not completely independent from the dependent variable.

Acknowledgements

GM, FJPC, CRA and MR were supported by the Sustainable Modernization of Traditional Agriculture (MasAgro) initiative from the Secretariat of Agriculture, Livestock, Rural Development, Fisheries and Food (SAGARPA) and the International Wheat Yield Partnership (IWYP) project. RJ, LG and AH were supported by funding from the BBSRC and IWYP. RJ, LG and AH were supported by funding from the BBSRC and IWYP (BB/N020871/1; BB/P016855/1).

We want to acknowledge Jose L. (Pancho) Crossa and Francisco Rodriguez Huerta from Genetic Resource Program in CIMMYT for their support with statistical analysis. We also acknowledge Luzie Wingen, John Innes Centre, for her support with SNP marker anchoring.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

AH, MR and GM conceptualised the project. GM, CRA and FJPC carried out phenotypic measurements. RJ, LG carried out genotyping and genetic analyses. GM and RJ wrote the manuscript. All the authors edited and approved the manuscript.

Data availability statement

Genotyping and phenotyping data will be made available on at the following repository: https://data.cimmyt.org/dataverse/iwypdvn

References

Acreche, M.M. and Slafer, G.A. (2006) Grain weight response to increases in number of grains in wheat in a Mediterranean area. *F. Crop. Res.* **98**, 52–59.

Ainsworth, E.A. and Long, S.P. (2005) What have we learned from 15 years of free-air CO2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO2. *New Phytol.* **165**, 351–372.

Aisawi, K. (2011) *Physiological processes associated with genetic progress in yield potential of wheat (Triticum aestivum L.).* PhD Thesis, Univ. Nottingham Sch. Biosci. Sutt. Bonningt. Campus, Leistershire, UK.

Aisawi, K.A.B., Reynolds, M.P., Singh, R.P. and Foulkes, M.J. (2015) The physiological basis of the genetic progress in yield potential of CIMMYT spring wheat cultivars from 1966 to 2009. *Crop Sci.* **55**, 1749.

Allen, A.M., Winfield, M.O., Burridge, A.J., Downie, R.C., Benbow, H.R., Barker, G.L.A., Wilkinson, P.A. *et al.* (2017) Characterization of a Wheat Breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (Triticum aestivum). *Plant Biotechnol. J.* **15**, 390–401.

Alvarado, G., López, M., Vargas, M., Pacheco, Á., Rodríguez, F., Burgueño, J. and Crossa, J. (2017) *META-R (Multi Environment Trail Analysis with R for Windows) Version 6.01*.

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., IWGSC whole-genome assembly principal investigators *et al.* (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, **361**, pii: eaar7191.

Assanga, S.O., Fuentealba, M., Zhang, G., Tan, C., Dhakal, S., Rudd, J.C., Ibrahim, A.M.H. *et al.* (2017) Mapping of quantitative trait loci for grain yield and its components in a US popular winter wheat TAM 111 using 90K SNPs. *PLoS ONE*, **12**, e0189669.

Barber, J. and Andersson, B. (1992) Too much of a good thing: light can be bad for photosynthesis. *Trends Biochem. Sci.* **17**, 61–66.

Bordes, J., Goudemand, E., Duchalais, L., Chevarin, L., Oury, F.X., Heumez, E., Lapierre, A. *et al.* (2014) Genome-wide association mapping of three important traits using bread wheat elite breeding populations. *Mol. Breed.* **33**, 755–768.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635.

Bustos, D.V., Hasan, A.K., Reynolds, M.P. and Calderini, D.F. (2013) Combining high grain number and weight through a DH-population to improve grain yield potential of wheat in high-yielding environments. *F. Crop. Res.* **145**, 106–115.

Carver, B.F. and Rayburn, A.L. (1994) Comparison of related wheat stocks possessing 1B or1RS.1BL chromosomes: Agronomic performance. *Crop Sci.* **34**, 1505–1510.

Cooper, M., DeLacy, I.H.H. and Basford, K.E.E. (1996) Relationships among analytical methods used to analyse genotypic adaptation in multi-environment trials. In *Plant Adaptation and Crop Improvement* (Cooper, M. and Hammer, G.L., eds), pp. 193–224. Wallingford, UK: CAB International.

Cossani, C.M. and Reynolds, M.P. (2015) Heat stress adaptation in elite lines derived from synthetic hexaploid wheat. *Crop Sci.* **55**, 2719.

Earl, D.A. and vonHoldt, B.M. (2012) Structure harvester: a website and program for visualizing structure output and implementing the evanno method. *Conserv. Genet. Resour.* **4**, 359–361.

Edae, E.A., Byrne, P.F., Manmathan, H., Haley, S.D., Moragues, M., Lopes, M.S. and Reynolds, M.P. (2013) Association mapping and nucleotide sequence variation in five drought tolerance candidate genes in spring wheat. *Plant Gen.* **6**, 1–13.

Edae, E.A., Byrne, P.F., Haley, S.D., Lopes, M.S. and Reynolds, M.P. (2014) Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. *Theor. Appl. Genet.* **127**, 791–807.

Endelman, J.B. and Jannink, J.-L. (2012) Shrinkage estimation of the realized relationship matrix. *G3 (Bethesda)* **2**, 1405–1413.

Evanno, G., Regnaut, S. and Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620.

FAO (2016) *FAOSTAT Crops production database*. Available at: www.fao.org/faostat/en/#data/QC (accessed 30 May 2018).

Fischer, R.A. (2007) Understanding the physiological basis of yield potential in wheat. *J. Agric. Sci.* **145**, 99–113.

Fischer, R.A., Rees, D., Sayre, K.D., Lu, Z.-M., Condon, A.G. and Saavedra, A.L. (1998) Wheat Yield Progress Associated with higher stomatal conductance and photosynthesis rate, and cooler canopies. *Crop Sci.* **38**, 1467–1475.

Foulkes, M.J., Snape, J.W., Shearman, V.J., Reynolds, M.P., Gaju, O. and Sylvester-Bradley, R. (2007) Genetic progress in yield potential in wheat: recent advances and future prospects. *J. Agric. Sci.* **145**, 17–29.

García, G.A., Hasan, A.K., Puhl, L.E., Reynolds, M.P., Calderini, D.F. and Miralles, D.J. (2013) Grain yield potential strategies in an elite wheat double-haploid population grown in contrasting environments. *Crop Sci.* **53**, 2577.

Gardiner, L.J., Joynson, R., Omony, J., Rusholme-Pilcher, R., Olohan, L., Lang, D., Bai, C. *et al.* (2018) Hidden variation in polyploid wheat drives local adaptation. *Genome Res.* **28**, 1319–1332.

Guan, P., Lu, L., Jia, L., Kabir, M.R., Zhang, J., Lan, T., Zhao, Y. *et al.* (2018) Global QTL analysis identifies genomic regions on chromosomes 4A and 4B harboring stable loci for yield-related traits across different environments in wheat (Triticum aestivum L.). *Front. Plant Sci.*, **9**, 529.

Haga, K. (2005) The rice COLEOPTILE PHOTOTROPISM1 gene encoding an ortholog of arabidopsis NPH3 Is required for phototropism of coleoptiles and lateral translocation of auxin. *Plant Cell Online.* **17**, 103–115.

Hassani-Pak, K., Castellote, M., Esch, M., Hindle, M., Lysenko, A., Taubert, J. and Rawlings, C. (2016) Developing integrated crop knowledge networks to advance candidate gene discovery. *Appl. Transl. Genomics*, **11**, 18–26.

Hawkesford, M.J., Araus, J.L., Park, R., Calderini, D., Miralles, D., Shen, T., Zhang, J. *et al.* (2013) Prospects of doubling global wheat yields. *Food Energy Secur.* **2**, 34–48.

Hill, W.G. and Robertson, A. (1968) Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**, 226–231.

Hutin, C., Nussaume, L., Moise, N., Moya, I., Kloppstech, K. and Havaux, M. (2003) Early light-induced proteins protect Arabidopsis from photooxidative stress. *Proc. Natl Acad. Sci.* **100**, 4921–4926.

Jakobsson, M. and Rosenberg, N.A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Lim, J.D., Hahn, S.J., Yu, C.Y. and Chung, I.M. (2005) Expression of the glutathione S-transferase gene (NT107) in transgenic Dianthus superbus. *Plant Cell. Tissue Organ Cult.* **80**, 277–286.

Liu, N., Bai, G., Lin, M., Xu, X. and Zheng, W. (2017) Genome-wide Association analysis of powdery mildew resistance in U.S. winter wheat. *Sci. Rep.*, **7**, 11743.

Long, S.P., Zhu, X.-G., Naidu, S.L. and Ort, D.R. (2006) Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* **29**, 315–330.

Lopes, M.S. and Reynolds, M.P. (2011) Drought adaptive traits and wide adaptation in elite lines derived from resynthesized hexaploid wheat. *Crop Sci.* **51**, 1617.

Lopes, M.S., El-Basyoni, I., Baenziger, P.S., Singh, S., Royo, C., Ozbek, K., Aktas, H. *et al.* (2015) Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *J. Exp. Bot.* **66**, 3477–3486.

Mason, R.E., Hays, D.B., Mondal, S., Ibrahim, A.M.H. and Basnet, B.R. (2013) QTL for yield, yield components and canopy temperature depression in wheat under late sown field conditions. *Euphytica*, **194**, 243–259.

Miralles, D.J. and Slafer, G. (1995) Individual grain weight response to genetic reduction in culm length in wheat as affected by source-sink manipulations. *F. Crop. Res.* **43**, 55–66.

Missihoun, T.D., Kotchoni, S.O. and Bartels, D. (2018) Aldehyde dehydrogenases function in the homeostasis of pyridine nucleotides in arabidopsis thaliana. *Sci. Rep.* **8**, 2936.

Monteith, J.L. (1977) Climate and the efficiency of crop production in Britain. *Philos. Trans. R. Soc. London Ser. B* **281**, 277–294.

Mujeeb-Kazi, A., Rosas, V. and Roldan, S. (1996) Conservation of the genetic variation of Triticum tauschii (Coss.) Schmalh. (Aegilops squarrosa auct. non L.) in synthetic hexaploid wheats (T. turgidum L. s.lat. x T. tauschii; 2n=6x=42, AABBDD) and its potential utilization for wheat improvement. *Genet. Resour. Crop Evol.* **43**, 129–134.

Niyogi, K.K. (1999) Photoprotection revisited: genetic and molecular approaches. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 333–359.

Ochagavía, H., Prieto, P., Savin, R., Griffiths, S. and Slafer, G. (2018) Dynamics of leaf and spikelet primordia initiation in wheat as affected by Ppd-1a alleles under field conditions. *J. Exp. Bot.* **69**, 2621–2631.

Parry, M.A.J., Reynolds, M., Salvucci, M.E., Raines, C., Andralojc, P.J., Zhu, X.-G., Price, G.D. *et al.* (2011) Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency. *J. Exp. Bot.* **62**, 453–467.

Pask, A., Pietragalla, J., Mullan, D. and Reynolds, M.P. (2012) *Physiological breeding II : a field guide to wheat phenotyping*.

Pritchard, J.K., Stephens, M. and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Quarrie, S., Pekic Quarrie, S., Radosevic, R., Rancic, D., Kaminska, A., Barnes, J.D., Leverington, M. *et al.* (2006) Dissecting a wheat QTL for yield present in a range of environments: from the QTL to candidate genes. *J. Exp. Bot.* **57**, 2627–2637.

Quintero, A., Molero, G., Reynolds, M.P. and Calderini, D.F. (2018) Trade-Off between grain weight and grain number in wheat depends on GXE interaction : a case study of an elite CIMMYT Panel (CIMCOG). *Eur. J. Agron.* **92**, 1–23.

R Core Team (2017) *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. URL https//www.R-project.org/

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

Ray, D.K., Mueller, N.D., West, P.C. and Foley, J.A. (2013) Yield trends are insufficient to double global crop production by 2050. *PLoS ONE*, **8**, e66428.

Reynolds, M.P. and Langridge, P. (2016) Physiological breeding. *Curr. Opin. Plant Biol.* **31**, 162–171.
Reynolds, M.P., van Ginkel, M. and Ribaut, J.M. (2000) Avenues for genetic modification of radiation use efficiency in wheat. *J. Exp. Bot.* **51**, 459–473.
Reynolds, M.P., Dreccer, F. and Trethowan, R.M. (2007) Drought-adaptive traits derived from wheat wild relatives and landraces. *J. Exp. Bot.* **58**, 177–186.
Reynolds, M., Foulkes, M.J., Slafer, G.A., Berry, P., Parry, M.A.J., Snape, J.W. and Angus, W.J. (2009a) Raising yield potential in wheat. *J. Exp. Bot.* **60**, 1899–1918.
Reynolds, M., Manes, Y., Izanloo, A. and Langridge, P. (2009b) Phenotyping approaches for physiological breeding and gene discovery in wheat. *Ann. Appl. Biol.* **155**, 309–320.
Reynolds, M.P., Pask, A.J.D., Hoppitt, W.J.E., Sonder, K., Sukumaran, S., Molero, G., Saint Pierre, C. *et al.* (2017) Strategic crossing of biomass and harvest index—source and sink—achieves genetic gains in wheat. *Euphytica*, **213**, 257.
Sadras, V.O. (2007) Evolutionary aspects of the trade-off between seed size and number in crops. *F. Crop. Res.* **100**, 125–138.
SAS Institute Inc. (2004) *SAS System for Windows. Version 9.1*. SAS Institute, Inc., Cary, NC, USA.
Sayre, K.D., Rajaram, S. and Fischer, R.A. (1997) Yield potential progress in short bread wheats in northwest Mexico. *Crop Sci.* **37**, 36.
Shearman, V.J., Sylvester-Bradley, R., Scott, R.K. and Foulkes, M.J. (2005) Physiological processes associated with wheat yield progress in the UK. *Crop Sci.* **45**, 175–185.
Singh, R.P., Singh, P.K., Rutkoski, J., Hodson, D.P., He, X., Jørgensen, L.N., Hovmøller, M.S. *et al.* (2016) Disease impact on wheat yield potential and prospects of genetic control. *Annu. Rev. Phytopathol.* **54**, 303–322.
Singh, S., Vikram, P., Sehgal, D., Burgueño, J., Sharma, A., Singh, S.K., Sansaloni, C.P. *et al.* (2018) Harnessing genetic potential of wheat germplasm banks through impact-oriented-prebreeding for future food and nutritional security. *Sci. Rep.* **8**, 12527.
Slafer, G.A., Savin, R. and Sadras, V.O. (2014) Coarse and fine regulation of wheat yield components in response to genotype and environment. *F. Crop. Res.* **157**, 71–83.
Sukumaran, S., Reynolds, M.P., Lopes, M.S. and Crossa, J. (2015) Genome-wide association study for adaptation to agronomic plant density: a component of high yield potential in spring wheat. *Crop Sci.* **55**, 1–11.
Sukumaran, S., Reynolds, M.P. and Sansaloni, C. (2018) Genome-wide association analyses identify QTL hotspots for yield and component traits in durum wheat grown under yield potential, drought, and heat stress environments. *Front. Plant Sci.* **9**, 81.
Sun, C., Zhang, F., Yan, X., Zhang, X., Dong, Z., Cui, D. and Chen, F. (2017) Genome-wide association study for 13 agronomic traits reveals distribution of superior alleles in bread wheat from the Yellow and Huai Valley of China. *Plant Biotechnol. J.* **15**, 953–969.
Trethowan, R.M.M. and Mujeeb-Kazi, A. (2008) *Novel Germplasm Resources for Improving Environmental Stress Tolerance of Hexaploid Wheat.* **48**, 1255–1265.
Valluru, R., Reynolds, M.P., Davies, W.J. and Sukumaran, S. (2017) Phenotypic and genome-wide association analysis of spike ethylene in diverse wheat genotypes under heat stress. *New Phytol.* **214**, 271–283.
Vargas, M., Combs, E., Alvarado, G., Atlin, G., Mathews, K. and Crossa, J. (2013) Meta: a suite of sas programs to analyze multienvironment breeding trials. *Agron. J.* **105**, 11–19.
Villareal, R.L., del Toro, E., Mujeeb-Kazi, A. and Rajaram, S. (1995) The 1BL/1RS chromosome translocation effect on yield characteristics in a Triticum aestivum L. cross. *Plant Breed.* **114**, 497–500.
Villareal, R.L., Banuelos, O., Mujeeb-Kazi, A. and Rajaram, S. (1998) Agronomic performance of chromosomes 1B and 1BL.1RS near-isolines in the spring bread wheat Seri M82. *Euphytica*, **103**, 195–202.
Wang, X., Xu, Y., Han, Y., Bao, S., Du, J., Yuan, M., Xu, Z. *et al.* (2006) Overexpression of RAN1 in rice and Arabidopsis alters primordial meristem, mitotic progress, and sensitivity to auxin. *Plant Physiol.* **140**, 91–101.
Wang, L., Ge, H., Hao, C., Dong, Y. and Zhang, X. (2012) Identifying loci influencing 1,000-kernel weight in wheat by microsatellite screening for evidence of selection during breeding. *PLoS ONE*, **7**, e29432.
Zadoks, J.C., Chang, T.T. and Konzak, C.F. (1974) A decimal code for the growth stages of cereals. *Weed Res.* **14**, 415–421.
Zhu, X.-G., Long, S.P. and Ort, D.R. (2010) Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol.* **61**, 235–261.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Histograms of the distribution of the phenotypic values of plant height, anthesis date (DTA) and days to maturity (DTM).
Figure S2 Boxplots of the best linear estimated predictions (BLUEs) of the main traits measured in HiBAP during 2 years of evaluation (Y16&Y17).
Figure S3 Linkage Disequilibrium Decay depicted as a scatter plot of pairwise SNP LD (R^2) and pairwise physical distance across the hexaploid wheat genome.
Figure S4 GWAS results using 9267 SNPs markers in HiBAP for yield traits based on BLUEs means obtained from the combined analysis from Y16 and Y17.
Figure S5 Venn diagram exhibiting the number of total and common MTA's detected for different traits.
Table S1 Growing conditions and main agronomical characteristics of the trial grown for 2 years in northeast Mexico under full irrigation conditions.
Table S2 Phenotypic correlations among the 31 traits presented in this study.
Table S3 Genetic correlations among the 31 traits presented in this study.
Table S4 The distribution of the 35K Axiom Wheat Breeders Array loci on the Refseq1.0 Chinese Spring wheat physical map.
Table S5 The distribution of 35K Axiom array SNPs that were polymorphic in the HiBAP panel.
Table S6 Linkage Disequilibrium/Decay statistics for the HiBAP panel.
Table S7 Summary of GWAS results from the trial evaluated during 2 years in northeast Mexico under full irrigation conditions.
Table S8 List of selected candidate genes found for the evaluated traits using GWA mapping.

