PAN-GENOME LEVEL GENOTYPE AND PHENOTYPE PREDICTION: ADVANCES IN PRECISION AGRICULTURE

EDITED BY: Penghao Wang, Washington Gapare and Jian Ma

PUBLISHED IN: Frontiers in Genetics and Frontiers in Veterinary Science

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-704-0
DOI 10.3389/978-2-88971-704-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

PAN-GENOME LEVEL GENOTYPE AND PHENOTYPE PREDICTION: ADVANCES IN PRECISION AGRICULTURE

Topic Editors:
Penghao Wang, Murdoch University, Australia
Washington Gapare, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
Jian Ma, Sichuan Agricultural University, China

Citation: Wang, P., Gapare, W., Ma, J., eds. (2021). Pan-Genome Level Genotype and Phenotype Prediction: Advances in Precision Agriculture. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-704-0

Table of Contents

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 August 2020
doi: 10.3389/fgene.2020.01012

Major QTL for Seven Yield-Related Traits in Common Wheat (*Triticum aestivum* L.)

Jingjing Jin[1,2†], Dan Liu[3,4†], Yongzhi Qi[1,2], Jun Ma[5] and Wenchao Zhen[2,6*]*

[1] College of Plant Protection, Hebei Agricultural University, Baoding, China, [2] State Key Laboratory of North China Crop Improvement and Regulation, Baoding, China, [3] Neijiang Academy of Agricultural Sciences, Neijiang, China, [4] School of Life Sciences and Engineering, Southwest University of Science and Technology, Mianyang, China, [5] College of Agronomy and Biotechnology, China Agricultural University, Beijing, China, [6] College of Agronomy, Hebei Agricultural University, Baoding, China

Flag leaves, plant height (PH), and spike-related traits are key determinants contributing to yield potential in wheat. In this study, we developed a recombinant inbred line (RIL) population with 94 lines derived from the cross between 'AS985472' and 'Sumai 3.' A genetic map spanned 3553.69 cM in length were constructed using 1978 DArT markers. Severn traits including flag leaf width (FLW), flag leaf length (FLL), PH, anthesis date (AD), spike length (SL), spikelet number spike (SNS), and spike density (SD) were evaluated against this RIL population under three different environments. Combined phenotypic data and genetic map, we identified quantitative trait loci (QTL) for each trait. A total of four major and stably expressed QTLs for FLW, PH, and SD were detected on chromosomes 2D and 4B. Of them, the major QTLs individually explained 10.10 – 30.68% of the phenotypic variation. QTLs with pleiotropic effects were identified on chromosomes 4A and 6D as well. Furthermore, the genetic relationships between seven yield-related traits were detected and discussed. A few genes related to leaf growth and development at the interval of a major locus for FLW on chromosome 2D were predicated. Overall, the present study provided useful information for understanding the genetic basis of yield-related traits and will be useful for marker-assisted selection in wheat breeding.

Keywords: wheat, recombinant inbred line, yield-related traits, quantitative trait loci, candidate genes

OPEN ACCESS

Edited by:
Jian Ma,
Sichuan Agricultural University, China

Reviewed by:
Linyi Qiao,
Shanxi Agricultural University, China
Feng Wang,
University of Helsinki, Finland

****Correspondence:***
Jun Ma
Junma@cau.edu.cn
Wenchao Zhen
wenchao@hebau.edu.cn

†These authors share first authorship

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *26 July 2020*
Accepted: *10 August 2020*
Published: *28 August 2020*

Citation:
Jin J, Liu D, Qi Y, Ma J and Zhen W (2020) Major QTL for Seven Yield-Related Traits in Common Wheat (Triticum aestivum L.). Front. Genet. 11:1012. doi: 10.3389/fgene.2020.01012

INTRODUCTION

Ninety-five percent of the energy in nature comes from photosynthesis (Zhai et al., 2002) and the leaves are the major photosynthetic organ in plants. Flag leaf, the first leaf under the spike of wheat (*Triticum aestivum* L.), contributes to photosynthesis and provides water and nutrients to the spikes for grain filling (Yang et al., 2016). Other agronomic traits like anthesis date (AD), spike length (SL), spikelet number per spike (SNS), and spike density (SD) are also key determinants of the plant architecture and yield potential. It is known that grain yield is closely correlated with AD (Woodruff and Tonks, 1983), plant height (PH) (Hedden, 2003), SL (Liu et al., 2018b), SNS

(Hai et al., 2008), and SD (Li et al., 2016). Thus, a comprehensive understanding of the genetic mechanism for flag leaf width (FLW), flag leaf length (FLL), PH and spike-related traits is critical for increasing grain yield.

Agronomic traits are usually controlled by multiple genes and numerous quantitative trait loci (QTL) for them have been reported on A, B, and D genomes in wheat. For instance, in hexaploid wheat, major QTLs for FLL, FLW, flag leaf area (FLA), the ratio of length/width of flag leaf (FLR), flag leaf angle (FLANG), fag leaf opening angle (FLOA) and fag leaf bend angle (FLBA) were mapped to chromosomes 2D, 5B, 4B (Ma et al., 2020). Liu et al. (2018c) detected QTLs for FLL, FLW, FLA and FLANG on chromosomes 1B, 2B, 3A, 3D, 4B, 5A, 6B, 7B, and 7D using a recombinant inbred line (RIL) population derived from 'ND3331' and 'Zang1817.' Hu et al. (2020) identified 161 QTLs for yield-related traits including grain yield per plant (GYP), spike number per plant (SN), kernel number per spike (KPS), SL, SNS, FLL, FLW, FLA, PH, AD and heading date (HD) on 21 chromosomes except 2D, 3D, and 6D. Although studies on traits related to flag leaf and spike have made great progress, there are still many novel loci that can be excavated and utilized from different germplasm resources.

Significant correlations between agronomic traits of wheat were observed in numerous studies. For example, a study of phenotypic correlations showed that SNS was significantly and positively correlated with SL, AD, and KPS (Ma et al., 2019a). Furthermore, QTLs or genes with pleiotropic effects on agronomic traits in wheat have been previously verified. For example, QTLs with pleiotropic effects to SN, SL, and KPS were identified on chromosomes 1B, 4B, and 5A (Deng et al., 2011). Similarly, Ma et al. (2020) detected two pleiotropic QTLs associated with FLL and FLR on chromosomes 5B, two pleiotropic QTLs for FLOA and FLBA on chromosomes 2D, and three pleiotropic QTLs for FLL, FLW and FLA on chromosomes 2D, and they shared the same or overlapped physical intervals on 'Chinese Spring' (CS) genomes. Additionally, these pleiotropic QTLs exhibited significant associations in Pearson correlation analysis. Thus, pleiotropic or linked loci could benefit improving breeding efficiency for multiple elite traits.

TABLE 1 | Phenotypic variation and heritability (H^2) for seven yield-related traits of the 'AS985472'/ 'Sumai 3' (AS) population in different environments.

Trait	Environment	Parents		AS985472/Sumai 3			
		AS985472	Sumai 3	Min-Max	Mean	STD	H^2
FLW (cm)	2019CZ	1.85	2.03*	1.59–2.38	1.93	0.18	
	2018CZ	1.70	1.88*	1.52–2.23	1.82	0.20	
	2017CZ	2.08	2.30**	1.70–2.90	2.11	0.24	
	BLUP	1.94	2.00	1.68–2.40	1.95	0.16	0.88
FLL (cm)	2019CZ	20.90	25.57**	19.41–31.05	24.05	2.43	
	2018CZ	22.89	30.25**	20.24–30.41	24.26	2.43	
	2017CZ	25.53	29.70**	20.00–30.40	24.62	2.63	
	BLUP	23.95	26.76	22.73–28.75	25.09	1.14	0.94
AD (d)	2019CZ	154.00	144.00	141.00–158.00	151.94	3.67	
	2018CZ	142.00	136.00	136.00–162.00	140.49	3.99	
	2017CZ	153.00	141.00	139.00–158.00	148.05	5.10	
	BLUP	149.11	141.76	140.27–157.83	145.21	3.02	0.75
PH (cm)	2019CZ	88.70	113.80**	62.10–129.90	96.46	15.94	
	2018CZ	85.33	109.67**	61.67–128.57	99.00	15.02	
	2017CZ	80.17	118.00**	70.33–131.17	104.34	14.56	
	BLUP	85.43	113.11	68.10–126.73	99.12	13.46	0.95
SL (cm)	2019CZ	9.67	13.99**	8.69–14.28	11.19	1.29	
	2018CZ	9.97	13.42**	8.40–14.60	11.13	1.22	
	2017CZ	10.77	15.63**	8.20–15.63	11.28	1.41	
	BLUP	10.16	14.26	8.51–14.44	11.17	1.18	0.97
SNS	2019CZ	19.20	21.40**	17.25–24.75	20.32	1.65	
	2018CZ	19.33	22.00*	17.00–25.00	20.39	1.61	
	2017CZ	19.67	22.00*	17.33–24.33	20.24	1.56	
	BLUP	19.44	21.51	17.42–24.49	20.27	1.43	0.92
SD	2019CZ	2.00**	1.54	1.49–2.15	1.83	0.16	
	2018CZ	1.94*	1.56	1.51–2.14	1.84	0.14	
	2017CZ	1.83**	1.37	1.37–2.15	1.81	0.17	
	BLUP	1.92	1.52	1.54–2.08	1.83	0.13	0.93

*FLW, flag leaf width; FLL, flag leaf length; AD, anthesis date; PH, plant height; SL, spike length; SNS, spikelet number spike; SD, spike density; STD, standard deviation; H^2, the broad-sense heritability; BLUP, best linear unbiased prediction;**significance level at P = 0.01; *significance level at P = 0.05; CZ, Chongzhou.*

The present study focused on detecting QTLs controlling flag leaf traits including FLL and FLW, and spike-related traits including SNS, SL and SD, and AD and PH in a RIL population developed from the cross between 'AS985472' and 'Sumai 3,' and evaluating their genetic correlations. This study will provide valuable information to understand the genetic basis of yield-related traits and help to accelerate molecular assisted breeding in wheat.

MATERIALS AND METHODS

Plant Materials

A total of 94 F_8 RILs generated from the cross of 'AS985472'/'Sumai 3' (AS) were used in the present study. 'AS985472' is an advance wheat line; 'Sumai 3' is an excellent germplasm resource with high resistance to Fusarium head blight (Xie et al., 2007).

Field Trials and Phenotypic Evaluation

From 2017 to 2019, 94 AS RILs and the two parents were planted at Chongzhou (CZ, 103° 38′ E, 30° 32′ N) of Sichuan Province with a random block design. Each line was in a single 1.5-m row with 30-cm apart between rows, and 15 seeds were planted in each row with 10-cm space between individuals.

Anthesis date (d) was calculated from the sowing date to date when more than 50% of the plants of a line flowered. After anthesis, FLW (cm) was determined by the widest section of the flag leaf, FLL (cm) was measured as the length from

FIGURE 1 | Morphology of the flag leaf **(a)**, spike **(b)**, and plant architecture **(c)** of 'Sumai 3' and 'AS985472'. (Scale bar = 10.5, 5, and 20 cm, respectively).

the base to the top of the flag leaf. The measurements of PH (cm), SL (cm), SNS and SD were carried out as described by Ma et al. (2019a). PH was obtained by measuring the height from the base to the top of the main spike excluding the awns. SL was the length of the main spike of an individual plant (excluding awns). SNS was determined by the number of spikelets of the spike for the main tiller and SD was SNS divided by SL.

Statistical Analysis and QTLs Detection

Data obtained from 2017 to 2019 in CZ were subjected to combined analysis. The mean values and the Student's *t*-test ($P < 0.05$) of the parental lines were calculated. For each RIL, the maximum and minimum values, mean values and standard deviation were analyzed using the SPSS Statistic 25 (IBM SPSS, Armonk, NY, United States).

To estimate random effects in statistics, the best linear unbiased prediction (BLUP) for seven yield-related traits in different environments were calculated using SAS version 8.0 (SAS Institute, Cary, NC, United States). The BLUP for the phenotypic values were calculated according to the model: $Y_i = X_i f + a_i + e_i$, where f = a vector of fixed effects, Xi = an incidence vector, e_i = the environmental deviation, and a_i = the phenotypic value (Goddard, 1992). The broad-sense heritability (H^2) was estimated using the following formula: $H^2 = V_G/(V_G + V_{GE}/r + V_E)$. Where V_G = genetic variance, V_{GE} = genotype × environment variance, r = the number of replicates, and V_E = environ mental variance (Smith et al., 1998).

The genetic linkage map was constructed according to a previous study (Liu et al., submitted). To retain high confidence markers, minor allele frequency (<0.3) of the markers were excluded using the BIN function of Icimapping 4.1 and the linkage group for AS population was integrated using the software Joinmap 4.0 according to Liu et al. (2018a). The linkage map was 3553.69 cM in length containing 31 linkage groups and 1978 DArT markers (Liu et al., submitted). Then, the putative QTLs were detected with a minimum LOD (log-of-odds) value of 2.5 by the BIP (Biparental Populations) module (Lin et al., 1996) and ICIM (inclusive composite interval

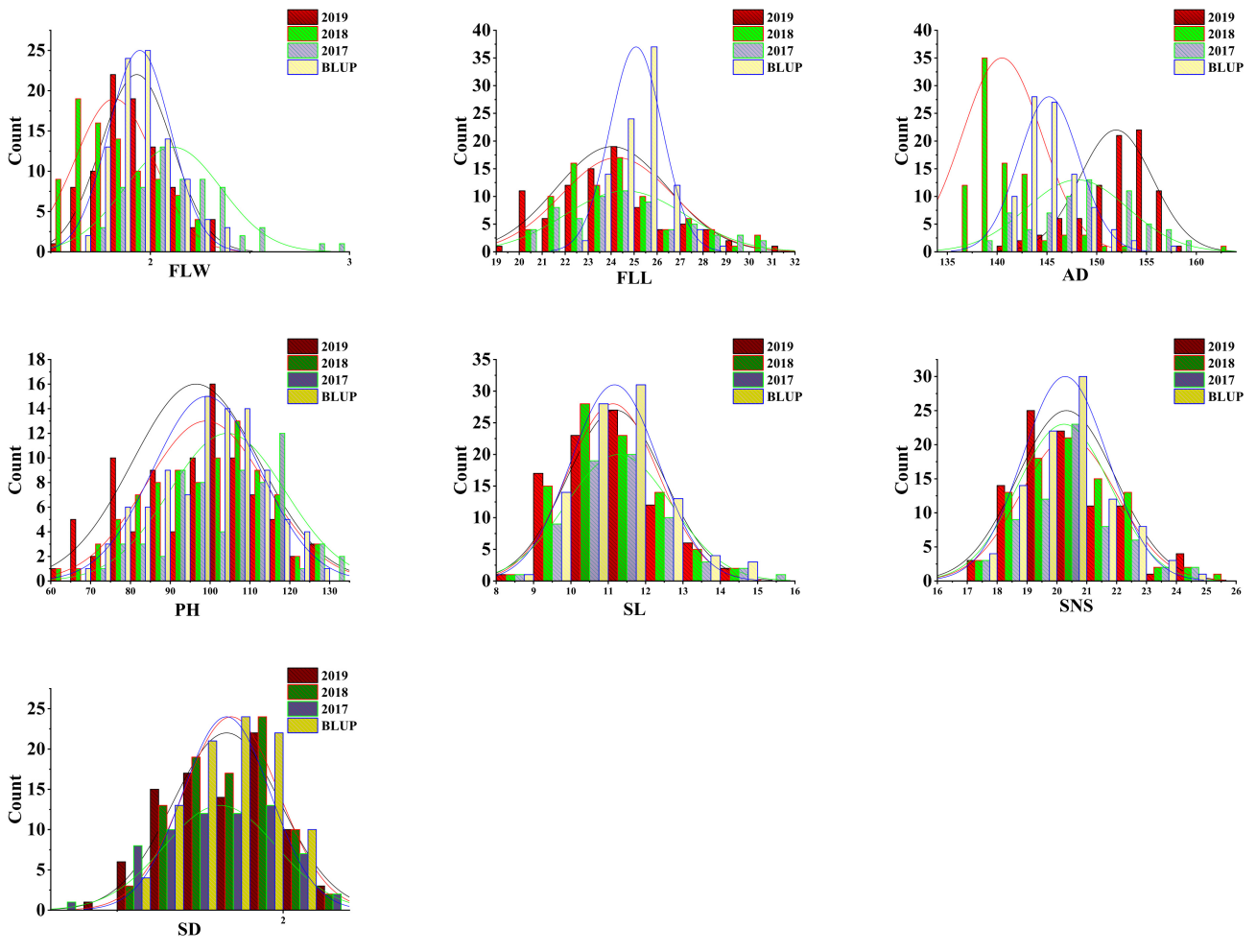

FIGURE 2 | Phenotypic distribution of wheat flag leaf width (FLW), flag leaf length (FLL), anthesis date (AD), plant height (PH), spike length (SL), spikelet number per spike (SNS) and spike density (SD) in the 'AS985472'/'Sumai 3' or AS recombinant inbred population.

mapping) method from Icimapping 4.1. Among the detected QTLs, those with >10% of phenotypic variation and could be detected in at least two tested environments as well as the BLUP dataset were treated as major stable QTLs and those with a common flanking marker were considered as a single QTL.

Comparison of QTLs Related to FLW on 2D

The physical positions of the major QTLs on the genome assembly of *T. aestivum* cv. Chinese Spring or CS (IWGSC RefSeq v1.0)[1] (The International Wheat Genome Sequencing Consortium, 2018), *Ae. tauschii* (Aet V4.0[2]) (Luo et al., 2017), and *T. turgidum*[3] (Avni et al., 2017) and the analysis of candidate genes within the interval between the flanking markers on CS genome were obtained according to previous studies (Ma et al., 2020). Additionally, the physical positions of the major loci detected in the present study were compared with the reported QTLs or genes.

RESULTS

Phenotypic Performance of the ES RILs

The analyses of phenotypic variation showed that significant differences existed between 'AS985472' and 'Sumai 3' ($P < 0.05$, **Table 1** and **Figure 1**). 'Sumai 3' had wider FLW, longer FLL, higher PH, more SNS, longer spike but lower SD, later AD than 'AS985472.' In addition, the continuous distributions with ranges from 1.52 to 2.90 cm in FLW, 19.41 to 31.05 cm for FLL, 136 to 162 d for AD, 61.67 to 131.17 cm for PH, 8.2 to 15.63 cm for SL, 17 to 25 for SNS and 1.37 to 2.15 for SD (**Table 1**) and transgressive segregation across all environments as well as in the BLUP datasets (**Figure 2**) indicated that the RILs were suitable for QTL analysis. The estimated H^2 of FLW, FLL, AD, PH, SL, SNS, and SD for ES RILs were ranged from 0.75 to 0.97, SL had the highest H^2 (0.97), followed by PH (0.95), and SD had the lowest H^2 (0.75), implicating that these traits were mainly controlled by genetic factors.

Relationships Among Seven Agronomic Traits

In the three environments as well as the BLUP dataset, for FLW, AD, FLL, PH, SL, SNS, and SD, significant correlations with coefficients ranging from 0.444 to 0.978 were detected ($P < 0.05$; **Table 2**). Correlation coefficients between seven agronomic traits using the BLUP dataset were presented in **Table 3**. FLW was significantly correlated to SNS ($r = 0.230$, $P < 0.05$), whereas there were no significant differences between FLW and other yield-related traits. For AD, positive and significant relationships ($0.269 \leq r \leq 0.446$, $P < 0.01$) between AD and FLL, SL and SNS were observed. FLL was significantly associated with AD, SL, and SNS ($0.269 \leq r \leq 0.455$, $P < 0.01$). PH was positively and significantly associated with SL and SNS ($0.237 \leq r \leq 0.446$, $P < 0.05$). Regarding SL, positive and

TABLE 2 | Phenotypic correlations of seven yield-related traits in different environments.

Trait	Environment	2019CZ	2018CZ	2017CZ	BLUP
FLW (cm)	2019CZ	1			
	2018CZ	0.744**	1		
	2017CZ	0.782**	0.676**	1	
	BLUP	0.921**	0.895**	0.923**	1
FLL (cm)	2019CZ	1			
	2018CZ	0.910**	1		
	2017CZ	0.897**	0.954**	1	
	BLUP	0.810**	0.819**	0.780**	1
AD (d)	2019CZ	1			
	2018CZ	0.444**	1		
	2017CZ	0.633**	0.654**	1	
	BLUP	0.659**	0.891**	0.950**	1
PH (cm)	2019CZ	1			
	2018CZ	0.852**	1		
	2017CZ	0.863**	0.887**	1	
	BLUP	0.956**	0.957**	0.958**	1
SL (cm)	2019CZ	1			
	2018CZ	0.921**	1		
	2017CZ	0.926**	0.934**	1	
	BLUP	0.973**	0.975**	0.978**	1
SNS	2019CZ	1			
	2018CZ	0.869**	1		
	2017CZ	0.863**	0.874**	1	
	BLUP	0.950**	0.960**	0.960**	1
SD	2019CZ	1			
	2018CZ	0.831**	1		
	2017CZ	0.591**	0.597**	1	
	BLUP	0.946**	0.938**	0.668**	1

*FLW, flag leaf width; FLL, flag leaf length; AD, anthesis date; PH, plant height; SL, spike length; SNS, spikelet number per spike; SD, spike density; **Significance level at P = 0.01.*

TABLE 3 | Phenotypic correlations between seven yield-related traits with BLUP data in 'AS985472'/'Sumai 3' (AS) population.

	AD	FLW	FLL	PH	SL	SNS	SD
AD	1						
FLW	−0.144	1					
FLL	0.269**	0.129	1				
PH	0.198	−0.047	0.143	1			
SL	0.446**	0.111	0.455**	0.466**	1		
SNS	0.385**	0.230**	0.404**	0.237**	0.668**	1	
SD	−0.251**	0.072	−0.229**	−0.409**	-0.724**	0.019	1

*AD, anthesis date; FLW, flag leaf width; FLL, flag leaf length; PH, plant height; SL, spike length; SNS, spikelet number per spike; SD, spike density; **Significance level at P = 0.01.*

[1]https://urgi.versailles.inra.fr/download/iwgsc/

[2]https://www.ncbi.nlm.nih.gov/assembly/GCA_002575655.1/#/def_asm_Primary_Assembly

[3]https://www.dropbox.com/sh/3dm05grokhl0nbv/AAC3wvlYmAher8fY0srX3gX9a?dl=0%2

significant correlations between SL and AD, FLL, PH and SNS were detected ($0.446 \leq r \leq 0.668$, $P < 0.01$). SNS was positively and significantly correlated with AD, FLW, FLL, PH, and SL ($0.230 \leq r \leq 0.668$, $P < 0.05$). For SD, SD was negatively and significantly related to AD, FLL, PH and SL ($-0.724 \leq r \leq -0.229$, $P < 0.05$).

TABLE 4 | Quantitative trait loci (QTL) analysis for seven yield-related traits with single environment method and BLUP data.

Trait	QTLs	Environments	Interval	Flanking maker	LOD	PVE (%)	Add
FLW (cm)	*QFlw.hebau-2D*	2017CZ	14.39 – 17.20	1128324| F| 0 ~ 100004655| F| 0	3.63	20.90	−0.11
		2019CZ	17.20 – 18.84	100004655| F| 0 ~ 1081989| F| 0	5.07	24.88	−0.09
		2018CZ	17.20 - 18.84	100004655| F| 0 ~ 1081989| F| 0	4.24	20.31	−0.09
		BLUP	17.20 – 18.84	100004655| F| 0 ~ 1081989| F| 0	4.16	19.30	−0.08
	QFlw.hebau-3D	2017CZ	48.40 – 49.45	2323109| F| 0 ~ 100005360| F| 0	2.65	15.58	−0.09
FLL (cm)	*QFll.hebau-5A*	BLUP	75.90 – 78.26	1136364| F| 0~3028423| F| 0	6.54	0.67	0.62
PH (cm)	*QPh.hebau-2D.1*	2019CZ	12.68 – 16.32	1116536| F| 0~2247268| F| 0	3.86	12.69	5.45
		2018CZ	12.68 – 16.32	1116536| F| 0~2247268| F| 0	5.79	10.10	5.38
		BLUP	12.68 – 16.32	1116536| F| 0~2247268| F| 0	8.25	11.30	5.36
	QPh.hebau-2D.2	2019CZ	7.82 – 12.03	3029203| F| 0~1119134| F| 0	3.63	11.74	−5.40
	QPh.hebau-4A.1	BLUP	76.74 – 77.42	2290028| F| 0~1121845| F| 0	7.10	9.41	−4.94
	QPh.hebau-4A.2	2018CZ	79.41 – 80.26	1372725| F| 0~1115816| F| 0	7.38	13.26	−6.19
	QPh.hebau-4B	2019CZ	82.24 – 85.62	1123959| F| 0~1123635| F| 0	6.84	24.74	−7.68
		2018CZ	82.24 – 85.62	1123959| F| 0~1123635| F| 0	12.18	25.51	−8.67
		2017CZ	82.24 – 85.62	1123959| F| 0~1123635| F| 0	4.82	30.68	−8.16
		BLUP	82.24 – 85.62	1123959| F| 0~1123635| F| 0	17.10	29.85	−8.86
	QPh.hebau-6A	2018CZ	60.43 – 64.22	1124209| F| 0~1283575| F| 0	7.50	13.52	6.25
		BLUP	60.43 – 64.22	1124209| F| 0~1283575| F| 0	10.02	14.23	6.03
SL (cm)	*QSl.hebau-2A*	2017CZ	110.73 – 113.18	2254084| F| 0~3026394| F| 0	2.84	13.64	−0.54
	QSl.hebau-2D	2019CZ	6.43 – 8.93	1090962| F| 0~1064588| F| 0	3.33	10.83	0.45
	QSl.hebau-4A	2019CZ	90.26 – 91.09	1115627| F| 0~100035209| F| 0	4.18	13.89	−0.51
	QSl.hebau-5A.1	2018CZ	20.01 – 21.26	3064643| F| 0~3029299| F| 0	3.29	14.41	−0.46
	QSl.hebau-5A.2	2019CZ	76.83 – 89.00	1664450| F| 0~1142113| F| 0	3.96	13.53	−0.51
	QSl.hebau-6D.1	2017CZ	9.46 – 13.31	100002192| F| 0~1132651| F| 0	4.46	22.52	−0.70
	QSl.hebau-6D.2	2018CZ	15.80 – 16.88	100007946| F| 0~1096139| F| 0	2.65	11.55	−0.41
	QSl.hebau-6D.3	BLUP	19.19 – 20.98	1138521| F| 0~1127306| F| 0	3.35	12.59	−0.44
SNS	*QSns.hebau-2D*	2019CZ	3.76 – 6.23	1101681| F| 0~3033925| F| 0	3.99	16.53	−0.66
		BLUP	3.76 – 6.23	1101681| F| 0~3033925| F| 0	3.04	14.69	−0.56
	QSns.hebau-3B	20118CZ	112.75 – 113.74	3034270| F| 0~1769222| F| 0	3.37	8.58	0.51
		2019CZ	112.75 – 113.74	3034270| F| 0~1769222| F| 0	2.55	10.79	0.52
	QSns.hebau-4A	2017CZ	90.26 – 91.09	1115627| F| 0~100035209| F| 0	3.41	22.96	−0.72
		2018CZ	90.26 – 91.09	1115627| F| 0~100035209| F| 0	5.29	13.50	−0.63
	QSns.hebau-5A	2019CZ	65.32 – 73.19	2275311| F| 0~1110394| F| 0	3.47	14.14	−0.59
	QSns.hebau-5D	2018CZ	0 – 3.55	2245326| F| 0~1269099| F| 0	4.52	11.11	−0.57
	QSns.hebau-7A	2018CZ	208.95 – 212.16	1067518| F| 0~1115252| F| 0	3.08	7.55	−0.47
	QSns.hebau-7B	2018CZ	96.84 – 97.69	1202000| F| 0~1229729| F| 0	6.41	17.22	0.71
SD	*QSd.hebau-2D*	2019CZ	3.95 – 5.17	1136748| F| 0~1111273| F| 0	2.81	10.19	−0.05
		2018CZ	3.95 – 5.17	1136748| F| 0~1111273| F| 0	2.95	12.83	−0.05
		BLUP	3.95 – 5.17	1136748| F| 0~1111273| F| 0	3.74	14.91	−0.05
	QSd.hebau-3D	2018CZ	33.84 – 46.08	100003200| F| 0~1203665| F| 0	2.62	11.22	0.04
	QSd.hebau-4B	2019CZ	60 – 80.99	1241081| F| 0~100003066| F| 0	2.77	12.36	0.06
	QSd.hebau-6D.1	BLUP	9.46 – 13.31	100002192| F| 0~1132651| F| 0	3.67	14.58	0.05
	QSd.hebau-6D.2	2019CZ	14.63 – 15.80	100035082| F| 0~100007946| F| 0	3.43	12.71	0.06
		2018CZ	14.63 – 15.80	100035082| F| 0~100007946| F| 0	2.84	12.32	0.05
AD (d)	*QAd.hebau-2A*	2018CZ	20.56 – 21.68	100004846| F| 0~2361439| F| 0	3.29	16.07	1.58
		BLUP	20.56 – 21.68	100004846| F| 0~2361439| F| 0	2.81	11.07	0.88
	QAd.hebau-2D.1	2017CZ	0 – 3.95	1084630| F| 0~1136748| F| 0	3.54	21.70	2.47
	QAd.hebau-2D.2	2019CZ	16.32 – 21.36	1020115| F| 0~1115866| F| 0	3.76	18.59	1.59
		BLUP	16.32 – 21.36	1020115| F| 0~1115866| F| 0	3.63	15.23	1.03

QTL, quantitative trait loci; PVE, Phenotypic variance explained; LOD, logarithm of odds; Add, Additive effect of a QTL; positive values: alleles from 'AS985472' are increasing the trait scores; negative values: alleles from 'Sumai 3' are increasing the scores; FLW, flag leaf width; FLL, flag leaf length; AD, anthesis date; PH, plant height; SL, spike length; SNS, spikelet number spike; SD, spike density; CZ, Chongzhou; BLUP, best linear unbiased prediction.

QTLs Analysis

QTLs conferring FLW, FLL, AD, PH, SL, SNS, and SD were detected in the AS population (**Table 4**). A total of two putative FLW QTLs (*QFlw.hebau-2D* and *QFlw.hebau-3D*) were identified on chromosomes 2D and 3D which individually explained 19.30 – 24.88% and 15.58% of the phenotypic variance, respectively; and the positive alleles of both were contributed by 'Sumai 3'. Among them, *QFlw.hebau-2D*, a major stable QTL flanked by *1128324| F| 0* and *1081989| F| 0* (**Table 4** and **Figure 3**), was detected in three environments and was confirmed by the BLUP dataset. For FLL, only a minor QTL (*QFll.hebau-5A-1*) was detected. With respect to AD, three QTLs with > 10% of the phenotypic variance were identified on chromosomes 2A and 2D, while they were detected in only one environment. For PH, six putative QTLs were mapped to chromosomes 2D, 4A, 4B and 6A. Two major and stable QTLs (*QPh.hebau-2D.1*, and *QPh.hebau-4B*) controlling PH accounted for the phenotypic variance up to 12.69 and 30.68%, and they were flanked by *1116536| F| 0 – 2247268| F| 0*, and *1123959| F| 0 - 1123635| F| 0*, respectively. 'AS985472' contributed the major alleles for increased PH at *QPh.hebau-2D.1*, and 'Sumai 3' contributed alleles for increased PH at *QPh.hebau-4B*. For SL, we detected eight loci, and two of the QTLs were co-localized with the major QTLs for SNS (*QSns.hebau-4A*) and for SD (*QSd.hebau-6D.1*). For SNS, seven QTLs with the 7.55 – 20.96% of the phenotypic variance were identified on chromosomes 2D, 3B, 4A, 5A, 5D, 7A, and 7B. Five QTLs for SD were mapped to chromosomes 2D, 3D, 4B, and 6D, of these, the major and stably expressed QTL *QSd.hebau-2D* accounted for 10.19–14.91% of the phenotypic variance and 'Sumai 3' contributed the major alleles at this locus.

The 94 AS RILs were divided into two groups according to the genotypes of the two flanking markers for major loci *QFlw.hebau-2D*, *QPh.hebau-2D.1*, *QPh.hebau-4B,* and *QSd.hebau-2D*. Student's *t*-test showed that RILs with the increased alleles from 'Sumai 3' significantly increased FLW in different environments as well as the BLUP dataset ($P < 0.01$, **Figure 4A**); and the lines with the positive alleles at *QPh.hebau-2D* from 'AS985472' were higher ($P < 0.05$, **Figure 4B**) than those from 'Sumai 3' in all environments except 2017CZ. In contrast, the lines with the increased alleles at *QPh.hebau-4B* from 'Sumai 3' were significantly ($P < 0.01$, **Figure 4C**) higher than those without major alleles. However, there were no differences between the lines with and without increasing alleles from *QSd.hebau-2D* (**Figure 4D**).

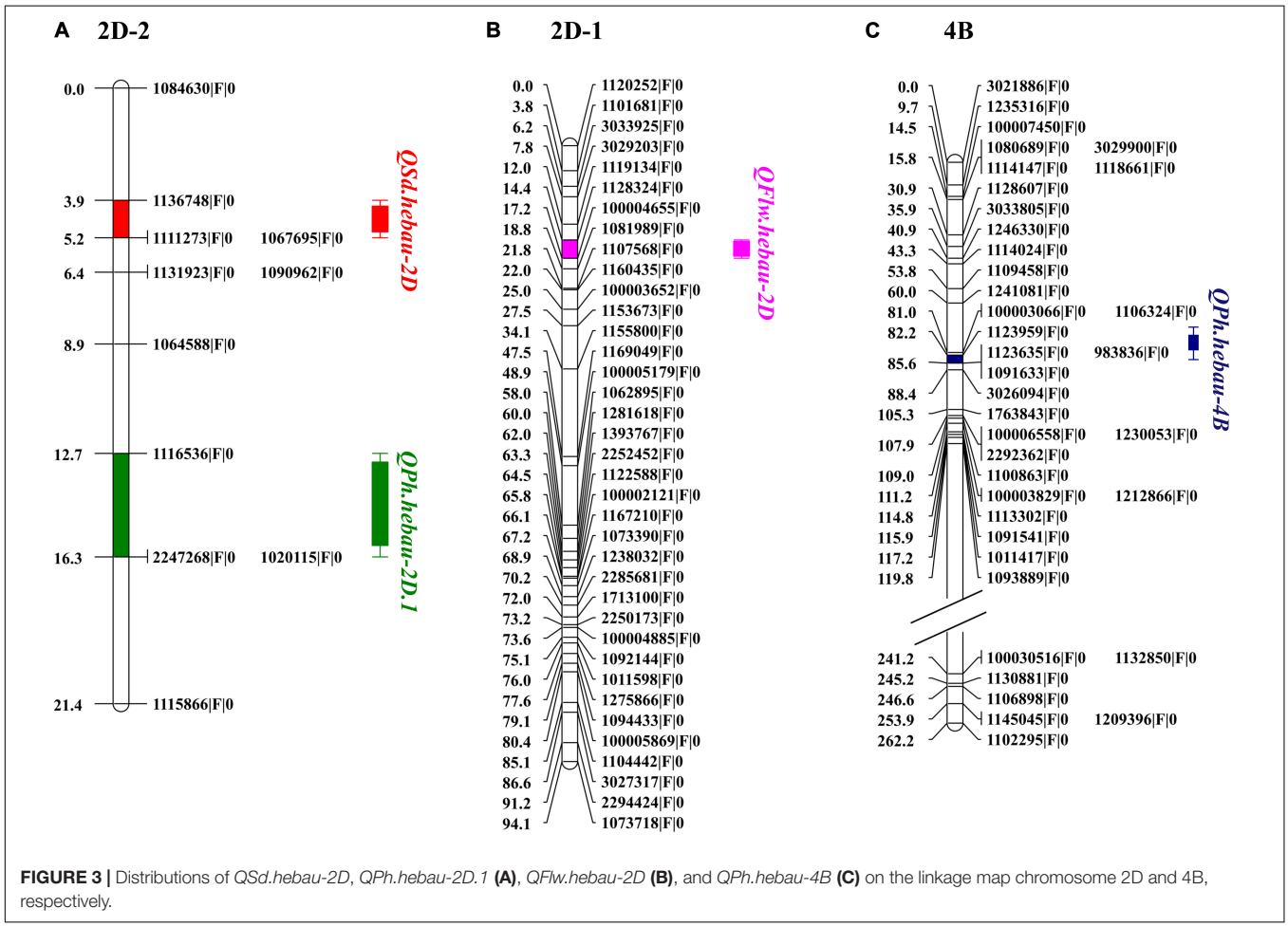

FIGURE 3 | Distributions of *QSd.hebau-2D*, *QPh.hebau-2D.1* **(A)**, *QFlw.hebau-2D* **(B)**, and *QPh.hebau-4B* **(C)** on the linkage map chromosome 2D and 4B, respectively.

DISCUSSION

Many QTL identified in this study were closely located to the chromosome region of known QTL. For example, *QPh.hebau-2D.1* for PH in the present study were located at 19.04 – 20.35 Mbp on CS 2DS and 19.80 – 21.14 Mbp on *Ae. tauschii* 2DS (**Figure 5A**), whereas *QPh.hebau-4B* located within the interval of 291.95 – 607.04 Mbp on CS 4BL and 276.11 – 586.46 Mbp on wild emmer 4BL (**Figure 5B**). The location of *QPh.hebau-4B* (291.95 – 607.04 Mbp) is far from RhtB1 (30.861–30.863 Mbp). These two QTL were physically located at the similar or overlapped positions as those reported previously by Wu et al. (2010) and Zhai et al. (2016), respectively. It is interesting that the height reducing allele of *QPh.hebau-2D.1* came from the tall parent Sumai 3. This result suggested that genotypes with a relatively poorer performance on height may still carry the beneficial allele that can be used for genetic improvement of the trait. Similar results was obtained in previous study. It has been shown that susceptible parent contributed the resistance alleles to various wheat diseases (Poole et al., 2012; Ma et al., 2019b). In addition to height QTL, a major and stably expressed QTL conferring SD designated as *QSd.hebau-2D*, which was located at 16.16 – 17.82 Mbp on CS and 16.39 – 18.36 Mbp on *Ae. tauschii* 2DS (**Figure 5A**), was in the overlapped physical region as the QTL reported by Heidari et al. (2011). The stable expression of these height and SD QTL under multiple genetic backgrounds in different studies emphasized their value for further fine mapping studies.

Two putative QTLs for FLW were detected on chromosomes 2D and 3D including *QFlw.hebau-2D* and *QFlw.hebau-3D*. *QFlw.hebau-2D*, located in the interval of *1128324| F| 0 – 1081989| F| 0*, was the major and stable locus identified in the present study. For chromosome 2D, numerous putative loci for flag-related traits and yield-related traits in wheat were identified (Fan et al., 2015; Ma et al., 2019a, 2020). In the present study, *QFlw.hebau-2D* was located in a 4.45 cM interval and physically mapped between 588.35 and 630.74 Mbp

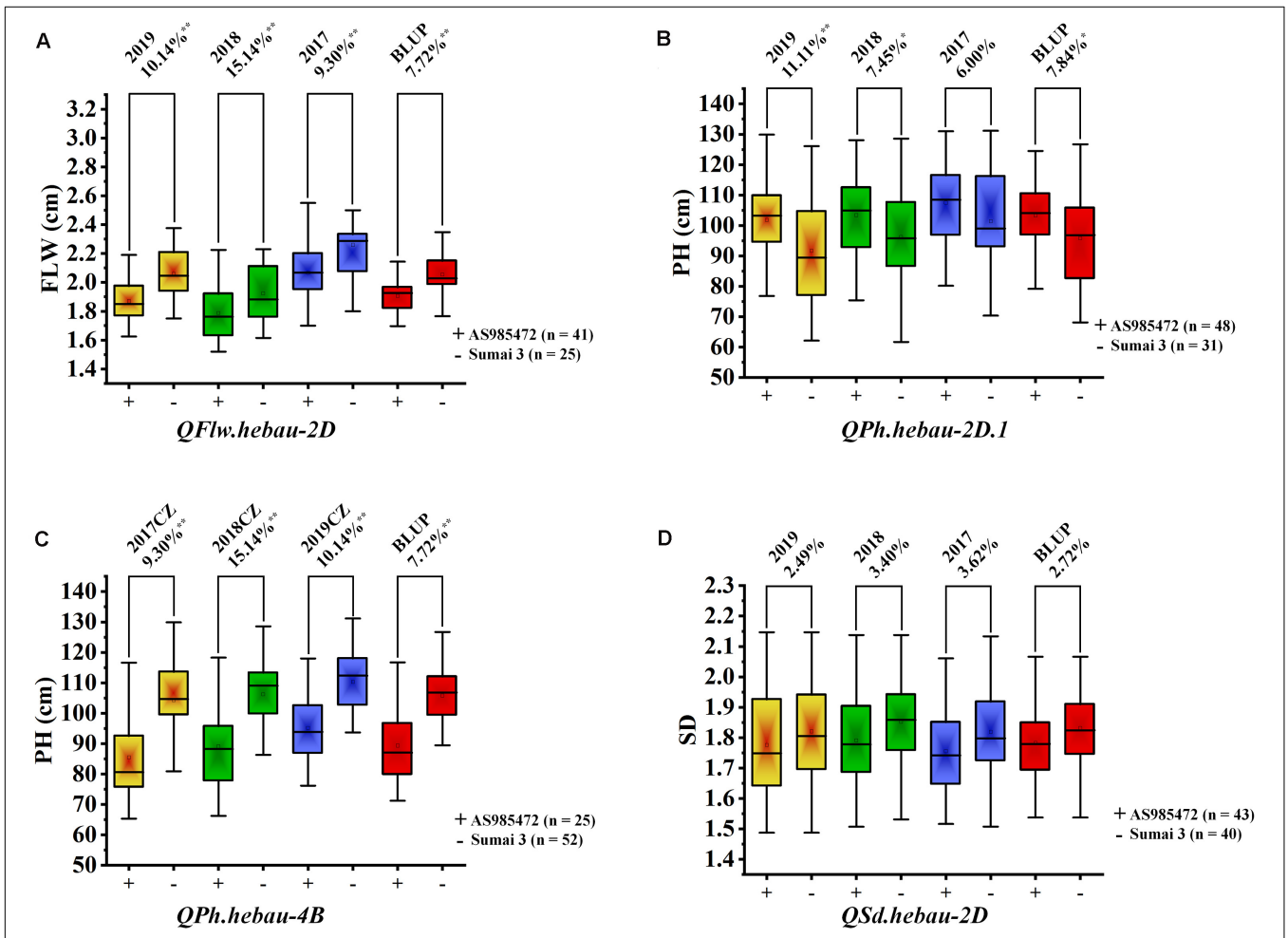

FIGURE 4 | The effects of the QTL *QFlw.hebau-2D* on flag leaf width **(A)**, *QPh.hebau-2D.1* on plant height **(B)**, *QPh.hebau-4B* on plant height **(C)** and *QSd.hebau-2D* on spike density **(D)**. Box plots represent RILs with and without major alleles which are grouped according to the flanking markers of the major QTLs. '+' indicates the homozygous lines from 'AS985472' and '–' indicates the homozygous lines from 'Sumai 3'. **Significant at $P = 0.01$, *Significant at $P = 0.05$.

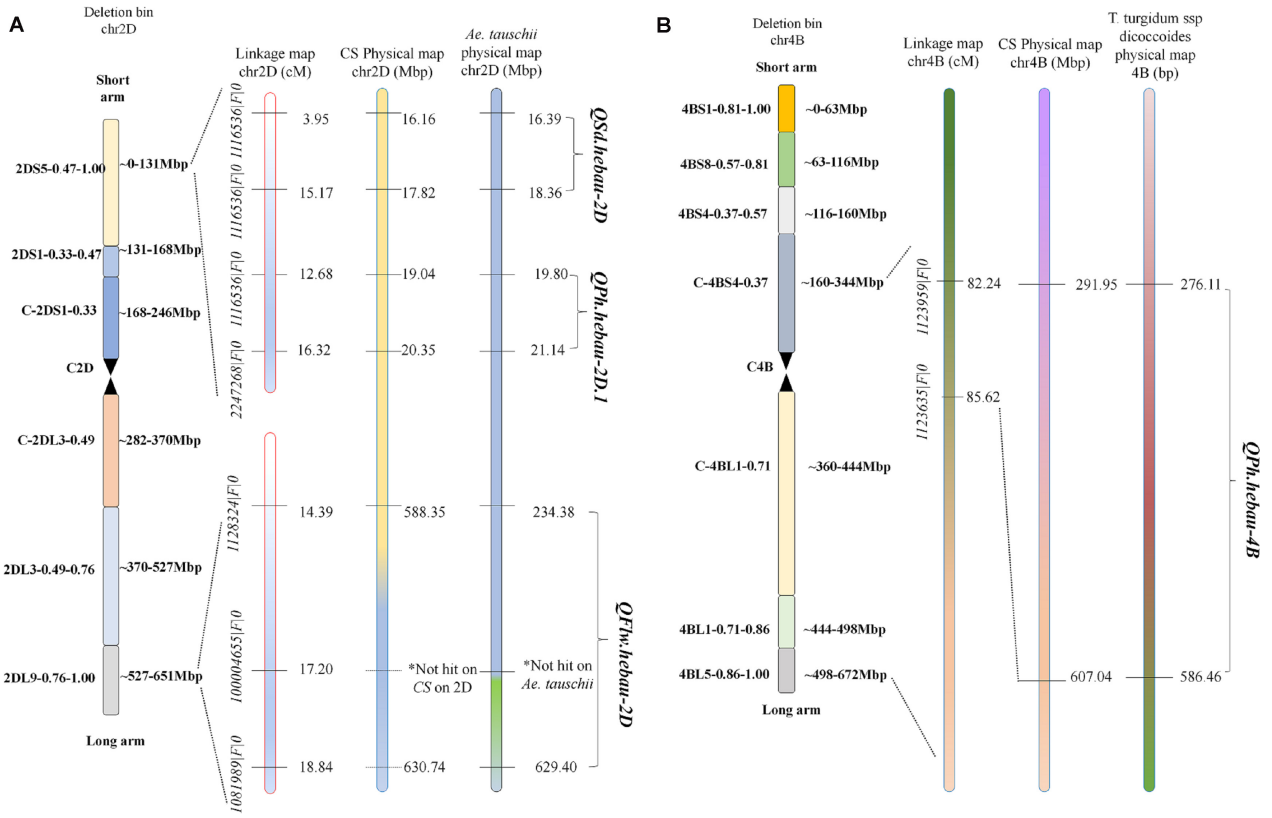

FIGURE 5 | The maps of *QFlw.hebau-2D*/*QPh.hebau-2D.1*/*QSd.hebau-2D* **(A)** and *QPh.hebau-4B* **(B)**.

on CS 2DL and 234.38 and 629.40 Mbp on *Ae. tauschii* 2DL (**Figure 5A**). Combined with the results from Ma et al. (2020), we reviewed the recently published articles related to FLW loci and compared the major FLW QTL detected in the present study with previous studies. Comparison analysis showed that it was overlapped with *qFlw-2D.2* (Fan et al., 2015) that was detected in an individual environment with minor effect. The large PVE (24.88) and LD (5.07) of *QFlw.hebau-2D* in our study suggested that value of this locus for FLW improvement.

Six hundred and eighty-nine putative candidate genes on CS genome were predicted at the interval of *QFlw.hebau-2D* (**Supplementary Table S1**). The results of the annotation indicated that several genes associated with plant growth and development. Additionally, there are numerous genes encoding the same protein. For example, auxin response factors, encoded by *TraesCS2D01G491200* and *TraesCS2D01G548900*, bind to TGTCTC auxin response elements in promoters of early auxin response genes (Tiwari et al., 2003), and thus regulating the development of plant tissues and organs. Twenty-two candidate genes including *TraesCS2D01G524300* encode an F-box protein which regulates floral development and the F-box protein TIR1 is an auxin receptor that mediates Aux/IAA degradation and auxin-regulated transcription (Dharmasiri et al., 2005). Flowering Locus T-like protein, encoded by *TraesCS2D01G538000* and *TraesCS2D01G538100*, is involved in the switch to flowering and thus aiding in grain set and dispersal.

MADS-box transcription factors are involved in various processes of plant growth and development (Ma et al., 2017). *TraesCS2D01G529700* encodes MADS-box transcription factor 8. Twelve genes encode SAUR-like auxin-responsive family proteins, which regulate cell expansion or division thus leading to leaf growth (Ren and Gray, 2015). These results showed that some of the putative candidate genes analyzed in the present study have key importance in understanding the genetic mechanism of flag leaf growth and development in wheat.

Significant genetic associations among seven yield-related traits were detected. SD was determined by the SNS divided by SL, thus SD was positively related with SNS and negatively related with SL, which was consistent with the phenotypic correlations in this study (**Table 3**). Significantly, several QTLs for SL with pleiotropic effects to SNS and SD that are the key components of grain yield in wheat were identified on chromosomes 4A and 6D, indicating that these traits may be controlled by the same locus. Major loci *QSd.hebau-2D* and *QPh.hebau-2D.1* were physically located at 16.16–17.82 Mbp and 19.04–20.35 Mbp, respectively, suggesting the intrinsic correlations may exist between SD and PH. Similarly, it is previously reported that various yield-related traits like biomass and SD (Kirigwi et al., 2007), SD and SL (Heidari et al., 2011), are controlled by the same QTLs or genes. The pleiotropic effects of these

loci further indicated their potential value for further researches and application in wheat breeding programs.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

JM and WZ designed the research. JJ, DL, and YQ performed the experiments. JJ and DL analyzed the data and wrote the manuscript. JM and JJ revised the manuscript. All authors read and approved the manuscript.

FUNDING

This research was supported by the National Key Research and Development Program of China (2017YFD0300906 and 2018YFD0300501) and State Key Laboratory of North China Crop Improvement and Regulation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.01012/full#supplementary-material

TABLE S1 | Predicated genes in the interval of the major QTL *QFlw.hebau-2D.*

REFERENCES

Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S., Gundlach, H., et al. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* 357, 93–97. doi: 10.1126/science.aan0032

Deng, S. M., Wu, X. Y., Wu, Y. Y., Zhou, R. H., Wang, H. G., Jia, J. Z., et al. (2011). Characterization and precise mapping of a QTL increasing spike number with pleiotropic effects in wheat. *Theor. Appl. Genet.* 122, 281–289. doi: 10.1007/s00122-010-1443-1

Dharmasiri, N., Dharmasiri, S., and Estelle, M. (2005). The F-box protein TIR1 is an auxin receptor. *Nature* 435, 441–445. doi: 10.1038/nature03543

Fan, X. L., Cui, F., Zhao, C. H., Zhang, W., Yang, L. J., Zhao, X. Q., et al. (2015). QTLs for flag leaf size and their influence on yield-related traits in wheat (*Triticum aestivum* L.). *Mol. Breed.* 35:24. doi: 10.1007/s11032-015-0205-9

Goddard, M. E. (1992). A mixed model for analyses of data on multiple genetic markers. *Theor. Appl. Genet.* 83, 878–886. doi: 10.1007/BF00226711

Hai, L., Guo, H. J., Wagner, C., Xiao, S. H., and Friedt, W. (2008). Genomic regions for yield and yield parameters in Chinese winter wheat (*Triticum aestivum* L.) genotypes tested under varying environments correspond to QTL in widely different wheat materials. *Plant Sci.* 175, 226–232. doi: 10.1016/j.plantsci.2008.03.006

Hedden, P. (2003). The genes of the green revolution. *Trends Genet.* 19, 5–9. doi: 10.1016/s0168-9525(02)00009-4

Heidari, B., Sayedtabatabaei, B. E., Saeidi, G., Kearsey, M. J., and Suenaga, K. (2011). Mapping QTL for grain yield, yield components, and spike features in a doubled haploid population of bread wheat. *Genome* 54, 517–527. doi: 10.1139/g11-017

Hu, J. M., Wang, X. Q., Zhang, G. X., Jiang, P., Chen, W. Y., Hao, Y. C., et al. (2020). QTL mapping for yield-related traits in wheat based on four RIL populations. *Theor. Appl. Genet.* 133, 917–933. doi: 10.1007/s00122-019-03515-w

Kirigwi, F. M., Van Ginkel, M., Brownguedira, G., Gill, B. S., Paulsen, G. M., and Fritz, A. K. (2007). Markers associated with a QTL for grain yield in wheat under drought. *Mol. Breed.* 20, 401–413. doi: 10.1007/s11032-007-9100-3

Li, C. L., Bai, G. H., Carver, B. F., Chao, S., and Wang, Z. L. (2016). Mapping quantitative trait loci for plant adaptation and morphology traits in wheat using single nucleotide polymorphisms. *Euphytica* 208, 299–312. doi: 10.1007/s10681-015-1594-x

Lin, H. X., Qian, H. R., Zhuang, J. Y., Lu, J., Min, S., Xiong, Z., et al. (1996). RFLP mapping of QTLs for yield and related characters in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 92, 920–927. doi: 10.1007/BF00224031

Liu, J. J., Luo, W., Qin, N. N., Ding, P. Y., Zhang, H., Yang, C. C., et al. (2018a). A 55 K SNP array-based genetic map and its utilization in QTL mapping for productive tiller number in common wheat. *Theor. Appl. Genet.* 131, 2439–2450. doi: 10.1007/s00122-018-3164-9

Liu, K., Sun, X. X., Ning, T. Y., Duan, X. X., Wang, Q. L., Liu, T. T., et al. (2018b). Genetic dissection of wheat panicle traits using linkage analysis and a genome-wide association study. *Theor. Appl. Genet.* 131, 1073–1090. doi: 10.1007/s00122-018-3059-9

Liu, K., Xu, H., Liu, G., Guan, P., Zhou, X., Peng, H., et al. (2018c). QTL mapping of flag leaf-related traits in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 131, 839–849. doi: 10.1007/s00122-017-3040-z

Luo, M. C., Gu, Y. Q., Puiu, D., Wang, H., Twardziok, S., Deal, K. R., et al. (2017). Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* 551, 498–502. doi: 10.1038/nature24486

Ma, J., Ding, P. Y., Liu, J. J., Li, T., Zou, Y. Y., Habib, A., et al. (2019a). Identification and validation of a major and stably expressed QTL for spikelet number per spike in bread wheat. *Theor. Appl. Genet.* 132, 3155–3167. doi: 10.1007/s00122-019-03415-z

Ma, J., Qin, N. N., Cai, B., Chen, G. Y., Ding, P. Y., Zhang, H., et al. (2019b). Identification validation of a novel major QTL for all-stage stripe rust resistance on 1BL in the winter wheat line 20828. *Theor. Appl. Genet.* 132, 1363–1373. doi: 10.1007/s00122-019-03283-7

Ma, J., Tu, Y., Zhu, J., Luo, W., Liu, H., Li, C., et al. (2020). Flag leaf size and posture of bread wheat: genetic dissection, QTL validation and their relationships with yield-related traits. *Theor. Appl. Genet.* 133, 297–315. doi: 10.1007/s00122-019-03458-2

Ma, J., Yang, Y. J., Luo, W., Yang, C. C., Ding, P. Y., Liu, Y. X., et al. (2017). Genome-wide identification and analysis of the MADS-box gene family in bread wheat (*Triticum aestivum* L.). *PLoS One* 12:e0181443. doi: 10.1371/journal.pone.0181443

Poole, G. J., Smiley, R. W., Paulitz, T. C., Walker, C. A., Carter, A. H., See, D. R., et al. (2012). Identification of quantitative trait loci (QTL) for resistance to fusarium crown rot (*Fusarium pseudograminearum*) in multiple assay environments in the pacific northwestern us. *Theor. Appl. Genet.* 125, 91–107. doi: 10.1007/s00122-012-1818-6

Ren, H., and Gray, W. M. (2015). SAUR proteins as effectors of hormonal and environmental signals in plant growth. *Mol. Plant* 8, 1153–1164. doi: 10.1016/j.molp.2015.05.003

Smith, S. E., Kuehl, R. O., Ray, I. M., Hui, R., and Soleri, D. (1998). Evaluation of simple methods for estimating broad-sense heritability in stands of randomly planted genotypes. *Crop Sci.* 38, 1125–1129. doi: 10.2135/cropsci1998.0011183X003800050003x

The International Wheat Genome Sequencing Consortium [IWGSC] (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191

Tiwari, S. B., Hagen, G., and Guilfoyle, T. J. (2003). The roles of auxin response factor domains in auxin-responsive transcription. *Plant Cell* 15, 533–543. doi: 10.1105/tpc.008417

Woodruff, D. R., and Tonks, J. (1983). Relationship between time of anthesis and grain yield of wheat genotypes with differing developmental pattern. *Crop Pasture Sci.* 34, 1–11. doi: 10.1071/ar9830001

Wu, X. S., Wang, Z. H., Chang, X. P., and Jing, R. L. (2010). Genetic dissection of the developmental behaviours of plant height in wheat under diverse water regimes. *J. Exp. Bot.* 61, 2923–2937. doi: 10.1093/jxb/erq 117

Xie, G. Q., Zhang, M. C., Chakraborty, S., and Liu, C. (2007). The effect of 3BS locus of Sumai 3 on fusarium head blight resistance in Australian wheats. *Aust. J. Exp. Agri.* 47, 603–607. doi: 10.1071/EA05250

Yang, D. L., Liu, Y., Cheng, H. B., Chang, L., Chen, J. J., Chai, S. X., et al. (2016). Genetic dissection of flag leaf morphology in wheat (*Triticum aestivum* L.) under diverse water regimes. *BMC Genet.* 17:94. doi: 10.1186/s12863-016-0399-9

Zhai, H. J., Feng, Z. Y., Li, J., Liu, X. Y., Xiao, S. H., Ni, Z. F., et al. (2016). QTL analysis of spike morphological traits and plant height in winter wheat (*Triticum aestivum* L.) using a high-density SNP and SSR-based linkage map. *Front. Plant Sci.* 7:1617. doi: 10.3389/fpls.2016. 01617

Zhai, H. Q., Cao, S. Q., Wan, J. M., Zhang, R. X., Lu, W., Li, L. B., et al. (2002). Relationship between leaf photosynthetic function at grain filling stage and yield in super high-yielding hybrid rice (*Oryza sativa*. L). *Sci. China life Sci.* 45:637. doi: 10.1007/BF02879752

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 04 September 2020
doi: 10.3389/fgene.2020.580452

Novel QTL Conferring Phosphorus Acquisition and Utilization Efficiencies in Barley

*Shangqing Gao[1†], Jiaqi Xia[1†], Shu Yuan[1], Youjie Shen[1], Xinting Zhong[1], Senfeng Zhang[1], Yuhang Li[1], Deyi Hu[1], Jian Zeng[1], Ting Lan[1], Yaxi Liu[2] and Guangdeng Chen[1]**

[1] College of Resources, Sichuan Agricultural University, Chengdu, China, [2] Triticeae Research Institute, Sichuan Agricultural University, Chengdu, China

Phosphorus (P) deficiency in agricultural soil is a major constraint for crop production and increasing P acquisition efficiency (PAE) of plants is considered as one of the most cost-effective solutions for yield increase. The objective of this study was to detect quantitative trait loci (QTL) controlling (PAE) and P utilization efficiency (PUE) in barley under applied (+P) and non-applied P (−P) conditions. Based on the analysis of a recombinant inbred lines (RILs) population derived from a cross between a malting barley variety and a wild barley accession, 17 QTL controlling PAE, PUE and yield traits were detected. The phenotypic variation explained by each of these QTL ranges from 11.0 to 24.7%. Significant correlation was detected between most of P-related traits and yield traits. Five QTL clusters were identified on four different chromosomes (1H, 3H, 5H, and 7H). Two of the QTL clusters, located on chromosome 1H (for GPUP/PUP) and 7H (for SPUE/SPC), respectively, are novel. Fourteen genes located in the interval harboring the major QTL were identified as candidates associated with P efficiency. The stable QTL for PAE, PUE and yield-related traits could be important for breeding P-efficient barley varieties.

Keywords: barley, phosphorus deficiency, phosphorus acquisition, phosphorus utilization, quantitative trait loci

OPEN ACCESS

Edited by:
Penghao Wang,
Murdoch University, Australia

Reviewed by:
Meixue Zhou,
University of Tasmania, Australia
Tingqiang Li,
Zhejiang University, China

***Correspondence:**
Guangdeng Chen
gdchen@sicau.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *06 July 2020*
Accepted: *12 August 2020*
Published: *04 September 2020*

Citation:
Gao S, Xia J, Yuan S, Shen Y, Zhong X, Zhang S, Li Y, Hu D, Zeng J, Lan T, Liu Y and Chen G (2020) Novel QTL Conferring Phosphorus Acquisition and Utilization Efficiencies in Barley. Front. Genet. 11:580452. doi: 10.3389/fgene.2020.580452

INTRODUCTION

Phosphorus (P) is one of the most important mineral nutrient elements for plant development and it plays an irreplaceable role in agricultural productions (Su et al., 2006; Wang and Yan, 2010; Noack et al., 2014). Although agronomic inputs of P fertilizer and manure collectively exceeded P removal by harvested crops at the global scale, P deficits covered almost 30% of the global cropland area (Macdonald et al., 2011). The application of P fertilizers is one of the most effective methods to alleviate soil P deficiency (Shen et al., 2011). However, most of the applied P may be immobilized by calcium (Ca) and magnesium (Mg) in alkaline soils or by ferrum (Fe) and aluminum (Al) in acid soils (Holford, 1997; Yang et al., 2011). Thus only 10–20% P could be absorbed in the year of application (Yang et al., 2011). The mineral phosphate is non-renewable (Sattari et al., 2012) and the un-absorbed P will run-off into surface water to cause eutrophication (Carpenter, 2008). It is widely believed that developing cultivars with high-efficiency P acquisition and utilization in P-deficient soils is one of the most economical and sustainable solutions in crop breeding programs worldwide (Yan et al., 2006; Liang et al., 2010).

It has been reported that P efficiency in crops was affected by a number of quantitative trait loci (QTL) (Yang et al., 2011). Based on mechanisms developed by plants to acquire and utilize P from the soils (Yang et al., 2011), the mapped QTL could be classified into two major types: for P acquisition efficiency (PAE) and for P utilization efficiency (PUE) (Yang et al., 2011). QTL for P efficiency have been identified in several crops including common wheat (*Triticum aestivum*) (Su et al., 2006), maize (*Zea mays*) (Cai et al., 2012), and rice (*Oryza sativa*) (Nishida et al., 2018). In wheat, a large number of P efficiency-related QTL have been detected. For example, Su et al. (2006) detected several QTL on seven different chromosomes (3B, 4B, 5A, 5D, 6A, 6B, and 7A, respectively) for PUE under P deficient and sufficient conditions. Four important QTL clusters controlling PAE and PUE were found at both seedling and mature stages of plant development (Yuan et al., 2017) (Six QTL for PAE were co-located with the QTL for zinc concentration or content in wheat grains (Shi et al., 2008). Therefore, genome-wide scanning for QTL controlling PAE and PUE could be an important work in crop breeding programs.

Barley (*Hordeum vulgare* L.) is the fourth largest cereal crop worldwide, and it is widely used as animal feed and in brewing industry (Schulte et al., 2009). Numerous QTL or genes for important traits of barley have been mapped, fine mapped or even cloned, and they showed great potential in MAS (Peighambari et al., 2005; Tavakol et al., 2015; Li et al., 2016). However, only a limited number of QTL for P efficiency have been reported (Gong et al., 2016; Guo et al., 2017), especially from mature plants. In the present study, the whole genome linkage map of a recombinant inbred lines (RIL) population derived from the cross between a wild barley accession and a cultivar was used to detect QTL for PAE, PUE and yield at maturity stage under both P applied and non-applied soil conditions. Overall, the objective of our study was to focus on excavating the major and stable QTL or QTL clusters that could provide available information for the barley breeding programs.

TABLE 1 | Physicochemical property of the experimental soil.

Classification	Values	Units
Soil pH	6.89	–
Organic content	15.8	g kg^{-1}
Total nitrogen (N)	0.4	g kg^{-1}
Alkali-hydrolyzable N	44.68	mg kg^{-1}
Available P	5.14	mg kg^{-1}
Rapidly available kalium (K)	23.69	mg kg^{-1}
Ca_2−P	7.25	mg kg^{-1}
Ca_8−P	3.97	mg kg^{-1}
Ca_{10}−P	230.67	mg kg^{-1}
Al−P	16.2	mg kg^{-1}
Fe−P	76.85	mg kg^{-1}
Organic−P	100.54	mg kg^{-1}
Active phytate P	2.25	mg kg^{-1}
Secondary active phytate P	145.12	mg kg^{-1}
Secondary stable phytate P	39.75	mg kg^{-1}
High stable phytate P	8.08	mg kg^{-1}

TABLE 2 | The investigated traits and the measurements in this study.

Type	Trait	Abbreviation	Method	Unit
PAE	Grain P concentration	GPC	Measure	g $plant^{-1}$GY
	Straw P concentration	SPC	Measure	g $plant^{-1}$SY
	Plant P concentration	PC	PUP/DM	g $plant^{-1}$DM
	Grain P uptake	GPUP	GY × GPC	g $plant^{-1}$
	Straw P uptake	SPUP	SY × SPC	g $plant^{-1}$
	Plant P uptake	PUP	GPUP+SPUP	g $plant^{-1}$
PUE	Grain P utilization efficiency	GPUE	GY/PUP	g GY g^{-1}P
	Straw P utilization efficiency	SPUE	SY/PUP	g SY g^{-1}P
	Plant P utilization efficiency	PUE	DM/PUP	g DW g^{-1}P
Yield trait	Grain yield	GY	Measure	g $plant^{-1}$
	Straw yield	SY	Measure	g $plant^{-1}$
	Dry matter	DM	GY + SY	g $plant^{-1}$

PAE, P acquisition efficiency; PUE, P use efficiency.

MATERIALS AND METHODS

Plant Materials

An RIL population consisting of 128 $F_{7:9}$ lines derived from a cross between Baudin, a high yielding malting variety adjust to a longer season, higher rainfall cropping region and parts of the medium rainfall cropping region of Western Australia and a wild barley (*H. spontaneum*) accession, CN4027 was used in this study.

Experimental Design

The calcareous alluvial soil used in pot trials was collected from Shuangbai village, Dujiangyan city in Sichuan, China. The physicochemical properties of the soil were shown in **Table 1**. Two pot trials were carried out (one from November 2016 to June 2017 and the other from November 2017 to June 2018) in the rainproof screenhouses of Sichuan Agricultural University. Each of the trials consisted of two treatments [P deficiency (−P, without P application) and P sufficiency (+P, 30 mg phytate P was applied per kilogram soil)] with three replications. Split plot arrangement was used in these trials. Twelve kg air-dried soil with 1.5 g N and 1.5 g K was crushed and uniformly mixed.

For each replication, 10 uniformly sized seeds of each of RILs as well as the parents were surface-sterilized by soaking in a 10% solution of hydroperoxide (H_2O_2) for 30 min followed by washing in deionized water. The disinfected seeds were placed in a chamber with constant temperature humidity (20°C, 60% humidity) for germination. Five germinated seeds were planted in each of the pots. Seedlings were thinned at 3-leaf stage and three seedlings were retained in each pot and used for further experiments.

Phenotypic Evaluation

Data for grain yield (GY), straw yield (SY), and dry matter (DM) were collected at maturity (**Table 2**). Harvested kernels and straws were placed in an over at 105°C for 30 min and then dried at 75°C until constant weights were reached. The dried materials were weighed and grounded to powder

for determining phosphorus content with the H_2SO_4-H_2O_2-molybdenum antimony colorimetric method (Guo et al., 2017). The evaluated traits were listed in **Table 2**.

QTL Mapping

Means of the traits for each RIL from the three replications were used to detect QTL. The genetic linkage map obtained from this population in an earlier study (Guo et al., 2017) was used for QTL mapping. Linkage analysis was carried out using the computer package JoinMap®4.0 (Van Ooijen, 2006). Segregation ratios of assessed markers were tested by Chi-square goodness-of-fit to a 1:1 ratio at the significance level of $p = 0.01$. LOD thresholds from 3 to 10 were tested and an optimum threshold was obtained. The Kosambi mapping function was used to convert recombination ratios to map distances. MapQTL® 5.0 (Van Ooijen, 2004) was used for QTL analysis. QTL were named following recommendations from the International Rules of Genetic Nomenclature[1].

Identification of Candidate Genes

To identify candidate genes for P-related loci, sequences of DArT markers linked closely to QTL were selected from the DArT genotyping provided by the Triticarte Pty. Ltd[2]. The database Ensembl Plants[3] was exploited to determine the physical positions and contigs of the P-related loci. Candidate genes were then further retrieved using physical position and contigs by database BARLEX searches (the Barley Draft Genome Explorer[4]). Orthologous genes for the candidate genes in other cereals and *Arabidopsis* were obtained from the PGSB database[5].

RESULTS

Phenotypic Variation of Assessed Traits

Phenotype values for each trait were significantly influenced by the application of P. GY, SY, and DM of the parents at −P were significantly lower than those under +P (**Table 3**). At the same P condition, significant differences between parents were detected for GY, SY, and DM. The cultivated barley Baudin exhibited higher values for each of the traits compared with those for the wild barley genotype CN4027 (**Table 3**). The coefficient of variation (*CV*) for the yield-related traits between the two treatments ranged from 28.30 to 52.99% (**Table 3**). Transgressive segregation in the RIL population was observed for all three traits (**Table 3**). Continuous variations with approximately normal distributions were detected for these traits.

Compared with those at −P, higher values were obtained for grain P concentration (GPC), straw P concentration (SPC), grain P uptake (GPUP), straw P uptake (SPUP) and plant P uptake (PUP) at +P. However, the values for grain P utilization efficiency (GPUE), straw P utilization efficiency (SPUG), and plant P utilization efficiency (PPUE) were higher at −P (**Table 3**). There were significant differences in these traits between the two parents under the two different treatments. Compared with the wild barley genotype CN4027, Baudin showed higher values for GPC, SPC, GPUP, SPUP, and PUP but lower values for GPUE, SPUG, and PUE (**Table 3**). The *CV* of the seven PAE- and PUE-related traits in the population between the two treatments also showed a wide distribution ranging from 31.80 to 62.54%. The transgressive segregation and approximately normal distributions could be also detected for the P-related traits (**Table 3**).

Correlations Between P-Related and Yield Traits

Phenotypic correlations between P-related and yield traits under the two treatments were presented in **Table 4**. Significant correlations were detected between each of the three yield traits and most of the P-related traits under both P conditions ($P < 0.01$ or 0.05). P-concentrations related traits, including GPC, SPC, and PC, were significantly and negatively correlated with the three yield traits (GY, SY, and DM) under the −P treatment except SPC, PC in trial 2. P-uptake related traits, including GPUP, SPUP and PUP, were significantly and positively correlated with the three yield-related traits (GY, SY, and DM) except GPUP in trial 1 and GPUP and PUP in trial 2 (**Table 4**). Traits associated with PUE, including SPUE and PUE, showed a significantly positive correlation with two of the yield-related traits (SY and DM) with coefficients ranging from 0.205 to 0.508 ($P < 0.05$). PC was significantly and negatively correlated with GY and SY under +P treatment. Traits related with P-uptake, including GPUP, SPUP, and PUP, were significantly and positively correlated with DM in both trials with coefficients ranging from 0.362 to 0.748 ($P < 0.01$).

Detection of QTL

A total of 17 QTL for P- and yield-related traits were identified. Phenotypic variations explained by these loci varied from 11.0 to 24.7% (**Table 5** and **Figure 1**). LOD values for these loci ranged from 3.03 to 7.31 (**Table 5**). The 17 QTL were distributed on 4 chromosomes, including 1H (2 QTL), 3H (9 QTL), 5H (2 QTL), and 7H (4 QTL). Positive alleles for eight of these QTL were contributed by Baudin, while the remaining nine were contributed by CN4027. In addition, nine of these 17 QTL were detected in two trials, and three of them were detected at the two P conditions.

One QTL (*Qgpc.sau-3H*) for GPC was detected on 3H, and its positive allele was derived from Baudin (**Figure 1** and **Table 5**). *Qgpc.sau-3H* was detected under the two different P conditions in Trial 1 but only under −P condition in Trial 2. Phenotypic variations explained by this locus varied from 13.4 – 14.7 %. Two QTL for SPC were detected and they were mapped on chromosomes 2H and 7H, respectively. The positive alleles for both loci were contributed by Baudin (**Figure 1** and **Table 5**). *Qspc.sau-3H* was detected at both +P and −P conditions in Trial 1, explaining 17.9 and 11.0% of the phenotypic variation, respectively. *Qspc.sau-7H* was detected at −P condition in

[1]http://wheat.pw.usda.gov/ggpages/wgc/98/Intro.htm

[2]https://www.diversityarrays.com/

[3]http://plants.ensembl.org/index.html

[4]https://apex.ipk-gatersleben.de/apex/f?p=284:10

[5]https://www.helmholtz-muenchen.de/pgsb

TABLE 3 | Variations of evaluated traits for the RIL population and their parents at maturity.

Types	Traits	Treatments	Trial 1						Trial 2					
			Parents		RILs				Parents		RILs			
			Baudin	CN4027	Mean + SD	Min	Max	*CV*%	Baudin	CN4027	Mean + SD	Min	Max	*CV*%
PAE	GPC	−P	2.108	1.620	2.835 ± 1.025	0.902	5.385	36.15	2.361	1.814	3.050 ± 1.182	0.876	6.031	38.75
		+P	3.831	2.890	3.188 ± 1.019	1.459	6.104	31.96	3.534	3.054	3.308 ± 1.052	1.337	6.466	31.80
	SPC	−P	0.981	0.467	0.963 ± 0.296	0.568	1.756	30.73	1.100	0.610	0.901 ± 0.348	0.412	1.832	38.62
		+P	1.521	0.820	1.145 ± 0.396	0.614	3.056	34.58	1.302	1.000	1.190 ± 0.430	0.529	2.678	36.13
	PC	−P	1.571	1.171	1.951 ± 0.671	0.866	3.706	34.39	1.747	1.211	2.033 ± 0.836	0.850	4.137	41.12
		+P	2.697	1.991	2.201 ± 0.839	1.069	4.775	38.11	2.443	2.166	2.280 ± 0.895	1.026	4.919	39.25
	GPUP	−P	8.527	5.176	12.199 ± 4.821	2.813	28.472	39.51	8.786	5.333	12.464 ± 5.274	2.937	32.129	42.31
		+P	18.508	11.387	16.341 ± 10.559	1.478	73.363	64.61	19.292	13.597	18.308 ± 11.451	5.119	78.005	62.54
	SPUP	−P	3.610	0.951	3.677 ± 1.119	1.845	6.535	30.43	3.886	1.803	3.482 ± 1.483	1.719	7.156	42.59
		+P	7.086	2.481	5.136 ± 2.349	1.713	10.805	45.73	6.794	3.389	5.911 ± 2.720	1.777	11.709	46.01
	PUP	−P	12.137	6.127	15.877 ± 5.571	6.037	30.826	35.08	12.672	7.136	16.093 ± 6.763	6.608	34.861	42.02
		+P	25.594	13.868	21.493 ± 11.306	4.700	79.112	52.60	26.086	16.986	24.273 ± 12.558	8.031	85.152	51.73
PUE	GPUE	−P	0.474	0.617	0.400 ± 0.189	0.186	1.109	47.25	0.424	0.551	0.381 ± 0.193	0.166	1.008	50.65
		+P	0.261	0.346	0.331 ± 0.109	0.164	0.685	32.93	0.283	0.327	0.321 ± 0.115	0.155	0.692	35.82
	SPUE	−P	1.019	2.141	1.096 ± 0.327	0.569	1.762	29.83	0.909	1.639	1.175 ± 0.447	0.530	2.426	38.04
		+P	0.657	1.220	0.938 ± 0.303	0.327	1.630	32.30	0.768	1.000	0.913 ± 0.323	0.373	1.892	35.37
	PUE	−P	0.636	0.854	0.559 ± 0.205	0.270	1.155	36.67	0.572	0.826	0.537 ± 0.230	0.242	1.176	42.83
		+P	0.371	0.502	0.495 ± 0.181	0.209	0.936	36.56	0.409	0.462	0.474 ± 0.180	0.203	0.974	37.97
Yield trait	GY	−P	4.045	3.195	4.371 ± 1.327	2.013	7.913	30.35	3.721	2.939	4.140 ± 1.237	1.982	7.280	29.87
		+P	4.831	3.940	5.086 ± 2.861	0.515	16.607	56.25	5.459	4.452	5.521 ± 2.926	1.140	16.766	52.99
	SY	−P	3.680	2.037	3.933 ± 1.197	1.378	6.663	30.43	3.533	2.956	3.929 ± 1.215	2.006	6.397	30.92
		+P	4.659	3.026	4.544 ± 1.832	1.708	9.529	40.31	5.218	3.389	5.084 ± 2.079	1.913	10.672	40.89
	DM	−P	7.725	5.232	8.305 ± 2.384	4.277	13.087	28.70	7.254	5.895	8.070 ± 2.284	4.566	12.268	28.30
		+P	9.490	6.966	9.651 ± 3.629	3.437	20.228	37.60	10.677	7.841	10.603 ± 4.008	3.862	20.742	37.80

PAE, P acquisition efficiency; PUE; for P use efficiency; GPC, Grain P concentration; SPC, Straw P concentration; PC, Plant P concentration; GPUP, Grain P uptake; SPUP, Straw P uptake; PUP, Plant P uptake; GPUE, Grain P utilization efficiency; SPUE, Straw P utilization efficiency; PUE, Plant P utilization efficiency; GY, Grain yield; SY, Straw yield; DM, Dry matter.

TABLE 4 | Correlations between P- and yield-related traits in the RIL population at maturity in barley.

Trial	Traits	Treatments	GPC	SPC	PC	GPUP	SPUP	PUP	GPUE	SPUE	PUE
T1	GY	−P	−0.218*	−0.432**	−0.186*	0.528**	0.241**	0.523**	0.105	0.427**	0.150
		+P	0.034	0.046	0.401**	0.882**	−0.077	0.839**	−0.032	−0.040	−0.436**
	SY	−P	−0.309**	−0.513**	−0.423**	0.172	0.527**	0.253**	0.149	0.494**	0.337**
		+P	−0.246**	−0.112	−0.436**	−0.193*	0.783**	−0.024	0.187*	0.037	0.421**
	DM	−P	−0.288**	−0.520**	−0.330**	0.397**	0.417**	0.437**	0.139	0.508**	0.264**
		+P	−0.105	−0.020	0.123	0.693**	0.362**	0.748**	0.074	−0.016	−0.162
T2	GY	−P	−0.195**	−0.134	−0.090	0.511**	0.296**	0.559**	0.079	0.148	0.069
		+P	0.001	0.069	0.394**	0.857**	−0.034	0.826**	0.017	−0.063	−0.378**
	SY	−P	−0.354**	−0.202*	−0.449**	0.032	0.617**	0.618	0.268**	0.213*	0.391**
		+P	−0.250**	−0.181	−0.493**	−0.178	0.710**	0.016	0.170	0.173	0.510**
	DM	−P	−0.313**	−0.191*	−0.307**	0.310**	0.519**	0.414**	0.197*	0.205*	0.262**
		+P	−0.149	−0.048	0.047	0.639**	0.396**	0.709**	0.117	0.049	−0.024

*PAE, P acquisition efficiency; PUE, for P use efficiency; GPC, Grain P concentration; SPC, Straw P concentration; PC, Plant P concentration; GPUP, Grain P uptake; SPUP, Straw P uptake; PUP, Plant P uptake; GPUE, Grain P utilization efficiency; SPUE, Straw P utilization efficiency; PUE, Plant P utilization efficiency; GY, Grain yield; SY, Straw yield; DM, Dry matter; *Significant at $P \leq 0.05$ level. **Significant at $P \leq 0.01$ level.*

Trial 1 and it explained 14.4% of the phenotypic variation. One significant QTL (*Qpc.sau-3H*) for PC was detected on chromosome 3H (**Figure 1** and **Table 5**) under both P conditions in both trials conducted. The phenotypic variation explained by these QTL ranged from 17.0 to 21.0% (**Table 5**). The positive allele of *Qpc.sau-3H* was contributed by Baudin.

TABLE 5 | QTL for P- and yield traits at maturity in barley.

Traits	QTL	Ch.[a]	Detection environment	Marker interval	LOD	PVE(%)[b]	Additive[c]
GPC	*Qgpc.sau-3H*	3H	T1−P	3264976–6283337	4.32	14.7	0.338
		3H	T2−P	3264976–3931069	3.90	13.4	0.380
		3H	T1+P	3433408–3264976	3.98	13.6	0.302
SPC	*Qspc.sau-3H*	3H	T1−P	4169758–4000155	5.01	17.9	0.100
		3H	T1+P	3264074–6248874	3.17	11.0	0.111
	Qspc.sau-7H	7H	T1−P	4186071–5241092	4.23	14.4	0.090
PC	*Qpc.sau-3H*	3H	T1−P	3264976–6283337	5.22	17.5	0.236
		3H	T2−P	3265461–4000155	5.98	21.0	0.272
		3H	T1+P	3264074–3264111	5.92	20.6	0.311
		3H	T2+P	3433483–4000155	4.67	17.0	0.277
GPUP	*Qgpup.sau-1H*	1H	T2+P	3272157–3395878	3.19	11.4	3.638
SPUP	*Qspup.sau-3H*	3H	T2+P	3257547–3926168	3.11	10.9	−0.811
PUP	*Qpup.sau-1H*	1H	T2+P	3272157–3395878	3.03	11.0	3.683
GPUE	*Qgpue.sau-3H*	3H	T1+P	3258653–3931069	4.82	16.5	−0.033
SPUE	*Qspue.sau-3H*	3H	T1−P	4006691–3266050	5.59	20.0	−0.120
	Qspue.sau-7H	7H	T1−P	3918068–5241092	3.98	13.6	−0.099
PUE	*Qpue.sau-3H*	3H	T1−P	3264976–3264111	3.68	12.7	−0.063
		3H	T2−P	3264976–3256099	4.22	15.3	−0.068
		3H	T1+P	3264074–6283337	7.31	24.7	−0.072
		3H	T2+P	3258653–3264111	4.88	17.9	−0.056
GY	*Qgy.sau-5H*	5H	T1−P	3266971–5241415	3.81	13.1	0.374
		5H	T2−P	3266971–3276370	3.54	12.6	0.338
	Qgy.sau-7H	7H	T1−P	3273337–4012713	3.55	12.3	−0.364
		7H	T2−P	3273337–3255382	3.28	11.4	−0.324
SY	*Qsy.sau-3H*	3H	T1−P	3264976–3263403	4.59	16.0	−0.378
		3H	T2−P	3264976–3263403	4.65	16.6	−0.390
		3H	T1+P	5250378–3257547	4.69	16.1	−0.651
		3H	T2+P	3433408–3257547	4.66	15.8	−0.735
DM	*Qdm.sau-3H*	3H	T1−P	4169758–4000155	3.95	14.3	−0.677
		3H	T2−P	3264976–4000155	3.64	13.7	−0.623
	Qdm.sau-5H	5H	T1−P	3266971–5241415	3.57	12.4	0.629
		5H	T2−P	3266971–5241415	3.74	13.0	0.608
	Qdm.sau-7H	7H	T1−P	3273337–4012713	3.57	12.6	−0.643
		7H	T2−P	3265420-3255382	3.41	11.8	−0.585

[a]Chromosome. [b]Percentage of the phenotypic variation explained by the QTL. [c]Additive effect of a QTL. Positive values of additive effect indicate that alleles from Baudin are increasing the trait scores, and negative values indicate that alleles from CN4027 are increasing the score.

One QTL (*Qgpup.sau-1H*) for GPUP was detected on 1H under the +P condition from Trial 2, and its positive allele was derived from Baudin (**Figure 1** and **Table 5**). This locus explained 11.4% of the phenotypic variation. One QTL (*Qspup.sau-3H*) for SPUP was detected on 3H, and the positive allele of it was derived from CN4027 (**Figure 1** and **Table 5**). *Qspup.sau-3H* was detected at +P condition in Trial 1, and it explained 10.9% of the phenotypic variation. One QTL (*Qpup.sau-1H*) for PUP was detected on 1H under the +P condition from Trial 2, and its positive allele was also derived from Baudin (**Figure 1** and **Table 5**). This locus explained 11.0% of the phenotypic variation.

One QTL (*Qgpue.sau-3H*) for GPUE was located on 3H under the +P condition from Trial 1 (**Figure 1** and **Table 5**). This locus explained 16.5% of the phenotypic variation and its positive allele was derived from CN0427. Two QTL for SPUE were located, on 3H and 7H chromosome, respectively. Positive alleles for both QTL were derived from CN0427 (**Figure 1** and **Table 5**). They explained 20.0 and 13.6% of the phenotypic variation, respectively. One stable QTL for PUE (*Qpue.sau-3H*) was detected on 3H under both P conditions from both Trial 1 and Trial 2 (**Figure 1** and **Table 5**). It explained 12.7 – 24.7% of the phenotypic variation. The positive allele of this locus was derived from CN0427.

Three QTL (*Qdm.sau-3H*, *Qdm.sau-7H*, and *Qdm.sau-7H*) for DM were detected under the −P condition from the two trials. They were mapped on chromosomes 3H, 5H and 7H, respectively (**Figure 1** and **Table 5**). The phenotypic variation explained by these QTL ranged from 11.8 to 14.3% (**Table 5**). The positive alleles of *Qdm.sau-3H* and *Qdm.sau-7H* were contributed by CN4027 and that of *Qdm.sau-5H* was contributed by Baudin. Two QTL (*Qgy.sau-5H* and *Qgy.sau-7H*) for GY were detected and they were mapped on chromosomes 5H and 7H, respectively

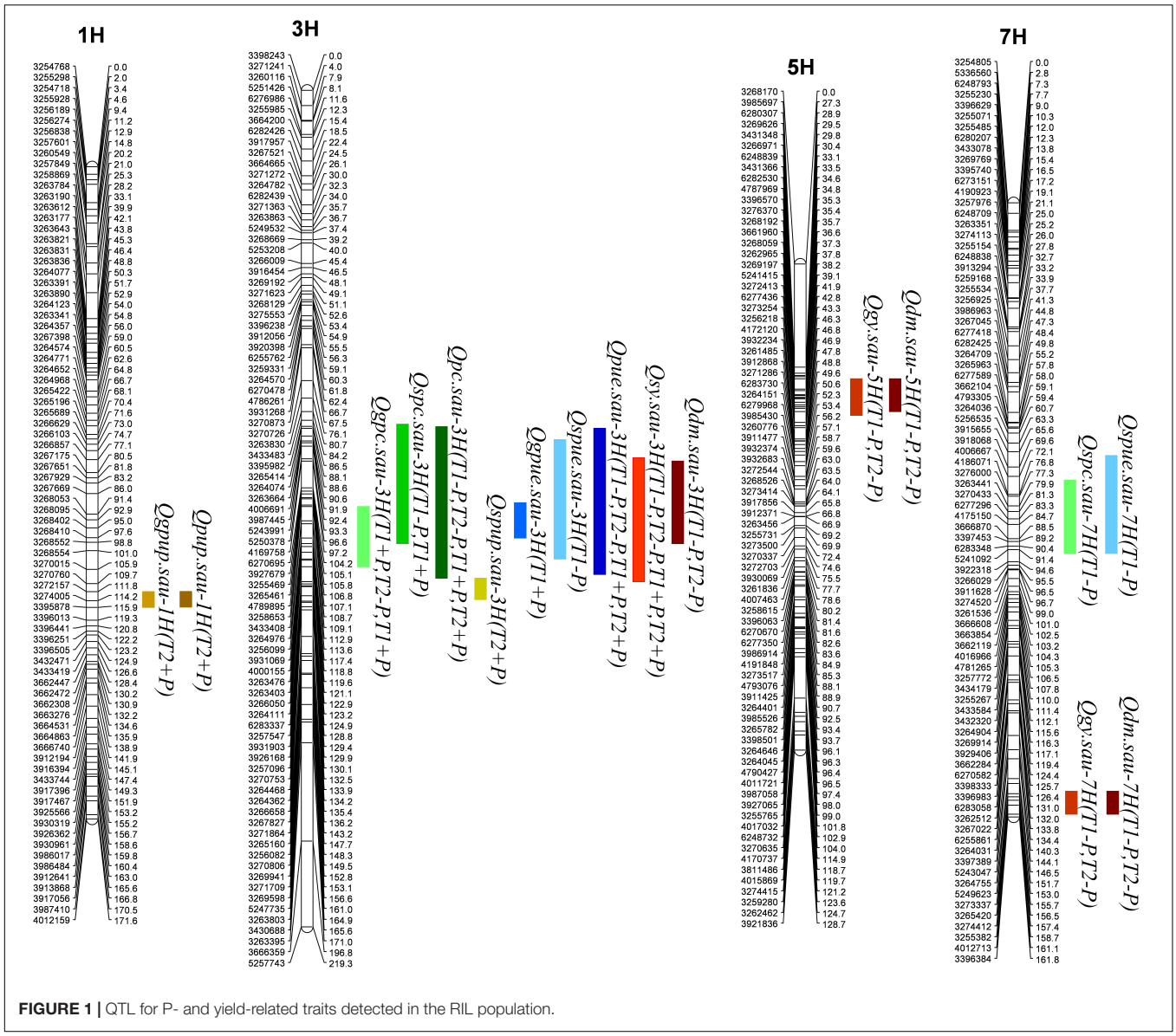

FIGURE 1 | QTL for P- and yield-related traits detected in the RIL population.

(**Figure 1** and **Table 5**). *Qgy.sau-5H* explained 12.6 and 13.1% of the phenotypic variation, respectively, and its positive alleles were contributed by Baudin. *Qgy.sau-7H* explained 11.4 and 12.3% of the phenotypic variation, respectively, and its positive allele was contributed by CN4027. A stable QTL (*Qsy.sau-3H*) for SY was detected on chromosome 3H under both the −P and +P conditions from both trials (**Figure 1** and **Table 5**). It was derived from CN4027, and explained 15.8 – 16.6% of the phenotypic variation.

Candidate Genes for the P-Related Loci

A total of fourteen candidate genes located in intervals harboring the P-related loci were detected by searching the BARLEX database. These candidate genes could be divided into four categories: acid phosphatase, phosphate transporter, acid phosphatase/vanadium-dependent haloperoxidase-related protein, and phospholipid metabolism (**Table 6**). The acid phosphatase gene (*AK354580*) and phosphate transporter gene (*MLOC_61737.2*) were identified in the intervals harboring *Qspue.sau-7H* and *Qspc.sau-7H* for PUE and PC. The candidate genes for the other two categories were located on three chromosomes and they confer PUE, PUP, and PC, respectively. One, one and two genes encoding acid phosphatase/vanadium-dependent haloperoxidase-related proteins were identified on 1H, 3H, and 7H, respectively. Two, two, and three genes associated with phospholipid metabolisms were identified on 1H, 3H, and 7H, respectively.

DISCUSSION

P is one of the macroelements for plants, and it was non-substitutable in many physiological and biochemical

TABLE 6 | Candidate genes or proteins in chromosomal intervals containing the various P-related loci at maturity in barley.

Stable QTL	Chr	Gene Name	Functional Annotation	*Oryza sativa*	*Zea mays*	*Arabidopsis thaliana*	Functional Annotation
Qpup.sau-1H *Qgpup.sau-1H*	1H	*MLOC_69370.3*	Acid phosphatase/ vanadium-dependent haloperoxidase related	*LOC_Os05g47530.1*	*GRMZM2G177617_T05*	\	
		MLOC_16149.3	Digeranylgeranylglyceryl phosphate synthase	*LOC_Os07g38850.1*	*GRMZM2G113476_T03*	*AT3G11945.2*	Homogentisate prenyltransferase
		AK356092	Putative phosphatidylinositol transfer protein	*LOC_Os01g50616.1* *LOC_Os05g46720.1* *LOC_Os02g04020.1*	*GRMZM2G073571_T03* *GRMZM2G171354_T01* *GRMZM2G157043_T01* *GRMZM2G174990_T03* *GRMZM2G355610_T01*	*AT1G19650.1* *AT1G75370.2* *AT2G21520.2* *AT4G39170.1*	Phosphatidylinositol/ phosphatidylcholine transfer protein SFH4 Sec14p-like phosphatidylinositol transfer family protein Sec14p-like phosphatidylinositol transfer family protein Phosphatidylinositol/ phosphatidylcholine transfer protein SFH4
Qgpc.sau-3H *Qspc.sau-3H* *Qpc.sau-3H* *Qspup.sau-3H*	3H	*MLOC_56200.1*	Acid phosphatase/ vanadium-dependent haloperoxidase related protein	*LOC_Os01g67560.1*	*GRMZM2G091435_T01*	\	
Qgpue.sau-3H *Qspue.sau-3H* *Qpue.sau-3H*		*MLOC_53886.2*	2-phosphoglycerate kinase-related protein	*LOC_Os02g57400.1* *LOC_Os09g39870.1*	*GRMZM2G017334_T01* *GRMZM2G342327_T03* *GRMZM2G123544_T01*	*AT5G60760.1* *AT3G45090.1*	P-loop NTPase domain-containing protein LPA1 homolog 1
		AK356601	Phosphatidylinositol transfer protein SFH5	*LOC_Os05g35460.1* *LOC_Os01g65380.1*	*GRMZM2G033641_T01* *GRMZM2G081652_T01* *GRMZM2G033649_T01*	*AT4G09160.1* *AT1G72160.1*	Patellin-5 Patellin-3
		MLOC_19234.6	Phosphatidylinositol-4-phosphate 5-kinase	*LOC_Os12g13440.1* *LOC_Os09g10650.1* *LOC_Os08g01390.1* *LOC_Os04g59540.1*	*GRMZM2G343218_T01* *GRMZM2G428386_T02* *GRMZM2G059179_T01* *GRMZM2G040296_T01*	*AT1G34260.1*	Putative 1-phosphatidylinositol-3-phosphate 5-kinase FAB1D
Qspue.sau-7H *Qspc.sau-7H*	7H	*AK354580*	Acid phosphatase 1	*LOC_Os06g36400.1*	*GRMZM2G103526_T01*	*AT4G29260.1* *AT4G29270.1*	Acid phosphatase-like protein Acid phosphatase-like protein
		MLOC_69490.1	Acid phosphatase/ vanadium-dependent haloperoxidase related protein	*LOC_Os08g03370.1*	*GRMZM2G057258_T01*	*AT1G24350.3* *AT1G67600.1*	Acid phosphatase/ vanadium-dependent haloperoxidase-related protein
		MLOC_38965.4	Acid phosphatase/ vanadium-dependent haloperoxidase-related protein	*LOC_Os06g33930.1*	*GRMZM2G071638_T01*	*AT3G12685.1*	Acid phosphatase/ vanadium-dependent haloperoxidase-related protein
		MLOC_61737.2	Phosphate transporter 1;8	*LOC_Os06g21950.1*	\	*AT1G20860.1* *AT1G76430.1*	phosphate transporter 1;8 Putative phosphate transporter
		AK362615	Phospholipase DDHD1	*LOC_Os08g01920.1*	*GRMZM2G023335_T01* *GRMZM2G318860_T02*	*AT1G31480.1*	Phospholipase SGR2
		MLOC_22194.1	1-phosphatidylinositol-3-phosphate 5-kinase	*LOC_Os04g59540.1* *LOC_Os08g01390.1* *LOC_Os09g10650.1* *LOC_Os12g13440.1*	*GRMZM2G040296_T01* *GRMZM2G059179_T01* *GRMZM2G428386_T02* *GRMZM2G343218_T01*	*AT1G34260.1*	Putative 1-phosphatidylinositol-3-phosphate 5-kinase FAB1D
		AK367170	1-phosphatidylinositol-3-phosphate 5-kinase	*LOC_Os03g28140.1* *LOC_Os06g14750.1* *LOC_Os08g34950.1* *LOC_Os09g23740.1* *LOC_Os08g33200.1*	*GRMZM2G066876_T01* *GRMZM2G092595_T01* *GRMZM2G111208_T01* *GRMZM2G132373_T01* *GRMZM2G153722_T01*	*AT1G71010.1* *AT3G14270.1* *AT4G33240.1*	Putative 1-phosphatidylinositol-3-phosphate 5-kinase FAB1C 1-phosphatidylinositol-3-phosphate 5-kinase FAB1B 1-phosphatidylinositol-3-phosphate 5-kinase FAB1A

metabolisms. Plant production could be reduced or even fail completely when soil P is deficient. As most of the applied P cannot be absorbed by plants, improving P uptake and use could offer a better sustainable method than only relying on fertilizer application (Gong et al., 2016). To explore desirable genes for P efficiency in barley, we investigated several P-related traits based on a RIL population derived from a cross between the wild barley CN4027 and the barley cultivar Baudin under −P and +P conditions. A total of 17 QTL, forming five clusters, were detected on chromosomes 1H, 3H, 5H, and 7H under the two different P conditions. Two of the QTL clusters, located on 1H (for GPUP/PUP) and on 7H (for SPUE/SPC), respectively, are novel as no other QTL conferring P-relative traits has ever been reported on these chromosomes.

How PAE and PUE Affect P Efficiency in Barley

The two parents of the mapping population used in this study showed relatively large differences in each of the investigated traits at maturity under both P conditions studied. The wild barley genotype CN4027 showed higher P utilization efficiency (GPUE, SPUE, PUE), while Baudin showed higher values in P acquisition efficiency traits (including GPC, SPC, PC, GPUP, SPUP, and PUP). The results from the phenotypic analysis were consistent with those from QTL identification in this study. QTL mapping revealed that positive alleles for most of the loci of PAE were derived from Baudin, indicating that this genotype had higher P acquisition efficiency than that of CN4027. Furthermore, positive alleles for QTL conferring GPUE and PUE were contributed by CN4027, implying that this genotype showed greater P utilization efficiency than that of Baudin.

The yield traits including GY, SY, and DM were significantly and positively correlated with PAE (GPUP, SPUP, and PUP) and PUE (GPUE, SPUE, and PUE) at both the −P and +P conditions (**Table 4**). This finding was highly consistent with those obtained at seeding stage in this population (Guo et al., 2017). And a similar result was observed in wheat (Su et al., 2009) and *Brassica napus* (Yang et al., 2011). While P concentrations including GPC, SPC, and PC were significantly and negatively correlated with most of three yield traits (**Table 4**). Thus, we think it will be challengeable to develop a cultivar with improved both PAE (PC and PUP) and PUE.

The QTL for PAE and PUE

In this study, a total of 17 QTL for PAE, PUE, and yield traits were detected on five QTL clusters under two P conditions. A novel QTL cluster for SPUE/SPC was located on 7H under −P condition. Various candidate genes located in this QTL cluster were detected using database BARLEX searching as described in sorghum (Mahmoud et al., 2018). The acid phosphatase (AK354580) and phosphate transporter (MLOC_61737.2) genes located in this QTL cluster were identified in the interval of *Qspue.sau-7H* and *Qspc.sau-7H* (**Table 6**). The phosphate transporter 1;8 was a high affinity phosphate transporter which was reportedly a kind of phosphate transporter induced by low phosphorus (Raghothama, 2000). The acid phosphatase 1 was also induced by low phosphorus (Baldwin et al., 2001; Zhang et al., 2014). Thus, these two genes were likely important candidates for the QTL cluster for SPUE/SPC on 7H.

Three QTL clusters containing seven QTL for PAE and four QTL for PUE were identified on chromosomes 1H, 3H, and 7H. Candidate genes related to the acid phosphatase/vanadium-dependent haloperoxidase-related protein and phospholipid metabolism were located on these three QTL clusters. As an important phosphorus component in plants, phospholipid played a major role in phosphorus metabolic process. The expression of phospholipid metabolism genes was significantly different under the different P treatments (Pariascatanaka et al., 2009; Ren et al., 2011). However, the mechanisms of PAE and PUE regulated by phospholipid and acid phosphatase/vanadium-dependent haloperoxidase-related protein have not yet been reported, providing valuable clues for further dissecting their molecular mechanisms for P efficiency in barley.

It was reported that high P efficiency in plants could be achieved through improving PAE or PUE (Parentoni and Junior, 2008). Some scientists held the view that P efficiency was mainly determined by PAE (Parentoni and Junior, 2008; Richardson et al., 2009). While Veneklaas et al. (2012) hypothesized that PUE might play a major role in P efficiency. And it was reported that PUE and PAE were intrinsically linked (Su et al., 2006). The identified QTL clusters for several different traits might explain their significant correlations. For example, PAE and PUE showed significant correlation to three yield traits, and the QTL for these traits were all located in the same interval on 3H, indicating that they were linked closely or even be controlled by a same gene. Additionally, several QTL for PAE and PUE have been detected in the same region on 3H at seeding stage in barley (Guo et al., 2017). The QTL for PAE were also detected on 3A and 3B of bread wheat (Shi et al., 2008; Su et al., 2009). Chromosome 3H of barley was homologous to 3A, 3B and 3D of wheat (Islam and Shepherd, 1992), and the genes were highly conserved between wheat and barley (Devos, 2005). These results further verified the existence of a QTL cluster for P efficiency on 3H.

The Challenge to Improve P Efficiency

An ideal P efficient genotype is usually characterized by high capacity to acquire more P in the P-deficient environment (i.e., PAE) and/or by high ratio of biomass and P content (i.e., PUE) (Guo et al., 2012). Results from the correlation analysis and QTL mapping indicated that enhancing PAE (including PUP, GPUP, and SPUP) and PUE would improve yield of barley under both +P and −P conditions. However, we observed that GPC and SPC would reduce yield at both +P and −P conditions. This means that it is not easy to simultaneously improve PAE and PUE. This finding is consistent with the results from Su et al. (2009) who reported that PAE and PUE were negatively correlated in wheat. A large number of QTL for P-efficiency have been reported in the last decade. However, few researches were utilized in crop breeding. We thus need to accelerate identifying

major and stable QTL for PAE or PUE and developing their linked markers for MAS in barley breeding.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

SG, JX, and SY developed the RIL population. SG, JX, YS, XZ, and GC performed the pot trials and tested yield traits. YuL, DH, JZ, TL, and YaL determined the P content of plant materials and soils. SG, JX, SY, and GC analyzed the data. SG, JX, and GC wrote the manuscript. GC conceived and designed the experiments. All the authors read and approved the final manuscript.

FUNDING

The research reported in this article was partially funded by the National Natural Science Foundation of China (No. 31971944), Science and Technology Project of Sichuan Province (No. 2019YFN0150), Project of Sichuan Province Youth Science and Technology Innovation Team (20CXTD0062), and the Key Science and Technology Program of Chengdu (2019-YF09-00082-SN).

REFERENCES

Baldwin, J. C., Karthikeyan, A. S., and Raghothama, K. G. (2001). LEPS2, a phosphorus starvation-induced novel acid phosphatase from tomato. *Plant Physiol*. 125, 728–737. doi: 10.1104/pp.125.2.728

Cai, H., Chu, Q., Gu, R., Yuan, L., Liu, J., Zhang, X., et al. (2012). Identification of QTLs for plant height, ear height and grain yield in maize (*Zea mays* L.) in response to nitrogen and phosphorus supply. *Plant Breed*. 131, 502–510. doi: 10.1111/j.1439-0523.2012.01963.x

Carpenter, S. R. (2008). Phosphorus control is critical to mitigating eutrophication. *Proc. Natl. Acad. Sci. U.S.A*. 105, 11039–11040. doi: 10.1073/pnas.0806112105

Devos, K. M. (2005). Updating the 'crop circle'. *Curr. Opin. Plant Blot*. 8, 155–162. doi: 10.1016/j.pbi.2005.01.005

Gong, X., Wheeler, R., Bovill, W. D., and Mcdonald, G. K. (2016). QTL mapping of grain yield and phosphorus efficiency in barley in a Mediterranean-like environment. *Theor. Appl. Genet*. 129, 1657–1672. doi: 10.1007/s00122-016-2729-8

Guo, J., Chen, G., Zhang, X., Li, T., Yu, H., and Chen, H. (2017). Mapping quantitative trait loci for tolerance to phosphorus-deficiency at the seedling stage in barley. *Euphytica* 213:114. doi: 10.1007/s10681-017-1907-3

Guo, Y., Kong, F., Xu, Y., Zhao, Y., Liang, X., Wang, Y. Y., et al. (2012). QTL mapping for seedling traits in wheat grown under varying concentrations of N, P and K nutrients. *Theor. Appl. Genet*. 124, 851–865. doi: 10.1007/s00122-011-1749-7

Holford, I. C. R. (1997). Soil phosphorus: its measurement, and its uptake by plants. *Soil Res*. 35, 227–240. doi: 10.1071/s96047

Islam, A. K. M. R., and Shepherd, K. W. (1992). Substituting ability of individual barley chromosomes for wheat chromosomes 1. substitutions involving barley chromosomes 1, 3 and 6. *Plant Breed*. 109, 141–150. doi: 10.1111/j.1439-0523.1992.tb00164.x

Li, K., Hegarty, J., Zhang, C., Wan, A., Wu, J., Guedira, G. B., et al. (2016). Fine mapping of barley locus Rps6 conferring resistance to wheat stripe rust. *Theor. Appl. Genet*. 129, 845–859. doi: 10.1007/s00122-015-2663-1

Liang, Q., Cheng, X., Mei, M., Yan, X., and Liao, H. (2010). QTL analysis of root traits as related to phosphorus efficiency in soybean. *Ann. Bot*. 106, 223–234. doi: 10.1093/aob/mcq097

Macdonald, G. K., Bennett, E. M., Potter, P., and Ramankutty, N. (2011). Agronomic phosphorus imbalances across the world's croplands. *Proc. Natl. Acad. Sci. U.S.A*. 108, 3086–3091. doi: 10.1073/pnas.1010808108

Mahmoud, A. F., Abouelwafa, S. F., and Shehzad, T. (2018). Identification of charcoal rot resistance QTLs in sorghum using association and in silico analyses. *J. Appl. Genet.* 59, 243–251. doi: 10.1007/s13353-018-0446-5

Nishida, S., Dissanayaka, D. M. S. B., Honda, S., Tateishi, Y., Chuba, M., Maruyama, H., et al. (2018). Identification of genomic regions associated with low phosphorus tolerance in japonica rice (Oryza sativa L.) by QTL-Seq. *Soil Sci. Plant Nutr*. 64, 278–281. doi: 10.1080/00380768.2017.1412238

Noack, S. R., Mclaughlin, M. J., Smernik, R. J., Mcbeath, T. M., and Armstrong, R. (2014). Phosphorus speciation in mature wheat and canola plants as affected by phosphorus supply. *Plant Soil* 378, 125–137. doi: 10.1007/s11104-013-2015-3

Parentoni, S., and Junior, C. (2008). Phosphorus acquisition and internal utilization efficiency in tropical maize genotypes. *Pesqui. Agropecu. Bras*. 43, 893–901. doi: 10.1590/s0100-204x2008000700014

Pariascatanaka, J., Satoh, K., Rose, T. J., Mauleon, R., and Wissuwa, M. (2009). Stress response versus stress tolerance: a transcriptome analysis of two rice lines contrasting in tolerance to phosphorus deficiency. *Rice* 2, 167–185. doi: 10.1007/s12284-009-9032-0

Peighambari, S. A., Samadi, B. Y., Nabipour, A., Charmet, G., and Sarrafi, A. (2005). QTL analysis for agronomic traits in a barley doubled haploids population grown in Iran. *Plant Sci*. 169, 1008–1013. doi: 10.1016/j.plantsci.2005.05.018

Raghothama, K. G. (2000). Phosphate transport and signaling. *Curr. Opin. Plant Biol*. 3, 182–187. doi: 10.1016/s1369-5266(00)00062-5

Ren, F., Chen, L., Guo, Q., Zhong, H., Wu, Y., Chang, Y.-Y., et al. (2011). Identification and expression analysis of genes induced by phosphate starvation in leaves and roots of *Brassica napus*. *Plant Growth Regul*. 65, 65–81. doi: 10.1007/s10725-011-9575-0

Richardson, A. E., Hocking, P. J., Simpson, R. J., and George, T. S. (2009). Plant mechanisms to optimise access to soil phosphorus. *Crop Pasture Sci*. 60, 124–143. doi: 10.1071/CP07125

Sattari, S. Z., Bouwman, A. F., Giller, K. E., and Van Ittersum, M. K. (2012). Residual soil phosphorus as the missing piece in the global phosphorus crisis puzzle. *Proc. Natl. Acad. Sci. U.S.A*. 109, 6348–6353. doi: 10.1073/pnas.1113675109

Schulte, D., Close, T. J., Graner, A., Langridge, P., Matsumoto, T., Muehlbauer, G., et al. (2009). The international barley sequencing consortium—at the threshold of efficient access to the barley genome. *Plant Physiol*. 149, 142–147. doi: 10.1104/pp.108.128967

Shen, J., Yuan, L., Zhang, J., Li, H., Bai, Z., Chen, X., et al. (2011). Phosphorus dynamics: from soil to plant. *Plant Physiol*. 156, 997–1005. doi: 10.1104/pp.111.175232

Shi, R., Li, H., Tong, Y., Jing, R., Zhang, F., and Zou, C. (2008). Identification of quantitative trait locus of zinc and phosphorus density in wheat (*Triticum aestivum* L.) grain. *Plant Soil* 306, 95–104. doi: 10.1007/s11104-007-9483-2

Su, J., Xiao, Y., Ming, L., Liu, Q., Li, B., Tong, Y., et al. (2006). Mapping QTLs for phosphorus-deficiency tolerance at wheat seedling stage. *Plant Soil* 281, 25–36. doi: 10.1007/s11104-005-3771-5

Su, J., Zheng, Q., Li, H., Li, B., Jing, R., Tong, Y.-P., et al. (2009). Detection of QTLs for phosphorus use efficiency in relation to agronomic performance of wheat grown under phosphorus sufficient and limited conditions. *Plant Sci*. 176, 824–836. doi: 10.1016/j.plantsci.2009.03.006

Tavakol, E., Okagaki, R. J., Verderio, G., Shariati, J. V., Hussien, A., Bilgic, H., et al. (2015). The barley uniculme4 gene encodes a BLADE-ON-PETIOLE-like protein that controls tillering and leaf patterning. *Plant Physiol*. 168, 164–174. doi: 10.1104/pp.114.252882

Van Ooijen, J. W. (2004). *MapQTL 5, Software for the Mapping of Quantitative Trait Loci in Experimental Populations*. Netherlands: Kyazma B.V.

Van Ooijen, J. W. (2006). *JoinMap® 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Netherlands: Kyazma B.V.

Veneklaas, E. J., Lambers, H., Bragg, J. G., Finnegan, P. M., Lovelock, C. E., Plaxton, W. C., et al. (2012). Opportunities for improving phosphorus−use efficiency in crop plants. *New Phytol*. 195, 306–320. doi: 10.1111/j.1469-8137.2012.04190.x

Wang, X., and Yan, X. (2010). Genetic improvement for phosphorus efficiency in soybean: a radical approach. *Ann. Bot*. 106, 215–222. doi: 10.1093/aob/mcq029

Yan, X., Wu, P., Ling, H., Xu, G., Xu, F., and Zhang, Q. (2006). Plant nutriomics in china: an overview. *Ann. Bot*. 98, 473–482. doi: 10.1093/aob/mcl116

Yang, M., Ding, G., Shi, L., Xu, F., and Meng, J. (2011). Detection of QTL for phosphorus efficiency at vegetative stage in Brassica napus. *Plant Soil*. 339, 97–111. doi: 10.1007/s11104-010-0516-x

Yuan, Y., Gao, M., Zhang, M., Zheng, H., Zhou, X., Guo, Y., et al. (2017). QTL mapping for phosphorus efficiency and morphological traits at seedling and maturity stages in wheat. *Front. Plant Sci*. 8:614. doi: 10.3389/fpls.2017.00614

Zhang, D., Song, H., Cheng, H., Hao, D., Wang, H., Kan, G., et al. (2014). The acid phosphatase-encoding gene GmACP1 contributes to soybean tolerance to low-phosphorus stress. *PLoS Genet*. 10:e1004061. doi: 10.1371/journal.pgen.1004061

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 25 September 2020
doi: 10.3389/fgene.2020.580782

Genetic Diversity and Population Structure of Asian and European Common Wheat Accessions Based on Genotyping-By-Sequencing

Xiu Yang[1,2†], Binwen Tan[1,2†], Haijiao Liu[1,2], Wei Zhu[1,2], Lili Xu[1], Yi Wang[1,2], Xing Fan[1,2], Lina Sha[1,2], Haiqin Zhang[1,2], Jian Zeng[3], Dandan Wu[1,2], Yunfeng Jiang[1,2], Xigui Hu[4], Guoyue Chen[1,2], Yonghong Zhou[1,2] and Houyang Kang[1,2]*

[1] State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University, Chengdu, China, [2] Triticeae Research Institute, Sichuan Agricultural University, Chengdu, China, [3] College of Resources, Sichuan Agricultural University, Chengdu, China, [4] Center of Wheat Research, Henan Institute of Science and Technology, Xinxiang, China

OPEN ACCESS

Edited by:
Penghao Wang,
Murdoch University, Australia

Reviewed by:
Fa Cui,
University of Chinese Academy of Sciences, China
Quan Xie,
Nanjing Agricultural University, China

***Correspondence:**
Houyang Kang
houyang.kang@sicau.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 07 July 2020
Accepted: 03 September 2020
Published: 25 September 2020

Citation:
Yang X, Tan B, Liu H, Zhu W, Xu L, Wang Y, Fan X, Sha L, Zhang H, Zeng J, Wu D, Jiang Y, Hu X, Chen G, Zhou Y and Kang H (2020) Genetic Diversity and Population Structure of Asian and European Common Wheat Accessions Based on Genotyping-By-Sequencing. Front. Genet. 11:580782. doi: 10.3389/fgene.2020.580782

Obtaining information on the genetic diversity and population structure of germplasm facilitates its use in wheat breeding programs. Recently, with the development of next-generation sequencing technology, genotyping-by-sequencing (GBS) has been used as a high-throughput and cost-effective molecular tool for examination of the genetic diversity of wheat breeding lines. In this study, GBS was used to characterize a population of 180 accessions of common wheat originating from Asia and Europe between the latitudes 30° and 45°N. In total, 24,767 high-quality single-nucleotide polymorphism (SNP) markers were used for analysis of genetic diversity and population structure. The B genome contained the highest number of SNPs, followed by the A and D genomes. The polymorphism information content was in the range of 0.1 to 0.4, with a mean of 0.26. The distribution of SNPs markers on the 21 chromosomes ranged from 243 on chromosome 4D to 2,337 on chromosome 3B. Structure and cluster analyses divided the panel of accessions into two subgroups (G1 and G2). G1 principally consisted of European and partial Asian accessions, and G2 comprised mainly accessions from the Middle East and partial Asia. Molecular analysis of variance showed that the genetic variation was greater within groups (99%) than between groups (1%). Comparison of the two subgroups indicated that G1 and G2 contained a high level of genetic diversity. The genetic diversity of G2 was slightly higher as indicated by the observed heterozygosity (H_o) = 0.23, and unbiased diversity index (*uh*) = 0.34. The present results will not only help breeders to understand the genetic diversity of wheat germplasm on the Eurasian continent between the latitudes of 30° and 45°N, but also provide valuable information for wheat genetic improvement through introgression of novel genetic variation in this region.

Keywords: genetic diversity, population structure, genotyping-by-sequencing, single nucleotide polymorphisms, common wheat

INTRODUCTION

Wheat (*Triticum aestivum* L.) is an important staple food crop for more than one-third of the world's population and provides about 20% of calories consumed by humans (Marcussen et al., 2014; Bhatta et al., 2017). As a result of ongoing population growth and climate change, it has been estimated that wheat production must increase by 50% by 2050 (Grassini et al., 2013; Ray et al., 2013; Marcussen et al., 2014). Thus, it seems that wheat production cannot meet demand. Therefore, a challenge for wheat breeders is to improve the stability of grain production and grain yield to meet the growing demand, and to improve resistance and tolerance to biotic and abiotic stresses (Winfield et al., 2018). Analysis of plant genetic diversity is an important aspect of plant breeding, inheritance, conservation, and evolution (Peterson et al., 2014). However, domestication and strong selection pressure by humans, and the use of modern breeding techniques, have increasingly narrowed the gene pool of wheat (Tanksley and McCouch, 1997; Haudry et al., 2007). Therefore, it is essential to enrich wheat germplasm resources by introducing favorable mutations into the cultivated gene pool (Tanksley and McCouch, 1997; Haudry et al., 2007; Zhang et al., 2017).

Morphological traits and molecular markers are two distinct tools for assessment of genetic diversity. However, molecular markers have gained substantial attention because morphological traits are often influenced by the environment (Huang et al., 2002). A variety of molecular markers have been used to study the genetic diversity of wheat, such as randomly amplified polymorphic DNA (Joshi and Nguyen, 1993), random fragment length polymorphisms (Siedler et al., 1994; Kim and Ward, 2000), amplified fragment length polymorphisms (Barrett and Kidwell, 1998; Burkhamer et al., 1998), sequence-tagged sites (Chen et al., 1994) and inter-sequence simple repeats (Nagaoka and Ogihara, 1997). Single-nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs) are the most commonly used molecular markers for evaluation of genetic diversity among wheat accessions (Huang et al., 2002; Eltaher et al., 2018). Furthermore, SNPs are not only the most abundant type of polymorphism in animal genomes but also exhibit a large number of sequence variants in plant genomes (Ganal et al., 2009, 2012; Rimbert et al., 2018). At present, SNPs are the marker of choice for plant research and plant breeding, such as analyses of marker–trait association, population structure, genomic selection, quantitative trait loci mapping, and research on plant breeding that particularly requires numerous markers (Kumar et al., 2012). Use of high-throughput sequencing technology to discover a large number of SNPs has proved to be not time-consuming and cost-effective (He et al., 2014). With the rapid development of next-generation sequencing (NGS) technologies, an approach for genotyping-by-sequencing (GBS) have been widely used in plant breeding programs (Elshire et al., 2011). GBS enormously reduces the complexity of a large genomes of species by choosing appropriate restriction enzymes (REs) (Poland et al., 2012a), such as wheat with large and complex genomes. Poland et al. (2012b) developed a GBS protocol using two REs (*Pst*I/*Msp*I), which can reduce complexity to a greater extent and achieve a more unified sequencing library than a one-enzyme protocol (Elshire et al., 2011). GBS has been used for genotyping in an increasing number of crops, such as maize, wheat, barley, rice, potato, and cassava. Romay et al. (2013) genotyped a set of 2,815 maize inbred accessions using 681,257 SNPs and observed that some SNPs were related to known candidate genes, involving kernel color, sweetness, and flowering time. Lam et al. (2010) obtained 205,614 SNPs by resequencing 31 soybean genotypes, which offered a precious genomic resource for soybean breeding programs. The GBS protocol was utilized to analyze genetic diversity of 369 Iranian hexaploid wheat accessions, in which a total of 566,439,207 sequence reads were generated and 133,039 SNPs were identified (Alipour et al., 2017). A set of 38,412 GBS-SNPs were identified after sequencing 365 soft winter wheat varieties and F_5-derived advanced breeding lines originating from multiple crosses in the Cornell University Wheat Breeding Program using a GBS procedure to analyze genetic diversity (Heslot et al., 2013).

The principal region of common wheat cultivation is located between the latitudes of 30°–60°N and 27°–40°S, mainly concentrated in the 30°–45°N region (Nuttonson, 1955). The 180 common wheat accessions used in the present study were collected from 16 countries between the latitudes of 30° and 45°N. The germplasm in this region not only provides novel sources of resistance to biotic and abiotic stresses, but also can enhance the biodiversity of breeding materials. To allow comparison between geographic origin and genotype data, the accessions were grouped into three broad geographical regions, namely Asia, the Middle East, and Europe. The main purpose of this study was to use GBS to evaluate the genetic diversity of the accessions from 16 countries between the latitudes of 30° and 45°N and, in addition, to explore the genetic relationship and population structure of these accessions from different regions.

MATERIALS AND METHODS

Plant Materials

A total of 180 common wheat accessions from 16 countries situated between the latitudes of 30° and 45°N were used in this study (**Supplementary Table 1**). The seeds were kindly provided by the Triticeae Research Institute, Sichuan Agricultural University, Sichuan, China, the United States Department of Agriculture–Agricultural Research Service (USDA-ARS)–National Plant Germplasm System, United States, and the Xinjiang Academy of Agricultural Sciences, Xinjiang, China.

Genotyping-By-Sequencing

Total genomic DNA was extracted from fresh young leaves of approximately 2-week-old seedlings using the Hi-DNAsecure Plant Kit DP350 (TIANGEN, Beijing, China). GBS libraries were constructed following the protocol of Poland et al. (2012b). A total of 180 samples were used for genome sequencing on an Illumina HiSeq PE150 platform. SNP calling was performed using TASSEL v. 5.2.40 (Glaubitz et al., 2014). The GBS analysis pipeline used the default parameters (Glaubitz et al., 2014). Paired-end reads were mapped

to "Chinese Spring" reference genome with Burrows-Wheeler Aligner (Version: 0.7.8) (Li and Durbin, 2009). The wheat "Chinese Spring" reference genome assembly made available by the International Wheat Genome Sequencing Consortium (IWGSC; RefSeq V1.0) in 2017 was used. The SNPs were filtered based on the criteria minor allele frequency (MAF) threshold <5% and missing values >10 (Li et al., 2015; Saint et al., 2016; Vikram et al., 2016). The detailed information of SNP scores in each of the 180 accessions are available in **Supplementary Table 2**.

Population Structure Analysis

Evolutionary relationships among the 180 wheat accessions were determined using the unweighted pair group method with arithmetic mean (UPGMA) based on genetic distances computed with TASSEL. Dendrograms were constructed using the dendrogram function, and then customized using the dendextend package (Galili, 2015) and circlize package (Gu et al., 2014) in R. Principal component analysis (PCA) was performed based on genetic distances among the lines computed with TASSEL (Bradbury et al., 2007). Principal components (PCs) were generated using the covariance method. Eigenvalues were generated to determine the proportion of variation explained by each PC. The first and second PCs were plotted using R.

The population structure of all accessions was evaluated using the Bayesian model-based clustering method in STRUCTURE 2.3.4 software (Pritchard et al., 2000). The STRUCTURE analysis was run five times, with K ranging from 1 to 10, using the admixture model, with burn-in of 100,000 generations and a Markov chain Monte Carlo of 100,000 generations (Chen et al., 2012; Zori´c et al., 2012). To identify the best fit for the number of clusters (K), the Evanno method was utilized (Evanno et al., 2005) using STRUCTURE HARVESTER software (Earl and vonHoldt, 2012). After selection of the optimal K, membership (the proportion of the population assigned to each cluster), mean population differentiation (F_{ST}), and H_e (Nei, 1978) were determined for each subpopulation identified. The F_{ST} value (Nei, 1977) of each subpopulation provides an estimate of the degree of fixation of alleles within the subpopulation. The H_e (analogs to allelic variation in a random mating population) (Nei, 1978) describes the average distance between individuals within the same population, where values close to 0 indicate that the individuals within the population are genetically identical.

Statistical Analysis

Basic summary statistics were calculated using PowerMarker 3.25 software, comprising gene diversity (GD), polymorphism information content (PIC), MAF, and observed heterozygosity (H_o) (Liu and Muse, 2005). The SNP distribution on each chromosome was counted with 5 Mb as a step, and all SNPs were mapped to IWGSC RefSeq v1.0. The heat map of SNP distribution was plotted using R. On each chromosome, the SNP markers with a PIC value between 0.21 and 0.33 were selected and a total of 7,461 SNPs were used for AMOVA. The number of subpopulations determined on the basis of a STRUCTURE analysis was used for AMOVA. Genetic indices, consisting of number of alleles (N_a), number of effective alleles (N_e), observed heterozygosity (H_o), diversity index (h), unbiased diversity index (uh), and Shannon's information index (I) were calculated. The AMOVA and estimation of genetic indices were performed using GeneAlEx 6.41 (Peakall and Smouse, 2006).

RESULTS

Chromosome Distribution of SNPs

A total of 24,767 SNPs were identified in the A, B, and D genomes. The highest number of SNPs were identified in the B genome (12,028), followed by the A genome (9,741), and the D genome had the lowest number of polymorphic markers with 2,998 (**Figure 1A**, **Supplementary Table 3**). In the A genome, chromosome 2A had the highest number of SNPs (1,761), and chromosome 6A harbored the lowest number (1,154); in the B genome, the highest and lowest number of SNPs were detected on chromosomes 3B and 4B (2,337 and 1,130, respectively); in the D genome, chromosome 2D had the highest number of SNPs (597), and chromosome 4D harbored the lowest number (243) (**Figure 1B**, **Supplementary Table 3**). The lowest and highest numbers of SNPs identified on an individual chromosome were 243 and 2,337 on chromosomes 4D and 3B, respectively (**Figure 1B**, **Supplementary Table 3**). The ratio of number of SNPs in the B to A genomes was 1.23, and that in the B to D and A to D genomes was 4.01 and 3.25, respectively. Thus, the number of SNPs in the A and B genomes exceedingly the number in the D genome, and the number of SNPs in the A genome was only slightly lower than that in the B genome. To characterize the distribution of SNPs in more detail, we used 5 Mb as a step

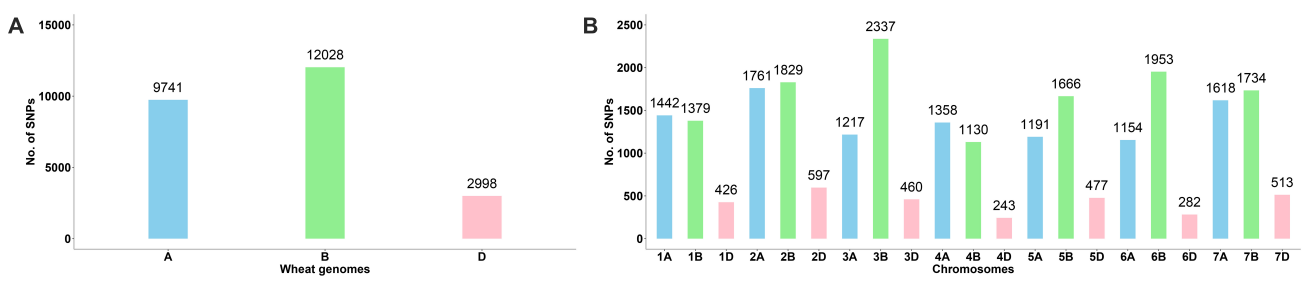

FIGURE 1 | (A, B) Chromosomal distribution of SNP markers on all chromosomes in the wheat genomes A, B, and D.

to map all SNPs to the IWGSC RefSeq v1.0, and drew the heat map of SNP distribution on each chromosome (**Figure 2**). For example, on the 2D chromosome, the physical segment with the highest number of SNPs was 520–525 Mb (**Figure 2B**). However, the physical segment of 630–635 Mb on the 5B chromosome had the highest number of SNPs (**Figure 2E**).

Population Structure

The 180 common wheat accessions were divided into three broad geographical regions (**Figure 3**, **Table 1**). Using PCA, the relationship between the wheat accessions based on the broad geographical regions was determined. In the PCA plot the accessions showed a loose distribution (**Figure 4**). The accessions from Asia and the Middle East were distributed evenly on PC1 but were less evenly distributed on PC2. The majority of Asian accessions were placed in the positive (upper) portion of the plot. European accessions were mainly clustered towards the right side (positive values) of PC1. And European accessions were divided into two parts by PC2. STRUCTURE analysis was used to study the population structure of the 180 accessions, and delta *K* values obtained were used to determine the optimal number of subpopulations. To determine the optimal *k* value, the number of clusters (*K*) was plotted. At $k = 2$ (**Figure 5**), a distinct peak was observed, indicating the presence of two subpopulations (**Figure 6**). Group 1 contained 137 accessions; Group 2 consisted of 43 accessions (**Table 2**). The degree of genetic differentiation and average distance (H_e) in each subpopulation (**Table 2**) suggested that the highest degree of genetic differentiation was detected in Group 2 ($F_{st} = 0.40$), whereas the lowest value was observed in Group 1 ($F_{st} = 0.13$). On the other hand, the lowest H_e was observed in Group 2 and the highest H_e was detected in Group 1. The results of STRUCTURE analyses (**Figure 6**), PCA, and the UPGMA cluster analysis (**Figure 7**, **Supplementary Figure 1**) showed a high degree of similarity. It was observed that individuals in Group 2 in the STRUCTURE analysis were separated from individuals in Group 1 on PC1. Individual accessions in Group 2 mainly originated from Asia and the Middle East; Group 1

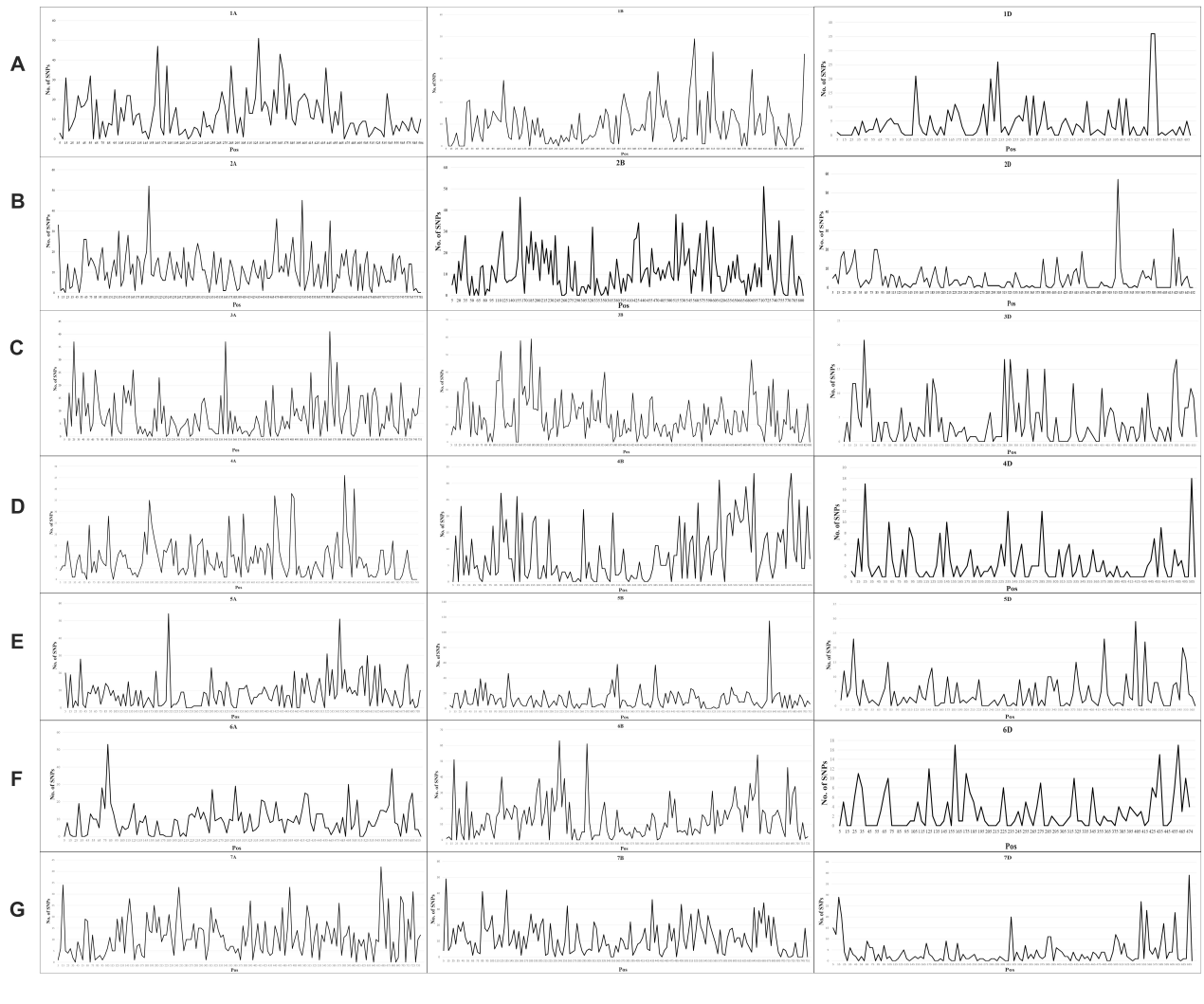

FIGURE 2 | (A–G) Heat map of SNPs distribution on each chromosome.

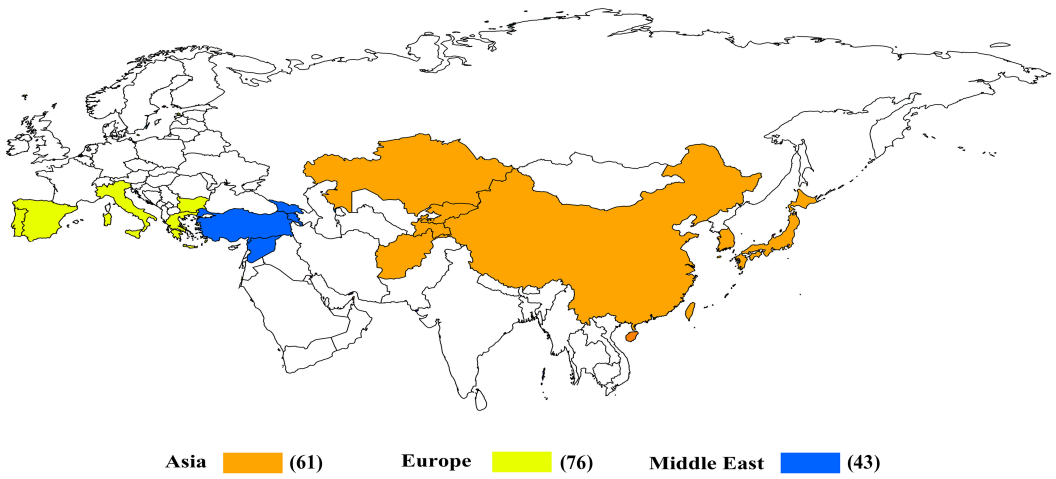

FIGURE 3 | Distribution of common wheat accessions used in this study across Asia and Europe.

principally consisted of Asian and European accessions. To further understand the clustering relationships among countries, we took the country of origin into consideration. Accessions from Spain were divided into two clusters: one portion was clustered with European accessions, and the other portion clustered with Asian/Middle Eastern accessions. Accessions from Kyrgyzstan, Kazakhstan, China, and Japan tended to cluster with accessions from Europe. Half of the accessions from Afghanistan were clustered with European accessions, and half was grouped with Asian/Middle Eastern accessions. The Middle Eastern accessions originating in Armenia mainly clustered separate from the European accessions.

The percentage apportioning to ancestral groups for each country was determined to examine the geographic distribution of the two STRUCTURE groups, which were projected onto a world map (**Figure 8**). The accessions from Europe were predominantly assigned to Group 1 (blue segments in **Figure 8**), and the accessions from Bulgaria (100%) and Portugal (100%) were assigned to Group 1 (**Figure 9**). For the four countries in the Middle East, except for accessions from Armenia (67% accessions in Group 2), the majority of accessions were assigned to Group 1 (**Figure 9**). Half of the accessions from Afghanistan were assigned to Group 1 and half were assigned to Group 2. In addition, 53% lines from Korea were placed in Group 2 and 47% were placed in Group 1 (**Figure 9**).

TABLE 1 | Provenance of the 180 common wheat accessions used in the study.

Region	Country	No. Acc.
Asia	Afghanistan	12
	Kyrgzstan	4
	Kazakhstan	4
	Tajikistan	4
	China	12
	Korea	15
	Japan	10
Middle East	Turkey	10
	Syria	10
	Georgia	14
	Armenia	9
Europe	Bulgaria	20
	Greece	14
	Italy	17
	Spain	13
	Portugal	12

Genetic Diversity

The genetic diversity analysis of the 180 accessions revealed that the mean GD and PIC were 0.32 and 0.26, respectively. The GD ranged from 0.1 to 0.5 and PIC ranged from 0.1 to 0.4 (**Figures 10A,B**). The H_o values ranged from 0 to 0.9, but for a considerable number of markers the H_o value was 0.1 (**Figure 10C**). The average MAF was 0.24 (**Figure 10D**). Intra-population genetic diversity analysis revealed that mean observed (N_a) and effective (N_e) allele numbers were 2.00 and 1.52 for the two subpopulations, respectively. The value of N_e in Group 2 (1.52) was higher than that in Group 1 (1.51). The mean values of I, H_o, h, and uh were 0.51, 0.23, 0.33, and 0.34, respectively. However, the Group 2 population showed slightly higher genetic diversity (H_o = 0.23 and uh = 0.34) (**Table 3**).

AMOVA and genetic diversity indices for the two subpopulations were calculated based on the results of the STRUCTURE analysis. The AMOVA revealed much greater variation within populations (99%) than among the populations (1%). High haploid N_m was observed (28.12), suggesting extremely high gene flow among subpopulations (**Table 4**). These results revealed low genetic differentiation among the subpopulations, but high genetic differentiation within subpopulations. The UPGMA cluster analysis based on 7,461 markers (**Supplementary Figure 2**) indicated two subgroups, which was consistent with the population structure analysis (based on 24,767 SNPs).

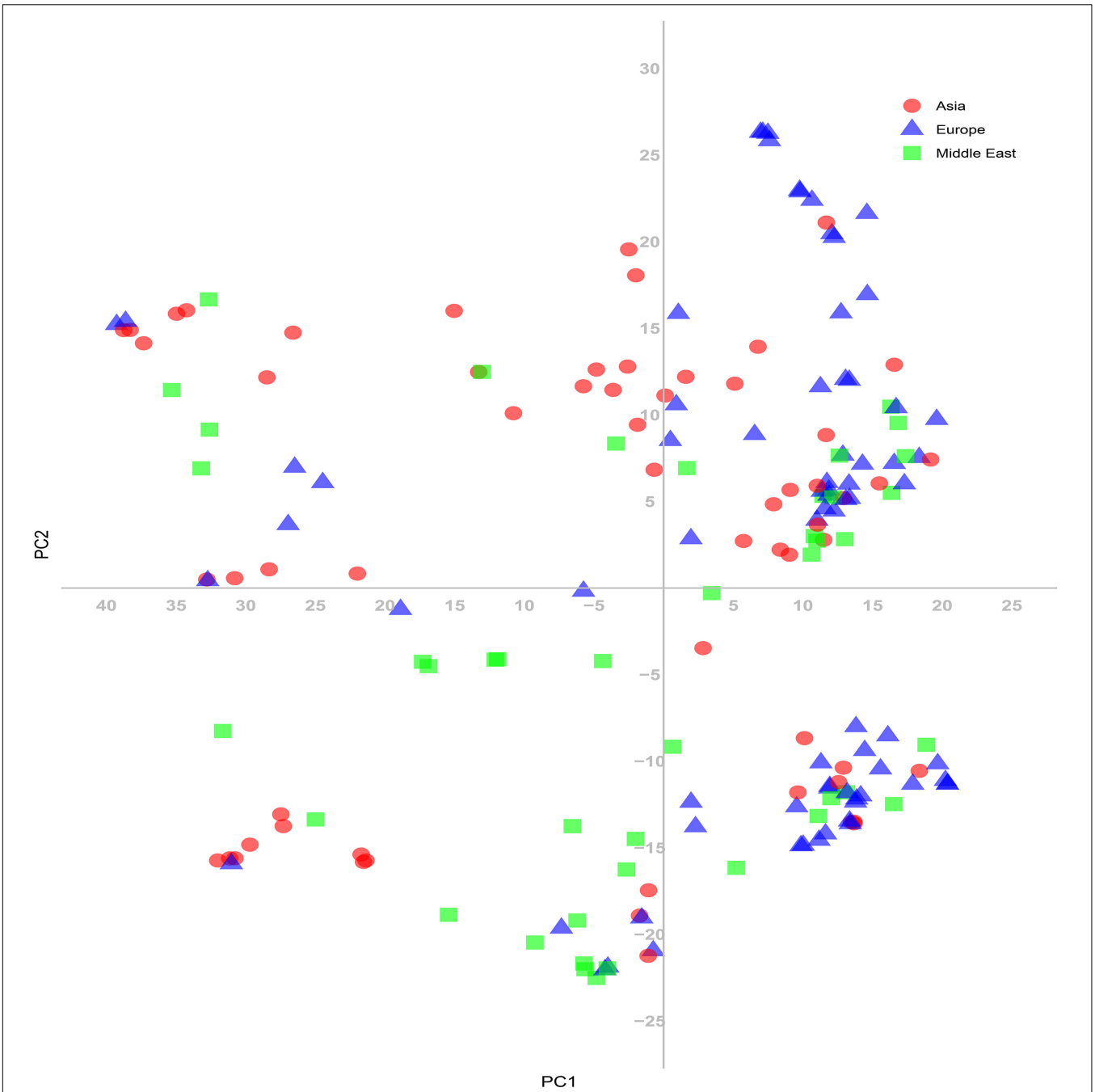

FIGURE 4 | Principal coordinate analysis (PCA) based on genetic distances among the common wheat accessions for the SNP markers.

DISCUSSION

Wheat germplasm resources are extremely important for breeders. One main wheat-producing area of the world is located between 30° and 45°N latitude, which is rich in wheat germplasm resources. A prerequisite for making full use of these germplasm resources is to assess their genetic diversity (Hawkes, 1981). We used GBS technology to discover a large number of SNPs for genotyping hexaploid wheat derived from diverse provenances in this study. In the present study, we obtained 24,767 SNPs markers and observed the lowest frequency of SNPs in the D genome, whereas the B genome contained the highest frequency of polymorphic markers, which is in agreement with the results of previous studies (Chao et al., 2009; Akhunov et al., 2010; Poland et al., 2012a; Berkman et al., 2013; Würschum et al., 2013; Marcussen et al., 2014; Shavrukov et al., 2014; Edae et al., 2015; Alipour et al., 2017; Eltaher et al., 2018; Rufo et al., 2019). D genome is the youngest one among the three

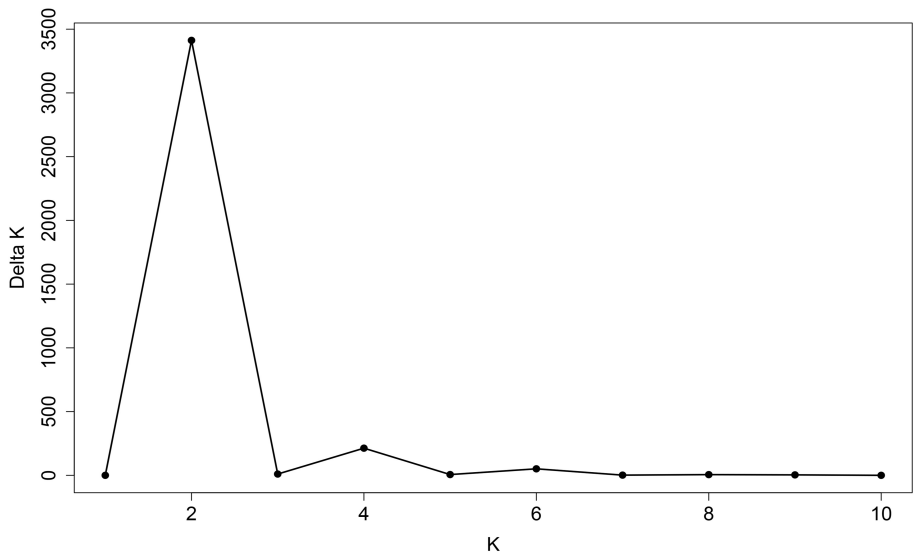

FIGURE 5 | Line graph of delta *K* over *K* from 1 to 10. The highest peak was observed at delta *K* = 2, which suggested the common wheat germplasm comprised two subgroups.

genomes in wheat evolutionary history (Talbert et al., 1998; Caldwell et al., 2004). It is likely that older genomes underwent gene duplication and accumulated more mutations that led to sequence polymorphism (Dvorak et al., 2006). Substantial early gene flow could have occurred between *T. aestivum* and its tetraploid progenitor *T. turgidum* (AABB) but not between the hexaploid and *Aegilops tauschii* (DD) (Caldwell et al., 2004; Dvorak et al., 2006). This could have resulted in greater sequence diversity in the A and B genomes than in D genome (Talbert et al., 1998; Caldwell et al., 2004; Dvorak et al., 2006; Berkman et al., 2013). Furthermore, the fewest SNP markers were located on chromosome 4D, whereas the highest number of SNP markers were located on chromosome 3B, as reported by Saintenac et al. (2013) and Alipour et al. (2017). Eltaher et al. (2018) obtained 25,566 SNPs by GBS for 270 $F_{3:6}$ Nebraska winter wheat accessions, and observed that the highest number of SNPs were located on chromosome 3B, whereas chromosome 3D carried the lowest number of SNPs. Bhatta et al. (2017) reported that chromosomes 2B and 4D had the highest and lowest numbers of SNPs, respectively. Chromosome 4D had the lowest number of markers and chromosome 1B had the highest number of markers in the study by Sukumaran et al. (2015). Allen et al. (2017) used 35,143 SNPs and reported that chromosome 2B had the highest number of markers and chromosome 4D had the lowest number of markers. In contrast, the present study showed that chromosome 3B harbored the highest number of SNPs and chromosome 4D had the lowest number. Meanwhile, we found some SNP hot spot regions in heat map of each chromosome harboring important QTLs. In the 160–170 Mb of chromosome 2B (**Figure 2B**), Boukhatem et al. (2002) analyzed a set of 98 F_8 recombinant inbred (RI) lines and found a QTL (*QYR-2B.2*) which was associated with yellow rust resistance. Similarly, in the 680–690 Mb of chromosome 1B (**Figure 2A**), a QTL (*QTgw.ipk-1B-FS4*) which was associated with TKW was identified (Nezhad et al., 2012).

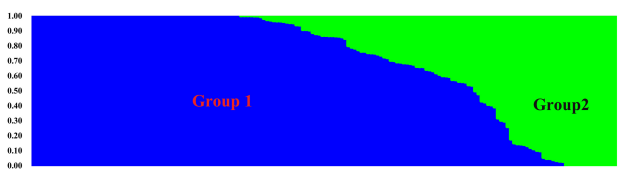

FIGURE 6 | Population structure of 180 common wheat accessions based on 24,767 SNPs markers. The population is divided into two color-coded subgroups. Each bar represents a single accession, and the colored segments within each bar reflect the proportional contributions of each subgroup to that accession.

The PIC contributes to a detailed understanding of the level of polymorphism between genotypes. On the basis of previous reports, the PIC can be divided into three categories: (1) when PIC > 0.5, the marker is considered to be highly polymorphic, (2) when 0.25 < PIC < 0.5, the marker is a moderately informative, and (3) when PIC < 0.25, the marker is a low-information marker (Botstein et al., 1980). Lopes et al. (2015) observed a PIC value of 0.27 using the 9K SNP array to genotype the WAMI population, and showed that spring wheat contained moderate levels of polymorphism. Novoselović et al. (2016)

TABLE 2 | Results of STRUCTURE analysis of 180 wheat accessions for the fixation index (F_{st}; indicating significant divergence), average distances (expected heterozygosity), and number of genotypes in each subpopulation.

Population	F_{st}	Exp. hetero	No. of Genotypes
G1	0.13	0.31	137
G2	0.40	0.24	43

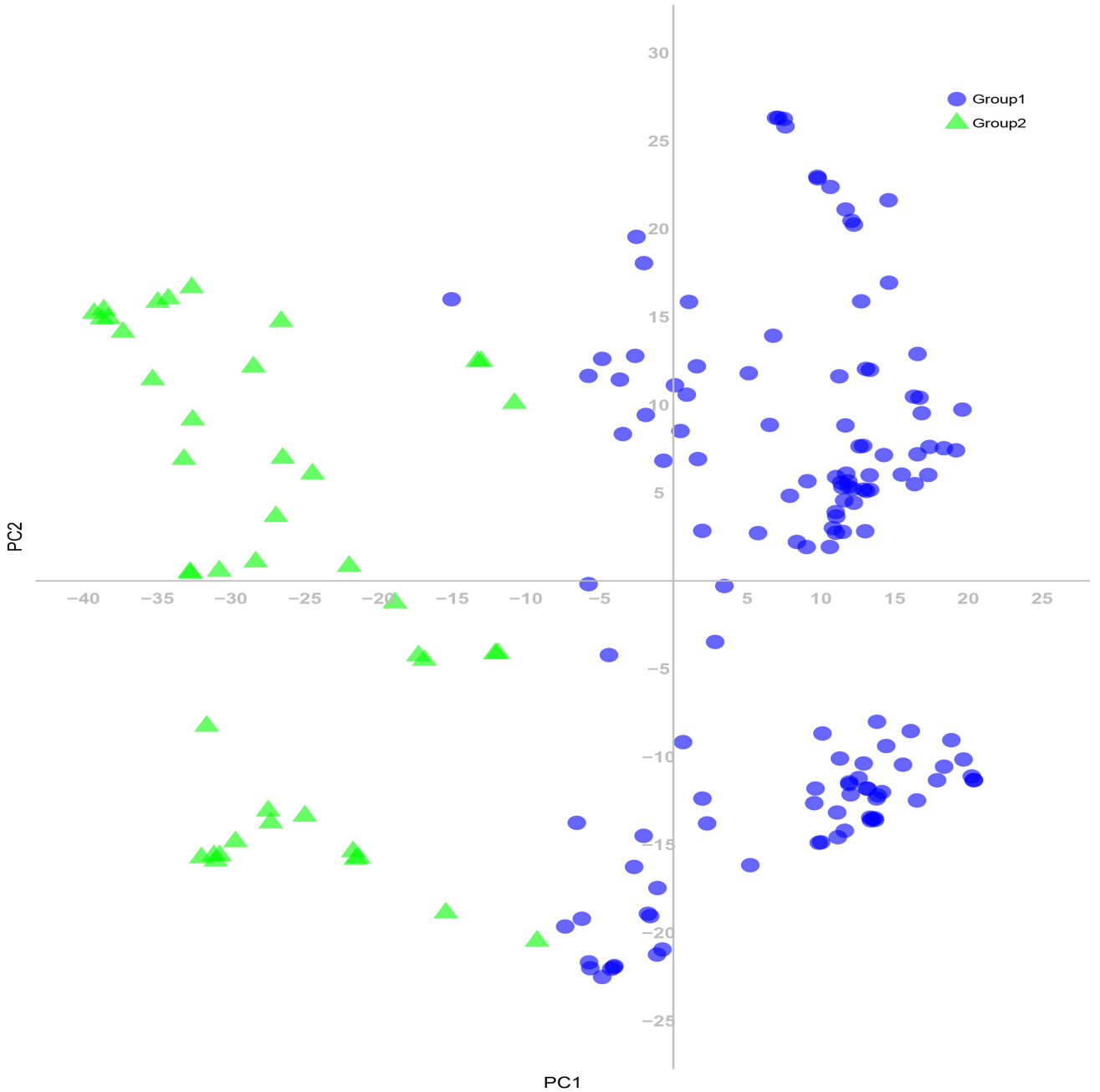

FIGURE 7 | Principal coordinate analysis (PCA) of 180 common wheat accessions based on 24,767 SNP markers. Colors reflect groupings derived from structure analysis.

genotyped a Croatian panel using a set of 1,229 Diversity Arrays Technology (DArT) markers and obtained an average PIC value of 0.30 among the populations, which indicated that the accessions from Croatia exhibited moderate polymorphism. Eltaher et al. (2018) analyzed 270 $F_{3:6}$ Nebraska winter wheat accessions, and observed a PIC value of 0.25, which indicated that the population contained moderate genetic diversity. El-Esawi et al. (2018) used 1,052 DArT markers to genotype Australian and Belgian wheat accessions, and obtained PIC values of 0.33 and 0.29, respectively, which demonstrated that Australian and Belgian wheat contained moderate genetic diversity. The present results showed that the mean PIC value (0.27) was in agreement with the above-mentioned studies, which indicated that the 180 accessions contained moderate polymorphism. On the other hand, Hao et al. (2011) genotyped 250 Chinese wheat accessions with 512 SSR markers and observed a PIC value of 0.65, which demonstrated that Chinese wheat showed high genetic polymorphism. Zhang et al. (2010) analyzed 205 elite

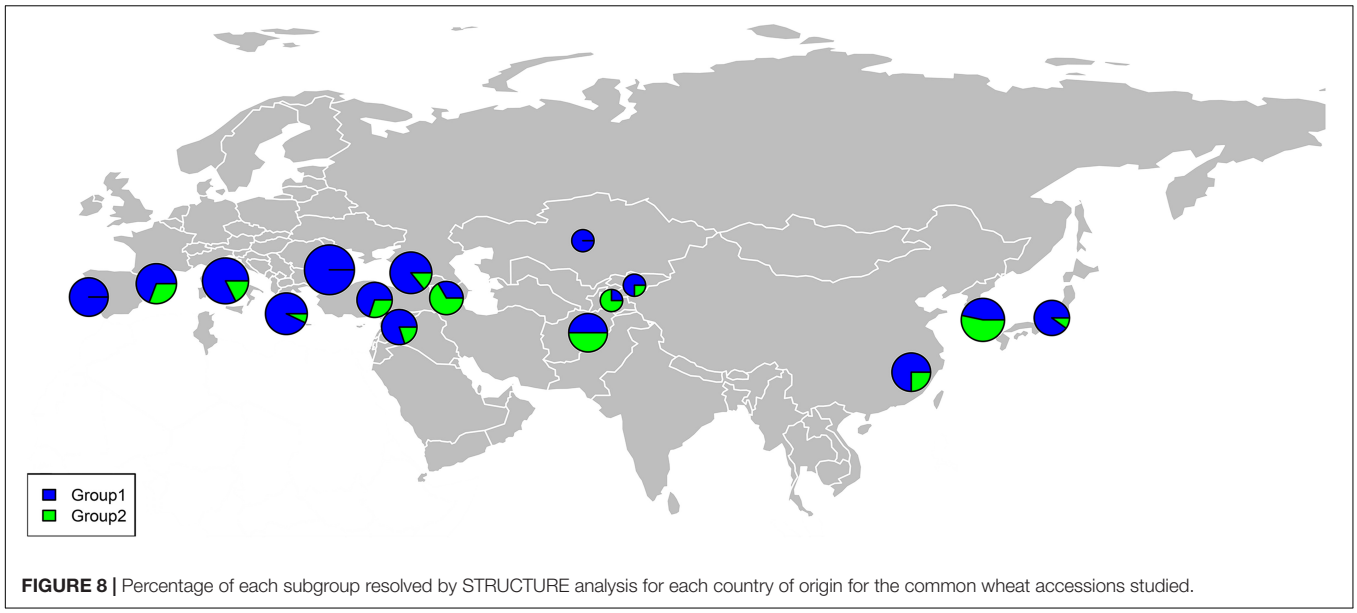

FIGURE 8 | Percentage of each subgroup resolved by STRUCTURE analysis for each country of origin for the common wheat accessions studied.

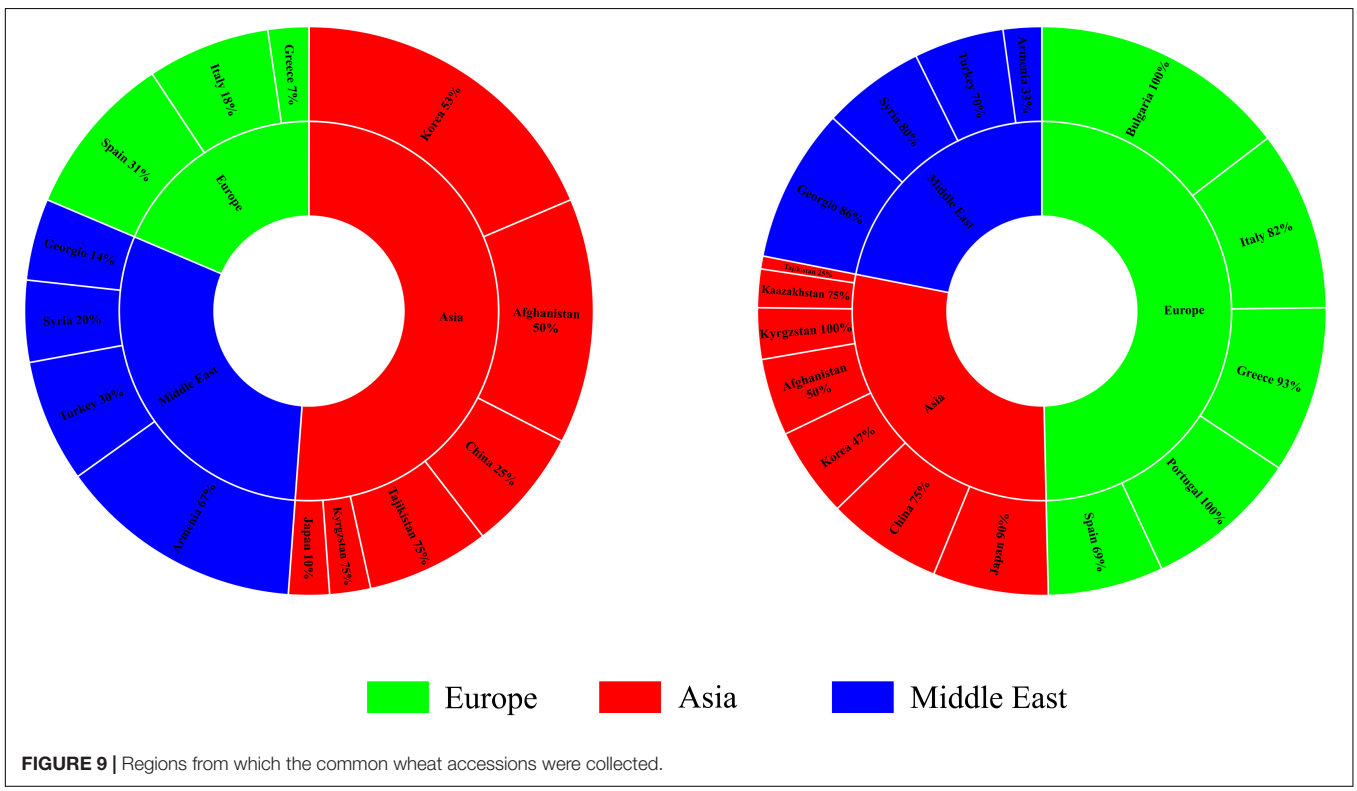

FIGURE 9 | Regions from which the common wheat accessions were collected.

wheat accessions in the United States, using 245 SSR markers, and obtained a PIC value of 0.54, which indicated that the accessions showed a high level of polymorphism. Relative to SSR markers, the lower PIC value of the SNP and DArT markers may be explained by their bi-allelic nature and slow mutation rate (Thuillet et al., 2002; Chesnokov and Artemyeva, 2015).

In the present study, we obtained meaningful information on genetic diversity indices in each subpopulation. High levels of genetic diversity were detected in Groups 1 and 2. The results of AMOVA showed that a high level of genetic diversity was observed within subpopulations, whereas the variation among subpopulations was extremely low (1%). This result may be caused by breeders selecting for specific traits, such as yield, stripe rust resistance, and herbicide tolerance. However, the low genetic variability among subpopulations is explained by the high gene flow (Arora et al., 2014). Wright (1965) showed that when N_m (haploid) values are less than 1, gene exchange among subpopulations is limited. In the present study we observed an

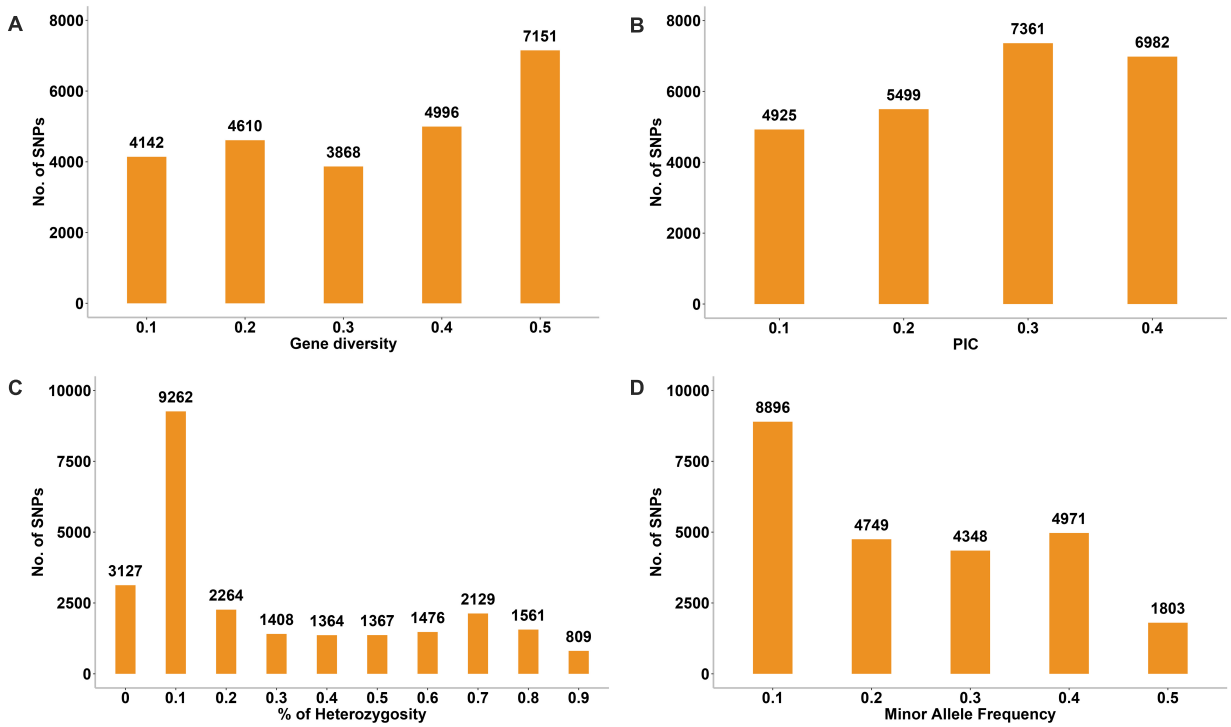

FIGURE 10 | Distribution of genetic diversity **(A)**, polymorphic information content (PIC) **(B)**, percentage of heterozygosity **(C)**, and minor allele frequency **(D)** for 24,767 SNP markers among the 180 common wheat accessions.

extremely high N_m value (28.12), indicating that high gene flow led to low genetic differentiation among subpopulations. The results of this study will not only help breeders to understand the genetic diversity of wheat germplasm on the Eurasian continent between the latitudes of 30° and 45°N, but also provides valuable information for genetic improvement of wheat through inclusion of novel genetic variation from China and certain other countries.

TABLE 3 | Means of genetic parameters for each subpopulation of the 180 wheat accessions. Number of alleles (N_a), number of effective allele (N_e), Shannon's index (*I*), observed heterozygosity (H_o), diversity index (*h*), and unbiased diversity index (*uh*).

Pop	N_a	N_e	*I*	H_o	*h*	*uh*
G1	2.00	1.51	0.51	0.22	0.33	0.33
G2	2.00	1.52	0.51	0.23	0.33	0.34
Mean	2.00	1.52	0.51	0.23	0.33	0.34

TABLE 4 | Analysis of molecular variance using 7461 SNP markers of genetic differentiation among and within two subpopulations of the 180 common wheat accessions.

Source	df	SS	MS	Est. Var.	%	*P* value
Among Pops	1	5,599.57	5,599.57	31.47	1%	0.001
Within Pops	178	630,097.27	3,539.87	3,539.87	99%	0.001
Total	179	635,696.84		3,571.34	100%	0.001
Nm			28.12			

The PCA revealed a degree of broad geographic partitioning of the accessions. A previous study by Winfield et al. (2018) used 32,443 polymorphic markers to genotype 804 hexaploid wheat accessions originating from more than 30 countries around the world, and observed that the majority of accessions from Europe clustered together, separated from the majority of Asian and Middle Eastern accessions. Similarly, in the study of Cavanagh et al. (2013), the European winter wheat population showed the strongest degree of genetic differentiation from the remaining populations. Balfourier et al. (2007) used a set of 38 SSR markers to analyze 3,942 accessions originating from 73 countries, and observed that accessions from several Near Eastern and Central Asian areas were grouped in the same subcluster and those from Far Eastern countries clustered together. Strelchenko et al. (2005) analyzed 78 wheat landraces originating from 22 countries and reported that the landraces were separated into European and Asian groups. Chen et al. (2019) reported that West Asian landraces, the majority of European landraces, several South and Central Asian landraces, and the majority of East Asian cultivars clustered together, whereas the majority of East Asian landraces were clustered with several West Asian landraces and the majority of South and Central Asian landraces. Lee et al. (2018) reported that many accessions from Afghanistan, Japan, and Korea were clustered in the same group, while germplasm from China, the Middle East, and Caucasus clustered in a separate group, and an intermediate group largely consisted mainly of accessions from Afghanistan, Japan, and Korea. In the present study,

although there was substantial overlap between clusters, the majority of accessions from Europe clustered together, whereas the accessions from Asia and the Middle East were distributed evenly on PC1 (**Figure 4**). However, the relationships of three overlapping subgroups was unclear, which raises the possibility of exchanging adapted germplasm. To obtain useful information on the genetic diversity and population structure of the accessions, they were divided into two subgroups on the basis of the population structure analysis (**Figure 6**). In the PCA (**Figure 7**), genotypes clustered consistent with the subpopulations identified in the STRUCTURE analysis. Moreover, the UPGMA cluster analysis (**Supplementary Figure 1**) was consistent with the results of the STRUCTURE analysis. The majority of European accessions were divided into Group 1, especially accessions from Bulgaria and Portugal (**Figure 9**), whereas portions of the Asian and Middle Eastern accessions were distributed in Groups 1 and 2, respectively. The accessions from Turkey, Syria, Georgia, Armenia, Afghanistan, Kyrgyzstan, and Tajikistan showed complex genetic backgrounds, which is not surprising. The area between the Black Sea and the Caspian Sea, and just south of this region (Iraq), is the assumed location of the center of origin of wheat domestication and seems to be a site of population consolidation. Chen et al. (2019) showed that Chinese wheat accessions were mainly derived from European landraces. In the present study, the accessions originating from China tended to cluster with European accessions (**Figure 9**). The improvement of Chinese wheat was based on hybridization programs, which included well-adapted landraces and introduced accessions. Furthermore, the introduction of foreign materials would promote the genetic improvement of Chinese wheat. Italian varieties including Villa Glori, Mentana, Funo, Abbondanza, St 2422/464, and Libellula were widely cultivated and utilized in the Yellow and Huai River valley winter wheat region, lower and middle Yangtze River valley winter wheat region, southwestern winter wheat region, and northwestern spring wheat region (Zhao et al., 2019). Varieties such as Ukraine 0246, New Ukraine, Red Star, Kavkaz, and Aurora from the former USSR were introduced and disseminated widely in Xinjiang (He et al., 2001). Therefore, the exchange and utilization of germplasm worldwide was an established way to expand the genetic basis of wheat breeding (Zhao et al., 2019).

CONCLUSION

In this study, a GBS protocol was used to investigate the population structure and genetic diversity of wheat accessions originating from the Eurasian continent between the latitudes of 30° and 45°N. The panel of 180 accessions was divided into two subgroups, which could be identified by their parentage and selection history. Group 1 principally consisted of European and a portion of Asian accessions, and Group 2 predominantly comprised Middle East and a portion of Asian accessions. Groups 1 and 2 showed high values for genetic diversity indices, which were higher for Group 2. The present results demonstrated that the 180 accessions represent high genetic diversity and can be used for future breeding programs to develop new wheat cultivars with desirable characteristics such as high yield potential, tolerance to biotic and abiotic stress, and good end-use quality, while being well-adapted to diverse environments in China and other countries.

DATA AVAILABILITY STATEMENT

We have uploaded our SNP sequencing data to FigShare repository (https://figshare.com/s/157a9613b935814cc2d5).

AUTHOR CONTRIBUTIONS

HK and YZ conceived and designed the research. XY, BT, and HL conducted the experiments. WZ, LX, YW, XF, LS, JZ, DW, YJ, and XH participated in the preparation of the reagents and materials. HZ and GC analyzed the data. XY and HK wrote the manuscript. All authors read and approved the manuscript.

FUNDING

This work was funded by the National Natural Science Foundation of China (Nos. 31771781 and 31971883), the National Key Research and Development Program of China (2016YFD0102000 and 2017YFD0100905), and the Science and Technology Bureau of Sichuan Province (2020YJ0348).

ACKNOWLEDGMENTS

We thank Prof. Jianjiang Li, Xinjiang Academy of Agricultural Sciences, Xinjiang, China, for kindly supplying the materials used in this study. We are also grateful to the reviewers for their valuable comments and recommendations. This manuscript has been released as a pre-print at https://www.researchsquare.com/article/rs-8272/v1, Plant Molecular Biology and Genetics, Population Genetics, doi: 10.21203/rs.2.17640/v1.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.580782/full#supplementary-material

SUPPLEMENTARY FIGURE 1 | UPGMA clustering dendrogram generated using 24,767 SNPs and 180 common wheat accessions. Colors reflect groups derived from STRUCTURE analysis.

SUPPLEMENTARY FIGURE 2 | Cluster analysis based on the genetic distances among the 180 accessions for 7,461 SNPs used for AMOVA.

SUPPLEMENTARY TABLE 1 | Details of the 180 common wheat accessions used in the study.

SUPPLEMENTARY TABLE 2 | The detailed information of SNP scores in each of the 180 accessions.

SUPPLEMENTARY TABLE 3 | Chromosomal distribution of the SNP markers used in the study.

REFERENCES

Akhunov, E. D., Akhunova, A. R., Anderson, O. D., Anderson, J. A., Blake, N., Clegg, M. T., et al. (2010). Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genom.* 11:702. doi: 10.1186/1471-2164-11-702

Alipour, H., Bihamta, M. R., Mohammadi, V., Peyghambari, S. A., Bai, G., Zhang, G., et al. (2017). Genotyping-by-sequencing (GBS) revealed MolGecular genetic diversity of iranian wheat landraces and cultivars. *Front. Plant Sci.* 8:1293. doi: 10.3389/fpls.2017.01293

Allen, A. M., Winfield, M. O., Burridge, A. J., Downie, R. C., Benbow, H. R., Barker, G. L. 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. doi: 10.1111/pbi.12635

Arora, A., Kundu, S., Dilbaghi, N., Sharma, I., and Tiwari, R. (2014). Population structure and genetic diversity among Indian wheat varieties using microsatellite (SSR) markers. *Aust. J. Crop. Sci.* 8, 1281–1289.

Balfourier, F., Roussel, V., Strelchenko, P., Exbrayat-Vinson, F., Sourdille, P., Boutet, G., et al. (2007). A worldwide bread wheat core collection arrayed in a 384-well plate. *Theor. Appl. Genet.* 114, 1265–1275. doi: 10.1007/s00122-007-0517-1

Barrett, B. A., and Kidwell, K. K. (1998). AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. *Crop Sci.* 38, 1261–1271. doi: 10.2135/cropsci1998.0011183X003800050025x

Berkman, P. J., Visendi, P., Lee, H. C., Stiller, J., Manoli, S., Lorenc, M. T., et al. (2013). Dispersion and domestication shaped the genome of bread wheat. *Plant Biotechnol. J.* 11, 564–571. doi: 10.1111/pbi.12044

Bhatta, M., Regassa, T., Rose, D. J., Baenziger, P. S., Eskridge, K. M., Santra, D. K., et al. (2017). Genotype, environment, seeding rate, and top-dressed nitrogen effects on end-use quality of modern Nebraska winter wheat. *J. Sci. Food Agric.* 97, 5311–5318. doi: 10.1002/jsfa.8417

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314–331. doi: 10.1016/0165-1161(81)90274-0

Boukhatem, N., Baret, P. V., Mingeot, D., and Jacquemin, J. M. (2002). Quantitative trait loci for resistance against yellow rust in two wheat derive recombinant inbred line populations. *Theor. Appl. Genet.* 104, 111–118. doi: 10.1007/s001220200013

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. doi: 10.1093/bioinformatics/btm308

Burkhamer, R. L., Lanning, S. P., Martens, R. J., Martin, J. M., and Talbert, L. E. (1998). Predicting progeny variance from parental divergence in hard red spring wheat. *Crop Sci.* 38, 243–248. doi: 10.2135/cropsci1998.0011183X003800010041x

Caldwell, K. S., Dvorak, J., Lagudah, E. S., Akhunov, E., Luo, M. C., Wolters, P., et al. (2004). Sequence polymorphism in polyploid wheat and their d-genome diploid ancestor. *Genetics* 167, 941–947. doi: 10.1534/genetics.103.016303

Cavanagh, C. R., Chao, S., Wang, S., Huang, B. E., Stephen, S., Kiani, S., et al. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl. Acad. Sci. U.S.A.* 10, 8057–8062. doi: 10.1073/pnas.1217133110

Chao, S., Zhang, W., Akhunov, E., Sherman, J., Ma, Y., Luo, M. C., et al. (2009). Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L.) cultivars. *Mol. Breed.* 23, 23–33. doi: 10.1007/s11032-008-9210-6

Chen, H., Jiao, C., Wang, Y., Wang, Y., Tian, C., Yu, H., et al. (2019). Comparative population genomics of bread wheat (*Triticum aestivum*) reveals its cultivation and breeding history in China. *bioRxiv[Preprint]*. doi: 10.1101/519587

Chen, H. B., Martin, J. M., Lavin, M., and Talbert, L. E. (1994). Genetic diversity in hard red spring wheat based on sequence-tagged-site PCR markers. *Crop Sci.* 34, 1628–1632. doi: 10.2135/cropsci1994.0011183X003400060037x

Chen, X., Min, D., Yasir, T. A., and Hu, Y. G. (2012). Genetic diversity, population structure and linkage disequilibrium in elite Chinese winter wheat investigated with SSR markers. *PLoS One* 7:e44510. doi: 10.1371/journal.pone.0044510

Chesnokov, Y. V., and Artemyeva, A. M. (2015). Evaluation of the measure of polymorphism information of genetic diversity. *Agricult. Biol.* 5, 571–578. doi: 10.15389/agrobiology.2015.5.571eng

Dvorak, J., Akhunov, E. D., Akhunov, A. R., Deal, K. R., and Luo, M. C. (2006). Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. *Mol. Biol. Evol.* 23, 1386–1396. doi: 10.1093/molbev/msl004

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. doi: 10.1007/s12686-011-9548-7

Edae, E. A., Bowden, R. L., and Poland, J. (2015). Application of population sequencing (POPSEQ) for ordering and imputing genotyping-by-sequencing markers in hexaploid wheat. *Genes Genomes Genet.* 5, 2547–2553. doi: 10.1534/g3.115.020362

El-Esawi, M. A., Witczak, J., Abomohra, A. E., Ali, H. M., Elshikh, M. S., and Ahmad, M. (2018). Analysis of the genetic diversity and population structure of Austrian and Belgian wheat germplasm within a regional context based on DArT markers. *Genes* 9:47. doi: 10.3390/genes9010047

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379

Eltaher, S., Sallam, A., Belamkar, V., Emara, H. A., Nower, A. A., Salem, K. F. M., et al. (2018). Genetic Diversity and Population Structure of F3:6 Nebraska Winter Wheat Genotypes Using Genotyping-By-Sequencing. *Front. Genet.* 9:76. doi: 10.3389/fgene.2018.00076

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. doi: 10.1111/j.1365-294X.2005.02553.x

Galili, T. (2015). Dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31, 3718–3720. doi: 10.1093/bioinformatics/btv428

Ganal, M. W., Altmann, T., and RoĖder, M. S. (2009). SNP identification in crop plants. *Curr. Opin. Plant. Biol.* 12, 211–217. doi: 10.1016/j.pbi.2008.12.009

Ganal, M. W., Polley, A., Graner, E. M., Plieske, J., Wieseke, R., Luerssen, H., et al. (2012). Large SNP arrays for genotyping in crop plants. *J. Biosci.* 37, 821–828. doi: 10.1007/s12038-012-9225-3

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., et al. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9:e90346. doi: 10.1371/journal.pone.0090346

Grassini, P., Eskridge, K. M., and Cassman, K. G. (2013). Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nat. Commun.* 4:2918. doi: 10.1038/ncomms3918

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812. doi: 10.1093/bioinformatics/btu393

Hao, C., Wang, L., Ge, H., Dong, Y., and Zhang, X. (2011). Genetic diversity and linkage disequilibrium in Chinese bread wheat (*Triticum aestivum* L.) *revealed by SSR markers*. *PLoS One* 6:e17279. doi: 10.1371/journal.pone.0017279

Haudry, A., Cenci, A., Ravel, C., Bataillon, T., Brunel, D., Poncet, C., et al. (2007). Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol. Biol. Evol.* 24, 1506–1517. doi: 10.1093/molbev/msm077

Hawkes, J. G. (1981). *Germplasm Collection, Preservation, and Use. In 2 Plant Breeding Symposium*. Iowa: Iowa State University Press.

He, J., Zhao, X., Laroche, A., Lu, Z. X., Liu, H. K., and Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front. Plant Sci.* 5:484. doi: 10.3389/fpls.2014.00484

He, Z. H., Rajaram, S., Xin, Z. Y., and Huang, G. Z. A. (2001). *History of Wheat Breeding in China*. Mexico, DF: Cimmyt.

Heslot, N., Rutkoski, J., Poland, J., Jannink, J. L., and Sorrells, M. E. (2013). Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. *PLoS One* 8:e74612. doi: 10.1371/journal.pone.0074612

Huang, X., Börner, A., Röder, M., and Ganal, M. (2002). Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. *Theor. Appl. Genet.* 105, 699–707. doi: 10.1007/s00122-002-0959-4

Joshi, C. P., and Nguyen, H. T. (1993). RAPD (random amplied polymorphic DNA) analysis based intervarietal genetic relationships among hexaploid wheats. *Plant Sci.* 93, 95–103. doi: 10.1016/0168-9452(93)90038-2

Kim, H. S., and Ward, R. W. (2000). Patterns of RFLP-based genetic diversity in germplasm pools of common wheat with different geographical or breeding program origins. *Euphytica* 115, 197–208. doi: 10.1023/A:1004022601879

Kumar, S., Banks, T. W., and Cloutier, S. (2012). SNP discovery through next-generation sequencing and its applications. *Int. J. Plant Genom.* 2012:831460. doi: 10.1155/2012/831460

Lam, H. M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F. L., et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42, 1053–1059. doi: 10.1038/ng.715

Lee, S., Choi, Y. M., Lee, M. C., Oh, S., and Jung, Y. (2018). Geographical comparison of genetic diversity in Asian landrace wheat (*Triticum aestivum* L.) germplasm based on high-molecular-weight glutenin subunits. *Genet. Resour. Crop Evol.* 65, 1591–1602. doi: 10.1007/s10722-018-0633-6

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Vikram, P., Singh, R. P., Kilian, A., Carling, J., Song, J., et al. (2015). A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC Genom.* 16:216. doi: 10.1186/s12864-015-1424-5

Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282

Lopes, M. S., Dreisigacker, S., Peña, R. J., Sukumaran, S., and Reynolds, M. P. (2015). Genetic characterization of the wheat association mapping initiative (WAMI) panel for dissection of complex traits in spring wheat. *Theor. Appl. Genet.* 128, 453–464. doi: 10.1007/s00122-014-2444-2

Marcussen, T., Sandve, S. R., Heier, L., Spannagl, M., Pfeifer, M., International Wheat Genome Sequencing Consortium, et al. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345:1250092. doi: 10.1126/science.1250092

Nagaoka, T., and Ogihara, Y. (1997). Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. *Theor. Appl. Genet.* 94, 597–602. doi: 10.1007/s001220050456

Nei, M. (1977). F-statistics and analysis of gene diversity in subdivided populations. *Ann. Hum. Gen.* 41, 225–233. doi: 10.1111/j.1469-1809.1977.tb01918.x

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89, 583–590. doi: 10.1007/BF00155576

Nezhad, K. Z., Weber, W. E., Röder, M. S., Sharma, S., Lohwasser, U., Meyer, R. C., et al. (2012). QTL analysis for thousand-grain weight under terminal drought stress in bread wheat (*Triticum aestivum* L.). *Euphytica* 186, 127–138. doi: 10.1007/s10681-011-0559-y

Novoselović, D., Bentley, A. R., Šimek, R., Dvojković, K., Sorrells, M. E., Gosman, N., et al. (2016). Characterizing Croatian wheat germplasm diversity and structure in a European context by DArT markers. *Front. Plant Sci.* 7:184. doi: 10.3389/fpls.2016.00184

Nuttonson, M. Y. (1955). Wheat-climatic relationships and the use of phenology in ascertaining the thermal and photothermal requirements of wheat. *Aibs Bull.* 83, 54–55. doi: 10.1097/00010694-195702000-00024

Peakall, R., and Smouse, P. E. (2006). GENALEX 6: genetic analysis in excel, Population genetic software for teaching and research. *Mol. Ecol. Notes.* 6, 288–295. doi: 10.1111/j.1471-8286.2005.01155.x

Peterson, G. W., Dong, Y., Horbach, C., and Fu, Y. B. (2014). Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. *Diversity* 6, 665–680. doi: 10.3390/d6040665

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. (2012b). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One* 7:e32253. doi: 10.1371/journal.pone.0032253

Poland, J. A., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., et al. (2012a). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5, 103–113. doi: 10.3835/plantgenome2012.06.0006

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

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. doi: 10.1371/journal.pone.0066428

Rimbert, H., Darrier, B., Navarro, J., Kitt, J., Choulet, F., Leveugle, M., et al. (2018). High throughput SNP discovery and genotyping in hexaploid wheat. *PLoS One* 13:e0186329. doi: 10.1371/journal.pone.0186329

Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., et al. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* 14:R55. doi: 10.1186/gb-2013-14-6-r55

Rufo, R., Alvaro, F., Royo, C., and Soriano, J. M. (2019). From landraces to improved cultivars: assessment of genetic diversity and population structure of Mediterranean wheat using SNP markers. *PLoS One* 14:e0219867. doi: 10.1371/journal.pone.0219867

Saint, P. C., Burgueno, J., Crossa, J., Dávila, G. F., Lopez, P. F., Moya, E. S., et al. (2016). Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. *Sci. Rep.* 6:27312. doi: 10.1038/srep27312

Saintenac, C., Jiang, D., Wang, S., and Akhunov, E. (2013). Sequence-based mapping of the polyploid wheat genome. *G3* 3, 1105–1114. doi: 10.1534/g3.113.005819

Shavrukov, Y., Suchecki, R., Eliby, S., Abugalieva, A., Kenebayev, S., and Langridge, P. (2014). Application of next-generation sequencing technology to study genetic diversity and identify unique SNP markers in bread wheat from Kazakhstan. *BMC plant Biol.* 14:258. doi: 10.1186/s12870-014-0258-7

Siedler, H., Messmer, M. M., Schachermayr, G. M., Winzeler, H., Winzeler, M., and Keller, B. (1994). Genetic diversity in European wheat and spelt breeding material based on RFLP data. *Theor. Appl. Genet.* 88, 994–1003. doi: 10.1007/BF00220807

Strelchenko, P., Street, K., Mitrofanova, O., Mackay, M., Balfourier, F., and Aleppo, S. (2005). Genetic diversity among hexaploid wheat landraces with different geographical origins revealed by microsatellites: comparison with AFLP, and RAPD data. *Proc. 4th Int. Crop Sci.* 2, 637-640.

Sukumaran, S., Dreisigacker, S., Lopes, M., Chavez, P., and Reynolds, M. P. (2015). Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. *Theor. Appl. Genet.* 128, 353–363. doi: 10.1007/s00122-014-2435-3

Talbert, L., Smith, L., and Blake, N. (1998). More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. *Genome* 41, 402–407. doi: 10.1139/g98-037

Tanksley, S. D., and McCouch, S. R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277, 1063–1066. doi: 10.1126/science.277.5329.1063

Thuillet, A. C., Bru, D., David, J., Roumet, P., Santoni, S., Sourdille, P., et al. (2002). Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. ssp durum desf. *Mol. Biol. Evol.* 19, 122–125. doi: 10.1093/oxfordjournals.molbev.a003977

Vikram, P., Franco, J., Burgue∼no-Ferreira, J., Li, H., Sehgal, D., Saint, P. C., et al. (2016). Unlocking the genetic diversity of Creole wheats. *Sci. Rep.* 6:23092. doi: 10.1038/srep23092

Winfield, M. O., Allen, A. M., Wilkinson, P. A., Burridge, A. J., Barker, G. L. A., Coghill, J., et al. (2018). High-density genotyping of the A.E. Watkins Collection of hexaploid landraces identifies a large molecular diversity compared to elite bread wheat. *Plant Biotechnol. J.* 16, 165–175. doi: 10.1111/pbi.12757

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating author (s): sewall wright reviewed work (s): published by: society for the study of evolution stable. *Evolution* 19, 395–420. doi: 10.1111/j.1558-5646.1965.tb01731.x

Würschum, T., Langer, S. M., Longin, C. F. H., Korzun, V., Akhunov, E., Ebmeyer, E., et al. (2013). Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. *Theor. Appl. Genet.* 126, 1477–1486. doi: 10.1007/s00122-013-2065-1

Zhang, D., Bai, G., Zhu, C., Yu, J., and Carver, B. F. (2010). Genetic diversity, population structure, and linkage disequilibrium in US elite winter wheat. *Plant Genome* 3, 117–127. doi: 10.3835/plantgenome2010.03.0004

Zhang, H. X., Zhang, F. N., Li, G. D., Zhang, S. N., Zhang, Z. G., and Ma, L. J. (2017). Genetic diversity and association mapping of agronomic yield traits in eighty six synthetic hexaploid wheat. *Euphytica* 213, 111. doi: 10.1007/s10681-017-1887-3

Zhao, J. J., Wang, Z. W., Liu, H. X., Zhao, J., Li, T., Hou, J., et al. (2019). Global status of 47 major wheat loci controlling yield, quality, adaptation and stress resistance selected over the last century. *BMC Plant Biol.* 19:5. doi: 10.1186/s12870-018-1612-y

Zori´c, M., Dodig, D., Kobiljski, B., Quarrie, S., and Barnes, J. (2012). Population structure in a wheat core collection and genomic loci associated with yield under contrasting environments. *Genetica* 140, 259–275. doi: 10.1007/s10709-012-9677-2

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 September 2020
doi: 10.3389/fgene.2020.583785

Transcriptomic Study for Identification of Major Nitrogen Stress Responsive Genes in Australian Bread Wheat Cultivars

Nigarin Sultana[1†], Shahidul Islam[1†], Angela Juhasz[1,2], Rongchang Yang[1], Maoyun She[1], Zaid Alhabbar[1], Jingjuan Zhang[1] and Wujun Ma[1]*

[1] State Agriculture Biotechnology Centre, College of Science, Health, Engineering and Education, Murdoch University, Perth, WA, Australia, [2] School of Science, Edith Cowan University, Joondalup, WA, Australia

High nitrogen use efficiency (NUE) in bread wheat is pivotal to sustain high productivity. Knowledge about the physiological and transcriptomic changes that regulate NUE, in particular how plants cope with nitrogen (N) stress during flowering and the grain filling period, is crucial in achieving high NUE. Nitrogen response is differentially manifested in different tissues and shows significant genetic variability. A comparative transcriptome study was carried out using RNA-seq analysis to investigate the effect of nitrogen levels on gene expression at 0 days post anthesis (0 DPA) and 10 DPA in second leaf and grain tissues of three Australian wheat (*Triticum aestivum*) varieties that were known to have varying NUEs. A total of 12,344 differentially expressed genes (DEGs) were identified under nitrogen stress where down-regulated DEGs were predominantly associated with carbohydrate metabolic process, photosynthesis, light-harvesting, and defense response, whereas the up-regulated DEGs were associated with nucleotide metabolism, proteolysis, and transmembrane transport under nitrogen stress. Protein–protein interaction and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis further revealed that highly interacted down-regulated DEGs were involved in light-harvesting and photosynthesis, and up-regulated DEGs were mostly involved in steroid biosynthesis under N stress. The common down-regulated genes across the cultivars included photosystem II 10 kDa polypeptide family proteins, plant protein 1589 of uncharacterized protein function, etc., whereas common up-regulated genes included glutamate carboxypeptidase 2, placenta-specific8 (PLAC8) family protein, and a sulfate transporter. On the other hand, high NUE cultivar Mace responded to nitrogen stress by down-regulation of a stress-related gene annotated as beta-1,3-endoglucanase and pathogenesis-related protein (PR-4, PR-1) and up-regulation of MYB/SANT domain-containing RADIALIS (RAD)-like transcription factors. The medium NUE cultivar Spitfire and low NUE cultivar Volcani demonstrated strong down-regulation of Photosystem II 10 kDa polypeptide family protein and predominant up-regulation of 11S globulin seed storage protein 2 and protein transport protein Sec61 subunit gamma. In grain tissue, most of the DEGs were related to nitrogen metabolism and proteolysis. The DEGs with high abundance in high NUE cultivar can be good candidates to develop nitrogen stress-tolerant variety with improved NUE.

Keywords: transcriptomics, nitrogen use efficiency, Australian wheat cultivars, nitrogen stress, RNA-seq

OPEN ACCESS

Edited by:
Jian Ma,
Sichuan Agricultural University, China

Reviewed by:
Genlou Sun,
Saint Mary's University, Canada
Quan Xie,
Nanjing Agricultural University, China

****Correspondence:***
Wujun Ma
W.MA@murdoch.edu.au

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *15 July 2020*
Accepted: *20 August 2020*
Published: *30 September 2020*

Citation:
Sultana N, Islam S, Juhasz A, Yang R, She M, Alhabbar Z, Zhang J and Ma W (2020) Transcriptomic Study for Identification of Major Nitrogen Stress Responsive Genes in Australian Bread Wheat Cultivars. Front. Genet. 11:583785. doi: 10.3389/fgene.2020.583785

INTRODUCTION

Over the past several decades, application of nitrogen fertilizer has been a practiced way to gain optimal crop yield. N fertilizer usage is predicted to reach 105 Tg N by 2030 and 135 Tg N by 2050 (Good et al., 2004). However, overuse of fertilizers can cause significant environmental issues such as erosion, soil quality depletion, and contamination of water supplies at local, regional, and global scales (Ahmad et al., 2008; Guo et al., 2010). Thus, it is important to develop new varieties with high nitrogen use efficiency (NUE). A better understanding of gene expression and regulation under nitrogen stressed conditions will help achieve this goal. Response to nitrogen scarcity in plants is controlled by changes in gene expression involved in different molecular mechanisms that are mainly related to plant developmental processes and yield (Zhang et al., 2006; Kant et al., 2011).

In particular, wheat grain production largely depends on the provision of N fertilizer and cultivars with high N uptake and utilization efficiency (Nyikako et al., 2014; Garnett et al., 2015; Cormier et al., 2016; Hitz et al., 2017). The biological pathways related to NUE are known to be strongly influenced by genetic variation as well as environmental factors such as N availability (Moll et al., 1982; Xu et al., 2012). Studies showed that N limitation can negatively affect wheat growth, morphology, and agronomic traits (Chandna and Ahmad, 2015; Curci et al., 2017; Wen et al., 2018; Wang J. et al., 2019).

Identifying key genes to improve stress tolerance in low N conditions is a feasible way to raise NUE. It is important to select cultivars that have contrasting NUEs for a comparative understanding of gene expression and regulation in response to N stressed conditions (Hirel et al., 2007; Kant et al., 2010). There are a number of approaches that have been undertaken by researchers to unravel how plants adapt to stressed conditions (Shrawat et al., 2008). In recent years, next-generation sequencing techniques have provided opportunities to study the gene expression and their regulations at the transcriptome level, and they have significantly enhanced the success rate of gene discovery (Diao et al., 2019). A number of studies also reported on transcriptome profiling by using Illumina's RNA-sequencing (Dai et al., 2015). Most of the studies demonstrated how a single genotype performed using contrasting environmental and growth conditions. In Arabidopsis, N response-related genes were identified using microarray analysis of gene expression changes in response to short-term and long-term treatments for nitrate with different concentrations (Wang et al., 2001; Price et al., 2004). Likewise, transcriptome study on different tissues with short-term N stress in rice also revealed a significant number of N responsive genes (Lian et al., 2006). Transcriptome study on long-term N stress was also reported in rice (Ym et al., 2009). However, a comprehensive transcriptome investigation by combining contrasting tissue, developmental stage, genotype, and N treatment is still lacking.

Nitrogen stress has a significant impact on the overall plant physiological process (Zhao et al., 2005) related to plant height, dry matter, grain yield (GY), and grain protein content (GPC) (Baligar et al., 2007; Wang W. et al., 2003). Nitrogen strongly influences photosynthesis through a large deposition of leaf N to ribulose bisphosphate carboxylase/oxygenase (Rubisco) and its involvement in stomatal opening (Evans, 1989). Approximately 75% of leaf N is allocated to chloroplasts, with about 27% of this utilized in Rubisco to ensure high photosynthetic activity (Evans, 1989; Makino, 2003). Nitrogen also influences photosynthesis via its impact on CO_2 assimilation and sugar partitioning (Drew et al., 1989; Foyer et al., 2011; Ishikawa-Sakurai et al., 2014). The decreased photosynthesis ultimately resulted in decreased biomass production and yield (Poorter and Evans, 1998; Long et al., 2006; Jin et al., 2015).

The regulation of plant photosynthetic activity is reported to be associated with brassinosteroids (BRs), a class of steroid hormones (Sakamoto et al., 2006; Komatsu et al., 2010). BRs are known to regulate stress responses and play important roles in regulating plant growth and development (Wang et al., 2012; Zhao and Li, 2012; Hayes, 2019). Several studies in Arabidopsis and rice showed involvement of BRs in controlling flowering, leaf senescence, chloroplast development, plant height, tiller numbers, and biomass, which are important agronomic traits affecting GY (Chono et al., 2003; Mussig et al., 2003; Sakamoto et al., 2006; Wu et al., 2008; Jeong et al., 2010). In wheat, BRs were also reported to be involved in promoting root growth and water stress tolerance (Hayes, 2019; Hou et al., 2019). However, the correlation of N stress on steroid biosynthesis has not been well studied. Thus, response to N stress is a rather complex process, and a better understanding of genes involved in different pathways is needed to develop stress-tolerant wheat varieties.

This study investigated three Australian bread wheat varieties, Mace, Spitfire, and Volcani, which are known to have high, medium, and low NUEs, respectively (Alhabbar et al., 2018b). Since gene expression in plants is controlled in a temporal and tissue-specific manner (Koltunow et al., 1990; Maizel and Weigel, 2004) and the N demand is subject to plant developmental stages, the current study used different tissues at different growth stages to unravel the broad picture of transcriptome profile with the objectives of identifying novel genes that are differentially expressed under long-term N stress compared to high N treatment, and by characterizing the underlying physiological and molecular mechanisms of tolerance to N stress.

MATERIALS AND METHODS

Plant Material, Growth Conditions, and Sample Collection

Three Australian wheat cultivars, Mace, Spitfire, and Volcani, were used in this study. Plants were grown in a glasshouse with a complete randomized block design (RCBD) including three replicates and using pots (190 mm height × 200 mm top diameter × 180 mm bottom diameter) without holes to avoid leaching. Plants were grown under controlled temperature and sunlight conditions [20/11°C (day/night)] for an 8 h light and 16 h dark photoperiod. The pots were watered manually based on soil water capacity. All plants were supplied with a basal nitrogen dose of 25 kg N ha^{-1} after 1 week of sowing.

Nitrogen-free Hoagland solution[1] was applied to all plants once every 2 weeks to meet the nutrient demand of plants except N. Two N rates—low (LN)/0 kg N ha^{-1} and high (HN)/100 kg N ha^{-1}—were applied at mid-tillering (Zadoks scale 22–25) and booting (Z43–Z45) stages for plants considered as low and high N treated, respectively. The timing for N applications was adjusted according to Zadoks (Z) decimal growth stage for wheat. Flexi-N (containing 50% urea, 25% nitrate, and 25% ammonium) was used as a source of N because of its high N content (42.2% N). Flexi-N was used since it contains nitrate that is directly available to plants while the urea and the ammonium become available more slowly, enabling a controlled release of N over an extended period (CSBP, 2012). Times for N application, recording of flowering time, measurement of chlorophyll content, and tissue collection were adjusted according to each cultivar's growth stage. For RNA extraction, the whole flag and second leaf samples were collected at the start of the flowering [0 days post anthesis (DPA)], 10, 20, and 30 DPA, while the developing grains were collected at 10, 20, and 30 DPA from the middle section of the main head, then snap-frozen in liquid nitrogen, and then stored at −80°C for later RNA extraction. Anthesis dates were estimated by the appearance of anthers on approximately 50% of all heads. Plant height was measured from soil surface to the top of the plant, and peduncle length was measured from the peduncle bottom to the joint with the stem. Chlorophyll content was measured using a handheld chlorophyll meter (IC-CCM-200—Chlorophyll Concentration Meter CCM-200 plus). One value per plant was taken from the flag leaf and second leaf on the main stem at four different growth stages: flowering (0 DPA), 10, 20, and 30 DPA. Each value was the average of three measurements recorded from the middle of the leaves. The main stem of each plant was individually labeled to ensure the same leaves were always measured. All plants in a pot (main stem plus tillers) were hand-harvested to measure yield components and the head number per plant counted. The heads were cut off and the seed number per head was counted. Grain samples were oven-dried in a forced-air circulating dryer at 60°C for 72 h. GPC was measured by near-infrared reflectance (NIR) spectroscopy using a FOSS NIR Systems model 5000 spinning cup. NIR data collection used DPIRD wheat calibrations developed over many years with the WinISI software (FOSS NIR Systems Inc., Laurel, MD, United States).

RNA Isolation, Library Construction, and Sequencing

Leaf and grain samples from three biological replicates were ground in liquid nitrogen, and the total RNA was extracted using a pre-chilled Trizol reagent (Invitrogen, Carlsbad, CA) following the manufacturer's directions, with some modifications. Proteins were removed with a protein extraction buffer (1 M Tris–HCl, 5 M NaCl, 10% SDS, 0.125 M EDTA, and 1 M DTT). After the protein removal, the acid phenol/chloroform/isopropanol (49:49:2), Trizol, and chloroform were added sequentially for the extraction of total RNA. Isopropanol was used for the precipitation of total RNA, which was subsequently treated with the Qiagen DNase kit to remove potential genomic DNA contamination. Concentration and purity were checked by Nanodrop, with 260/280 absorbance ratios of approximately 2.0, and the degradation and potential contamination was tested by agarose gel electrophoresis. RNA integrity was confirmed with an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). The mRNA was enriched using oligo (dT) beads and then fragmented randomly in a fragmentation buffer, followed by cDNA synthesis using random hexamers and reverse transcriptase. After first-strand synthesis, a custom second-strand synthesis buffer (Illumina) was added together with dNTPs, RNase H, and *Escherichia coli* polymerase I to generate the second strand by nick-translation. The final cDNA library was ready after a round of purification, terminal repair, A-tailing, ligation of sequencing adapters, size selection, and PCR enrichment. Library concentration was first estimated using a Qubit 2.0 fluorometer (Life Technologies) and then diluted to 1 ng μl^{-1} before checking the insert size on an Agilent 2100 Bioanalyzer. The concentration was then quantified at greater accuracy by quantitative PCR (Q-PCR) (library activity >2 nM). Each library with an individual barcode was sequenced by Illumina HiSeqTM PE125/PE150 (Illumina Inc., United States).

Transcriptome Analysis

A total of 90 different samples, including 30 each from three cultivars, Mace, Spitfire, and Volcani, were used for RNA-seq analysis. The samples were subjected to low and high nitrogen treatments to study a broad range of cell responses under nitrogen stress. For both treatment conditions, the replicates showed a high correlation coefficient ($r > 0.8$) between samples. For the RNA-seq downstream analysis, three samples (VAScLNR3, VEScHNR1, and SEScLNR2) were excluded due to sample quality. A total of 2070.85 million raw reads were filtered. A total of 1963.99 million clean reads were aligned against IWGSC RefSeq v1.0 gene models that produce 1750.09 million total mapped reads (TMRs), of which 128.89 million were mapped to multiple sites (MMR) and 1621.21 million were uniquely mapped. Among the TMRs, 810.66 million were mapped with a positive strand and 810.55 million were mapped with a negative strand (**Supplementary Table 1**). The average leaf Q20, Q30, and GC (Base G + Base C) contents were 96.93, 92.31, and 57.21%, respectively. Similarly, the average grain Q20, Q30, and GC (Base G + Base C) contents were 96.78, 92.16, and 57.79%, respectively. For both tissues, 95% of the total reads were filtered as cleaned reads, which confirms the fine quality of the sequencing results. Approximately, an average of 89% of clean reads were mapped for N-treated leaf samples, whereas 86% were mapped for grain tissue (**Supplementary Table 2**). For each sample, the percent of reads mapped to exon regions was above 90%, intron reads less than 5%, and intergenic reads less than 3%. The distribution of mapped reads of each sample in chromosome 3B was the highest, while the lowest reads were mapped in chromosome 6A. The gene expression level was measured by calculating the reads mapped to exons. Read count was proportional to the actual expression level as well as to the gene length and the sequencing depth. In order to make

[1]https://www.bioworld.com/productinfo/3_43_288_690/126289/Hoagland-Medium-Nitrogen-Free.html

comparable gene expression levels estimated from different genes and experiments, fragment per kilobase of transcript per million mapped reads (FPKM) was used for normalization. Considering the influence of various gene lengths and sequencing intensity, FPKM is commonly used to make comparison of gene expression levels among different samples.

Analysis of Differentially Expressed Genes

For the FPKM, a value of 1.0 was set as the threshold for determining whether a gene was expressed or not. HiSeq v0.6.1 (a Python package for high-throughput sequencing data analysis) was used to analyze gene expression levels in this experiment using the union mode. The correlation between samples was justified by the square of the Pearson correlation coefficient. The DESeq (version 1.10.1, R Bioconductor package) was used to conduct the differential expression analysis. The normalized data were fitted to a negative binomial generalized linear model. The threshold of the *p*-value after normalization (padj, q-value) was set as $\leq$0.05 for filtering accurate differentially expressed genes (DEGs). The clustering of DEGs was analyzed based on the FPKM value with the use of ggplot2 (version 2.1.0) and pheatmap (version 1.0.8) (Anders and Huber, 2010, 2012; Robinson et al., 2010; Trapnell et al., 2012). The DEGs were identified using the functional annotations of the IWGSC RefSeq v1.0 gene annotation.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Pathway Enrichment Analysis of DEGs

Gene ontology (GO) analysis was performed using ShinyGO v0.61[2]. GO with a false discovery rate (FDR) corrected at $p \leq 0.05$ was regarded as significant enrichment (Young et al., 2010). KOBAS (version 2.0[3], a web server for annotation and identification of enriched pathways and diseases, was applied for Kyoto Encyclopedia of Genes and Genomes (KEGG[4]) pathway enrichment analysis. Pathways with an FDR corrected at $p \leq 0.05$ were considered as significant enrichment (Mao et al., 2005; Kanehisa et al., 2007).

Protein–Protein Interaction Analysis

To predict the interaction of DEGs at the protein level under N stress and further confirmation of association of DEGs with biological pathways at the protein level, deduced amino acid sequences of DEGs were used to make a protein–protein network using the STRING (version 11.0) tool[5], a database for known and predicted protein interactions and functional associations predicted in common pathways. Due to the lack of detailed annotation of the wheat protein data available in STRING, we used two well-annotated species, rice, and Arabidopsis, as the reference to get protein–protein interaction information of the homologous wheat proteins. The global network graph of these interactions was constructed using the experimentally evident interacted proteins, and disconnected nodes (proteins) were removed to show the advanced view of highly connected proteins. MCL clustering using the inflation parameter 1.70 was used to show the association of clusters in KEGG pathways.

Hierarchical Cluster Analysis

Hierarchical cluster analysis was performed using the Morpheus package[6] Complete linkage analysis was performed using the Spearman rank correlation values.

Statistical Analysis

All data generated from the glasshouse experiments were analyzed by SPSS (version 24). Univariate analysis of variance (UNIANOVA) was used to determine the significance of different factors on each agronomic trait and protein parameter. The significant statistical difference was judged at $p \leq 0.05$.

RESULTS

Agronomic Performance of Wheat Cultivars Under Low and High Nitrogen Conditions

Under N stress (0 Kg N/ha), most of the agronomic traits were affected negatively in all three cultivars. In general, tiller number, GY, and chlorophyll content were mostly affected by N stress, whereas flowering days and GPC were less affected. A strong variation in grain weight per plant has been observed, which is considered as a yield component for small-scale glasshouse experiments. Grain weight per plant was dropped by 78% in Mace, 81% in Spitfire, and 80% in Volcani (**Figure 1A**) due to N stress. Similarly, under N stress, the tiller number (**Figure 1B**) was decreased by 72.4, 84.2, and 81.2%, and the chlorophyll content of both flag leaf and leaf 2 (**Figures 1C,D**) were decreased by approximately 85, 80, and 68% for Mace, Spitfire, and Volcani, respectively. In addition, a significant reduction in plant height (**Figure 1E**), main head length (**Figure 1F**), and spikelet number per head (**Figure 1G**) has also been observed under N stress. Flowering days (**Figure 1H**) were decreased by 3.2, 4.9, and 7.9%, and the GPC (**Figure 1I**) decreased by 4.1, 9.3, and 29.5% in Mace, Spitfire, and Volcani, respectively. A significant negative impact of low N on leaf area and peduncle length has also been noticed (**Figures 1J–L**). Overall, the influence of N stress on growth and agronomic traits was variable across the cultivars, where Spitfire and Volcani were more affected compared to Mace.

Overview of RNA-Seq Transcriptome Profile in Response to Nitrogen Stress

A total of 12,108 DEGs in leaf tissue and 276 DEGs in grain tissue were identified under the N stressed condition. Mace, Spitfire, and Volcani had 699, 10,535, and 1671 DEGs in second leaf and another 25, 252, and 16 DEGs in grain tissue, respectively. In

[2]http://bioinformatics.sdstate.edu/go/

[3]http://kobas.cbi.pku.edu.cn/

[4]http://www.genome.jp/kegg/

[5]https://string-db.org/

[6]https://software.broadinstitute.org/morpheus/

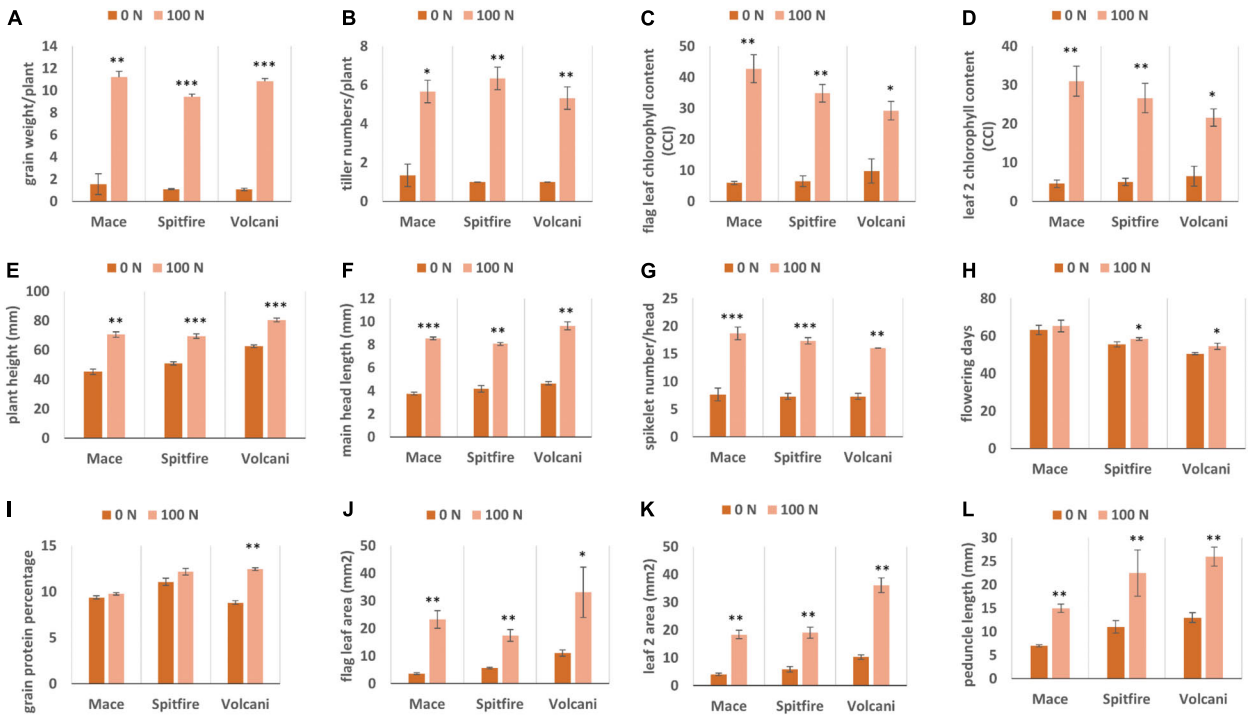

FIGURE 1 | Growth and agronomic traits of *Triticum aestivum* cultivars (Mace, Spitfire and Volcani) under low (0N) and high (100 N) treatments. **(A)** grain weight/plant **(B)** tiller numbers/plant **(C)** flag leaf chlorophyll content **(D)** leaf 2 chlorophyll content **(E)** plant height **(F)** main head length **(G)** spikelet number/head **(H)** flowering days **(I)** grain protein percentage **(J)** flag leaf area **(K)** leaf 2 area **(L)** peduncle length. The values are presented as means ± standard deviation (SD) of three independent biological repeats. Error bars were calculated from three biological replicates and one-way ANOVA was used to test for significance of nitrogen treatment effects on different parameters at $P \leq 0.05$ level. *, **, *** Significant at the 0.05, 0.01 and 0.001 probability level, respectively.

the second leaf tissue, under N stress, the total up- and down-regulated DEGs at two different time points were variable across the cultivars. In Mace, at 0 DPA, the down-regulated and up-regulated DEGs were 434 and 102, respectively. At 10 DPA, the up-regulated and down-regulated DEGs were 109 and 74, respectively. Similarly, in Volcani, the down-regulated DEGs at 0 and 10 DPA were counted as 753 and 430, respectively, whereas the up-regulated DEGs were 354 at 0 DPA and 261 at 10 DPA. Cultivar Spitfire showed 536 up-regulated and 39 down-regulated DEGs at 0 DPA, whereas it showed 6624 up-regulated and 3830 down-regulated DEGs at 10 DPA. On the other hand, in grain tissue at 10 DPA, the down-regulated DEGs were 5, 237, and 8, while the up-regulated DEGs were 0, 15, and 8 identified in Mace, Spitfire, and Volcani, respectively. Variation in up- and down-regulated genes across the cultivars can be related to the difference in their response to N stress (**Figure 2**).

Common DEGs Between Leaf and Grain

A total of 50 common DEGs were identified between the second leaf and grain tissue under the N stressed condition, of which 30 were down-regulated and 7 were up-regulated in both tissues. Thirteen DEGs showed inconsistent up- and down-regulation (**Supplementary Table 3**). Several stress-related genes have been identified among those common DEGs with >log2 fold change, including plasma membrane ATPase, Serine protease HtrA-like, transcription factor AS2/LOB, etc. Several transmembrane transport-related proteins including sulfate transporter, glycosyltransferase, and WAT1-related protein were also common in second leaf and grain tissues. On the other hand, NUE-related glutamine synthetase and glutamine dumper were significantly up-regulated in second leaf tissues but down-regulated in the grain tissue of Volcani, indicating their tissue-specific expression. Another gene related to amino acid metabolism, isoaspartyl peptidase/L-asparaginase, was up-regulated in the second leaf and grain tissue of Spitfire and Mace, indicating non-specific tissue expression. In general, the common down-regulating DEGs were largely involved in carbohydrate metabolic process (chitinase, trehalose-6-phosphate synthase) and oxidation–reduction process (aldehyde dehydrogenase, peroxidase, methyl sterol monooxygenase 1-2, gibberellin 20 oxidase 2, catalase). The up-regulating common DEGs are involved in N compound metabolic process (glutamine synthetase), sulfate transmembrane transport (sulfate transporter), and amino acid metabolism (aminotransferase like protein, isoaspartyl peptidase/L-asparaginase).

Common DEGs Between 0 and 10 DPA

Under N stress, some DEGs showed consistent up- or down-regulation at both 0 and 10 DPA (**Figure 3**) despite the fact that they were variable between the cultivars. For example, in Mace, a total of 28 DEGs were commonly expressed at both 0 and 10 DPA, and of them, 23 were down-regulated

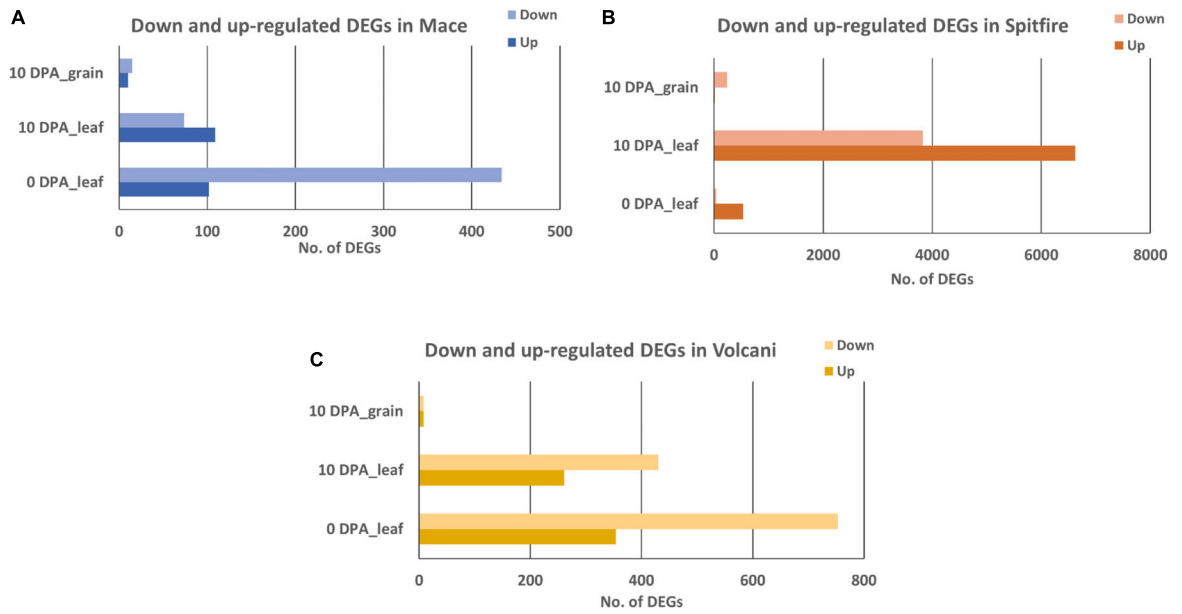

FIGURE 2 | Number of up- and down-regulated expressed at 0 and 10 DPA DEGs in wheat cultivars **(A)** Mace, **(B)** Spitfire, and **(C)** Volcani. Differentially expressed genes (DEGs), 0 days post anthesis (0 DPA), 10 days post anthesis (10 DPA).

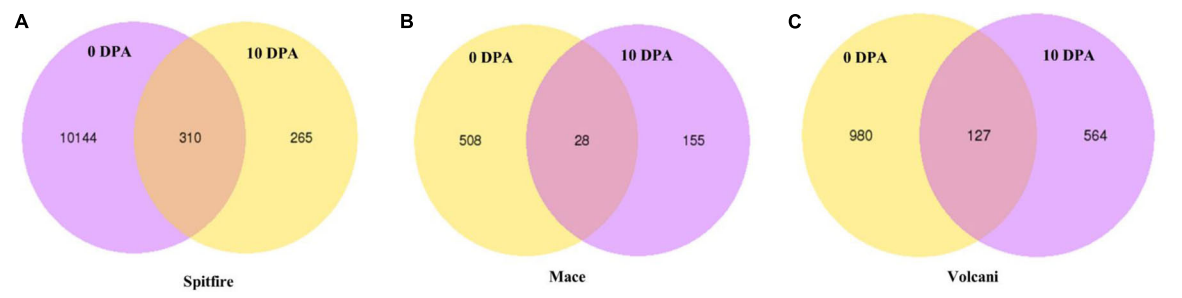

FIGURE 3 | Consistently expressed DEGs at two time points in the leaf tissue of wheat cultivars **(A)** Spitfire, **(B)** Mace, and **(C)** Volcani. Differentially expressed genes (DEGs), 0 days post anthesis (0 DPA), 10 days post anthesis (10 DPA).

and 5 were upregulated under N stress. Some of those DEGs showed high fold change ($> +2$ or < -2) including plant protein DUF1589 of uncharacterized protein function, gibberellin receptor GID1A, catalase, a two-component response regulator, and cytochrome P450 was down-regulated, whereas RADIALIS-like TF, glycosyltransferase, and receptor-like protein kinase were up-regulated. In contrast, in Spitfire, among the 310 commonly expressed DEGs at 0 and 10 DPA, 16 showed down-regulation and 294 showed up-regulation under N stress. Among the DEGs in Spitfire, the Dof zinc finger protein, two-component response regulator, glycine-rich protein-A3, and calcium-dependent protein kinase 15 were down-regulated (log2 fold change < -4.0), and the cinnamoyl CoA reductase, receptor-like kinase, protein kinase-like, translation initiation factor IF-2, aspartate-tRNA ligase, and a beta-glucosidase were up-regulated under N stress. In Volcani, a total of 127 DEGs were found to be expressed both at 0 and 10 DPA under N stress. Of them, 86 were down-regulated and 38 were up-regulated commonly at both DPA, whereas three DEGs were down-regulated at 0 DPA but up-regulated at 10 DPA. The top down-regulated DEGs included a chlorophyll a-b binding protein, methyltransferase, endo-1,3 beta-glucanase, and plant protein DUF1589 of uncharacterized protein function, whereas the top up-regulated DEGs included cinnamoyl CoA reductase, MYB TF, glycosyltransferase, and beta-glucosidase (**Supplementary Table 4**).

Common DEGs Among Cultivars

Venn diagram analysis was used to identify the number of common DEGs among the cultivars (Hulsen et al., 2008). In the second leaf tissue, down-regulated 4 DEGs at 0 and 10 DEGs at 10 DPA whereas only 3 up-regulated DEGs at 10 DPA were found in common. The common down-regulated DEGs were identified as glycine-rich protein A3, calcium-dependent protein kinase 15, etc. The common up-regulated DEGs were identified as sulfate transporter and L-allo-threonine aldolase, which is related to amino acid metabolism. However, in grain tissue, only two

TABLE 1 | List of DEGs common among wheat cultivars: Mace, Spitfire, and Volcani.

Tissue	Up-/down-regulation	Stage	Gene_id	Annotation	Log2 fold change		
					Volcani	Spitfire	Mace
Leaf	Down	0 DPA	TraesCS3A02G439500LC	Glycine-rich protein A3	−5.816	−4.5188	−5.5826
			TraesCS3D02G150400LC	Glycine-rich protein A3	−5.4414	−3.2827	−4.9171
			TraesCS4A02G245300	Protein DETOXIFICATION	−6.9194	−2.5502	−4.1621
			TraesCS3D02G150300LC	Calcium-dependent protein kinase 15	−4.3632	−4.3883	−5.9832
	Down	10 DPA	TraesCS2D02G555300	ARM repeat superfamily protein	−2.7152	−3.2171	−3.3102
			TraesCS2D02G259200	Two-component response regulator	−1.9826	−2.7094	−2.2783
			TraesCS1B02G388700	Methyltransferase	−6.0313	−7.4915	−4.9876
			TraesCS6B02G051800	Glycerol-3-phosphate acyltransferase	−2.6867	−2.1339	−2.0549
			TraesCS7D02G516800	Chaperone protein dnaJ	−1.8551	−3.0128	−2.0382
			TraesCS3D02G144900	Protein DJ-1	−3.295	−5.0529	−3.237
			TraesCS5A02G472500	Amino acid transporter, putative	−4.416	−2.385	−2.4211
			TraesCS2B02G277300	Two-component response regulator	−2.7784	−3.5708	−2.8394
			TraesCS3D02G316900LC	Nucleoside triphosphatase I	−4.4166	HN	−4.1874
			TraesCS7D02G388400	Tryptophan synthase beta chain	−2.9456	−1.4694	−1.9753
	Up	10 DPA	TraesCS7D02G084100	Sulfate transporter	2.0942	3.4823	2.2596
			TraesCS7B02G128800	Epoxide hydrolase 2	4.7394	1.3931	2.5759
			TraesCS2D02G379000	L-allo-threonine aldolase	3.7648	2.4275	2.1093
Grain	Down	10 DPA	TraesCS2A02G194500	LOB domain-containing protein, putative	−1.3364	−2.2599	−1.6826
			TraesCS2D02G193400	LOB domain-containing protein	−1.5747	−2.1909	−1.8922

Differentially expressed genes (DEGs), 0 days post anthesis (0 DPA), 10 days post anthesis (10 DPA).

down-regulated DEGs identified across the three cultivars were annotated as LOB-domain containing proteins (**Table 1**).

While considering the common DEGs between two cultivars, in most cases, the highest number of DEGs was common between Spitfire and Volcani among all combinations (**Figure 4**). The major common down-regulated DEGs between Spitfire and Volcani in second leaf were identified as methyltransferase, chlorophyll a-b binding protein, methyltransferase, and aquaporin, whereas the common up-regulated DEGs were identified as aminotransferase, early light-induced protein, F-box domain-containing protein, and glycosyltransferase. In grain tissues, among the four common down-regulated DEGs between Spitfire and Volcani, cysteine proteinase inhibitor and Ureide permease-like protein are related to N metabolism. The top down-regulated DEGs common between Mace and Volcani were identified as photosystem II 10 kDa polypeptide family proteins, chlorophyll a-b binding protein, and plant protein DUF1589 of uncharacterized protein function, whereas plasma membrane ATPase and glycosyltransferase were found as the common top up-regulated DEGs. Similarly, the common top up-regulated DEGs between Mace and Spitfire were identified as a vacuolar-processing enzyme, a boron transporter, a nuclease S1, and cytochrome P450 family protein, whereas down-regulated DEGs were identified as haloacid dehalogenase-like hydrolase (HAD) superfamily protein and thaumatin-like protein.

Gene Ontology Reflects the Function of DEGs in Response to Nitrogen Stress

The top 10 biological process GO terms characteristic to the DEGs are presented in **Figure 5**. The frequency of the GO term is shown as percentage of the genes compared to the total gene number related to the GO term.

In the case of the second leaf tissue, transmembrane transport GO term appeared as the top group within up-regulated DEGs in all three cultivars. Notably, another three top GO terms were common in Spitfire and Volcani, which were ion transport, lipid metabolic process, and metal ion transport, indicating that these two cultivars have some common physiological response mechanisms to N stress. In contrast, Mace did not have any other top 10 GO common with either cultivar. DNA metabolic process and organelle organization are the next top GO terms for cultivar Mace. On the other hand, genes showing decreased expression under N stress in Mace second leaf tissue were mostly involved in defense response and carbohydrate metabolic process. In Spitfire, decreasing gene expression was largely related to photosynthesis and light harvesting, organonitrogen compound biosynthesis process, and small molecule biosynthetic process. Similarly, in Volcani, genes with decreased expression patterns were also related to photosynthesis, carbohydrate metabolic process, and response to external stimulus.

In the grain tissue, the transmembrane transport process GO term was the top enriched group among the up-regulated DEGs in Mace and Spitfire. Mace also showed enrichment in carbohydrate metabolic process and ion transport. However, proteolysis and negative regulation of catalytic activity were common in Spitfire and Volcani among the top 10 enriched GO terms. Nitrogen compound transport appeared as the common GO term in all three cultivars among the down-regulated DEGs. Mace did not show enrichment for negative regulation of endopeptidase activity and proteolysis like Spitfire and Volcani.

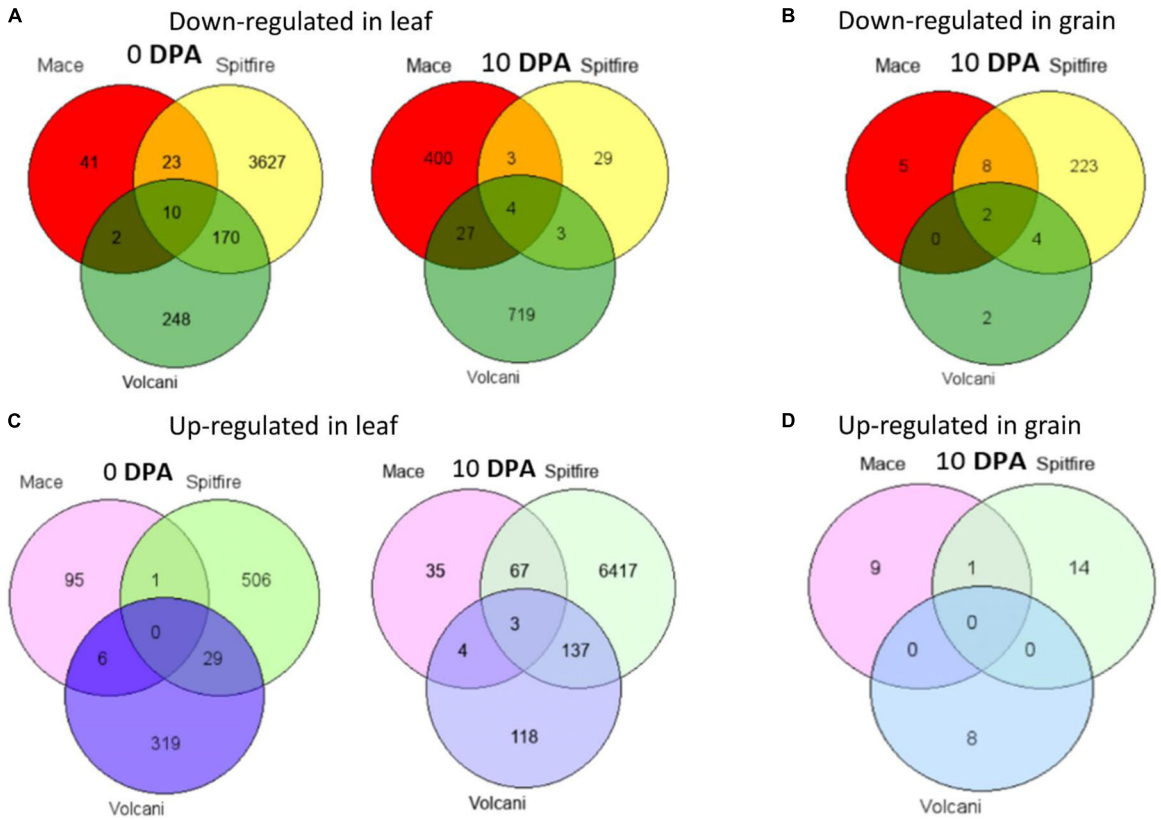

FIGURE 4 | Venn diagrams of DEGs shared among Mace, Spitfire, and Volcani in leaf and grain tissues at two development stages. **(A)** The number of down-regulated genes in leaf. **(B)** The number of down-regulated genes in grain. **(C)** The number of up-regulated genes in leaf. **(D)** The number of up-regulated genes in grain. Differentially expressed genes (DEGs), 0 days post anthesis (0 DPA), 10 days post anthesis (10 DPA).

However, in Mace and Spitfire, glycolysis process was the top enriched down-regulated GO term.

KEGG Analysis Spanned Function of DEGs in Response to Nitrogen Stress

Using the well-annotated rice genome as a reference, KEGG pathway enrichment analysis identified significantly enriched metabolic pathways and signal transduction pathways associated with DEGs. The top 20 most significantly enriched pathways were selected to produce the KEGG scatter plot (**Supplementary Figure 1**). Results for the KEGG pathway terms that were significant at adjusted *p*-value q-≤0.5 are shown in **Tables 2**, **3** for second leaf and grain, respectively. Under N stress, a total of 41 KEGG pathway terms were significantly associated with 12,108 DEGs in the second leaf, and 14 KEGG pathway terms were associated with 276 DEGs in the grain tissue. Among the 41 significant KEGG terms for the second leaf tissue, 3, 15, and 6 KEGG terms were specific to Mace, Spitfire, and Volcani, respectively, whereas one KEGG term was common between Mace and Spitfire, six KEGG terms between Mace and Volcani, and five KEGG terms between Spitfire and Volcani (**Table 2**). There were five KEGG terms common in all three cultivars under N stress, namely, phenylpropanoid biosynthesis, biosynthesis of secondary metabolites, flavonoid biosynthesis, metabolic pathways, and starch and sucrose metabolism. Among the 14 significant KEGG terms associated with DEGs in grain, eight, four, and one KEGG terms were specific to Mace, Spitfire, and Volcani, respectively (**Table 3**). The DEGs in the grain of all three cultivars were commonly associated with the KEGG pathway term glycolysis/gluconeogenesis. Among the significant KEGG terms for DEGs in the second leaf, zeatin biosynthesis, arginine and proline metabolism, and sulfur metabolism were specific to Mace, with terms like plant–pathogen interaction, photosynthesis, pentose phosphate pathway, porphyrin, and chlorophyll metabolism specific to Spitfire and beta-alanine metabolism, tryptophan metabolism, ubiquinone, and other terpenoid-quinone biosynthesis pathways found only in DEGs of Volcani. In the grain tissue, the KEGG pathways specific to Mace were glycerolipid metabolism, sphingolipid metabolism, porphyrin and chlorophyll metabolism, and galactose metabolism, whereas the pathways specific to Spitfire were alanine, aspartate and glutamate metabolism, glycine, serine, and threonine metabolism, and ribosome biogenesis in eukaryotes. The pathways specific to Volcani were cysteine and methionine metabolism. In addition, some KEGG pathways were common between two cultivars only, e.g., Mace and Spitfire had a MAPK signaling pathway common in the second leaf

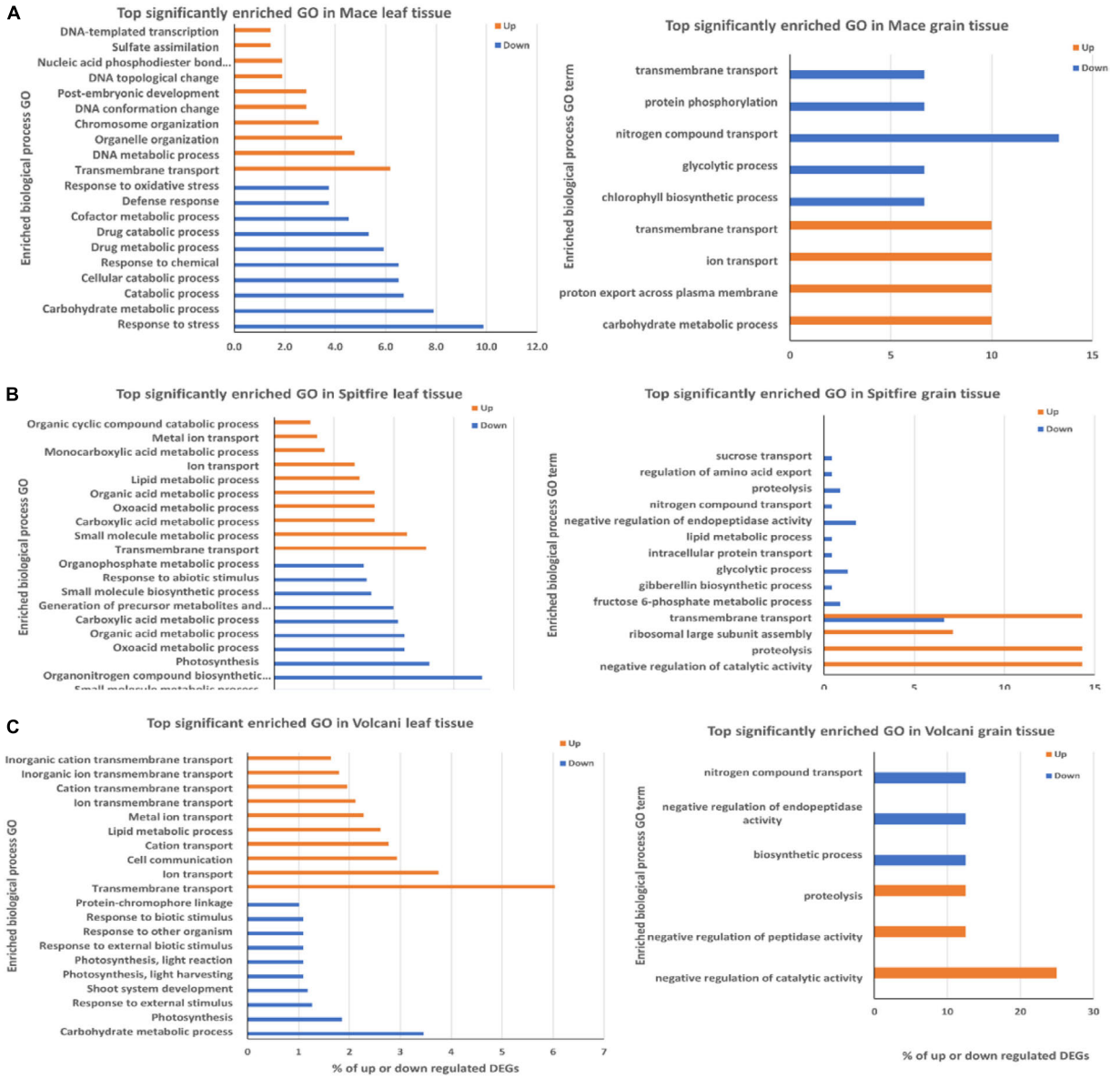

FIGURE 5 | Top biological process GO terms in leaf and grain tissue of wheat cultivars **(A)** Mace, **(B)** Spitfire, and **(C)** Volcani. The frequency of the GO term is shown as percentage of the genes related to the GO term. GO, gene ontology.

and glycolysis/gluconeogenesis common in the grain tissue. The KEGG terms common in Mace and Volcani included glutathione metabolism, galactose metabolism, and ABC transporters, whereas biosynthesis of amino acids, photosynthesis–antenna proteins, and circadian rhythm–plant pathways were common between Spitfire and Volcani.

Protein–Protein Interaction Network Analysis of DEGs

MCL clustering using the inflation parameter 1.70 was used to show the association of clusters in KEGG pathways (**Figures 6**, **7**). Networks showed that a large number of proteins were involved in multiple interactions and grouped into seven major clusters. Among the seven clusters, two large clusters were enriched in photosynthesis and steroid biosynthesis. All the interacting DEGs identified as photosynthesis-related, and photosynthesis antenna proteins were down-regulated while some DEGs related to steroid biosynthesis were up- or down-regulated. Among the other clusters, the majority of down-regulated DEGs were involved in carbohydrate metabolism, amino sugar, and nucleotide metabolism, whereas up-regulated DEGs were mostly related to amino acid metabolism and signaling. In the biosynthesis of secondary metabolites, both the up- and down-regulated DEGs were involved. Overall, the number of down-regulated DEGs was higher in the network and was mainly involved in photosynthesis and photosynthesis-antenna proteins.

TABLE 2 | Significantly (adjusted p-value ≤ 0.5) enriched KEGG pathways in Mace, Spitfire, and Volcani under nitrogen stress in second leaf tissue.

Cultivar	KEGG pathway	KEGG id
Mace	Zeatin biosynthesis	osa00908
	Arginine and proline metabolism	osa00330
	Sulfur metabolism	osa00920
Spitfire	Plant–pathogen interaction	osa04626
	Carbon metabolism	osa01200
	Glyoxylate and dicarboxylate metabolism	osa00630
	Photosynthesis	osa00195
	Glycine, serine, and threonine metabolism	osa00260
	Carbon fixation in photosynthetic organisms	osa00710
	Glycolysis/gluconeogenesis	osa00010
	Fructose and mannose metabolism	osa00051
	Alanine, aspartate, and glutamate metabolism	osa00250
	Taurine and hypotaurine metabolism	osa00430
	Pentose phosphate pathway	osa00030
	One carbon pool by folate	osa00670
	Porphyrin and chlorophyll metabolism	osa00860
	Histidine metabolism	osa00340
	Ascorbate and aldarate metabolism	osa00053
Volcani	Beta-alanine metabolism	osa00410
	Terpenoid backbone biosynthesis	osa00900
	Ubiquinone and other terpenoid-quinone biosynthesis	osa00130
	Tryptophan metabolism	osa00380
	Butanoate metabolism	osa00400
	Phenylalanine, tyrosine, and tryptophan biosynthesis	osa00651
Mace–Spitfire	MAPK signaling pathway—plant	osa4016
Mace–Volcani	Stilbenoid, diarylheptanoid, and gingerol biosynthesis	osa00945
	Plant hormone signal transduction	osa04075
	Phenylalanine metabolism	osa00360
	Galactose metabolism	osa00052
	ABC transporters	osa02010
	Glutathione metabolism	osa00480
Spitfire–Volcani	Biosynthesis of amino acids	osa01230
	Photosynthesis-antenna proteins	osa00196
	Circadian rhythm—plant	osa04712
	Cyanoamino acid metabolism	osa00460
	Cysteine and methionine metabolism	osa00270
Mace–Spitfire–Volcani	Phenylpropanoid biosynthesis	osa00940
	Biosynthesis of secondary metabolites	osa01110
	Flavonoid biosynthesis	osa00941
	Metabolic pathways	osa01100
	Starch and sucrose metabolism	osa00500

TABLE 3 | Significantly (adjusted p-value ≤ 0.5) enriched KEGG pathway in Mace, Spitfire, and Volcani under nitrogen stress in grain tissue.

Cultivar	KEGG pathway	KEGG id
Mace	Galactose metabolism	osa00052
	Oxidative phosphorylation	osa00190
	Glycerolipid metabolism	osa00561
	Sphingolipid metabolism	osa00600
	Glycosphingolipid biosynthesis—globo series	osa00603
	Porphyrin and chlorophyll metabolism	osa00860
	Biosynthesis of secondary metabolites	osa01110
	RNA degradation	osa03018
Spitfire	Alanine, aspartate, and glutamate metabolism	osa00250
	Alanine, aspartate, and glutamate metabolism	osa00250
	Glycine, serine, and threonine metabolism	osa00260
	Ribosome biogenesis in eukaryotes	osa03008
Volcani	Cysteine and methionine metabolism	osa00270
Mace–Spitfire	Glycolysis/gluconeogenesis	osa00010

Identification of Nitrogen Metabolism-Related Genes in Response to Nitrogen Stress

N metabolism is a vital biological process in plants that determines crop productivity and yield (Stitt et al., 2002; Cai et al., 2009). The DEGs involved in N uptake, transport, and assimilation were listed separately and are presented in **Table 4**. Most of the N metabolism-related DEGs showed up-regulation under N stress. Among the most significant DEGs (fold change >2.0), 65% were up-regulated and 35% were down-regulated (**Table 5**). Spitfire showed abundancy for N metabolism-related DEGs compared to Mace and Volcani. The top up-regulated N metabolism-related DEGs included amino acid permease, glutamate dehydrogenase, low-affinity nitrate transporter protein NRT1/PTR family 1.1, tyrosine aminotransferase, and high-affinity nitrate transporter, whereas the top down-regulated DEGs included amino acid transporter family protein, nitrate transporter 1.1, nitrate transporter 1.2, nitrate reductase, and tryptophan aminotransferase. Spitfire showed the most induction ratio for protein NRT1/PTR FAMILY 1.1 (log2 fold change 6.4) and tyrosine aminotransferase (log2 fold change 5.74). Mace showed up-regulation of cationic amino acid transporter and down-regulation of amino acid transporter family proteins, amino acid permease, and protein NRT1/PTR FAMILY 1.1. Volcani showed up-regulation of amino acid permease, nitrate transporter protein NRT1/PTR FAMILY 5.5, and ammonium transporter and down-regulation of isoaspartyl peptidase/L-asparaginase, nitrate transporter 1.1 and 1.2, and tryptophan aminotransferase.

Identification of Common Nitrogen Stress-Responsive Genes Across the Cultivars

Identification of the common DEGs between two N treatments included genes from 6 pair comparisons (2 developmental stages × 3 cultivars) for leaf and 3 pair comparisons (1 developmental stage × 3 cultivars) for grain tissue. In the second leaf, a total of 14 up-regulated and 42 down-regulated DEGs were identified that were common in all three cultivars (**Tables 6**, **7**). Among the 14 up-regulated common DEGs, aldo/keto reductase family protein, nuclease S1, alcohol dehydrogenase, putative, placenta-specific8 (PLAC8) family protein, and sulfate transporter showed relatively high (log2) fold

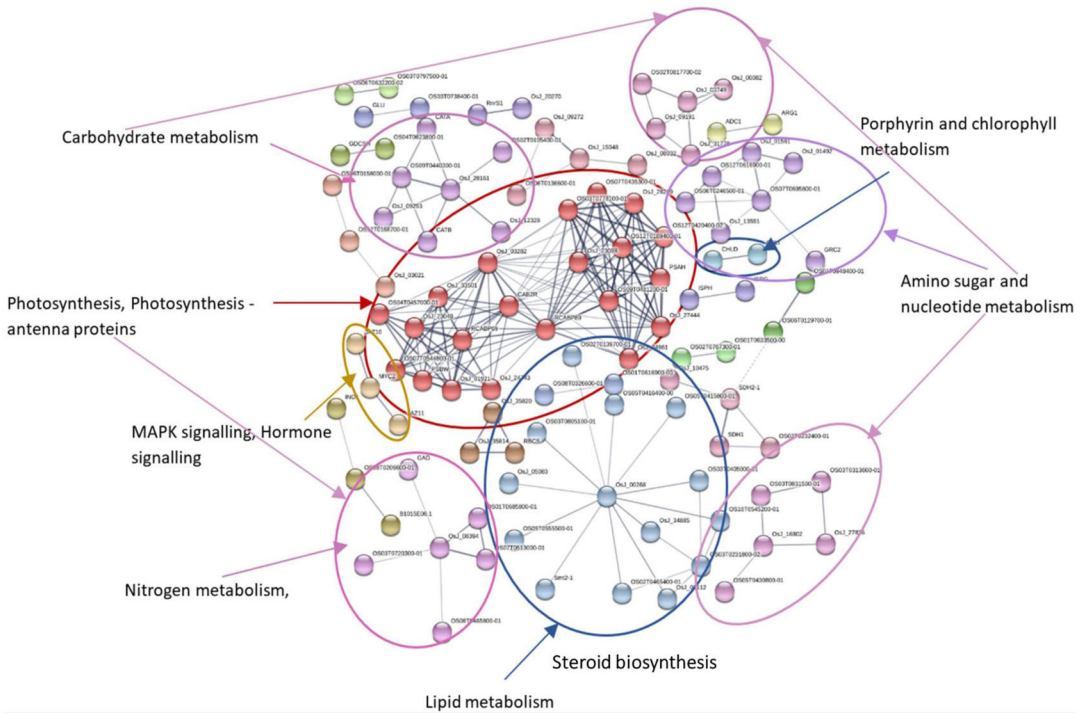

FIGURE 6 | Protein–protein interaction network analysis of DEGs under N stress using *Oryza sativa* as reference. The different highlighted color indicates the different clusters of DEGs involved in different KEGG pathways. Differentially expressed genes (DEGs), Kyoto Encyclopedia of Genes and Genomes (KEGG).

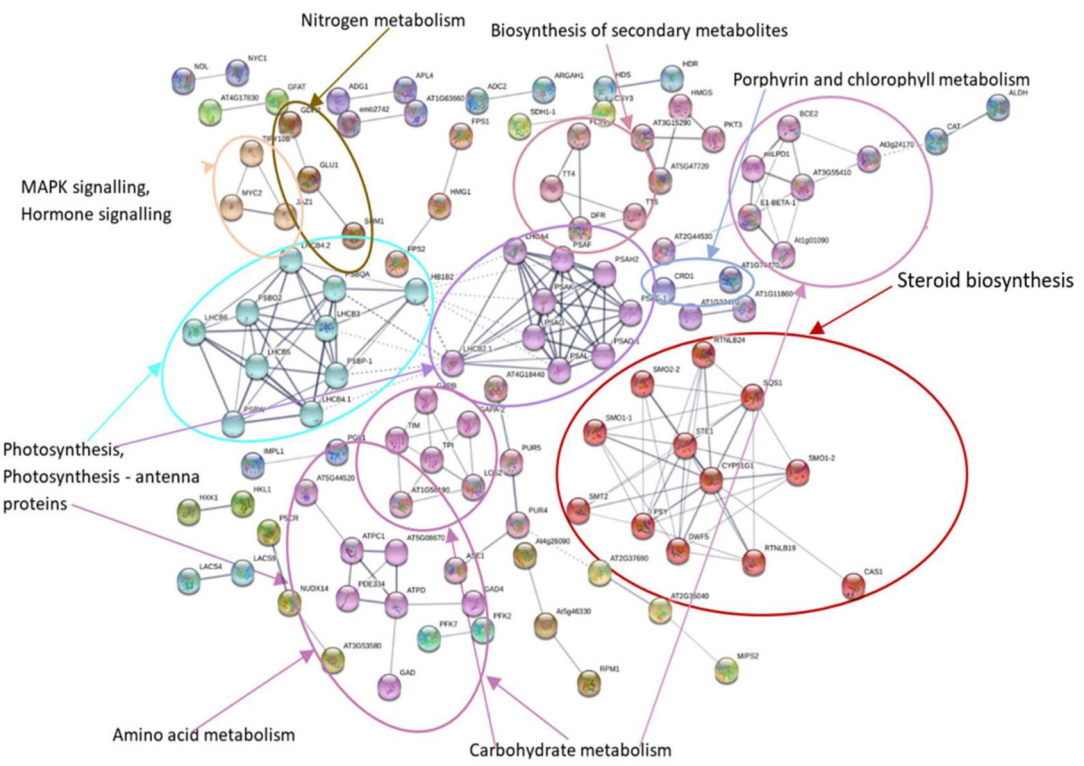

FIGURE 7 | Protein–protein interaction network analysis of DEGs under N stress using *Arabidopsis thaliana* as reference. The different highlighted color indicates the different clusters of DEGs involved in different KEGG pathways. Differentially expressed genes (DEGs), Kyoto Encyclopedia of Genes and Genomes (KEGG).

TABLE 4 | Differentially expressed genes (DEGs) involved in nitrogen uptake, transport, and assimilation.

Name	NUE class	DEG count (current study)	NUE related effect	Host	References
AAT	Amino acid transporter	57	Important for early seed development	Arabidopsis	Schmidt et al., 2007
AMT	Ammonium transporter	5	Enhanced ammonium permeability improves growth and yield	Rice	Ranathunge et al., 2014
CAT	Cationic amino acid transporter	12	Involved in intracellular amino acid storage and mobilization	Arabidopsis	Yang H. et al., 2015
CLC	Chloride channel protein	14	Enhanced N assimilation and tolerance to stress	Oilseed rape	Liao et al., 2018
LHT	Lysine/histidine transporter	1	Disruption of LHT lead to growth inhibition and low yield	Rice	Wang X. et al., 2019
NRT	Nitrate transporter	36	Suppression of NO_3^--starvation-induced leaf senescence	Arabidopsis	Meng et al., 2016
OPT	Oligopeptide transporter	5	Essential for embryo development	Arabidopsis	Stacey et al., 2002
AGT	Alanine: glyoxylate aminotransferase	2	Catalyze transamination reaction in peroxisome	Arabidopsis	Liepman and Olsen, 2001
ASN	Asparagine synthetase	1	Regulation of plant development and tiller outgrowth	Rice	Lu et al., 2018
AspAT	Aspartate aminotransferase	5	Overexpression related to increase amino acid content in seed	Rice	Zhou et al., 2009
GDH	Glutamate dehydrogenase	3	Played important role in nitrogen metabolism and plant growth, and grain yield	Rice	Abiko et al., 2010
GOGAT	Glutamate synthase	3	Increased ammonium assimilation in root	Arabidopsis	Konishi et al., 2014
GS	Glutamine synthetase	7	Knockdown negatively affect plant growth, spikelet no., grain weight	Rice	Tabuchi et al., 2005
NR	Nitrate reductase	2	Increase lateral root formation under partial nitrate nutrition	Rice	Sun et al., 2015
NiR	Nitrite reductase	3	Increased nitrite assimilation	Arabidopsis	Takahashi et al., 2001
TS	Threonine synthase	3	Inhibition related to high methionine biosynthesis	Potato	Zeh et al., 2001
TrP	Tryptophan aminotransferase	3	Improved grain yield	Wheat	Shao et al., 2017
TAT	Tyrosine aminotransferase	1	Differentially expressed between low and high nitrogen treatments	Wheat	Current study
GDU	Glutamine dumper	2	Involved in export of amino acids	Arabidopsis	Pilot et al., 2004

change. Eight of the 42 down-regulated DEGs in the leaf tissue showed high log2 fold change, including 3 photosystem II 10 kDa polypeptide family protein, 2 methyltransferases, chlorophyll a-b binding protein, cytoplasmic dynein 2 heavy chain 1, and plant protein 1589 of uncharacterized protein function. However, the only two down-regulated DEGs were commonly expressed in the grain tissue involved LOB domain-containing proteins. These common genes can be considered important N responsive genes.

To reveal the high N responsive genes, the top 10 up-regulated and top 10 down-regulated DEGs were selected in three cultivars. The log2 fold change value of each group is shown in **Supplementary Tables 5**, **6** for the second leaf and **Supplementary Tables 7**, **8** for grain. In the second leaf tissue, the stress-associated glutathione S-transferase (GST), RADIALIS-like TF, and plasma membrane ATPase were the most N responsive up-regulated DEGs in Mace. In Spitfire, the top N responsive up-regulated DEGs were isocitrate lyase, laccase, and 11S globulin seed storage protein 2 related to carbon metabolism, lignin metabolism, and nutrient reservoir, respectively, whereas in Volcani, 1-phosphatidylinositol-3-phosphate 5-kinase, caleosin, protein transport protein Sec61 subunit gamma, and elongation factor G have appeared on top. In the grain tissue, the top up-regulated DEGs in Mace were isoaspartyl peptidase/L-asparaginase, plasma membrane ATPase, and trypsin family protein. The up-regulated DEGs in Spitfire showing high responsiveness to N stress were mainly N metabolism-related and aminotransferase like protein and aspartic proteinase nepenthesin, whereas in Volcani, invertase/pectin methyl esterase inhibitor family protein, cysteine proteinase inhibitor, and glycosyltransferase that is mainly associated with proteolysis and negative regulation of proton export across plasma membrane were found more prominent. There was a prevalence of defense-related down-regulated DEGs detected in the second leaf tissue of Mace, whereas photosynthesis-related DEGs were abundant in both Spitfire and Volcani. In the grain tissue of Mace, Spitfire, and Volcani, the down-regulated DEGs were predominantly related to proteolysis and N metabolism.

To select the genes that can be related to the tolerance to N starvation in high NUE cultivars, further analysis was done for the top genes using hierarchical clustering (see footnote 6).

TABLE 5 | Up- and down-regulated nitrogen metabolism-related DEGs identified under nitrogen stress in three wheat cultivars (Mace, Spitfire, and Volcani).

Gene id	Annotation	Mace		Spitfire		Volcani	
		0 DPA	10 DPA	0 DPA	10 DPA	0 DPA	10 DPA
TraesCS3D02G402300	Amino acid permease				2.1348		
TraesCS3D02G219200	Amino acid permease				2.1584	2.4518	
TraesCS3A02G407000	Amino acid permease				2.2959		
TraesCS7B02G093200	Amino acid permease				2.4131		
TraesCS6B02G314200	Amino acid permease				2.4322	2.662	
TraesCS7A02G188200	Amino acid permease				2.9323		
TraesCS6A02G285300	Amino acid permease				3.162	3.1234	
TraesCS3B02G441100	Amino acid permease			4.0062	3.2429		
TraesCS7D02G189000	Amino acid permease				3.4394		
TraesCS2D02G304300	Amino acid permease				4.126		
TraesCS2A02G305900	Amino acid permease				4.4484		
TraesCS3A02G406800	Amino acid permease				4.9359		
TraesCS3A02G388000	Amino acid permease	2.535		3.2765			
TraesCS3D02G381500	Amino acid permease					2.2199	
TraesCS1D02G264700	Amino acid permease				−2.9099		
TraesCS3D02G479700	Amino acid permease				−2.5016		
TraesCS3D02G381400	Amino acid permease						−2.0337
TraesCS3D02G229500	Amino acid transporter family protein				3.5398		
TraesCSU02G023100	Amino acid transporter family protein				−4.7472	−5.0356	
TraesCS6B02G382000	Amino acid transporter family protein				−1.6934	−2.0961	
TraesCS6D02G331400	Amino acid transporter family protein	−4.2987			1.4516		
TraesCS2B02G065200	Amino acid transporter family protein					−4.3865	
TraesCS7A02G312800	Amino acid transporter, putative				2.4193	1.9703	
TraesCS7D02G309200	Amino acid transporter, putative				2.4433		
TraesCS3D02G363200	Amino acid transporter, putative				2.849		
TraesCS7B02G212500	Amino acid transporter, putative				2.9047		
TraesCS7D02G309400	Amino acid transporter, putative				2.9901		
TraesCS7B02G213000	Amino acid transporter, putative				2.9956		
TraesCS5A02G472600	Amino acid transporter, putative				3.2717	2.3759	
TraesCS3A02G370300	Amino acid transporter, putative				3.3707		
TraesCS2D02G595500	Amino acid transporter, putative				Inf		
TraesCS5A02G472500	Amino acid transporter, putative	−2.7256	−2.4211		−2.385		−4.416
TraesCS5D02G485100	Amino acid transporter, putative	−2.8538					
TraesCS3B02G031700	Amino acid transporter, putative					−2.6345	
TraesCS5D02G180000	Amino acid transporter, putative					−4.5107	
TraesCS4A02G352900	Ammonium transporter				2.7949		2.1202
TraesCS5D02G519400	Ammonium transporter				3.0723	2.6766	
TraesCS5B02G520200	Ammonium transporter				3.4314		
TraesCS2A02G035700	Arginase				2.4832		
TraesCS5D02G134500	Cationic amino acid transporter		2.7066		2.1466		
TraesCS5A02G126900	Cationic amino acid transporter		2.3413		3.0414		
TraesCS5B02G126000	Cationic amino acid transporter		2.7743		3.1056		
TraesCS5D02G031800	Cationic amino acid transporter, putative				2.0679		
TraesCS5A02G025400	Cationic amino acid transporter, putative				2.3862		
TraesCS5A02G375600	Cationic amino acid transporter, putative				3.2084		
TraesCS2A02G389900	Glutamate dehydrogenase				3.1899		
TraesCS2D02G388800	Glutamate dehydrogenase				3.7682		
TraesCS2B02G409300	Glutamate dehydrogenase				4.1625		
TraesCS4A02G063800	Glutamine synthetase				2.0725		
TraesCS4B02G240900	Glutamine synthetase				2.2295		
TraesCS6D02G065600	Glutamine synthetase, putative, expressed				−2.0463		

(Continued)

TABLE 5 | Continued

Gene id	Annotation	Mace		Spitfire		Volcani	
		0 DPA	10 DPA	0 DPA	10 DPA	0 DPA	10 DPA
TraesCS2A02G186400	GMP synthase (glutamine-hydrolyzing)				2.7531		
TraesCS6D02G193200	High affinity nitrate transporter				3.0699		
TraesCS3D02G067500	Isoaspartyl peptidase/L-asparaginase				2.3634		−1.7908
TraesCS3B02G088700LC	Nitrate reductase (NADH) 1				−3.1147		
TraesCS3B02G218400LC	Nitrate transporter 1				−2.3208		
TraesCS7B02G201900	Nitrate transporter 1.1				2.3595		
TraesCS1D02G032700	Nitrate transporter 1.1				−2.9002		−2.8223
TraesCS1D02G214300	Nitrate transporter 1.1				−2.4916	−2.4339	
TraesCS1A02G211000	Nitrate transporter 1.1					−3.7814	−2.7462
TraesCS1B02G225000	Nitrate transporter 1.1					−2.4696	−2.9974
TraesCS7D02G357300	Nitrate transporter 1.2				2.3233		
TraesCS5D02G067100	Nitrate transporter 1.2				−3.2591	−3.852	
TraesCS6B02G364600	Nitrite reductase				−2.0746		
TraesCS5D02G012500	NRT1/PTR family protein 2.2				2.2733		
TraesCS4A02G283900	NRT1/PTR family protein 2.2				2.2821		
TraesCS5B02G039100	NRT1/PTR family protein 2.2				−2.6357		
TraesCS1D02G256700	Protein NRT1/PTR FAMILY 1.1				6.4103		
TraesCS7A02G206400	Protein NRT1/PTR FAMILY 1.1	−2.8104			1.0641		
TraesCS6D02G260500	Protein NRT1/PTR FAMILY 5.1				2.4435		
TraesCS3D02G375800	Protein NRT1/PTR FAMILY 5.5				2.8288	3.6607	
TraesCS3A02G382400	Protein NRT1/PTR FAMILY 5.5				−2.735		
TraesCS3D02G375200	Protein NRT1/PTR FAMILY 5.5				−2.3129		
TraesCS3D02G375500	Protein NRT1/PTR FAMILY 5.5				−2.2639		
TraesCS5D02G498700	Protein NRT1/PTR FAMILY 5.5				−2.2198		
TraesCS3A02G382900	Protein NRT1/PTR FAMILY 5.5		−2.3531				
TraesCS3D02G093300	Tryptophan aminotransferase				−3.4654		−2.4512
TraesCS3A02G093000	Tryptophan aminotransferase				−3.1087		−3.515
TraesCS3D02G246700	Tryptophan aminotransferase						2.0486
TraesCS6D02G512700LC	Tyrosine aminotransferase				5.7401		

The green highlighted values stand for expressional change strength when log2 fold change >2.0 whereas brown when log2 fold change < −2.0. The intensity of color increases with the increase in degree strength. Inf is an indication of differential up-regulation strength when the respective gene only expressed at low N condition. Differentially expressed genes (DEGs), 0 days post anthesis (0 DPA), 10 days post anthesis (10 DPA).

The top genes that showed high read count at low N in a high NUE cultivar (Mace) can be related to its tolerance to N stress. In the second leaf tissue, the top up-regulated DEGs with high abundance were principally identified as RADIALIS-like TFs, GST, and PLAC8 family protein (**Figure 8**). Similarly, in the grain tissue, plasma membrane ATPase, isoaspartyl peptidase/L-asparaginase, and alpha-galactosidase were identified as the top up-regulated DEGs expressed abundantly in high NUE cultivar Mace (**Figure 9**).

DISCUSSION

To improve NUE, it is important to understand the plant response to N treatments, especially to N limitation at both physiological and transcriptome levels. Targeting improved GPC and GY, the present study aimed to explore the transcriptome response of wheat to long-term N stress and identify potential candidate genes that are differentially expressed with high relative abundance across different genotypes in common. According to previous study, the GY of Mace is higher than those of Spitfire and Volcani, whereas the GPC of Mace is relatively lower than those of Spitfire and Volcani. It was also reported that the GY and GPC of Spitfire are affected more negatively under N-limiting conditions (Alhabbar et al., 2018a). Thus, it is important to unravel the underlying genes that can contribute to N stress tolerance for further genetic manipulation study.

Inadequate supply of N negatively affects plant morphology, limits growth, and decreases biomass in wheat (van der Werf et al., 1993). Most plants exhibit prominent changes in their growth and development under N-stressed conditions. Previous studies reported that adaptations of plants with nutrient-stressed conditions are mainly dependent on morphological changes (Wang J. et al., 2019; Zhao et al., 2005). The results of this study also confirmed that low N stress inhibited wheat growth, with significant negative impact on different phenotypes (**Figure 1**). These results were consistent with the N stress studies in wheat (Curci et al., 2017), sorghum (Gelli et al., 2014), corn

TABLE 6 | Top common up-regulated DEGs identified among wheat cultivars Mace, Spitfire, and Volcani under nitrogen stress.

Tissue	Gene ID	Gene description	Mace		Spitfire		Volcani	
			0 DPA	10 DPA	0 DPA	10 DPA	0 DPA	10 DPA
Leaf	TraesCS2B02G356100	Aldo/keto reductase family protein		3.6628	6.9364			3.5198
	TraesCS2D02G507800	Nuclease S1		3.4537	5.5328	5.7459	5.7088	
	TraesCS6A02G369800	Alcohol dehydrogenase, putative	2.7501			2.5822		2.9478
	TraesCS6B02G406000	Alcohol dehydrogenase, putative	1.6613			2.579		2.2855
	TraesCS6B02G406300	Alcohol dehydrogenase, putative	3.37			3.7911		4.8851
	TraesCS6D02G353300	Alcohol dehydrogenase, putative	2.2798			1.7571		3.9739
	TraesCS7B02G128800	Epoxide hydrolase 2		2.5759		1.3931		4.7394
	TraesCS7D02G488800	Fatty acid hydroxylase superfamily	1.998			2.1748		3.9123
	TraesCS3A02G276800	Glutamate carboxypeptidase 2		2.2004	2.5489	1.6163	4.0203	
	TraesCS3B02G311000	Glutamate carboxypeptidase 2		2.1156	3.2494	1.6598	4.6158	
	TraesCS2D02G379000	L-allo-threonine aldolase		2.1093		2.4275		3.7648
	TraesCS3D02G496500	PLAC8 family protein		3.5915		2.1039	3.3148	
	TraesCS7D02G084100	Sulfate transporter		2.2596		3.4823		2.0942
	TraesCSU02G455900LC	Zinc finger CCCH domain-containing protein 8		1.9752		2.731	2.0028	

The green highlighted values stand for expressional change in strength when log2 fold change >2.0. The intensity of color increases with an increase in the degree of strength. No common up-regulated DEGs among the three cultivars were found in grain tissue. Differentially expressed genes (DEGs), 0 days post anthesis (0 DPA), 10 days post anthesis (10 DPA).

(Jin et al., 2015), and rice (Sinha et al., 2018). In Mace, GY and GPC were less affected by N stress compared to those in Spitfire and Volcani in glasshouse conditions, although under high N conditions, the GPC of Spitfire and Volcani was higher than that of Mace. In agreement with the previous study (Alhabbar et al., 2018b), this study further confirms that Mace is more tolerant of N stress. It was reported that the GY in maize was decreased by 38% with the change in N treatment from high to low (Gallais and Hirel, 2004), which can be associated with the interrupted synthesis of chlorophyll and photosynthesis performance (McCullough et al., 1994). Many studies also reported the influence of hormones and N metabolism- and nutrient stress-related genes on agronomic traits (Singh et al., 1973; Cai et al., 2009). Thus, it is predicted that under a N stress condition, many genes involved in different biological pathways are cross-talking in mitigating the adverse effect of stress instead of a single factor. However, the GPC and the number of days to flower were less affected by N stress, which explains that these parameters can be rather controlled by genotype.

Under N stress, the genes that expressed differentially were mostly leaf specific compared to grain. Also, the DEGs in the leaf were related to versatile functions, whereas a significant percentage of DEGs in the grain were related to transport and N metabolism. The 50 common DEGs between the second leaf and grain identified were mostly related to defense, amino acid metabolism, N metabolism, carbohydrate metabolism, and sulfate transport. It is known that in the plastid of the leaf, sulfate is converted to sulfide using the reducing power of photosynthesis and incorporated into amino acids that later remobilize to developing seeds (Gallardo et al., 2014; Jobe et al., 2019). Developing seeds requires sulfur amino acids to synthesize storage protein to secure germination for the next generation (Leustek et al., 2000; Saito, 2000). DEG analysis also showed a higher number of DEGs in Spitfire (10,535 in the second leaf and 252 in grain) in comparison to Volcani (1671 in the second leaf and 16 in grain) and Mace (699 in the second leaf and 25 in grain), which indicates that under N stress, Spitfire responds more actively, and that involves more signaling pathways than Volcani and Mace. Spitfire responded to N stress mostly by up-regulating, whereas Mace and Volcani responded by down-regulating the DEGs.

Precedence of any biological processes at a particular developmental stage is correlated with the changes in the expression pattern of corresponding genes involved. GO enrichment analysis is an effective method to understand the key biological processes participating in adapting stress. For instance, a N starvation study in durum wheat reported N compound metabolism, carbon metabolism, and photosynthesis as the top enriched biological processes (Curci et al., 2017). The oxidation–reduction process and metabolic process were top enriched biological processes in wheat seedlings in response to N limitation (Wang J. et al., 2019). The top enriched biological processes in rice have been reported to be associated with metabolic processes, cellular processes, and transport under N-starved conditions (Yang S. Y. et al., 2015). This study showed that the up-regulated DEGs were mainly associated with transmembrane transport, whereas the down-regulated DEGs were mainly associated with metabolic process and stress response, which supports that during grain filling, the plant increases its overall remobilization through protein degradation and transport (McCullough et al., 1994; Masclaux-Daubresse et al., 2008). Significant up-regulation of transmembrane transport, nitrogenous compound transport, and proteolysis was common in all three cultivars (**Figure 5**). However, in Mace, a greater percentage of up-regulated DEGs were related to DNA conformation change and sulfate assimilation, whereas in Spitfire and Volcani, DEGs were highly significant in the lipid metabolic process. The up-regulation of DNA metabolic

TABLE 7 | Top common down-regulated DEGs identified among wheat cultivars Mace, Spitfire, and Volcani under nitrogen stress.

Tissue	Gene ID	Gene description	Mace		Spitfire		Volcani	
			0 DPA	10 DPA	0 DPA	10 DPA	0 DPA	10 DPA
Leaf	TraesCS5A02G472500	Amino acid transporter, putative	−2.7256	−2.4211		−2.385		−4.416
	TraesCS4B02G363700	Apyrase	−2.0965	−2.7293		−4.2829	−2.2361	
	TraesCS2D02G555300	ARM repeat superfamily protein	−3.0653	−3.3102		−3.2171	−3.5676	−2.7152
	TraesCS3B02G191100	Branched-chain amino acid aminotransferase-like	−3.3945			−2.6334	−5.7474	
	TraesCS7A02G068600	Caffeoyl-CoA O-methyltransferase	−3.0217			−4.6356	−2.7735	
	TraesCS3D02G150300LC	calcium-dependent protein kinase 15	−5.9832		−4.3883	−4.219	−4.3632	
	TraesCS7D02G450300LC	CASP-like protein	−2.5774			−2.4122	−3.3004	
	TraesCS6D02G048300	Catalase	−3.7606			−1.9888		−2.5888
	TraesCS6B02G056800	Catalase	−5.2287			−2.4917		−2.5875
	TraesCS7D02G516800	Chaperone protein dnaJ		−2.0382		−3.0128		−1.8551
	TraesCS5D02G464800	Chlorophyll a-b binding protein, chloroplastic	−7.1186			−6.6902	−7.2124	−5.7018
	TraesCS5A02G350600	Chlorophyll a-b binding protein, chloroplastic	−2.0693			−3.4473	−2.5948	−2.6866
	TraesCS5B02G353200	Chlorophyll a-b binding protein, chloroplastic	−1.7397			−4.9486	−4.2498	
	TraesCS5D02G357600	Chlorophyll a-b binding protein, chloroplastic	−2.4064			−3.6449	−2.6951	−3.6498
	TraesCS4A02G099000	Cysteine-rich receptor-kinase-like protein		−2.7975		−2.3484	−5.2356	
	TraesCS7A02G271200	Cytochrome P450 family protein	−1.6183			−2.1064		−1.9589
	TraesCS7B02G455000	Cytokinin oxidase/dehydrogenase	−2.4797			−3.452		−2.9646
	TraesCS2B02G336700	Cytoplasmic dynein 2 heavy chain 1	−5.165			HN	HN	
	TraesCS2A02G143200	Gibberellin receptor GID1A	−5.7235	−4.3087		−2.9138	−5.0713	
	TraesCS2D02G146500	Gibberellin receptor GID1A	−2.5369	−1.8515		−2.4214	−3.4673	
	TraesCS3B02G022900	Glutamate decarboxylase		−2.0427		−2.9517	−2.1619	
	TraesCS6B02G051800	Glycerol-3-phosphate acyltransferase		−2.0549		−2.1339	−3.2053	−2.6867
	TraesCS3A02G439500LC	Glycine-rich protein A3	−5.5826		−4.5188	−5.2864	−5.816	−3.8105
	TraesCS3D02G150400LC	Glycine-rich protein A3	−4.9171		−3.2827	−5.2365	−5.4414	−4.3841
	TraesCS1D02G375700	Methyltransferase	−3.0195			−6.6682	−7.8387	−6.7881
	TraesCS1B02G388700	Methyltransferase		−4.9876		−7.4915	−8.3664	−6.0313
	TraesCS2D02G258800	N-succinylglutamate 5-semialdehyde dehydrogenase	−2.3577			−2.3847		−2.5793
	TraesCS3D02G316900LC	Nucleoside triphosphatase I		−4.1874		HN	−4.6189	−4.4166
	TraesCS6B02G412100	Photosystem II 10 kDa polypeptide family protein	−8.0008			−8.3988	−11.599	
	TraesCS6D02G358900	Photosystem II 10 kDa polypeptide family protein	−7.9982			−8.6648	−10.592	
	TraesCS6A02G374400	Photosystem II 10 kDa polypeptide family protein	−8.3938			−7.2319	−8.4213	
	TraesCS4D02G203800	Plant protein 1589 of uncharacterized protein function	−6.9524	HN		HN	HN	
	TraesCS1D02G286100	Plant protein 1589 of uncharacterized protein function	−4.1266			−3.3077	−4.1694	−4.9212
	TraesCS1A02G287200	Plant protein 1589 of uncharacterized protein function	−2.8192			−2.7703	−3.9601	−4.9727
	TraesCS4A02G245300	Protein DETOXIFICATION	−4.1621		−2.5502	−6.7866	−6.9194	
	TraesCS3D02G144900	Protein DJ-1		−3.237		−5.0529		−3.295
	TraesCS7B02G454000	RNAse THREE-like protein 3		−2.1126		−2.2225	−1.9664	
	TraesCS5A02G018000	Thaumatin-like protein	−2.6156	−2.7996		−4.5464	−4.4694	
	TraesCS7D02G388400	Tryptophan synthase beta chain	−2.7685	−1.9753		−1.4694	−3.3613	−2.9456
	TraesCS2B02G277300	Two-component response regulator		−2.8394		−3.5708		−2.7784
	TraesCS2D02G259200	Two-component response regulator	−2.4252	−2.2783		−2.7094		−1.9826
	TraesCS1B02G196200	U-box domain-containing protein 4	−1.8516			−1.8029		−3.0021
Grain	TraesCS2D02G193400	LOB domain-containing protein		−1.8922		−2.1909		−1.5747
	TraesCS2A02G194500	LOB domain-containing protein, putative		−1.6826		−2.2599		−1.3364

The brown highlighted values stand for expressional change in strength when log2 fold change < −2.0. The intensity of color increases with an increase in the degree of strength. Differentially expressed genes (DEGs), 0 days post anthesis (0 DPA), 10 days post anthesis (10 DPA).

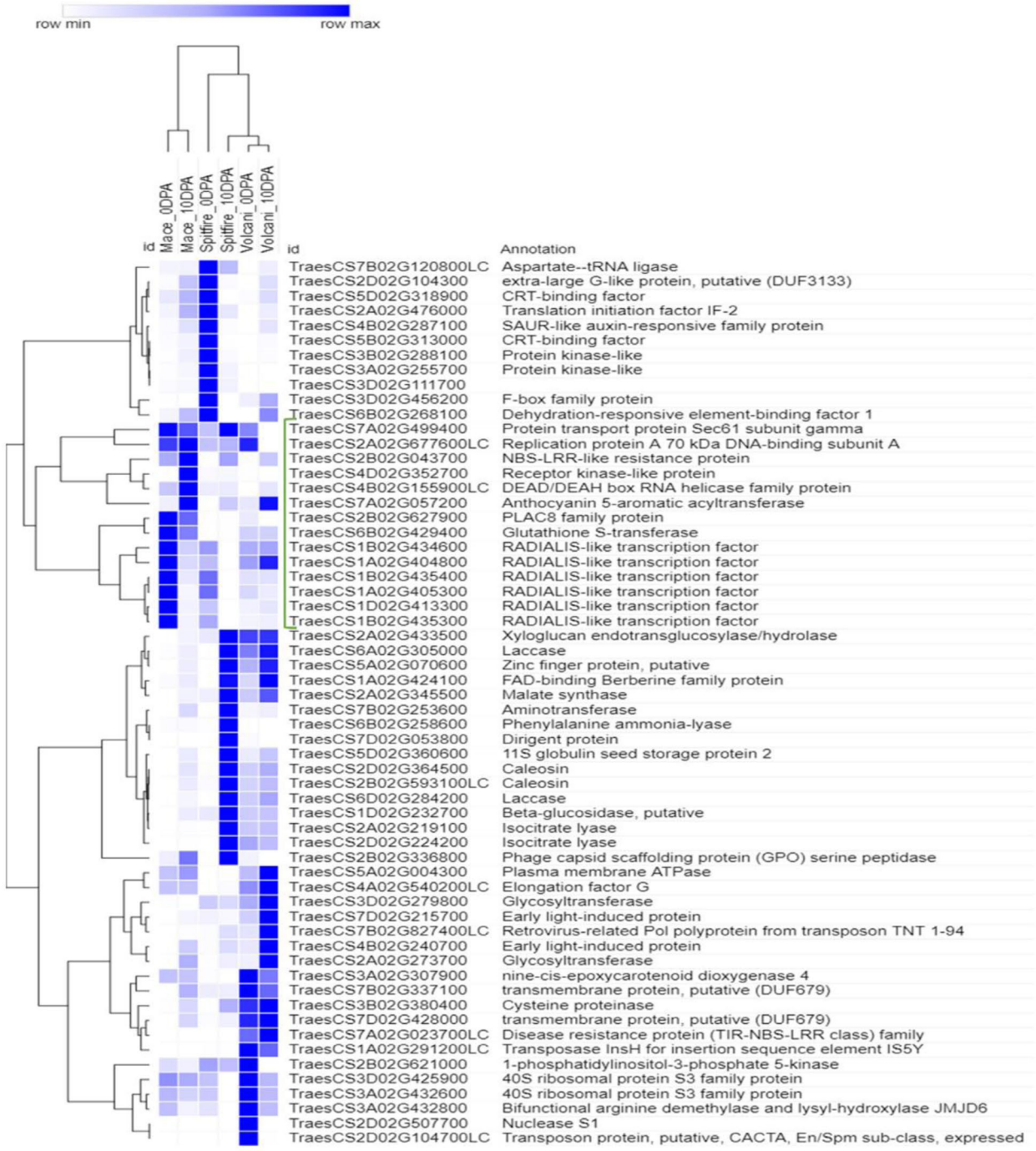

FIGURE 8 | Hierarchical clustering of top up-regulated DEGs in leaf tissue. The green bracketed genes are the highly expressed up-regulated genes in high NUE wheat cultivar Mace. Differentially expressed genes (DEGs), 0 days post anthesis (0 DPA), 10 days post anthesis (10 DPA).

process in Mace can be related to epigenetic change, which underlies its stability under N stress conditions. A long-term primed state of the epigenetic mechanism involves DNA conformation change such as change in chromatin structure, variation in composition and position of the nucleosome, and post-transcriptional modification to cope more efficiently with the subsequent stress (Chinnusamy and Zhu, 2009). Also, the increase in sulfate assimilation in Mace can be related to the synthesis of proteins rich in S-containing amino acids such as glutathione, which is a major component of the stress response (Yamaguchi et al., 1999; Kopriva and Rennenberg, 2004). In Spitfire and Volcani, the increase in lipid metabolism can be

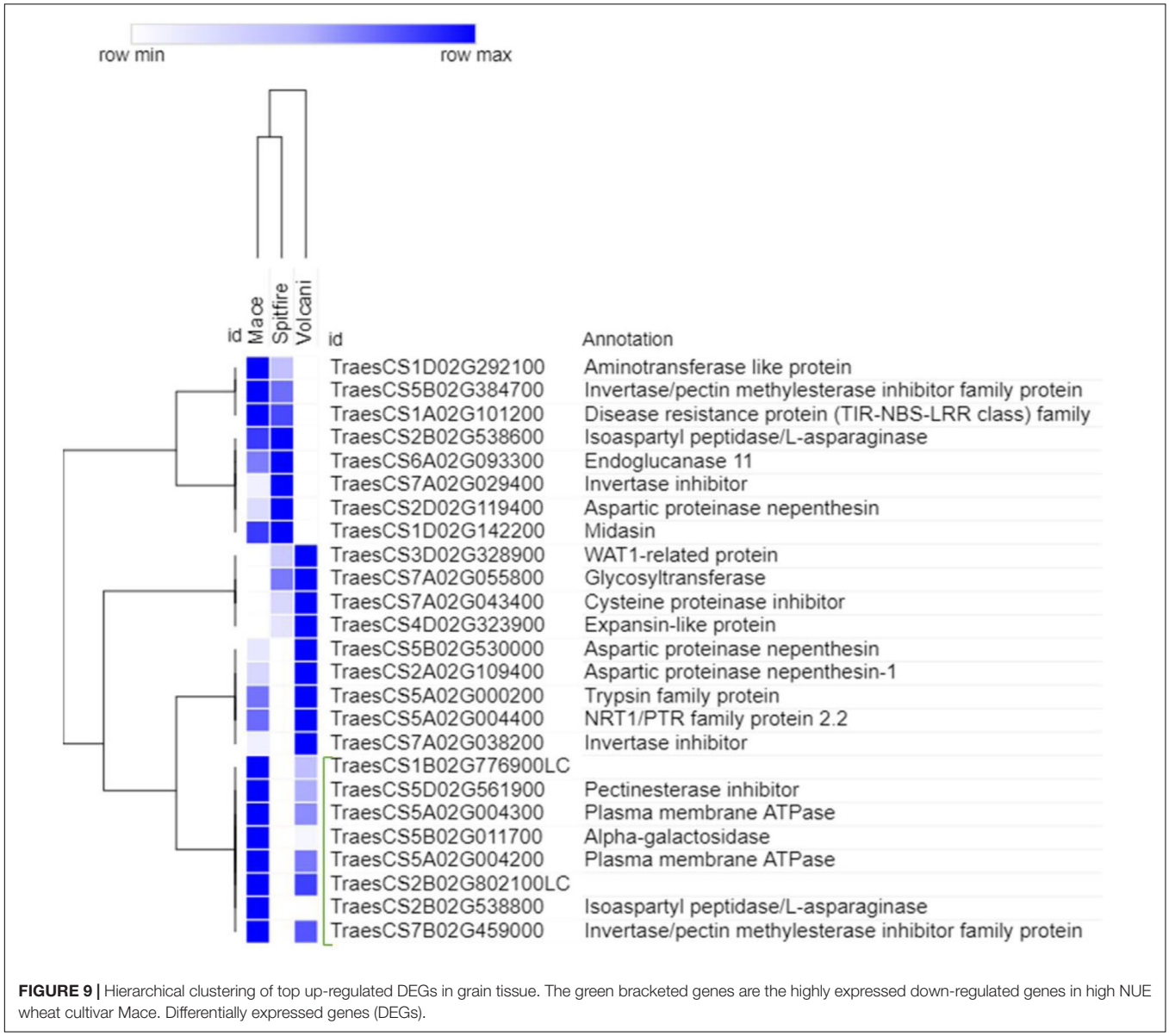

FIGURE 9 | Hierarchical clustering of top up-regulated DEGs in grain tissue. The green bracketed genes are the highly expressed down-regulated genes in high NUE wheat cultivar Mace. Differentially expressed genes (DEGs).

related to senescence (Xiao and Chye, 2011). It was reported that, during senescence, synthesis of phytyl-ester synthase is induced, which is associated with the synthesis of triglycerol and phytyl-esters of plastid fatty acids (Xiao and Chye, 2011; Troncoso-Ponce et al., 2013). In contrast to the up-regulated DEGs, a higher percentage of down-regulated DEGs in Mace were significantly related to the cellular catabolic process, which is known to be related to plant biotic and abiotic stress response (Tavladoraki et al., 2012), whereas in Spitfire and Volcani, the DEGs were more abundant in photosynthesis. The significantly decreased expression of photosynthesis-related DEGs can be related to decreased grain weight per plant in Spitfire and Volcani (Zhao et al., 2005; Boussadia et al., 2010). However, in all three cultivars, the down-regulated DEGs were more prominent in the carbohydrate metabolic process, which indicated that, regardless of genotypes, N stress can negatively affect plant carbohydrate metabolism (Rufty et al., 1988) and plants adapted to N stress by down-regulating the expression of many genes of this kind.

Kyoto Encyclopedia of Genes and Genomes analysis results also revealed that in all three cultivars, DEGs were involved in phenylpropanoid biosynthesis, biosynthesis of secondary metabolites, flavonoid biosynthesis, and sucrose and starch metabolism. The regulation of these genes in stress adaptation has been reported in several studies (Dixon and Paiva, 1995; Huang et al., 2010; Akula and Ravishankar, 2011; Petrussa et al., 2013). However, some cultivar-specific differences highlighted the importance of genetic variability in stress response (**Tables 2**, **3**). For example, DEGs were more abundantly related to MAPK signaling in Mace and Spitfire, plant hormone signal transduction, glutathione metabolism in Mace and Volcani, photosynthesis-antenna proteins, and circadian rhythm in Spitfire and Volcani. Also, some DEGs that were

significantly abundant in pathways related to zeatin biosynthesis in Mace, terpenoid biosynthesis in Volcani, and plant–pathogen interaction in Spitfire are important to identify the underlying genes related to biological pathways to develop stress-tolerant cultivars (Cheong et al., 2002; Vickers et al., 2009).

Protein–protein interaction analysis has been used to identify DEGs that are interacting in different biological processes such as photosynthesis, photosynthesis-antenna proteins, and steroid biosynthesis. Photosynthesis is the vital biological process by which plants absorb light energy and assimilate CO_2 to produce dry matter and comprises reactions that are regulated by proteins in the chloroplast (Chandna and Ahmad, 2015). Within this highly interactive and regulated system, change in one component can cause changes to other components. The strength of photosynthesis capacity is mainly dependent on the N content of chloroplasts in the leaf (Evans, 1989; Evans and Poorter, 2001; Ripullone et al., 2003). Numerous studies have reported that N significantly affects photosynthesis (Wei et al., 2016; Lin et al., 2017) through its association with the light reaction in the chloroplast and/or the dark reaction (Sage et al., 1988; Von Caemmerer, 2000). The light-harvesting complex (LHC) comprises chlorophylls a and b and the chlorophyll a-b binding protein and is closely associated with photosystem I and II. LHC plays an important role as a light receptor that captures and delivers the excitation energy between two photosystems and adjusts the distribution of excitation energy by being phosphorylated reversely under changing light conditions (Sage et al., 1988). The PSII outer antenna LHCB proteins are important components of the major LHC, and they consist of minor antenna complexes LHCB4 (CP29), LHCB5 (CP26), and LHCB6 (CP24) and major antenna complexes that comprise homo- and heterotrimers of LHCB1, LHCB2, and LHCB3 (Jansson, 1994, 1999). In the present study, all the chlorophyll a-b binding proteins that interacted with each other in adjusting N stress were down-regulated. In agreement with the study in rice seedlings in a water-stressed environment (Dalal and Tripathy, 2018), the current study identified significantly decreased expression of components of LHCs of both PSII and PSI (**Figure 10** and **Supplementary Table 9**). Moreover, the decreased photosynthesis rate and chlorophyll content under N-stressed condition (**Figure 1**) can be related to the differential expression of chlorophyll a-b binding proteins.

The rate of photosynthesis has an intense positive correlation with N status in soil (Makino et al., 2003; Nunes-Nesi et al., 2010). Under N stress, a plant might adapt by reduced chloroplast surface area and a decreased light energy absorption, which can affect photosynthesis negatively (Li et al., 2009, 2013; Muller et al., 2009; Georgieva et al., 2010). In the present study, many PSII and PSI subunits showed a decreased expression in low NUE cultivars Spitfire and Volcani under N-stressed condition (**Figure 11** and **Supplementary Table 10**), which can impede photosystem repair and photosynthetic electron transport chain function (Foyer and Shigeoka, 2011). Also, the expression of cyt559 had decreased, which binds most of the cofactors in the photocatalytic activity of photosystem II. Among the down-regulated DEGs of PSII components, the core components PsbO, PsbP, and PsbQ are known to be involved in the water oxidation and its optimization process (Bricker et al., 2012). PsaK is associated with the LHCI antenna system, and PsaO plays a role in the formation of the docking site for LHCII binding to PSI (Jensen et al., 2007). The down-regulated PetC provides resistance to photo-oxidative damages by contributing to the thermal intemperance of light energy and lumenal acidification and mediates electron transfer between PSII and PSI (Munekage et al., 2001). The photosynthetic electron transport component showed down-regulation for PetE that participates in electron transfer between P700 and the cytochrome b6-f complex in photosystem I and PetF (ferredoxins are iron–sulfur proteins) transfer electrons in a wide variety of metabolic reactions (Achard et al., 2008). PetH plays a significant role in balancing cyclic and noncyclic electron flow to supply the ATP and reducing power required by the plant (Claeys et al., 2012). Moreover, F-type ATPase gamma and a c subunit aid electron transport in both photosystems I and II were also significantly down-regulated. Significantly down-regulated DEGs in low NUE cultivars Spitfire and Volcani were found as the components of the LHC system and PSI and PSII, underlying their molecular basis of low GY mechanisms. Therefore, understanding N stress-responsive DEGs that participated in photosynthesis might provide a base to improve the photoprotection capacity to sustain photosynthesis as well as improving plant N-stress tolerance.

The up- and down-regulated DEGs in relation to steroid hormone biosynthesis, specifically the BR biosynthesis, in low NUE cultivars also lead to understanding the role of this hormone in N stress adaptation (**Figure 12** and **Supplementary Table 11**). Down-regulated DEGs were more prominent compared to up-regulated DEGs in this pathway, which indicates the declined BR hormone biosynthesis. Previous studies showed exogenous application of BR enhanced photosynthesis under stress conditions (Niu et al., 2016; Shu et al., 2016). Chlorophyll is an important parameter and is commonly used to measure photosynthetic activity. However, chlorophyll is highly sensitive and responds to stress by decreasing the chlorophyll a, b content in leaves (Rehman et al., 2016). In low NUE cultivar Spitfire and medium NUE cultivar Volcani, significantly decreased level of chlorophyll a-b binding protein can be associated with their reduced chlorophyll content compared to high NUE cultivar Mace (**Figure 1**). Previous studies also reported that BR can reducing the harmful effect of stress by activating the synthesis of antioxidants like glutathione reductase, catalase, peroxidase, etc., contributing to increase in yield and yield components (Hayat et al., 2000; Vardhini and Anjum, 2015; Anwar et al., 2018). In high NUE cultivar Mace, the absence of significant DEGs related to BR biosynthesis that are interacted at the protein level can be related to its increased tolerance to N stress, relatively high chlorophyll content, tiller number, and grain weight per plant (**Figure 1**). So far, no previous study has been reported on the putative role of BRs in wheat under N-stressed conditions. Thus, identifying the involvement of BR biosynthesis provides a suitable platform to explore the essential role of BR in N stress tolerance and further application of BRs to improve wheat production.

Through annotation of the transcriptome, several known and putatively N-metabolism-related genes were identified

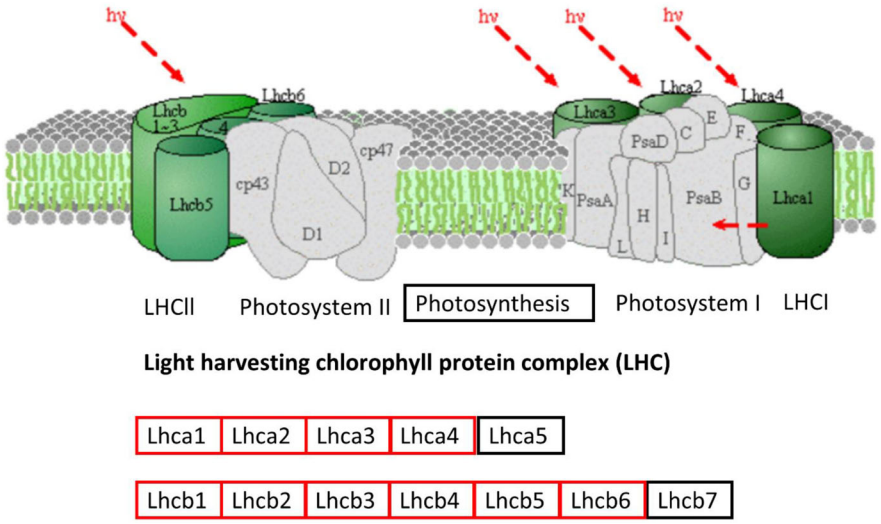

FIGURE 10 | Down-regulated protein orthologous (*Arabidopsis thaliana*) of differentially expressed genes in photosynthesis-antenna proteins pathways. The proteins in red are down-regulated. The pathway map was generated using KEGG (Kanehisa et al., 2016, 2017).

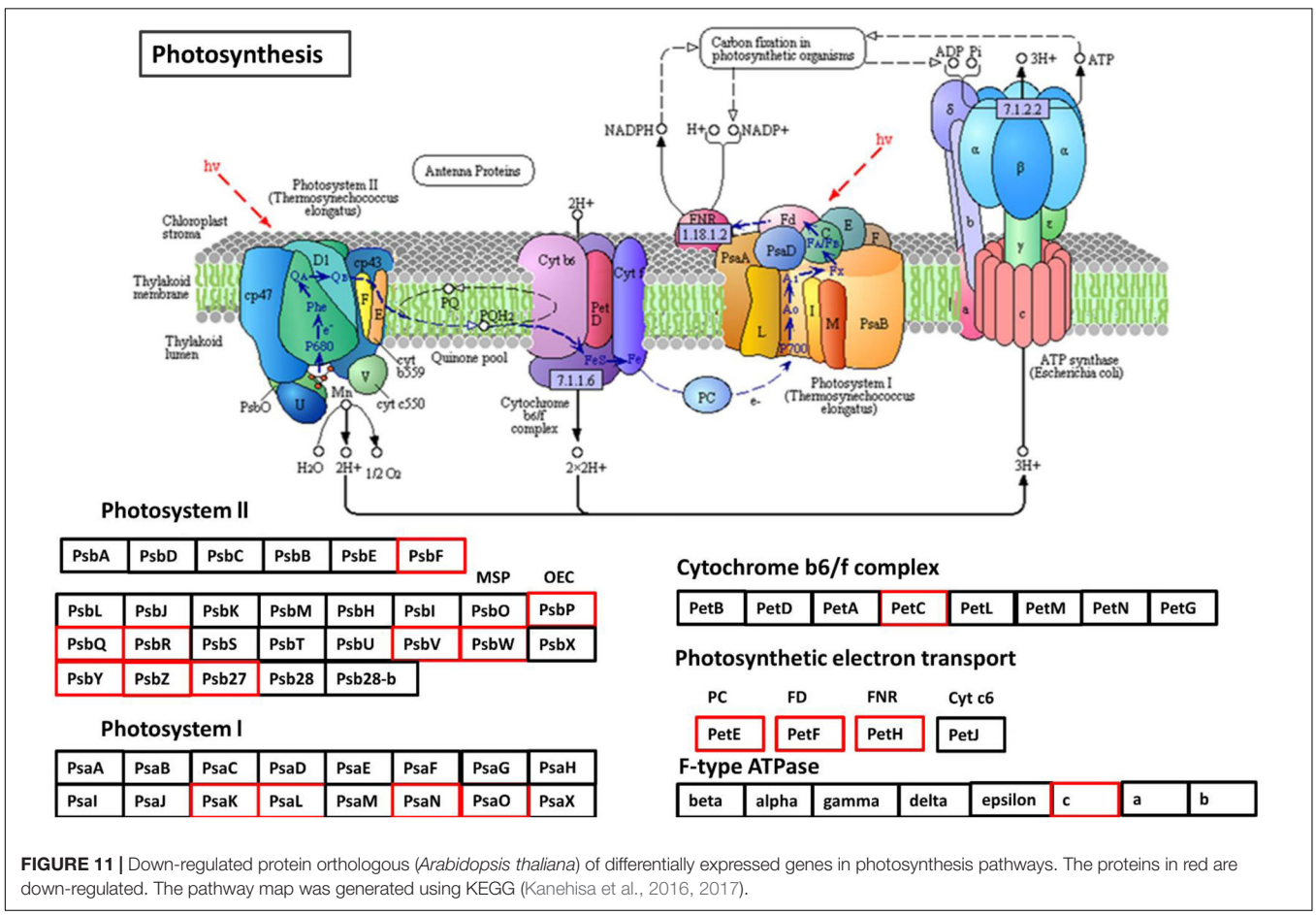

FIGURE 11 | Down-regulated protein orthologous (*Arabidopsis thaliana*) of differentially expressed genes in photosynthesis pathways. The proteins in red are down-regulated. The pathway map was generated using KEGG (Kanehisa et al., 2016, 2017).

both to be up- and down-regulated. Usually, N stress increases the expression of high-affinity transport systems for nitrate and ammonium (Crawford and Glass, 1998). Previous reports showed that high-affinity nitrate transporters were expressed in N-starved seedlings of Arabidopsis (Wang R. et al., 2003). In rice, the nitrate transporter (*OsNRT2.2*) in association

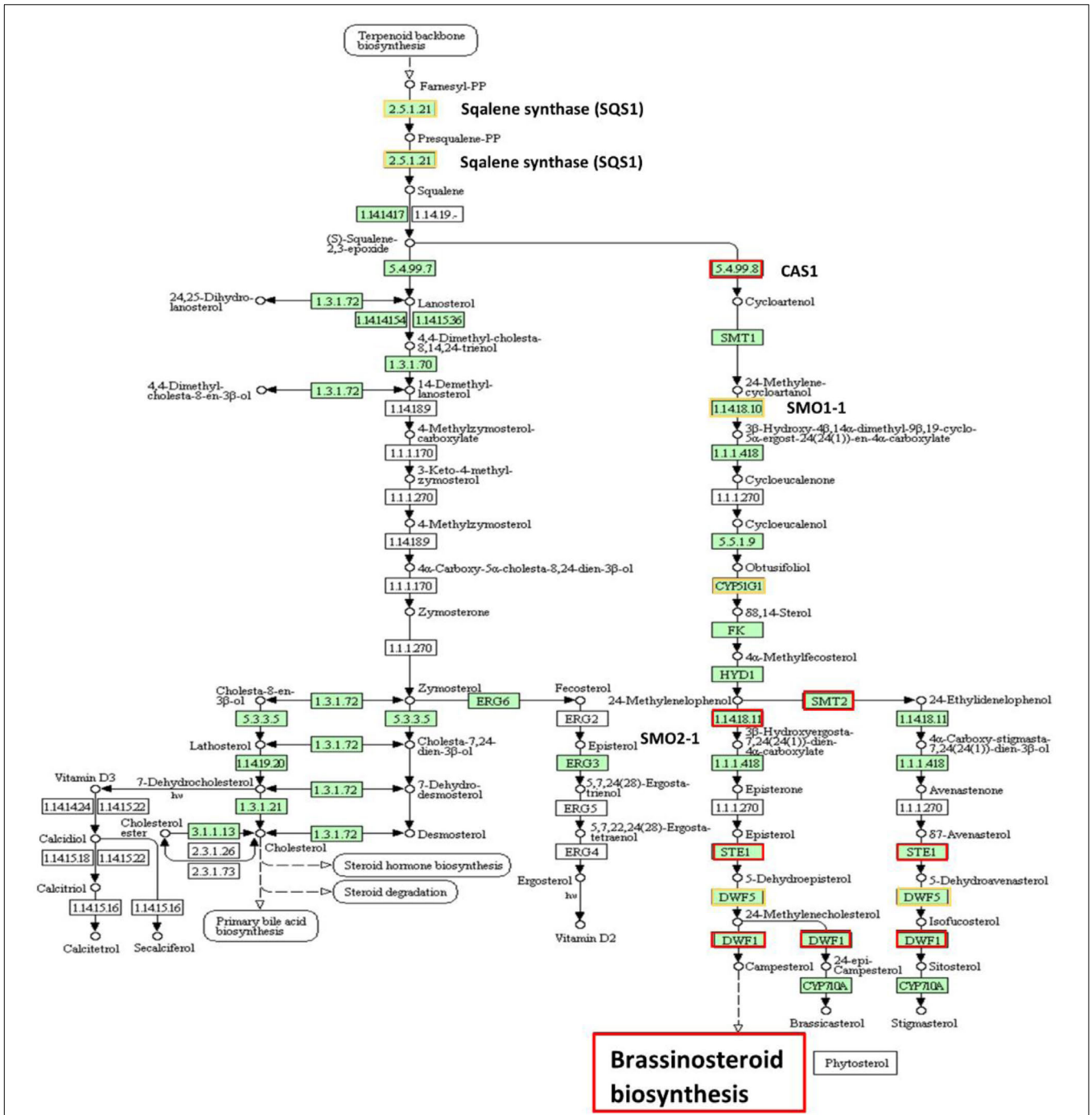

FIGURE 12 | Up- and down-regulated protein orthologous (*Arabidopsis thaliana*) of differentially expressed genes in steroid biosynthesis. The protein names bordered in yellow are up-regulated and red are down-regulated. The pathway map was generated using KEGG (Kanehisa et al., 2016, 2017).

with *OsNAR2.1* transports nitrate, which can promote the elongation of lateral roots (Feng et al., 2011; Li et al., 2007). In the current study, the expression of high-affinity nitrate transporters (NRT1/PTR family protein 2.2) was up-regulated under N stress. This indicates a more efficient N uptake under N-limited condition. On the other hand, the expression of most of the dual affinity nitrate transporters (like nitrate transporter 1.1) was decreased under N stress, which is known to regulate root and shoot growth (Mounier et al., 2014). The down-regulation of Protein NRT1/PTR FAMILY 5.5 and tryptophan aminotransferase can be related to the retarded growth and low GPC and GY of low N-treated plants (Won et al., 2011; Léran et al., 2015). Also, a decreased expression of nitrite reductase was observed, which is related to nitric oxide (NO) homeostasis

(Chamizo-Ampudia et al., 2017). NO can act as a signaling molecule in plant immune response, defense-related gene expression, and the hypersensitive response mechanism (Mur et al., 2012). Conversely, most of the N metabolism-related DEGs of the known and putative amino acid permease, amino acid transporter, ammonium transporter, glutamate dehydrogenase, glutamine synthase family and tyrosine aminotransferase, and tryptophan aminotransferase were up-regulated. Glutamine synthetase is a key enzyme (Cai et al., 2009) that catalyzes the conversion of glutamate (Glu) to glutamine (Gln). GOGAT is involved in the transfer of the amide group of Gln to a-ketoglutarate (2-OG) to subsequently produce Glu (Cren and Hirel, 1999). Gln is involved in the biosynthesis of organic nitrogenous compounds, such as amino acids, nucleotides, and chlorophyll, and plays a major role in regulating plant N assimilation in grain production (Martin et al., 2006; Gadaleta et al., 2011). In this study, most of the significant DEGs related to N metabolism were found in Spitfire, and the lowest number of DEGs with a significant change in expression was identified in Mace. The high abundance of a nitrate transporter and ammonium assimilatory gene abundance in low NUE cultivars Spitfire and Volcani can be related to adapt the N-stressed condition. Similar results were observed in a transcriptome study with sorghum, where N assimilator genes were abundant in sensitive and low NUE cultivars (Singh et al., 1973). The smaller number of N metabolism-related DEGs can be related to a better tolerance of Mace to low N conditions. These findings are also supported by similar outcomes in rice (Lian et al., 2006).

The common DEGs that were simultaneously induced or repressed under N stress across the three cultivars are also potentially important for N stress response. Among the common down-regulated DEGs (**Table 7**), photosystem II 10 kDa polypeptide family protein and chlorophyll a-b binding protein are related to photosynthesis and light harvesting, which are sensitive to stress (Rehman et al., 2016; Nowicka et al., 2018). The chlorophyll content was significantly decreased under the N-stressed condition in all three cultivars compared to high N. Down-regulation of stress-responsive DEGs like catalase, thaumatin-like protein, and cytochrome P450 family protein is also known to be related to stress adaptation (Cai et al., 2013; Alam and Ghosh, 2018). Also, the expression of phytohormone-related DEGs such as gibberellin receptor GID1A and cytokinin oxidase/dehydrogenase showed down-regulation common in all three cultivars. Reduced GA levels and signaling are known to be associated with restrained growth and development of plant by inducing accumulation of DELLA (Colebrook et al., 2014), known as positive regulators of N stress-induced anthocyanin accumulation (Zhang et al., 2017). Under salt stress, the DELLA mutant has been reported to be strongly correlated with plant growth, height, time to flowering, and stress tolerance (Achard et al., 2008). Other studies also showed that GA-induced DELLA has a positive effect on stress tolerance (Claeys et al., 2012). In this study, the reduced plant height and growth can be related to a reduction in GA. Other phytohormone cytokinins (CKs) can regulate plant developmental processes under stressed conditions (Rubio-Wilhelmi et al., 2011). Recent studies reported that CKs act as a long-distance messenger that signals the N status of the plant in regulating the nutrient uptake system (Rubio-Wilhelmi et al., 2011). Overexpression of CK degradation enzyme-CKX has been known to exhibit an increased drought and salinity tolerance (Schmülling et al., 2003; Nishiyama et al., 2012). In addition, cytokinin and gibberellin also influence photosynthesis under stressed conditions (Caers et al., 1985; Biemelt et al., 2004). Some other common down-regulated genes involved glycine-rich protein A3, which binds and stabilizes the stress-inducible transcripts (Sahi et al., 2007), methyltransferase related to epigenetic tolerance to stress through DNA methylation (Boyko and Kovalchuk, 2008), calcium-dependent protein kinase 15, which functions in long-term adaptive processes or plant development by facilitating cross-talk between different Ca^{2+}-mediated stress signaling pathways (Lee and Rudd, 2002; Schulz et al., 2013), and a two-component response regulator that plays a role in stress response by transducing extracellular signals to the cytoplasm through phosphotransfer between the two components (Urao et al., 1998). Amino acid metabolism-related genes like putative amino acid transporters and branched-chain amino acid aminotransferase-like proteins are common in the three cultivars and can also contribute to stress tolerance by down-regulating their expression (Good et al., 2007). Interestingly, three DEGs annotated as plant protein 1589 with uncharacterized function were all down-regulated, which are potentially important candidates for further study. The down-regulation of LOB domain-containing proteins that were common in the grain of the three cultivars was reported to be involved in lateral root formation (Liu et al., 2005; Yang W. et al., 2015). LOB domain-containing proteins are also known to control the BR hormone negatively in N metabolism as well as plant growth and development (Bell et al., 2012; Ma et al., 2017).

Some genes common in the second leaf of the three cultivars showed up-regulation (**Table 6**) that can facilitate tolerance to N stress to survive. Among the 14 up-regulated DEGs, 4 were annotated as putative alcohol dehydrogenase family proteins that were also reported to accumulate at an increased level under low-temperature stress in maize and rice (Christie et al., 1991). Two glutamate carboxypeptidase 2 were up-regulated that are known to negatively regulate drought and freezing stress and play a role in carbon and amino acid metabolism (Shi et al., 2013). Another up-regulated DEG annotated as PLAC8 family protein was reported to be involved in cadmium tolerance and accumulation, which can also be a good candidate to increase N stress tolerance (Wang F. et al., 2019). Also, a sulfate transporter was found to be up-regulated and has been previously reported as affected by N deficiency (Yu et al., 2018). As sulfur assimilation is important for the biosynthesis of S-containing amino acids that remobilize to develop seeds for storage protein synthesis, a sulfate transporter is worth further study under the context of NUE.

The DEGs that demonstrated a high expressional change due to N stress can be an important candidate for N stress response. Analyzing the expressional variation of genes across the cultivars with different NUE, this study proposed that the highly up-regulated genes expressed in high NUE cultivar Mace with high abundance can contribute to N stress tolerance. In the second leaf of Mace, the top abundantly expressed up-regulated DEGs (**Supplementary Table 9**) in Mace involved RADIALIS-like

(RADL) TFs, GST, and PLAC8 family protein. GST was reported to catalyze the glutathione-dependent detoxification reactions and the reduction of hydroperoxides. It also plays a role in protection against environmental stresses by binding and sequestrating secondary metabolites like flavonoids and phenolics (Tahkokorpi et al., 2007). In maize, *ZmGSTU1* can protect plant cells from oxidative stress damage through binding and conjugating porphyrinogens. Under stress conditions and during senescence, porphyrinogens leak from chloroplast to the cytosol and become oxidized to the lipophilic and phytotoxic protoporphyrin (Dixon et al., 2010; Lederer and Böger, 2005). Binding of GST to leaked porphyrinogens can prevent their auto-oxidation, protecting plant cells from oxidative stress (Lederer and Böger, 2003). RADL TFs are a subfamily of MYB-related genes containing a single SANT (*SWI3/ADA2/N-CoR/TFIIIB*)/MYB DNA-binding domain, which is highly homologous to the RADIALIS gene product of *Antirrhinum majus*. The Antirrhinum RADIALIS gene is involved in the regulation of floral asymmetry, and mutation of this gene results in a symmetrical (or radial) floral morphology (Baxter et al., 2007). In Arabidopsis, a RADL TF (*RSM1*) is implicated in controlling early photomorphogenesis (Hamaguchi et al., 2008). In rice, overexpression of RADL3 TF (*OsRL3*) exhibits a stay-green phenotype during dark-induced senescence in an ABA-dependent pathway (Park et al., 2018). A study in lady's bedstraw (*Galium verum*) to understand the genetic basis of morphological difference of its two variants showed that two short insertions in the promoter region of RADL1 in one variant can be related with its nonfunctionality and dwarfism (Jeong et al., 2014). In Barley, the effect of CENTRORADIALIS (CEN) on developmental timing and shoot and spike morphologies has been reported (Bi et al., 2019). No previous study has been conducted for the function of RADL TFs in wheat. In the current study, the expression of RADL TFs was different across the cultivars, and future study is necessary to reveal whether any genetic variation is present at the cultivar level of this gene family. Another top up-regulated abundantly expressed PLAC8 family protein can have conserved biochemical function due to its conserved core domain; however, specific functions of these family proteins are still unclear. In Arabidopsis, only PLAC8 domain-containing protein AtPCR and similar proteins in rice and other organisms (Song et al., 2011) are implicated in cadmium resistance. PLAC8-containing proteins can also control cell size and number (Frary et al., 2000) in plant. It is reported in yeast that cadmium tolerance can involve DNA repair (Di Vietro et al., 2014). In this study, as the PLAC8 family gene was up-regulated under N stress and high abundance was detected in a high NUE cultivar, it can be predicted that this gene may play an important role in N stress tolerance. Similarly, in the grain tissue of Mace, plasma membrane ATPase was identified as one of the top up-regulated DEGs (**Supplementary Table 7**), known to be induced in a condition that requires a greater transport activity and plays an important role in nutrient uptake (Janicka-Russak, 2011). Overexpression of plasma membrane ATPase is also associated with cadmium stress tolerance (Di Vietro et al., 2014). Another top up-regulated abundantly expressed DEG in Mace is a homolog of Arabidopsis (AT3G16150) K+-dependent L-asparaginase, which is associated with efficient metabolism of L-Asn under high metabolic demand of N (Bruneau et al., 2006). Its homolog in model legume *Lotus japonicus* has been reported to be involved in N remobilization and seed production (Credali et al., 2013). Alpha-galactosidase that highly up-regulated under N stress is a homolog of Arabidopsis AT5G08370, which plays an important role in leaf development by loosening and expanding cell wall (Chrost et al., 2007). It is also reported that alpha-galactosidase can contribute in completing energy-dependent senescence process and stress response in spite of severe decline in photosynthesis by maintaining the steady state of sugar supply through breakdown of wall polysaccharide (Pandey et al., 2017). To conclude, the above-mentioned genes, notably RADIALIS-like TFs, PLAC8 family proteins that are not characterized in wheat yet can be potential candidates to improve NUE and tolerance to N stress.

CONCLUSION

Identification of DEGs across bread wheat genotypes with contrasting stress tolerance facilitates a better understanding of the genetic bases of N stress tolerance. Here, the RNA-seq analysis using second leaf and grain tissues of low and high N treated wheat plants demonstrated that gene transcripts involved in lipid biosynthesis, transmembrane transport, cell communication, and small molecule biosynthesis were abundantly expressed in low NUE cultivars under N stress. Higher expression of these genes will enable low-NUE genotypes to thrive under stress conditions. The abundance of N metabolism-related genes in low NUE cultivars also contributes to N stress adaptation. The DEGs among the three cultivars showed variation in the magnitude of change in the expression, which indicates varying degrees of tolerance to N stress. Genes that were differentially expressed between low and high N treatments can also be indirectly involved in N metabolism. The DEGs across genotypes provide an understanding of how differently wheat genotypes encounter the N stress and how they adapt. Common N responsive genes across cultivars indicated that these genes are involved in common pathways under N stress. Moreover, the top DEGs with high expression in a high NUE cultivar would be the potential candidates to be explored for improving wheat NUE.

DATA AVAILABILITY STATEMENT

The RNA-seq data has been deposited in the NCBI. Submission details: NCBI Sequence Read Archive (SRA) submission: SUB7918524.

AUTHOR CONTRIBUTIONS

NS carried out the entire experiment, analysis, and writing up. SI contributed to experimental management and writing. AJ contributed to experimental design and data interpretation. RY, MS, and JZ contributed in carrying out laboratory and glasshouse

experiment. WM supervised the study and contributed to writing. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

This study is funded by GRDC project UMU00048 and Murdoch University Ph.D. scholarship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.583785/full#supplementary-material

Supplementary Figure 1 | KEGG enrichment pathways under N stress in three wheat cultivars Mace, Spitfire, and Volcani. **(A)** Mace leaf tissue; **(B)** Spitfire leaf tissue; **(C)** Volcani leaf tissue; **(D)** Mace grain tissue; **(E)** Spitfire grain tissue; **(F)** Volcani grain tissue.

Supplementary Table 1 | Detailed of RNA-seq data and read mapping.

Supplementary Table 2 | Summary of RNA-seq data and read mapping.

Supplementary Table 3 | Common differentially expressed genes (DEGs) between leaf and grain tissue independent of developmental stage and genotype.

Supplementary Table 4 | Consistently expressed DEGs between 0 and 10 DPA in second leaf tissue of of wheat cultivars Mace, Spitfire, and Volcani.

Supplementary Table 5 | Top significantly up-regulated DEGs in the second leaf tissue of wheat cultivars Mace, Spitfire, and Volcani.

Supplementary Table 6 | Top significantly down-regulated DEGs in the second leaf tissue of wheat cultivars Mace, Spitfire, and Volcani.

Supplementary Table 7 | Top significantly up-regulated DEGs in the grain tissue of wheat cultivars Mace, Spitfire, and Volcani.

Supplementary Table 8 | Top significantly down-regulated DEGs in the grain tissue of wheat cultivars Mace, Spitfire, and Volcani.

Supplementary Table 9 | List of highly connected down-regulated DEGs in photosynthesis-antenna proteins.

Supplementary Table 10 | List of highly connected down-regulated DEGs in photosynthesis.

Supplementary Table 11 | List of highly connected down-regulated DEGs in steroid biosynthesis.

REFERENCES

Abiko, T., Wakayama, M., Kawakami, A., Obara, M., Kisaka, H., Miwa T., et al. (2010). Changes in nitrogen assimilation, metabolism, and growth in transgenic rice plants expressing a fungal NADP (H)-dependent glutamate dehydrogenase (gdhA). *Planta* 232, 299–311. doi: 10.1007/s00425-010-1172-3

Achard, P., Gong, F., Cheminant, S., Alioua, M., Hedden, P., and Genschik, P. (2008). The cold-inducible CBF1 factor–dependent signaling pathway modulates the accumulation of the growth-repressing DELLA proteins via its effect on gibberellin metabolism. *Plant Cell.* 20, 2117–2129. doi: 10.1105/tpc.108.058941

Ahmad, A., Khan, I., Abrol, Y. P., and Iqbal, M. (2008). Genotypic variation of nitrogen use efficiency in Indian mustard. *Environ. Pollut.* 154, 462–466. doi: 10.1016/j.envpol.2007.10.007

Akula, R., and Ravishankar, G. A. (2011). Influence of abiotic stress signals on secondary metabolites in plants. *Plant Sign. Behav.* 6, 1720–1731. doi: 10.4161/psb.6.11.17613

Alam, N. B., and Ghosh, A. (2018). Comprehensive analysis and transcript profiling of Arabidopsis thaliana and Oryza sativa catalase gene family suggests their specific roles in development and stress responses. *Plant Physiol. Biochem.* 123, 54–64. doi: 10.1016/j.plaphy.2017.11.018

Alhabbar, Z., Islam, S., Yang, R., Diepeveen, D., Anwar, M., Balotf, S., et al. (2018a). Associations of NAM-A1 alleles with the onset of senescence and nitrogen use efficiency under Western Australian conditions. *Euphytica* 214:180. doi: 10.1007/s10681-018-2266-4

Alhabbar, Z., Yang, R., Juhasz, A., Xin, H., She, M., Anwar, M., et al. (2018b). NAM gene allelic composition and its relation to grain-filling duration and nitrogen utilisation efficiency of Australian wheat. *PLoS One.* 13:e0205448. doi: 10.1371/journal.pone.0205448

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Nat. Prec.* 11:R106. doi: 10.1186/gb-2010-11-10-r106

Anders, S., and Huber, W. (2012). *Differential expression of RNA-Seq data at the gene level–the DESeq package*. Heidelberg, Germany: European Molecular Biology Laboratory (EMBL).

Anwar, A., Liu, Y., Dong, R., Bai, L., Yu, X., and Li, Y. (2018). The physiological and molecular mechanism of brassinosteroid in response to stress: a review. *Biol. Res.* 51:46. doi: 10.1186/s40659-018-0195-2

Baligar, V. C., Fageria, N. K., and He, Z. L. (2007). Nutrient Use Efficiency in Plants. *Commun. Soil Sci. Plant Anal.* 32, 921–950. doi: 10.1081/CSS-100104098

Baxter, C. E., Costa, M. M. R., and Coen, E. S. (2007). Diversification and co-option of RAD-like genes in the evolution of floral asymmetry. *Plant J.* 52, 105–113. doi: 10.1111/j.1365-313X.2007.03222.x

Bell, E. M., Lin, W.-C., Husbands, A. Y., Yu, L., Jaganatha, V., Jablonska, B., et al. (2012). Arabidopsis lateral organ boundaries negatively regulates brassinosteroid accumulation to limit growth in organ boundaries. *Proc. Nat. Acad. Sci.* 109, 21146–21151. doi: 10.1073/pnas.1210789109

Bi, X., van Esse, W., Mulki, M. A., Kirschner, G., Zhong, J., Simon, R., et al. (2019). CENTRORADIALIS interacts with FLOWERING LOCUS T-like genes to control floret development and grain number. *Plant Physiol.* 180, 1013–1030. doi: 10.1104/pp.18.01454

Biemelt, S., Tschiersch, H., and Sonnewald, U. (2004). Impact of altered gibberellin metabolism on biomass accumulation, lignin biosynthesis, and photosynthesis in transgenic tobacco plants. *Plant Physiol.* 135, 254–265. doi: 10.1104/pp.103.036988

Boussadia, O., Steppe, K., Zgallai, H., El Hadj, S. B., Braham, M., Lemeur, R., et al. (2010). Effects of nitrogen deficiency on leaf photosynthesis, carbohydrate status and biomass production in two olive cultivars 'Meski'and 'Koroneiki'. *Sci. Horticult.* 123, 336–342. doi: 10.1016/j.scienta.2009.09.023

Boyko, A., and Kovalchuk, I. (2008). Epigenetic control of plant stress response. *Environ. Mole. Mutagen.* 49, 61–72. doi: 10.1002/em.20347

Bricker, T. M., Roose, J. L., Fagerlund, R. D., Frankel, L. K., and Eaton-Rye, J. J. (2012). The extrinsic proteins of Photosystem II. *Biochim. Biophys. Acta.* 1817, 121–142. doi: 10.1016/j.bbabio.2011.07.006

Bruneau, L., Chapman, R., and Marsolais, F. (2006). Co-occurrence of both L-asparaginase subtypes in Arabidopsis: At3g16150 encodes a K+-dependent L-asparaginase. *Planta* 224, 668–679. doi: 10.1007/s00425-006-0245-9

Caers, M., Rudelsheim, P., Van Onckelen, H., and Horemans, S. (1985). Effect of heat stress on photosynthetic activity and chloroplast ultrastructure in correlation with endogenous cytokinin concentration in maize seedlings. *Plant Cell Physiol.* 26, 47–52.

Cai, H., Xie, W., and Lian, X. (2013). Comparative analysis of differentially expressed genes in rice under nitrogen and phosphorus starvation stress conditions. *Plant Mole. Biol. Rep.* 31, 160–173. doi: 10.1007/s11105-012-0485-8

Cai, H., Zhou, Y., Xiao, J., Li, X., Zhang, Q., and Lian, X. (2009). Overexpressed glutamine synthetase gene modifies nitrogen metabolism and abiotic stress responses in rice. *Plant Cell Rep.* 28, 527–537. doi: 10.1007/s00299-008-0665-z

Chamizo-Ampudia, A., Sanz-Luque, E., Llamas, A., Galvan, A., and Fernandez, E. (2017). Nitrate reductase regulates plant nitric oxide homeostasis. *Trends Plant Sci.* 22, 163–174. doi: 10.1016/j.tplants.2016.12.001

Chandna, R., and Ahmad, A. (2015). Nitrogen stress-induced alterations in the leaf proteome of two wheat varieties grown at different nitrogen levels. *Physiol. Mole. Biol. Plants* 21, 19–33. doi: 10.1007/s12298-014-0277-8

Cheong, Y. H., Chang, H.-S., Gupta, R., Wang, X., Zhu, T., and Luan, S. (2002). Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis. *Plant Physiol.* 129, 661–677. doi: 10.1104/pp.002857

Chinnusamy, V., and Zhu, J.-K. (2009). Epigenetic regulation of stress responses in plants. *Curr. Opin. Plant Biol.* 12, 133–139. doi: 10.1016/j.pbi.2008.12.006

Chono, M., Honda, I., Zeniya, H., Yoneyama, K., Saisho, D., Takeda, K., et al. (2003). A semidwarf phenotype of barley uzu results from a nucleotide substitution in the gene encoding a putative brassinosteroid receptor. *Plant Physiol.* 133, 1209–1219. doi: 10.1104/pp.103.026195

Christie, P. J., Hahn, M., and Walbot, V. (1991). Low-temperature accumulation of alcohol dehydrogenase-1 mRNA and protein activity in maize and rice seedlings. *Plant Physiol.* 95, 699–706. doi: 10.1104/pp.95.3.699

Chrost, B., Kolukisaoglu, U., Schulz, B., and Krupinska, K. (2007). An α-galactosidase with an essential function during leaf development. *Planta* 225, 311–320. doi: 10.1007/s00425-006-0350-9

Claeys, H., Skirycz, A., Maleux, K., and Inzé, D. (2012). DELLA signaling mediates stress-induced cell differentiation in Arabidopsis leaves through modulation of anaphase-promoting complex/cyclosome activity. *Plant Physiol.* 159, 739–747. doi: 10.1104/pp.112.195032

Colebrook, E. H., Thomas, S. G., Phillips, A. L., and Hedden, P. (2014). The role of gibberellin signalling in plant responses to abiotic stress. *J. Exp. Biol.* 217, 67–75. doi: 10.1242/jeb.089938

Cormier, F., Foulkes, J., Hirel, B., Gouache, D., Moënne-Loccoz, Y., and Le Gouis, J. (2016). Breeding for increased nitrogen-use efficiency: a review for wheat (T. *aestivum L.). Plant Breed.* 135, 255–278. doi: 10.1111/pbr.12371

Crawford, N. M., and Glass, A. D. (1998). Molecular and physiological aspects of nitrate uptake in plants. *Trends Plant Sci.* 3, 389–395. doi: 10.1016/S1360-1385(98)01311-9

Credali, A., García-Calderón, M., Dam, S., Perry, J., Díaz-Quintana, A., Parniske, M., et al. (2013). The K+-dependent asparaginase, NSE1, is crucial for plant growth and seed production in Lotus japonicus. *Plant Cell Physiol.* 54, 107–118. doi: 10.1093/pcp/pcs156

Cren, M., and Hirel, B. (1999). Glutamine synthetase in higher plants regulation of gene and protein expression from the organ to the cell. *Plant Cell Physiol.* 40, 1187–1193. doi: 10.1093/oxfordjournals.pcp.a029506

CSBP (2012). The Flexi-N Range is a locally developed liquid fertilisers which apply nitrogen, sulphur and potassium evenly and accurately. Available online at: https://csbp-fertilisers.com.au

Curci, P. L., Aiese Cigliano, R., Zuluaga, D. L., Janni, M., Sanseverino, W., and Sonnante, G. (2017). Transcriptomic response of durum wheat to nitrogen starvation. *Sci. Rep.* 7:1176. doi: 10.1038/s41598-017-01377-0

Dai, Z., Plessis, A., Vincent, J., Duchateau, N., Besson, A., Dardevet, M., et al. (2015). Transcriptional and metabolic alternations rebalance wheat grain storage protein accumulation under variable nitrogen and sulfur supply. *Plant J.* 83, 326–343. doi: 10.1111/tpj.12881

Dalal, V. K., and Tripathy, B. C. (2018). Water-stress induced downsizing of light-harvesting antenna complex protects developing rice seedlings from photo-oxidative damage. *Sci. Rep.* 8:5955. doi: 10.1038/s41598-017-14419-4

Di Vietro, L., Daghino, S., Abbà, S., and Perotto, S. (2014). Gene expression and role in cadmium tolerance of two PLAC8-containing proteins identified in the ericoid mycorrhizal fungus Oidiodendron maius. *Fungal Biol.* 118, 695–703. doi: 10.1016/j.funbio.2014.04.011

Diao, J., Liu, H., Hu, F., Li, L., Wang, X., Gai, C., et al. (2019). Transcriptome analysis of immune response in fat greenling (Hexagrammos otakii) against Vibrio harveyi infection. *Fish Shellf. Immunol.* 84, 937–947. doi: 10.1016/j.fsi.2018.10.067

Dixon, D. P., Skipsey, M., and Edwards, R. (2010). Roles for glutathione transferases in plant secondary metabolism. *Phytochemistry* 71, 338–350. doi: 10.1016/j.phytochem.2009.12.012

Dixon, R. A., and Paiva, N. L. (1995). Stress-induced phenylpropanoid metabolism. *Plant Cell* 7:1085. doi: 10.2307/3870059

Drew, M. C., He, C.-J., and Morgan, P. W. (1989). Decreased ethylene biosynthesis, and induction of aerenchyma, by nitrogen-or phosphate-starvation in adventitious roots of Zea mays L. *Plant Physiol.* 91, 266–271. doi: 10.1104/pp.91.1.266

Evans, J., and Poorter, H. (2001). Photosynthetic acclimation of plants to growth irradiance: the relative importance of specific leaf area and nitrogen partitioning in maximizing carbon gain. *Plant. Cell Environ.* 24, 755–767. doi: 10.1046/j.1365-3040.2001.00724.x

Evans, J. R. (1989). Photosynthesis and nitrogen relationships in leaves of C 3 plants. *Oecologia* 78, 9–19. doi: 10.1007/BF00377192

Feng, H., Fan, X., Yan, M., Liu, X., Miller, A. J., and Xu, G. (2011). Multiple roles of nitrate transport accessory protein NAR2 in plants. *Plant Sign. Behav.* 6, 1286–1289. doi: 10.4161/psb.6.9.16377

Foyer, C. H., Noctor, G., and Hodges, M. (2011). Respiration and nitrogen assimilation: targeting mitochondria-associated metabolism as a means to enhance nitrogen use efficiency. *J. Exp. Bot.* 62, 1467–1482. doi: 10.1093/jxb/erq453

Foyer, C. H., and Shigeoka, S. (2011). Understanding oxidative stress and antioxidant functions to enhance photosynthesis. *Plant Physiol.* 155, 93–100. doi: 10.1104/pp.110.166181

Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Van Der Knaap, E., Cong, B., et al. (2000). fw2. 2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289, 85–88. doi: 10.1126/science.289.5476.85

Gadaleta, A., Nigro, D., Giancaspro, A., and Blanco, A. (2011). The glutamine synthetase (GS2) genes in relation to grain protein content of durum wheat. *Funct. Integr. Genom.* 11, 665–670. doi: 10.1007/s10142-011-0235-2

Gallais, A., and Hirel, B. (2004). An approach to the genetics of nitrogen use efficiency in maize. *J. Exp. Bot.* 55, 295–306. doi: 10.1093/jxb/erh006

Gallardo, K., Courty, P. E., Le Signor, C., Wipf, D., and Vernoud, V. (2014). Sulfate transporters in the plant's response to drought and salinity: regulation and possible functions. *Front. Plant Sci.* 5:580. doi: 10.3389/fpls.2014.00580

Garnett, T., Plett, D., Heuer, S., and Okamoto, M. (2015). Genetic approaches to enhancing nitrogen-use efficiency (NUE) in cereals: challenges and future directions. *Funct. Plant Biol.* 42, 921–941. doi: 10.1071/FP15025

Gelli, M., Duo, Y., Konda, A. R., Zhang, C., Holding, D., and Dweikat, I. (2014). Identification of differentially expressed genes between sorghum genotypes with contrasting nitrogen stress tolerance by genome-wide transcriptional profiling. *BMC Genomics.* 15:179. doi: 10.1186/1471-2164-15-179

Georgieva, K., Sarvari, E., and Keresztes, A. (2010). Protection of thylakoids against combined light and drought by a lumenal substance in the resurrection plant Haberlea rhodopensis. *Annal. Bot.* 105, 117–126. doi: 10.1093/aob/mcp274

Good, A. G., Johnson, S. J., De Pauw, M., Carroll, R. T., Savidov, N., Vidmar, J., et al. (2007). Engineering nitrogen use efficiency with alanine aminotransferase. *Botany* 85, 252–262. doi: 10.1139/B07-019

Good, A. G., Shrawat, A. K., and Muench, D. G. (2004). Can less yield more? Is reducing nutrient input into the environment compatible with maintaining crop production? *Trends Plant Sci.* 9, 597–605. doi: 10.1016/j.tplants.2004.10.008

Guo, J. H., Liu, X. J., Zhang, Y., Shen, J., Han, W., Zhang, W., et al. (2010). Significant acidification in major Chinese croplands. *Science* 327, 1008–1010. doi: 10.1126/science.1182570

Hamaguchi, A., Yamashino, T., Koizumi, N., Kiba, T., Kojima, M., Sakakibara, H., et al. (2008). A small subfamily of Arabidopsis RADIALIS-LIKE SANT/MYB genes: a link to HOOKLESS1-mediated signal transduction during early morphogenesis. *Biosci. Biotechnol. Biochem.* 72, 2687–2696. doi: 10.1271/bbb.80348

Hayat, S., Ahmad, A., Mobin, M., Hussain, A., and Fariduddin, Q. (2000). Photosynthetic rate, growth, and yield of mustard plants sprayed with 28-homobrassinolide. *Photosynthetica* 38, 469–471. doi: 10.1023/A:1010954411302

Hayes, S. (2019). BRacing for Water Stress: Brassinosteroid Signaling Promotes Drought Survival in Wheat. *Plant Physiol.* 180, 18–19. doi: 10.1104/pp.19.00314

Hirel, B., Le Gouis, J., Ney, B., and Gallais, A. (2007). The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. *J. Exper. Bot.* 58, 2369–2387. doi: 10.1093/jxb/erm097

Hitz, K., Clark, A. J., and Van Sanford, D. A. (2017). Identifying nitrogen-use efficient soft red winter wheat lines in high and low nitrogen environments. *Field Crops Res.* 200, 1–9. doi: 10.1016/j.fcr.2016.10.001

Hou, L., Zhang, A., Wang, R., Zhao, P., Zhang, D., Jiang, Y., et al. (2019). Brassinosteroid Regulates Root Development with Highly Redundant Genes in Hexaploid Wheat. *Plant Cell Physiol.* 60, 1761–1777. doi: 10.1093/pcp/pcz088

Huang, J., Gu, M., Lai, Z., Fan, B., Shi, K., Zhou, Y.-H., et al. (2010). Functional analysis of the Arabidopsis PAL gene family in plant growth, development, and response to environmental stress. *Plant Physiol.* 153, 1526–1538. doi: 10.1104/pp.110.157370

Hulsen, T., de Vlieg, J., and Alkema, W. (2008). BioVenn–a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics.* 9:488. doi: 10.1186/1471-2164-9-488

Ishikawa-Sakurai, J., Hayashi, H., and Murai-Hatano, M. (2014). Nitrogen availability affects hydraulic conductivity of rice roots, possibly through changes in aquaporin gene expression. *Plant Soil.* 379, 289–300. doi: 10.1007/s11104-014-2070-4

Janicka-Russak, M. (2011). Plant plasma membrane H+-ATPase in adaptation of plants to abiotic stresses. *Abiot. Str. Respon. Plants Physiol. Biochem. Gen. Perspect.* 1, 197–218. doi: 10.5772/24121

Jansson, S. (1994). The light-harvesting chlorophyll a/b-binding proteins. *Biochim. Biophys. Acta.* 1184, 1–19. doi: 10.1016/0005-2728(94)90148-1

Jansson, S. (1999). A guide to the Lhc genes and their relatives in Arabidopsis. *Trends Plant Sci.* 4, 236–240. doi: 10.1016/S1360-1385(99)01419-3

Jensen, P. E., Bassi, R., Boekema, E. J., Dekker, J. P., Jansson, S., Leister, D., et al. (2007). Structure, function and regulation of plant photosystem I. *Biochim. Biophys. Acta* 1767, 335–352. doi: 10.1016/j.bbabio.2007.03.004

Jeong, K., Pak, J.-H., and Kim, J. (2014). An evolutionary and developmental approach toward understanding of growth variations of two closely related dwarfism of Galium verum. *Biologia* 69, 1660–1667. doi: 10.2478/s11756-014-0479-0

Jeong, Y. J., Shang, Y., Kim, B. H., Kim, S. Y., Song, J. H., Lee, J. S., et al. (2010). BAK7 displays unequal genetic redundancy with BAK1 in brassinosteroid signaling and early senescence in Arabidopsis. *Mol. Cells* 29, 259–266. doi: 10.1007/s10059-010-0024-0

Jin, X., Yang, G., Tan, C., and Zhao, C. (2015). Effects of nitrogen stress on the photosynthetic CO2 assimilation, chlorophyll fluorescence, and sugar-nitrogen ratio in corn. *Sci. Rep*. 5:9311. doi: 10.1038/srep09311

Jobe, T. O., Zenzen, I., Rahimzadeh Karvansara, P., and Kopriva, S. (2019). Integration of sulfate assimilation with carbon and nitrogen metabolism in transition from C3 to C4 photosynthesis. *J. Exp. Bot.* 70, 4211–4221. doi: 10.1093/jxb/erz250

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2007). KEGG for linking genomes to life and the environment. *Nucl. Acids Res.* 36, D480–D484. doi: 10.1093/nar/gkm882

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucl. Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucl. Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

Kant, S., Bi, Y.-M., and Rothstein, S. J. (2010). Understanding plant response to nitrogen limitation for the improvement of crop nitrogen use efficiency. *J. Exp. Bot.* 62, 1499–1509.

Kant, S., Bi, Y. M., and Rothstein, S. J. (2011). Understanding plant response to nitrogen limitation for the improvement of crop nitrogen use efficiency. *J. Exp. Bot.* 62, 1499–1509. doi: 10.1093/jxb/erq297

Koltunow, A. M., Truettner, J., Cox, K. H., Wallroth, M., and Goldberg, R. B. (1990). Different temporal and spatial gene expression patterns occur during anther development. *Plant Cell* 2, 1201–1224. doi: 10.2307/3869340

Komatsu, T., Kawaide, H., Saito, C., Yamagami, A., Shimada, S., Nakazawa, M., et al. (2010). The chloroplast protein BPG2 functions in brassinosteroid-mediated post-transcriptional accumulation of chloroplast rRNA. *Plant J.* 61, 409–422. doi: 10.1111/j.1365-313X.2009.04077.x

Konishi, N., Ishiyama, K., Matsuoka, K., Maru, I., Hayakawa, T., Yamaya, T., et al. (2014). NADH-dependent glutamate synthase plays a crucial role in assimilating ammonium in the Arabidopsis root. *Physiol. Plant.* 152, 138–151. doi: 10.1111/ppl.12177

Kopriva, S., and Rennenberg, H. (2004). Control of sulphate assimilation and glutathione synthesis: interaction with N and C metabolism. *J. Exp. Bot.* 55, 1831–1842. doi: 10.1093/jxb/erh203

Lederer, B., and Böger, P. (2003). Binding and protection of porphyrins by glutathione S-transferases of Zea mays L. *Biochim. Biophys. Acta* 1621, 226–233. doi: 10.1016/S0304-4165(03)00073-4

Lederer, B., and Böger, P. (2005). A ligand function of glutathione S-transferase. *Zeitschr. für Naturforsch, C.* 60, 166–171. doi: 10.1515/znc-2005-3-403

Lee, J., and Rudd, J. J. (2002). Calcium-dependent protein kinases: versatile plant signalling components necessary for pathogen defence. *Trends Plant Sci.* 7, 97–98. doi: 10.1016/S1360-1385(02)02229-X

Léran, S., Garg, B., Boursiac, Y., Corratgé-Faillie, C., Brachet, C., Tillard, P., et al. (2015). AtNPF5. 5, a nitrate transporter affecting nitrogen accumulation in Arabidopsis embryo. *Sci. Rep.* 5:7962. doi: 10.1038/srep07962

Leustek, T., Martin, M. N., Bick, J.-A., and Davies, J. P. (2000). Pathways and regulation of sulfur metabolism revealed through molecular and genetic studies. *Annu. Rev. Plant Biol.* 51, 141–165. doi: 10.1146/annurev.arplant.51.1.141

Li, W., Wang, Y., Okamoto, M., Crawford, N. M., Siddiqi, M. Y., and Glass, A. D. (2007). Dissection of the AtNRT2. 1: AtNRT2. 2 inducible high-affinity nitrate transporter gene cluster. *Plant Physiol.* 143, 425–433. doi: 10.1104/pp.106.091223

Li, Y., Gao, Y., Xu, X., Shen, Q., and Guo, S. (2009). Light-saturated photosynthetic rate in high-nitrogen rice (Oryza sativa L.) *leaves is related to chloroplastic CO*2 concentration. *J. Exp. Bot.* 60, 2351–2360. doi: 10.1093/jxb/erp127

Li, Y., Ren, B., Ding, L., Shen, Q., Peng, S., and Guo, S. (2013). Does chloroplast size influence photosynthetic nitrogen use efficiency? *PLoS One.* 8:e62036. doi: 10.1371/journal.pone.0062036

Lian, X., Wang, S., Zhang, J., Feng, Q., Zhang, L., Fan, D., et al. (2006). Expression profiles of 10,422 genes at early stage of low nitrogen stress in rice assayed using a cDNA microarray. *Plant Mole. Biol.* 60, 617–631. doi: 10.1007/s11103-005-5441-7

Liao, Q., Zhou, T., Yao, J.-y., Han, Q.-f., Song, H.-x., Guan, C.-y., et al. (2018). Genome-scale characterization of the vacuole nitrate transporter *Chloride Channel* (*CLC*) genes and their transcriptional responses to diverse nutrient stresses in allotetraploid rapeseed. *PLoS One* 13:e0208648. doi: 10.1371/journal.pone.0208648

Liepman, A. H., and Olsen, L. J. (2001). Peroxisomal alanine: glyoxylate aminotransferase (AGT1) is a photorespiratory enzyme with multiple substrates in *Arabidopsis thaliana*. *Plant J.* 25, 487–498. doi: 10.1046/j.1365-313x.2001.00961.x

Lin, J., Wang, Y., Sun, S., Mu, C., and Yan, X. (2017). Effects of arbuscular mycorrhizal fungi on the growth, photosynthesis and photosynthetic pigments of Leymus chinensis seedlings under salt-alkali stress and nitrogen deposition. *Sci. Tot. Environ.* 576, 234–241. doi: 10.1016/j.scitotenv.2016.10.091

Liu, H., Wang, S., Yu, X., Yu, J., He, X., Zhang, S., et al. (2005). ARL1, a LOB-domain protein required for adventitious root formation in rice. *Plant J.* 43, 47–56. doi: 10.1111/j.1365-313X.2005.02434.x

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. doi: 10.1111/j.1365-3040.2005.01493.x

Lu, K., Wu, B., Wang, J., Zhu, W., Nie, H., Qian, J., et al. (2018). Blocking amino acid transporter *OsAAP3* improves grain yield by promoting outgrowth buds and increasing tiller number in rice. *Plant Biotechnol. J.* 16, 1710–1722. doi: 10.1111/pbi.12907

Ma, W., Wu, F., Sheng, P., Wang, X., Zhang, Z., Zhou, K., et al. (2017). The LBD12-1 transcription factor suppresses apical meristem size by repressing Argonaute 10 expression. *Plant Physiol.* 173, 801–811. doi: 10.1104/pp.16.01699

Maizel, A., and Weigel, D. (2004). Temporally and spatially controlled induction of gene expression in Arabidopsis thaliana. *Plant J.* 38, 164–171. doi: 10.1111/j.1365-313X.2004.02027.x

Makino, A. (2003). Rubisco and nitrogen relationships in rice: Leaf photosynthesis and plant growth. *Soil Sci. Plant Nutr.* 49, 319–327. doi: 10.1080/00380768.2003.10410016

Makino, A., Sakuma, H., Sudo, E., and Mae, T. (2003). Differences between maize and rice in N-use efficiency for photosynthesis and protein allocation. *Plant Cell Physiol.* 44, 952–956. doi: 10.1093/pcp/pcg113

Mao, X., Cai, T., Olyarchuk, J. G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21, 3787–3793. doi: 10.1093/bioinformatics/bti430

Martin, A., Lee, J., Kichey, T., Gerentes, D., Zivy, M., Tatout, C., et al. (2006). Two cytosolic glutamine synthetase isoforms of maize are specifically involved in the control of grain production. *Plant Cell* 18, 3252–3274. doi: 10.1105/tpc.106.042689

Masclaux-Daubresse, C., Reisdorf-Cren, M., and Orsel, M. (2008). Leaf nitrogen remobilisation for plant development and grain filling. *Plant Biol.* 10, 23–36. doi: 10.1111/j.1438-8677.2008.00097.x

Meng, S., Su, L., Li, Y., Wang, Y., Zhang, C., Zhao, Z. et al. (2016). Nitrate and ammonium contribute to the distinct nitrogen metabolism of *Populus simonii* during moderate salt stress. *PLoS One* 11:e0150354. doi: 10.1371/journal.pone.0150354

McCullough, D., Aguilera, A., and Tollenaar, M. (1994). N uptake, N partitioning, and photosynthetic N-use efficiency of an old and a new maize hybrid. *Cana. J. Plant Sci.* 74, 479–484. doi: 10.4141/cjps94-088

Moll, R., Kamprath, E., and Jackson, W. (1982). Analysis and interpretation of factors which contribute to efficiency of nitrogen utilization. *Agron. J.* 74, 562–564. doi: 10.2134/agronj1982.00021962007400030037x

Mounier, E., Pervent, M., Ljung, K., Gojon, A., and Nacry, P. (2014). Auxin-mediated nitrate signalling by NRT 1.1 *participates in the adaptive response of A rabidopsis root architecture to the spatial heterogeneity of nitrate availability*. *Plant Cell Environ.* 37, 162–174. doi: 10.1111/pce.12143

Muller, O., Oguchi, R., Hirose, T., Werger, M. J., and Hikosaka, K. (2009). The leaf anatomy of a broad-leaved evergreen allows an increase in leaf nitrogen content in winter. *Physiol. Plantar.* 136, 299–309. doi: 10.1111/j.1399-3054.2009.01224.x

Munekage, Y., Takeda, S., Endo, T., Jahns, P., Hashimoto, T., and Shikanai, T. (2001). Cytochrome b(6)f mutation specifically affects thermal dissipation of absorbed light energy in Arabidopsis. *Plant J.* 28, 351–359. doi: 10.1046/j.1365-313X.2001.01178.x

Mur, L. A. J., Sivakumaran, A., Mandon, J., Cristescu, S. M., Harren, F. J. M., and Hebelstrup, K. H. (2012). Haemoglobin modulates salicylate and jasmonate/ethylene-mediated resistance mechanisms against pathogens. *J. Exp. Bot.* 63, 4375–4387. doi: 10.1093/jxb/ers116

Mussig, C., Shin, G. H., and Altmann, T. (2003). Brassinosteroids promote root growth in Arabidopsis. *Plant Physiol.* 133, 1261–1271. doi: 10.1104/pp.103.028662

Nishiyama, R., Le, D. T., Watanabe, Y., Matsui, A., Tanaka, M., Seki, M., et al. (2012). Transcriptome analyses of a salt-tolerant cytokinin-deficient mutant reveal differential regulation of salt stress response by cytokinin deficiency. *PLoS One.* 7:e32124. doi: 10.1371/journal.pone.0032124

Niu , J.-H., Ahmad Anjum, S., Wang, R., Li, J.-H., Liu, M.-R., Song, J.-X., et al. (2016). Exogenous application of brassinolide can alter morphological and physiological traits of Leymus chinensis (Trin.) *Tzvelev under room and high temperatures*. *Chil. J. Agricul. Res.* 76, 27–33. doi: 10.4067/S0718-58392016000100004

Nowicka, B., Ciura, J., Szymanska, R., and Kruk, J. (2018). Improving photosynthesis, plant productivity and abiotic stress tolerance - current trends and future perspectives. *J. Plant Physiol.* 231, 415–433. doi: 10.1016/j.jplph.2018.10.022

Nunes-Nesi, A., Fernie, A. R., and Stitt, M. (2010). Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. *Mol. Plant.* 3, 973–996. doi: 10.1093/mp/ssq049

Nyikako, J., Schierholt, A., Kessel, B., and Becker, H. C. (2014). Genetic variation in nitrogen uptake and utilization efficiency in a segregating DH population of winter oilseed rape. *Euphytica* 199, 3–11. doi: 10.1007/s10681-014-1201-6

Pandey, J. K., Dash, S. K., and Biswal, B. (2017). Nitrogen-deficiency-induced loss in photosynthesis and modulation of β-galactosidase activity during senescence of Arabidopsis leaves. *Acta Physiol. Plantar.* 39:75. doi: 10.1007/s11738-017-2371-3

Park, D.-Y., Shim, Y., Gi, E., Lee, B.-D., An, G., Kang, K., et al. (2018). The MYB-related transcription factor RADIALIS-LIKE3 (OsRL3) functions in ABA-induced leaf senescence and salt sensitivity in rice. *Environ. Exp. Bot.* 156, 86–95. doi: 10.1016/j.envexpbot.2018.08.033

Petrussa, E., Braidot, E., Zancani, M., Peresson, C., Bertolini, A., Patui, S., et al. (2013). Plant flavonoids—biosynthesis, transport and involvement in stress responses. *Int. J. Mole. Sci.* 14, 14950–14973. doi: 10.3390/ijms140714950

Pilot, G., Stransky, H., Bushey, D. F., Pratelli, R., Ludewig, U., Wingate, V. P., et al. (2004). Overexpression of *GLUTAMINE DUMPER1* leads to hypersecretion of glutamine from hydathodes of Arabidopsis leaves. *Plant Cell* 16, 1827–1840. doi: 10.1105/tpc.021642

Poorter, H., and Evans, J. R. (1998). Photosynthetic nitrogen-use efficiency of species that differ inherently in specific leaf area. *Oecologia* 116, 26–37. doi: 10.1007/s004420050560

Price, J., Laxmi, A., Martin, S. K. S., and Jang, J.-C. (2004). Global transcription profiling reveals multiple sugar signal transduction mechanisms in Arabidopsis. *Plant Cell* 16, 2128–2150. doi: 10.1105/tpc.104.022616

Ranathunge, K., El-Kereamy, A., Gidda, S., Bi, Y.-M., and Rothstein, S. J. (2014). *AMT1;1* transgenic rice plants with enhanced NH_4^+ permeability show superior growth and higher yield under optimal and suboptimal NH_4^+ conditions. *J. Exp. Bot.* 65, 965–979. doi: 10.1093/jxb/ert458

Rehman, S. U., Bilal, M., Rana, R. M., Tahir, M. N., Shah, M. K. N., Ayalew, H., et al. (2016). Cell membrane stability and chlorophyll content variation in wheat (Triticum aestivum) genotypes under conditions of heat and drought. *Crop Pasture Sci.* 67, 712–718. doi: 10.1071/CP15385

Ripullone, F., Grassi, G., Lauteri, M., and Borghetti, M. (2003). Photosynthesis–nitrogen relationships: interpretation of different patterns between Pseudotsuga menziesii and Populus× euroamericana in a mini-stand experiment. *Tree Physiol.* 23, 137–144. doi: 10.1093/treephys/23.2.137

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Rubio-Wilhelmi, M., Sanchez-Rodriguez, E., Rosales, M., Begona, B., Rios, J., Romero, L., et al. (2011). Effect of cytokinins on oxidative stress in tobacco plants under nitrogen deficiency. *Environ. Exp. Bot.* 72, 167–173. doi: 10.1016/j.envexpbot.2011.03.005

Rufty, T. W., Huber, S. C., and Volk, R. J. (1988). Alterations in leaf carbohydrate metabolism in response to nitrogen stress. *Plant Physiol.* 88, 725–730. doi: 10.1104/pp.88.3.725

Sage, R. F., Sharkey, T. D., and Seemann, J. R. (1988). The in-vivo response of the ribulose-1,5-bisphosphate carboxylase activation state and the pool sizes of photosynthetic metabolites to elevated CO2 inPhaseolus vulgaris L. *Planta* 174, 407–416. doi: 10.1007/BF00959528

Sahi, C., Agarwal, M., Singh, A., and Grover, A. (2007). Molecular characterization of a novel isoform of rice (Oryza sativa L.) *glycine rich-RNA binding protein and evidence for its involvement in high temperature stress response*. *Plant Sci.* 173, 144–155. doi: 10.1016/j.plantsci.2007.04.010

Saito, K. (2000). Regulation of sulfate transport and synthesis of sulfur-containing amino acids. *Curr. Opin. Plant Biol.* 3, 188–195. doi: 10.1016/S1369-5266(00)00063-7

Sakamoto, T., Morinaka, Y., Ohnishi, T., Sunohara, H., Fujioka, S., Ueguchi-Tanaka, M., et al. (2006). Erect leaves caused by brassinosteroid deficiency increase biomass production and grain yield in rice. *Nat. Biotechnol.* 24, 105–109. doi: 10.1038/nbt1173

Schmidt, R., Stransky, H., and Koch, W. (2007). The amino acid permease AAP8 is important for early seed development in *Arabidopsis thaliana*. *Planta* 226, 805–813. doi: 10.1007/s00425-007-0527-x

Schmülling, T., Werner, T., Riefler, M., Krupková, E. Y., and Manns, I. B. (2003). Structure and function of cytokinin oxidase/dehydrogenase genes of maize, rice, Arabidopsis and other species. *J. Plant Res.* 116, 241–252. doi: 10.1007/s10265-003-0096-4

Schulz, P., Herde, M., and Romeis, T. (2013). Calcium-dependent protein kinases: hubs in plant stress signaling and development. *Plant Physiol.* 163, 523–530. doi: 10.1104/pp.113.222539

Shao, A., Ma, W., Zhao, X., Hu, M., He, X., Teng, W., et al. (2017). The auxin biosynthetic *TRYPTOPHAN AMINOTRANSFERASE RELATED TaTAR2.1-3A* increases grain yield of wheat. *Plant Physiol.* 174, 2274–2288. doi: 10.1104/pp.17.00094

Shi, Y., Wang, Z., Meng, P., Tian, S., Zhang, X., and Yang, S. (2013). The glutamate carboxypeptidase AMP 1 mediates abscisic acid and abiotic stress responses in A rabidopsis. *New Phytol.* 199, 135–150. doi: 10.1111/nph.12275

Shrawat, A. K., Carroll, R. T., DePauw, M., Taylor, G. J., and Good, A. G. (2008). Genetic engineering of improved nitrogen use efficiency in rice by the tissue-specific expression of alanine aminotransferase. *Plant Biotechnol. J.* 6, 722–732. doi: 10.1111/j.1467-7652.2008.00351.x

Shu, S., Tang, Y., Yuan, Y., Sun, J., Zhong, M., and Guo, S. (2016). The role of 24-epibrassinolide in the regulation of photosynthetic characteristics and nitrogen metabolism of tomato seedlings under a combined low temperature and weak light stress. *Plant Physiol. Biochem.* 107, 344–353. doi: 10.1016/j.plaphy.2016.06.021

Singh, T., Paleg, I., and Aspinall, D. (1973). Stress metabolism I. *Nitrogen metabolism and growth in the barley plant during water stress. Austr. J. Biol. Sci.* 26, 45–56. doi: 10.1071/BI9730045

Sinha, S. K., Sevanthi, V. A., Chaudhary, S., Tyagi, P., Venkadesan, S., Rani, M., et al. (2018). Transcriptome Analysis of Two Rice Varieties Contrasting for Nitrogen Use Efficiency under Chronic N Starvation Reveals Differences in Chloroplast and Starch Metabolism-Related Genes. *Genes* 9:208. doi: 10.3390/genes9040206

Song, W.-Y., Hörtensteiner, S., Tomioka, R., Lee, Y., and Martinoia, E. (2011). Common functions or only phylogenetically related? The large family of PLAC8 motif-containing/PCR genes. *Mole. Cells* 31, 1–7. doi: 10.1007/s10059-011-0024-8

Stacey, M. G., Koh, S., Becker, J., and Stacey, G. (2002). AtOPT3, a member of the oligopeptide transporter family, is essential for embryo development in Arabidopsis. *Plant Cell* 14, 2799–2811. doi: 10.1105/tpc.005629

Stitt, M., Müller, C., Matt, P., Gibon, Y., Carillo, P., Morcuende, R., et al. (2002). Steps towards an integrated view of nitrogen metabolism. *J. Exp. Bot.* 53, 959–970. doi: 10.1093/jexbot/53.370.959

Sun, H., Li, J., Song, W., Tao, J., Huang S., Chen, S., et al. (2015). Nitric oxide generated by nitrate reductase increases nitrogen uptake capacity by inducing lateral root formation and inorganic nitrogen uptake under partial nitrate nutrition in rice. *J. Exp. Bot.* 66, 2449–2459. doi: 10.1093/jxb/erv030

Tabuchi, M., Sugiyama, K., Ishiyama, K., Inoue, E., Sato, T., Takahashi, H., et al. (2005). Severe reduction in growth rate and grain filling of rice mutants lacking OsGS1;1, a cytosolic glutamine synthetase1;1. *Plant J.* 42, 641–651. doi: 10.1111/j.1365-313X.2005.02406.x

Tahkokorpi, M., Taulavuori, K., Laine, K., and Taulavuori, E. (2007). After-effects of drought-related winter stress in previous and current year stems of Vaccinium myrtillus L. *Environ. Exp. Bot.* 61, 85–93. doi: 10.1016/j.envexpbot.2007.03.003

Takahashi, M., Sasaki, Y., Ida, S., and Morikawa, H. (2001). Nitrite reductase gene enrichment improves assimilation of NO_2 in Arabidopsis. *Plant Physiol.* 126, 731–741. doi: 10.1104/pp.126.2.731

Tavladoraki, P., Cona, A., Federico, R., Tempera, G., Viceconte, N., Saccoccio, S., et al. (2012). Polyamine catabolism: target for antiproliferative therapies in animals and stress tolerance strategies in plants. *Amino. Acids.* 42, 411–426. doi: 10.1007/s00726-011-1012-1

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Proto.* 7, 562–578. doi: 10.1038/nprot.2012.016

Troncoso-Ponce, M. A., Cao, X., Yang, Z., and Ohlrogge, J. B. (2013). Lipid turnover during senescence. *Plant Sci.* 205, 13–19. doi: 10.1016/j.plantsci.2013.01.004

Urao, T., Yakubov, B., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998). Stress-responsive expression of genes for two-component response regulator-like proteins in Arabidopsis thaliana. *FEBS Lett.* 427, 175–178. doi: 10.1016/S0014-5793(98)00418-9

van der Werf, A., van Nuenen, M., Visser, A. J., and Lambers, H. (1993). Contribution of physiological and morphological plant traits to a species' competitive ability at high and low nitrogen supply : A hypothesis for inherently fast- and slow-growing monocotyledonous species. *Oecologia* 94, 434–440. doi: 10.1007/BF00317120

Vardhini, B. V., and Anjum, N. A. (2015). Brassinosteroids make plant life easier under abiotic stresses mainly by modulating major components of antioxidant defense system. *Front. Environ. Sci.* 2:67. doi: 10.3389/fenvs.2014.00067

Vickers, C. E., Gershenzon, J., Lerdau, M. T., and Loreto, F. (2009). A unified mechanism of action for volatile isoprenoids in plant abiotic stress. *Nat. Chem. Biol.* 5:283. doi: 10.1038/nchembio.158

Von Caemmerer, S. (2000). *Biochemical models of leaf photosynthesis*. Australia: CSIRO publishing. doi: 10.1071/9780643103405

Wang, Y.-H., Garvin, D. F., and Kochian, L. V. (2001). Nitrate-induced genes in tomato roots. *Array analysis reveals novel genes that may play a role in nitrogen nutrition. Plant Physiol.* 127, 345–359. doi: 10.1104/pp.127.1.345

Wang, R., Okamoto, M., Xing, X., and Crawford, N. M. (2003). Microarray analysis of the nitrate response in Arabidopsis roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. *Plant Physiol.* 132, 556–567. doi: 10.1104/pp.103.021253

Wang, W., Vinocur, B., and Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* 218, 1–14. doi: 10.1007/s00425-003-1105-5

Wang, Z. Y., Bai, M. Y., Oh, E., and Zhu, J. Y. (2012). Brassinosteroid signaling network and regulation of photomorphogenesis. *Ann. Rev. Gen.* 46, 701–724. doi: 10.1146/annurev-genet-102209-163450

Wang, F., Tan, H., Han, J., Zhang, Y., He, X., Ding, Y., et al. (2019). A novel family of PLAC8 motif-containing/PCR genes mediates Cd tolerance and Cd accumulation in rice. *Environ. Sci. Euro.* 31, 1–13. doi: 10.1186/s12302-019-0259-0

Wang, J., Song, K., Sun, L., Qin, Q., Sun, Y., Pan, J., et al. (2019). Morphological and Transcriptome Analysis of Wheat Seedlings Response to Low Nitrogen Stress. *Plants* 8:98. doi: 10.3390/plants8040098

Wang, X., Yang, G., Shi, M., Hao, D., Wei, Q., Wang, Z., et al. (2019). Disruption of an amino acid transporter *LHT1* leads to growth inhibition and low yields in rice. *BMC Plant Biol.* 19, 1–11. doi: 10.1186/s12870-019-1885-9

Wei, S., Wang, X., Shi, D., Li, Y., Zhang, J., Liu, P., et al. (2016). The mechanisms of low nitrogen induced weakened photosynthesis in summer maize (Zea mays L.) *under field conditions. Plant Physiol. Bio. Chem.* 105, 118–128. doi: 10.1016/j.plaphy.2016.04.007

Wen, D., Xu, H., Xie, L., He, M., Hou, H., Wu, C., et al. (2018). Effects of Nitrogen Level during Seed Production on Wheat Seed Vigor and Seedling Establishment at the Transcriptome Level. *Int. J. Mol. Sci.* 19:3417. doi: 10.3390/ijms19113417

Won, C., Shen, X., Mashiguchi, K., Zheng, Z., Dai, X., Cheng, Y., et al. (2011). Conversion of tryptophan to indole-3-acetic acid by TRYPTOPHAN AMINOTRANSFERASES OF ARABIDOPSIS and YUCCAs in Arabidopsis. *Proc. Nat. Acad. Sci.* 108, 18518–18523. doi: 10.1073/pnas.1108436108

Wu, C. Y., Trieu, A., Radhakrishnan, P., Kwok, S. F., Harris, S., Zhang, K., et al. (2008). Brassinosteroids regulate grain filling in rice. *Plant Cell* 20, 2130–2145. doi: 10.1105/tpc.107.055087

Xiao, S., and Chye, M.-L. (2011). New roles for acyl-CoA-binding proteins (ACBPs) in plant development, stress responses and lipid metabolism. *Progr. Lip. Res.* 50, 141–151. doi: 10.1016/j.plipres.2010.11.002

Xu, G., Fan, X., and Miller, A. J. (2012). Plant nitrogen assimilation and use efficiency. *Annu. Rev. Plant Biol.* 63, 153–182. doi: 10.1146/annurev-arplant-042811-105532

Yamaguchi, Y., Nakamura, T., Harada, E., Koizumi, N., and Sano, H. (1999). Differential accumulation of transcripts encoding sulfur assimilation enzymes upon sulfur and/or nitrogen deprivation in Arabidopsis thaliana. *Biosci. Biotechnol. Biochem.* 63, 762–766. doi: 10.1271/bbb.63.762

Yang, S. Y., Hao, D. L., Song, Z. Z., Yang, G. Z., Wang, L., and Su, Y. H. (2015). RNA-Seq analysis of differentially expressed genes in rice under varied nitrogen supplies. *Gene* 555, 305–317. doi: 10.1016/j.gene.2014.11.021

Yang, H., Stierhof, Y.-D., and Ludewig, U. (2015). The putative Cationic Amino Acid Transporter 9 is targeted to vesicles and may be involved in plant amino acid homeostasis. *Front. Plant Sci.* 6:212. doi: 10.3389/fpls.2015.00212

Yang, W., Yoon, J., Choi, H., Fan, Y., Chen, R., and An, G. (2015). Transcriptome analysis of nitrogen-starvation-responsive genes in rice. *BMC Plant Biol.* 15:31. doi: 10.1186/s12870-015-0425-5

Ym, B. I., Kant, S., Clark, J., Gidda, S., Ming, F., Xu, J., et al. (2009). Increased nitrogen-use efficiency in transgenic rice plants over-expressing a nitrogen-responsive early nodulin gene identified from rice expression profiling. *Plant Cell Environ.* 32, 1749–1760. doi: 10.1111/j.1365-3040.2009.02032.x

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11:R14. doi: 10.1186/gb-2010-11-2-r14

Yu, Z., Juhasz, A., Islam, S., Diepeveen, D., Zhang, J., Wang, P., et al. (2018). Impact of mid-season sulphur deficiency on wheat nitrogen metabolism and biosynthesis of grain protein. *Scientif. Rep.* 8:2499. doi: 10.1038/s41598-018-20935-8

Zeh, M., Casazza, A. P., Kreft, O., Roessner, U., Bieberich, K., Willmitzer L., et al. (2001). Antisense inhibition of threonine synthase leads to high methionine content in transgenic potato plants. *Plant Physiol.* 127, 792–802.

Zhang, D., Zhang, Y., Yang, W., and Miao, G. (2006). Biological response of roots in different spring wheat genotypes to low nitrogen stress. *Zuo wu xue bao.* 32, 1349–1354.

Zhang, Y., Liu, Z., Liu, J., Lin, S., Wang, J., Lin, W., et al. (2017). GA-DELLA pathway is involved in regulation of nitrogen deficiency-induced anthocyanin accumulation. *Plant Cell Rep.* 36, 557–569. doi: 10.1007/s00299-017-2102-7

Zhao, B., and Li, J. (2012). Regulation of brassinosteroid biosynthesis and inactivation. *J. Integr. Plant Biol.* 54, 746–759. doi: 10.1111/j.1744-7909.2012.01168.x

Zhao, D., Reddy, K. R., Kakani, V. G., and Reddy, V. R. (2005). Nitrogen deficiency effects on plant growth, leaf photosynthesis, and hyperspectral reflectance properties of sorghum. *Europ. J. Agron.* 22, 391–403. doi: 10.1016/j.eja.2004.06.005

Zhou, Y., Cai, H., Xiao, J., Li X., Zhang, Q., and Lian, X. (2009). Over-expression of aspartate aminotransferase genes in rice resulted in altered nitrogen metabolism and increased amino acid content in seeds. *Theor. Appl. Genet.* 118, 1381–1390. doi: 10.1007/s00122-009-0988-3

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 16 October 2020
doi: 10.3389/fgene.2020.583245

Differential Expression of Genes Involved in Saikosaponin Biosynthesis Between *Bupleurum chinense* DC. and *Bupleurum scorzonerifolium* Willd

Ma Yu[1,2], Hua Chen[2], Shi-Hang Liu[3], Yu-Chan Li[2], Chun Sui[1], Da-Bin Hou[2] and Jian-He Wei[1]**

[1] Institute of Medicinal Plant Development (IMPLAD), Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2] School of Life Sciences and Engineering, Southwest University of Science and Technology, Mianyang, China, [3] Triticeae Research Institute, Sichuan Agricultural University, Chengdu, China

Radix Bupleuri (roots of *Bupleurum* spp.) is an important medicinal herb. Triterpenoid saponins of saikosaponins generally constitute the main class of secondary metabolites of plants in the *Bupleurum* genus. However, the molecular regulatory mechanism underlying their biosynthesis remains elusive. In this study, we observed significantly different saikosaponin biosynthesis between *Bupleurum chinense* and *Bupleurum scorzonerifolium* at the seedling stage. The sequential and expression characterization of 232 genes in the triterpenoid saponin biosynthetic pathway, which includes the mevalonate (MVA) pathway and methylerythritol phosphate (MEP) pathway, between *B. chinense* and *B. scorzonerifolium* was also investigated. Sixty of these genes may be involved in saikosaponin biosynthesis. Manipulation of these genes, especially those of the β-AS, P450, and UGT families, may improve saikosaponin production.

Keywords: *Bupleurum*, saikosaponin biosynthesis, transcriptome analysis, gene expression, P450

OPEN ACCESS

Edited by:
Penghao Wang,
Murdoch University, Australia

Reviewed by:
Feng Wang,
University of Helsinki, Finland
Junfeng Chen,
Shanghai University of Traditional Chinese Medicine, China

****Correspondence:***
Jian-He Wei
jhwei@implad.ac.cn
Da-Bin Hou
yuwen0073@126.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *14 July 2020*
Accepted: *16 September 2020*
Published: *16 October 2020*

Citation:
Yu M, Chen H, Liu S-H, Li Y-C, Sui C, Hou D-B and Wei J-H (2020) Differential Expression of Genes Involved in Saikosaponin Biosynthesis Between Bupleurum chinense DC. and Bupleurum scorzonerifolium Willd. Front. Genet. 11:583245. doi: 10.3389/fgene.2020.583245

INTRODUCTION

Radix Bupleuri (roots of *Bupleurum* spp.) is one of the most important medicinal herbs in Eurasia and North Africa used as a treatment for fever, chronic hepatitis, nephrotic syndrome, inflammatory diseases, menstrual disorders, and digestive ulcers (Pistelli et al., 1996; Guo et al., 2000; Ikegami et al., 2006; Mabberley, 2008). In addition, this herb has been used for more than 2,000 years in China (Tan et al., 2008). In the Chinese Pharmacopoeia, the official botanical origin of Bupleuri Radix is the roots of *Bupleurum chinense* DC. or *Bupleurum scorzonerifolium* Willd. (Chinese Pharmacopoeia Commission, 2015). Several groups of secondary metabolites have been isolated from *Bupleurum* species, including triterpenoid saponins (saikosaponins), steroidal saponins, lignans, essential oils, and polysaccharides (Ashour and Wink, 2011). Among these, saikosaponins generally represent the main class of secondary metabolites and constitute to up to 7% of the total dry weight of roots of plants in the *Bupleurum* genus (Ashour and Wink, 2011). Owing to their wide range of pharmacological activities, including their immunomodulatory activity, anti-inflammatory activity, antioxidant and hepatoprotective activity, cytotoxicity, antitumor activity, and antiviral activity, these triterpenoid saponins, especially

saikosaponins a and d, are the most important pharmacological constituent in *Bupleurum* root extracts (Zhao and Xiao, 2007; Lin et al., 2013; Liang et al., 2014).

To date, more than 130 glycosylated oleanane-type and ursane-type saponins have been isolated from the genus *Bupleurum* L. (Ashour and Wink, 2011; Wang et al., 2017b; Chelghoum et al., 2018). A previous study on the species *B. chinense*, *B. scorzonerifolium*, and *Bupleurum falcatum* L. demonstrated that saikosaponins are mainly distributed in the tissues of the cork and cortex of roots (Liang et al., 2014). Similar results were reported based on histochemical studies (Tan et al., 2008). In addition, the synthesis and accumulation of saikosaponins is strongly influenced by intrinsic factors, including growth stage, developmental phase, and root structure, and by environmental conditions, such as drought, fertility, and light deficiency (Tan et al., 2008; Zhu et al., 2009a,b; Gong et al., 2017). The combination of cultural practices together with manipulation of the expression of genes involved in triterpenoid saponin biosynthesis may be a more effective way to improve the total yield of saikosaponins.

In the biosynthetic pathway of triterpenoid saponins in higher plants, the mevalonate (MVA) pathway in the cytosol and the methylerythritol phosphate (MEP) pathway in the plastids are essential biosynthetic processes for formation of the triterpenoid backbone of the five-carbon intermediates isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). In transcriptome studies, genes involved in the biosynthetic pathway of saikosaponins were identified in *Bupleurum kaoi*, *B. chinense*, and *B. scorzonerifolium* (Chen et al., 2007; Sui et al., 2011; Sui et al., 2015). Moreover, genes involved in the biosynthesis of saikosaponins such as squalene epoxidase, β-amylase, cytochrome P450, and uridine diphosphate glycosyl transferases have been cloned in *B. kaoi*, *B. falcatum*, and *B. chinense*, and their expression profiles have been identified (Lin et al., 2006; Kim et al., 2011; Gao et al., 2016). Overexpression of the *BcbZIP134* gene in *B. chinense* and *BfSS1* in *B. falcatum* has been reported (Kim et al., 2011; Xu et al., 2019). Additionally, cytochrome P450 monooxygenase (P450) of *CYP716Y1* from *B. falcatum* was combined with oxidosqualene cyclase, P450, and glycosyltransferase genes to construct a synthetic biological platform for the production of bioactive triterpene sapo(ge)nins in yeast (Moses et al., 2014). However, the molecular regulatory mechanism underlying triterpenoid saponin biosynthesis remains elusive.

In this study, we investigated the sequential and expression characterization of 232 genes in the MVA pathway and MEP pathway in *B. chinense* and *B. scorzonerifolium* to identify putative genes involved in the biosynthesis of saikosaponins in *Bupleurum* L.

MATERIALS AND METHODS

Plant Materials

The two experimental materials, the commercial varieties Chuanbeichai No. 1 (CBC1) and Chuanhongchai No. 1 (CHC1), which are varieties of *B. chinense* and *B. scorzonerifolium*, respectively, were used. All of the plants were bred via systemic selection and purifying selection from farmholding populations by Dr. Jianhe Wei from the Institute of Medicinal Plant Development (IMPLAD), Chinese Academy of Medical Sciences and Peking Union Medical College, and Dr. Da-bin Hou from Southwest University of Science and Technology.

For each genotype, seeds were placed on moist filter paper before germination. Germinated seedlings were then grown in modified Hoagland's nutrient solution. Five-day-old and 15-day-old CBC1 and CHC1 plants were utilized for isoform sequencing (iso-seq) analysis, transcriptome analysis, and saikosaponin a (SS a) and d (SS d) content assays. For 5-day-old seedlings, whole fresh roots were harvested (S1). For 15-day-old seedlings, 5 mm of the root tip without the region of differentiation (S2) and with the region of differentiation (S3) were harvested separately.

Extraction of Saikosaponins and HPLC Analysis

Samples S1, S2, and S3 (three replications each) were dried for 72 h using a freeze-drier (LGJ-18, Beijing Songyuan Huaxing Technology Development Co., Ltd., China). The SS a and SS d content was determined using a Waters HPLC (high-performance liquid chromatography) system (Waters 1525 Binary HPLC Pump, United States) and an ASB-vensil C18 column (250 mm × 4.6 mm, 5 μm). Reference standards of SS a and SS d were purchased from the National Institutes for Food and Drug Control, Beijing, China. The methods and conditions for determination have been reported previously (Xu et al., 2019).

Iso-seq and Transcriptome Analyses

The leaf and root samples were mixed and utilized for iso-seq library construction. The iso-seq analysis followed the method published by Wang et al. (2017a). Transcriptome analyses of samples S1, S2, and S3 were performed on an Illumina HiSeq 2500 platform (Illumina, San Diego, CA, United States) as previously described (Yang et al., 2018). Both iso-seq and transcriptome analyses were performed at the Novogene Bioinformatics Institute (Novogene, Beijing, China). Three replications were included in this study.

Candidate Gene Selection

Gene sequences of 20 families involved in the MVA pathway and MEP pathway were selected from iso-seq and transcriptome data with annotated gene names. These genes include glycosyltransferases (UGT), P450, β-amyrin synthase (β-AS), squalene synthase (SS), squalene epoxidase (SE), farnesyl diphosphate synthase (FPS), mevalonate-5-pyrophosphate decarboxylase (MVD), phosphomevalonate kinase (PMK), mevalonate kinase (MK), 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR), 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS), acetyl-CoA C-acetyltransferase (AACT), 1-deoxy-D-xylulose-5-phosphate synthase (DXS), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), 2-C-methyl-D-erythritol 4-phosphate cytidylyl transferase (CMS), 4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol kinase (CMK), 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (MCS),

4-hydroxy-3-methyl but-2-(E)-enyl diphosphate synthase (HDS), 4-hydroxy-3-methyl but-2-(E)-enyl diphosphate reductase (IDS), isopentenyl diphosphate isomerase (IDI), and geranyl diphosphate synthase (GPS). CAP3 software was used to identify overlaps between different sequences and to remove redundant fragments[1]. The conserved domains of each family were analyzed by using SMART online software[2] and used as a query to search the NCBI non-redundant protein database via BLASTX to further validate the potential candidate genes[3]. The isoelectric points (PIs) and molecular weights (MWs) of deduced proteins were calculated using the ExPASy Compute pI/Mw tool[4], and the subcellular localizations were predicted using Cell-PLoc 2.0[5]. Differential expression analyses were performed by the DESeq2 R package (Anders and Huber, 2010) to identify differentially expressed genes (DEGs) among S1, S2, and S3 within each species and between the two species. The differentially expressed unigenes were further filtered based on their count number (at least one stage was greater than 1) and the log2(fold change) [the log2(fold change) between two stages was greater than 2].

RESULTS

SS a and SS d were detected in the roots of *B. chinense* plants at the 1-day-old stage, whereas they significantly accumulated in the region of differentiation at the 15-day-old stage (**Table 1**). In *B. scorzonerifolium*, no peaks of SS a and SS d were identified in any of the samples during the HPLC analysis. Therefore, the genes showed significantly different expression in S3 in *B. chinense*, but insignificant or the opposite expression in *B. scorzonerifolium* would be interesting.

A total of 223 genes with complete open reading frames (ORFs) were identified from the transcriptome database of *B. chinense* and *B. scorzonerifolium* (**Supplementary Table S1**). One to 96 genes were identified for a single gene family whose members are involved in triterpenoid saponin biosynthesis, although the MCS family had zero genes. In the differential expression analysis of *B. chinense*, 86 genes showed significantly different expression in the S3 sample (**Figures 1**, **2**). Among these genes, 1 to 36 genes were identified for gene families including UGT, P450, β-AS, SS, FPS, GPS, PMK, HMGCR, HMGS, AACT, IDS, HDS, CMS, DXR, and DXS. No gene with significantly different expression was detected for SE, MVD, IDI, MK, and CMK. Five genes involved the gene families β-AS, SS, CMS, HDS, and IDS; these genes exhibited significantly similar expression in *B. scorzonerifolium*. Seven genes in the FPS, HMGCR, DXS, and P450 gene families showed significantly reversed expression. The *BcDXS15975* gene in the DXS family was identified only in *B. chinense* and was not detected in any of the samples of *B. scorzonerifolium*.

[1]http://doua.prabi.fr/software/cap3

[2]http://smart.embl-heidelberg.de/

[3]http://www.ncbi.nlm.nih.gov/

[4]http://web.expasy.org/compute_pi/

[5]http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/

TABLE 1 | Contents of saikosaponin a and saikosaponin d in the roots of *B. chinense* and *B. scorzonerifolium*.

Species	Sample	SS a content (μg/g)	SD	SS d content (μg/g)	SD
B. chinense	S1	0	0	0	0
	S2	0	0	0	0
	S3	479.77	154.95	270.64	154.34
B. scorzonerifolium	S1	0	0	0	0
	S2	0	0	0	0
	S3	0	0	0	0

DISCUSSION

Triterpenoids are derived from C5 isoprene units of IPP and DMAPP through a "head-to-tail" connection (Hillier and Lathe, 2019). Both of these triterpenoid backbones can be synthesized through the MVA pathway in the cytoplasm or the MEP pathway in the plastids. For the MVA pathway, three HMGS genes and one HMGCR gene (*BcHMDH15256*) showed significantly upregulated expression in S3 of the *B. chinense* transcriptome, whereas one AACT gene, seven HMGCR genes, and one PMK gene showed significantly downregulated expression. For these genes, the expression differences among S1, S2, and S3 were insignificant in *B. scorzonerifolium*, with the exception of the *BcHMDH1723* gene (**Figure 1**). In addition, *BcHMDH1723* showed the opposite expression trend between *B. chinense* and *B. scorzonerifolium*. We did not detect genes with significantly different expression in the MK family or the MVD family.

For the MEP pathway, 12 unigenes encoding enzymes (five DXS genes, three CMS genes, two HDS genes, and two IDS genes) showed significantly downregulated expression in S3 of the *B. chinense* transcriptome, whereas the expression of two DXS genes, one DXR gene, one CMS gene, one HDS gene, and six IDS genes significantly increased. The CMS gene *BcISPD108768*, the HDS gene *BcISPG26774*, and the IDS gene *BcISPH33774* showed similar trends in the *B. scorzonerifolium* transcriptome, whereas the DXS gene *BcDXS14138* showed the opposite trend. The DXS gene *BcDXS15975* was detected only in the *B. chinense* transcriptome.

Both the MVA and MEP pathways produce the C5 unit IPP, which can be transformed into its isomer DMAPP by IDI (Xue et al., 2019). In the present study, we did not find that the IDI gene showed significant expression differences among S1, S2, and S3. IPP and DMAPP are assembled into GPP and FPP by GPS and FPS prenyltransferases, respectively. Expression of the GPS gene *BcGGPPS5581* increased in S3 of the *B. chinense* transcriptome, and expression of the FPS gene *BcFPPS50324* decreased. SS catalyzes the joining of two units of FPP in a "tail-to-tail" fashion of SQ. Kim et al. (2011) found that overexpression of *BfSS1* in *B. falcatum* more powerfully regulates downstream genes than does MeJA treatment in triterpene and phytosterol biosynthesis. In the present study, three candidate genes in the SS family showed significant expression differences in S3 of the *B. chinense* transcriptome, and the expression of all these candidate genes among the three samples in the *B. scorzonerifolium* transcriptome

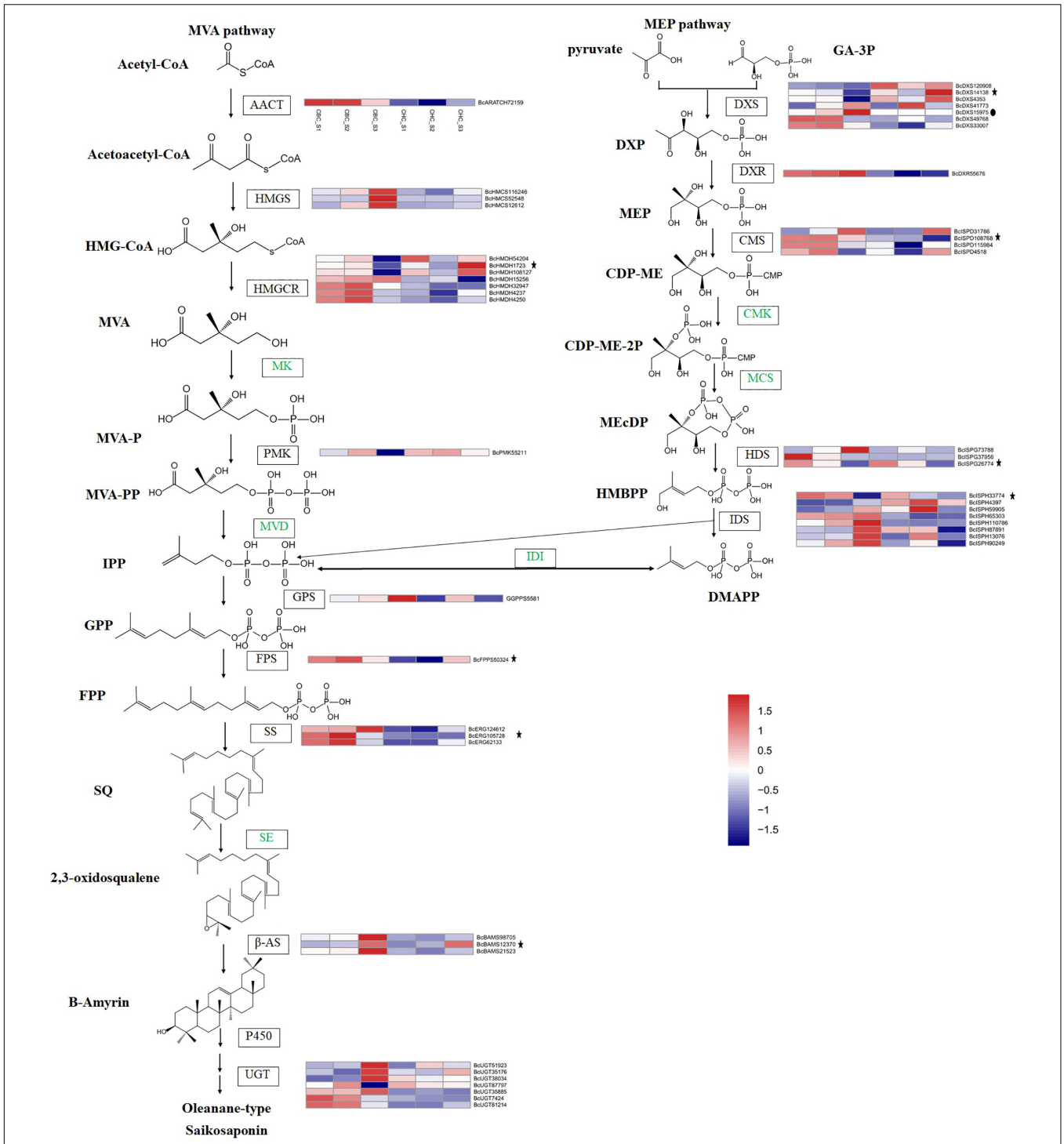

FIGURE 1 | Putative triterpenoid backbone biosynthesis pathway (MVA and MEP) in *Bupleurum chinense* DC. and *Bupleurum scorzonerifolium* Willd. and gene expression profiles of the key enzymes. All genes outside of these pathways showed significantly differential expression in S3 of *B. chinense*. The genes with asterisks also showed significantly differential expression in S3 of *B. scorzonerifolium*. The genes with solid dots indicate those detected only in *B. chinense* DC. The samples include S1 (F), S2 (T), and S3 (M) in *B. chinense* (CBC) and *B. scorzonerifolium* (CHC).

was not significant. SQ is oxidized by SE to give rise to 2,3-oxidosqualene. We cloned the full-length SE gene in *B. chinense* (Gao et al., 2016). However, no candidate gene was identified in the differential expression analysis.

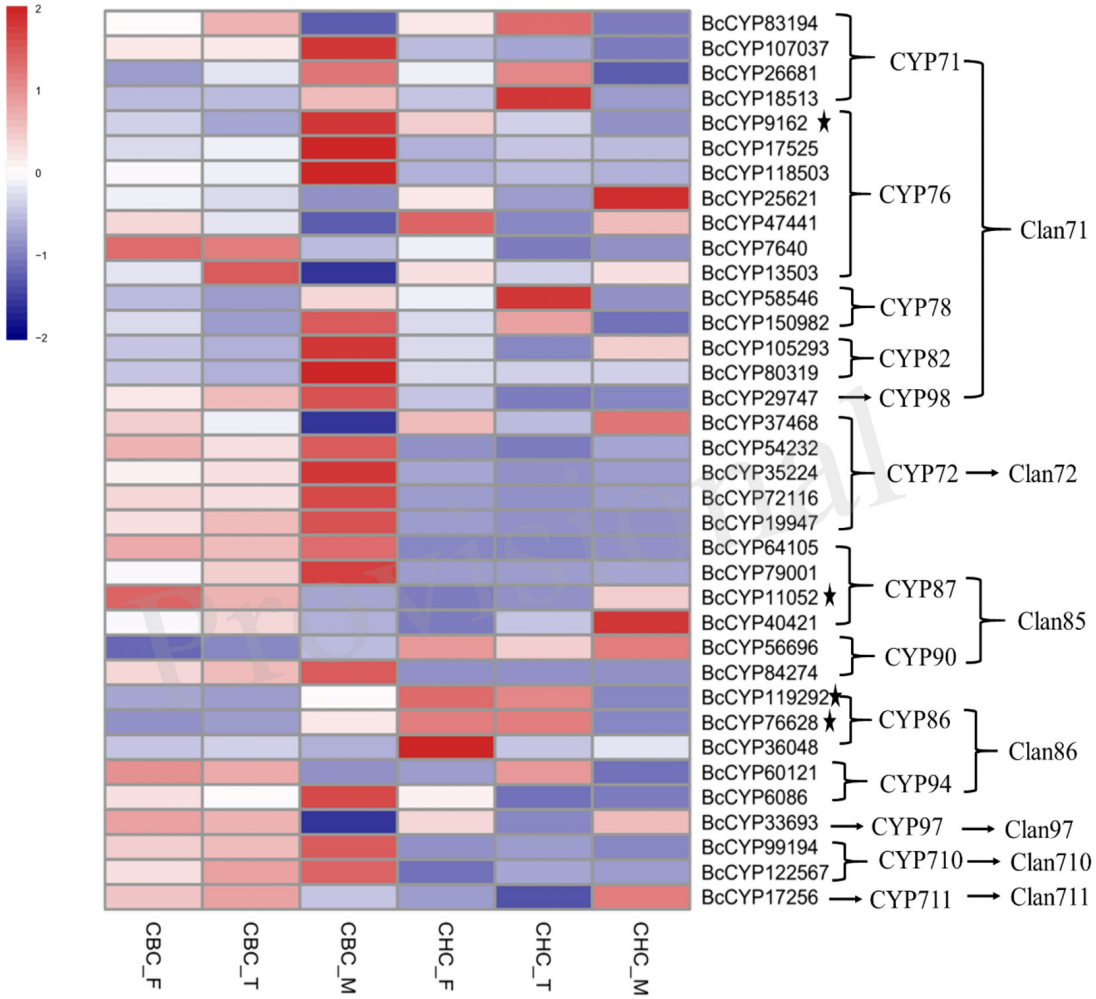

FIGURE 2 | Gene expression profiles of candidate P450s involved in the biosynthesis of triterpenoid saponins. All these genes showed significantly differential expression in S3 of *Bupleurum chinense* DC. The genes with asterisks also showed significantly differential expression in S3 of *Bupleurum scorzonerifolium* Willd. The sample information is listed in **Figure 1**.

In *Bupleurum*, the first committed step in the synthesis of triterpenoid saponins involves the cyclization of 2,3-oxidosqualene by β-AS. Therefore, β-AS is presumed to be the enzyme that catalyzes the first committed step in saikosaponin biosynthesis. A previous study indicated that β-AS in *B. kaoi* and *B. chinense* exhibits tissue-specific responses to both MeJA and PEG (Lin et al., 2013; Zhang et al., 2016). We have also cloned the promoter of β-AS in *B. chinense* and evaluated its activity previously (Gao et al., 2015). We found three β-AS genes (*BcBAMS98705*, *BcBAMS12370*, and *BcBAMS21523*) with significantly increased expression in S3 of *B. chinense*. The *BcBAMS12370* gene showed a similar expression trend in *B. scorzonerifolium*, whereas the expression of the other two genes was non-significant.

For the production of saikosaponins in *Bupleurum*, P450 enzymes catalyze the oxidation of β-amyrin to form 13,28-epoxy and the C11/C12 double-bond structure or two double bonds at C11/C12 and C13/C18 together with hydroxylation at C16 and C23. Previous studies have indicated that the CYP90, CYP72, CYP710, and CYP711 subfamilies of the P450 family may involve triterpenoid-metabolizing enzymes (Ohnishi et al., 2006; Hamberger and Bak, 2013; Yu et al., 2017). We found that 37 P450 genes showed significant expression in the three samples of *B. chinense*, and four of these genes exhibited a significant opposite trend in *B. scorzonerifolium* (**Figure 2**). These genes are members of 13 families of 7 clans in the P450 family. Among these genes, two unigenes belong to the CYP90 family, five unigenes belong to the CYP72 family, two unigenes belong to the CYP710 family, and the unigene *BcCYP17256* belongs to the CYP711 family. Members of the CYP716 family contribute to the diversification of eudicot triterpenoid biosynthesis (Miettinen et al., 2017). Moses et al. (2014) found that the *CYP716Y1* gene from *B. falcatum* catalyzes the C-16α hydroxylation of triterpenes in yeast. However, we did not identify any CYP716 gene that showed significantly different expression in S3 of *B. chinense*.

At the last step of saikosaponin biosynthesis, UGT catalyzes the glycosylation of hydroxylated β-amyrin at C3. Glycosylation contributes to the highly diverse nature of terpenoids in plants (Field and Osbourn, 2008). Therefore, UGT is presumed to be the key enzyme involved in the modification of saikosaponins. A transcriptome analysis of *B. chinense* revealed that, among 196 UGT genes, three were the most likely candidates involved in saikosaponin biosynthesis (Sui et al., 2011). Three *BkUGT85A* proteins that contain a highly conserved region of a motif of glycosyltransferases, which are involved in the production of secondary metabolites in plants, were identified in *B. kaoi* by Lin et al. (2013). In the present study, four UGT genes showed significantly increased expression in S3 of *B. chinens*e, whereas four genes showed the opposite expression trend. The expression differences among all genes in *B. scorzonerifolium* were non-significant.

CONCLUSION

In this study, unigenes showed significantly different expression in the S3 sample in *B. chinense*, while insignificant or the opposite expression in *B. scorzonerifolium* may be involved in saikosaponin biosynthesis. Manipulation of these genes, especially those of the β-AS, P450, and UGT families, may improve saikosaponin production. The combined expression of these genes and reconstituting the synthesis of monoglycosylated saponins in yeast may provide a platform for the production of bioactive saikosaponins.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in NCBI PRJNA645610 for *B. chinense* (SRR12223465-SRR12223473) and PRJNA662700 for B. scorzonerifolium (SRR12647620-SRR12647628).

AUTHOR CONTRIBUTIONS

MY wrote the manuscript. HC performed the data analysis. S-HL seeded Chuanbeichai No.1 and Chuanhongchai No.1. Y-CL performed the HPLC analysis. CS performed the transcriptome analysis. J-HW bred out the commercial varieties of Chuanbeichai No.1 and Chuanhongchai No.1. J-HW and D-BH funded the whole project and helped MY to complete the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China under grant number 81603223, the China Agriculture Research System under grant number CARS-21, the CAMS Innovation Fund for Medical Sciences (CIFMS) (2016-I2M-2-003), the National Transgenic Major Project of China (2019ZX08010004-005), the Natural Science Foundation of Science and Technology Department of Sichuan Province (2019YJ0443 and 2019YFH0072), and Young Elite Scientists Sponsorship Program by CACM (2019-QNRC2-C15).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.583245/full#supplementary-material

REFERENCES

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. doi: 10.1186/gb-2010-11-10-r106

Ashour, M. L., and Wink, M. (2011). Genus : a review of its phytochemistry, pharmacology and modes of action. *J. Pharm. Pharmacol.* 63, 305–321. doi: 10.1111/j.2042-7158.2010.01170.x

Chelghoum, M., Smati, D., Mitaine-Offer, A.-C., Paululat, T., and Lacaille-Dubois, M.-A. (2018). Four new triterpene saponins from *Bupleurum rigidum* L. *Phytochem. Lett.* 27, 223–228. doi: 10.1016/j.phytol.2018.07.025

Chen, L.-R., Chen, Y.-J., Lee, C.-Y., and Lin, T.-Y. (2007). MeJA-induced transcriptional changes in adventitious roots of *Bupleurum kaoi*. *Plant Sci.* 173, 12–24. doi: 10.1016/j.plantsci.2007.03.013

Chinese Pharmacopoeia Commission (2015). *The Pharmacopoeia of the People's Republic of China, 2015 Edition Part I*. Beijing: China Medical Science Press.

Field, B., and Osbourn, A. E. (2008). Metabolic diversification—independent assembly of operon-like gene clusters in different plants. *Science* 320, 543–547. doi: 10.1126/science.1154990

Gao, K., Wu, S.-R., Wang, L., Xu, Y.-H., Wei, J.-H., and Sui, C. (2015). Cloning and analysis of β-amyrin synthase gene in *Bupleurum chinense*. *Genes Genom.* 37, 767–774. doi: 10.1007/s13258-015-0307-0

Gao, K., Xu, J.-S., Sun, J., Xu, Y.-H., Wei, J.-H., and Sui, C. (2016). Molecular cloning and expression of squalene epoxidase from a medicinal plant. *Bupleurum chinense*. *Chinese Herb. Med.* 8, 67–74. doi: 10.1016/s1674-6384(16)60010-2

Gong, J., Liu, M., Xu, S., Jiang, Y., Pan, Y., Zhai, Z., et al. (2017). Effects of light deficiency on the accumulation of saikosaponins and the ecophysiological characteristics of wild *Bupleurum chinense* DC. in China. *Ind. Crops Products* 99, 179–188. doi: 10.1016/j.indcrop.2017.01.040

Guo, Y., Matsumoto, T., Kikuchi, Y., Ikejima, T., Wang, B., and Yamada, H. (2000). Effects of a pectic polysaccharide from a medicinal herb, the roots of *Bupleurum falcatum* L. on interleukin 6 production of murine B cells and B cell lines. *Immunopharmacology* 49, 307–316. doi: 10.1016/s0162-3109(00)00245-9

Hamberger, B., and Bak, S. (2013). Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philos. Trans. R. Soc. B-Biol. Sci.* 368:20120426. doi: 10.1098/rstb.2012.0426

Hillier, S. G., and Lathe, R. (2019). Terpenes, hormones and life: isoprene rule revisited. *J. Endocrinol.* 242, R9–R22. doi: 10.1530/JOE-19-0084

Ikegami, F., Sumino, M., Fujii, Y., Akiba, T., and Satoh, T. (2006). Pharmacology and toxicology of *Bupleurum* root-containing Kampo medicines in clinical use. *Hum. Exp. Toxicol.* 25, 481–494. doi: 10.1191/0960327106het654oa

Kim, Y. S., Cho, J. H., Park, S., Han, J.-Y., Back, K., and Choi, Y.-E. (2011). Gene regulation patterns in triterpene biosynthetic pathway driven by overexpression of squalene synthase and methyl jasmonate elicitation in *Bupleurum falcatum*. *Planta* 233, 343–355. doi: 10.1007/s00425-010-1292-9

Liang, Z., Oh, K., Wang, Y., Yi, T., Chen, H., and Zhao, Z. (2014). Cell type-specific qualitative and quantitative analysis of saikosaponins in three *Bupleurum* species using laser microdissection and liquid chromatography–quadrupole/time of flight-mass spectrometry. *J. Pharm. Biomed. Anal.* 97, 157–165. doi: 10.1016/j.jpba.2014.04.033

Lin, T.-Y., Chiou, C.-Y., and Chiou, S.-J. (2013). Putative genes involved in saikosaponin biosynthesis in *Bupleurum* species. *Int. J. Mol. Sci.* 14, 12806–12826. doi: 10.3390/ijms140612806

Lin, W.-Y., Peng, P.-H., and Lin, T.-Y. (2006). "Cloning and characterization of beta-amyrin synthase from *Bupleurum kaoi*," in *Proceedings of the 8th International Congress of Plant Molecular Biology*, (Adelaide, SA: Book of Abstracts, ISPMB).

Mabberley, D. J. (2008). *[Plant-book]; Mabberley's plant-book: a portable dictionary of plants, their classifications and uses; utilizing Kubitzki's" The families and genera of vascular plants" (1990-) and current botanical literature; arranged according to the principles of molecular systematics*. Cambridge: Cambridge University Press.

Miettinen, K., Pollier, J., Buyst, D., Arendt, P., Csuk, R., Sommerwerk, S., et al. (2017). The ancient CYP716 family is a major contributor to the diversification of eudicot triterpenoid biosynthesis. *Nat. Commun.* 8, 1–13. doi: 10.1038/ncomms14153

Moses, T., Pollier, J., Almagro, L., Buyst, D., Van Montagu, M., Pedreño, M. A., et al. (2014). Combinatorial biosynthesis of sapogenins and saponins in Saccharomyces cerevisiae using a C-16α hydroxylase from *Bupleurum falcatum*. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1634–1639. doi: 10.1073/pnas.1323369111

Ohnishi, T., Szatmari, A.-M., Watanabe, B., Fujita, S., Bancos, S., Koncz, C., et al. (2006). C-23 hydroxylation by Arabidopsis CYP90C1 and CYP90D1 reveals a novel shortcut in brassinosteroid biosynthesis. *Plant cell* 18, 3275–3288. doi: 10.1105/tpc.106.045443

Pistelli, L., Bertoli, A., Bilia, A. R., and Morelli, I. (1996). Minor constituents from *Bupleurum fruticosum* roots. *Phytochemistry* 41, 1579–1582. doi: 10.1016/0031-9422(95)00749-0

Sui, C., Chen, M., Xu, J., Wei, J., Jin, Y., Xu, Y., et al. (2015). Comparison of root transcriptomes and expressions of genes involved in main medicinal secondary metabolites from *Bupleurum chinense* and *Bupleurum scorzonerifolium*, the two Chinese official Radix bupleuri source species. *Physiol. Plant.* 153, 230–242. doi: 10.1111/ppl.12254

Sui, C., Zhang, J., Wei, J., Chen, S., Li, Y., Xu, J., et al. (2011). Transcriptome analysis of *Bupleurum chinense* focusing on genes involved in the biosynthesis of saikosaponins. *BMC Genom.* 12:539. doi: 10.1186/1471-2164-12-539

Tan, L. L., Cai, X., Hu, Z. H., and Ni, X. L. (2008). Localization and dynamic change of saikosaponin in root of *Bupleurum chinense*. *J. Int. Plant Biol.* 50, 951–957. doi: 10.1111/j.1744-7909.2008.00668.x

Wang, T., Wang, H., Cai, D., Gao, Y., Zhang, H., Wang, Y., et al. (2017a). Comprehensive profiling of rhizome-associated alternative splicing and alternative polyadenylation in moso bamboo (Phyllostachys edulis). *Plant J.* 91, 684–699. doi: 10.1111/tpj.13597

Wang, Y., Guo, Q., Cheng, Z., Zeng, K., Liang, H., Tu, P., et al. (2017b). New saikosaponins from the roots of *Bupleurum chinense*. *Phytochem. Lett.* 21, 183–189. doi: 10.1016/j.phytol.2017.06.005

Xu, J., Wu, S.-R., Xu, Y.-H., Ge, Z.-Y., Sui, C., and Wei, J.-H. (2019). Overexpression of BcbZIP134 negatively regulates the biosynthesis of saikosaponins. *Plant Cell Tissue Organ Culture (PCTOC)* 137, 297–308. doi: 10.1007/s11240-019-01571-0

Xue, L., He, Z., Bi, X., Xu, W., Wei, T., Wu, S., et al. (2019). Transcriptomic profiling reveals MEP pathway contributing to ginsenoside biosynthesis in Panax ginseng. *BMC Genom.* 20:383. doi: 10.1186/s12864-019-5718-x

Yang, L., Jin, Y., Huang, W., Sun, Q., Liu, F., and Huang, X. (2018). Full-length transcriptome sequences of ephemeral plant Arabidopsis pumila provides insight into gene expression dynamics during continuous salt stress. *BMC Genom.* 19:717. doi: 10.1186/s12864-018-5106-y

Yu, J., Tehrim, S., Wang, L., Dossa, K., Zhang, X., Ke, T., et al. (2017). Evolutionary history and functional divergence of the cytochrome P450 gene superfamily between *Arabidopsis thaliana* and Brassica species uncover effects of whole genome and tandem duplications. *Bmc Genom.* 18:733. doi: 10.1186/s12864-017-4094-7

Zhang, Y., Zhou, Z., Xia, P., Liang, Z., Liu, S., and Liu, Z. (2016). Expression of key enzyme genes and content of saikosaponin in saikosaponin biosynthesis under drought stress in *Bupleurum chinense*. *China J. Chinese Mater. Med.* 41, 643–647.

Zhao, Z., and Xiao, P. (2007). *Encyclopedia on Contemporary Medicinal Plants*. Shang Hai: World Publishing Corporation.

Zhu, Z., Liang, Z., and Han, R. (2009a). Growth and saikosaponin production of the medicinal herb *Bupleurum chinense* DC. under different levels of nitrogen and phosphorus. *Ind. Crops Products* 29, 96–101. doi: 10.1016/j.indcrop.2008.04.010

Zhu, Z., Liang, Z., and Han, R. (2009b). Saikosaponin accumulation and antioxidative protection in drought-stressed *Bupleurum chinense* DC. plants. *Environ. Exp. Botany* 66, 326–333. doi: 10.1016/j.envexpbot.2009.03.017

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 October 2020
doi: 10.3389/fgene.2020.602495

Identification and Validation of a Novel Major Quantitative Trait Locus for Plant Height in Common Wheat (*Triticum aestivum* L.)

*Zhiqiang Wang[1†], Haiyan Hu[2†], Xiaojun Jiang[1], Yang Tao[3], Yu Lin[1], Fangkun Wu[1], Shuai Hou[1], Shihang Liu[1], Caixia Li[1], Guangdeng Chen[4] and Yaxi Liu[1,5]**

[1] Triticeae Research Institute, Sichuan Agricultural University, Chengdu, China, [2] School of Life Sciences and Technology, Henan Institute of Science and Technology, Xinxiang, China, [3] Rice Research Institute, Sichuan Agricultural University, Chengdu, China, [4] College of Resources, Sichuan Agricultural University, Chengdu, China, [5] State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Chengdu, China

OPEN ACCESS

Edited by:
Penghao Wang,
Murdoch University, Australia

Reviewed by:
Hai Long,
Chinese Academy of Sciences, China
Zehou Liu,
Sichuan Academy of Agricultural Sciences, China
Zitong Yu,
Kansas State University, United States

***Correspondence:**
Yaxi Liu
liuyaxi@sicau.edu.cn
orcid.org/0000-0001-6814-7218

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: 03 September 2020
Accepted: 02 October 2020
Published: 22 October 2020

Citation:
Wang Z, Hu H, Jiang X, Tao Y, Lin Y, Wu F, Hou S, Liu S, Li C, Chen G and Liu Y (2020) Identification and Validation of a Novel Major Quantitative Trait Locus for Plant Height in Common Wheat (Triticum aestivum L.). Front. Genet. 11:602495. doi: 10.3389/fgene.2020.602495

Plant height (PH) plays a pivotal role in plant morphological architecture and is associated with yield potential in wheat. For the quantitative trait locus (QTL) analysis, a recombinant inbred line population was developed between varieties differing significantly in PH. Two major QTL were identified on chromosomes 4B (*QPh.sicau-4B*) and 6D (*QPh.sicau-6D*) in multiple environments, which were then validated in two different backgrounds by using closely linked markers. *QPh.sicau-4B* explained 10.1–21.3% of the phenotypic variance, and the location corresponded to the dwarfing gene *Rht-B1*. *QPh.sicau-6D* might be a novel QTL for PH, explaining 6.6–13.6% of the phenotypic variance and affecting spike length, thousand-kernel weight, and spikelet compactness. Three candidate genes associated with plant growth and development were identified in the physical interval of *QPh.sicau-6D*. Collectively, we identified a novel stable and major PH QTL, *QPh.sicau-6D*, which could aid in the development of closely linked markers for marker-assisted breeding and cloning genes underlying this QTL.

Keywords: wheat, plant height, quantitative trait locus, validation, candidate gene

INTRODUCTION

Bread wheat (*Triticum aestivum* L.) is an important staple crop, ranking the third after maize and rice in terms of yield in China (Edae et al., 2014; Liu et al., 2018). According to the Food and Agriculture Organization of the United Nations[1], the global wheat grain yield in 2017 was 771.7 million tons, contributing to approximately 20% of the calories consumed by humans. Plant height (PH) is an important yield component trait associated with plant morphological architecture and other yield-related traits, such as spike length, spikelet number per spike, spikelet compactness (SC), and thousand-kernel weight (TKW), thus affecting the yield potential (Sakamoto and Matsuoka, 2004; Gao et al., 2015; Kowalsk et al., 2016; Guan et al., 2018). To develop high grain yield lines, Donald (1968) proposed the idea of breeding crop ideotypes with a relatively short PH, single culm, strong stem, and large and erect ear.

[1] http://www.fao.org/faostat/en/#data

In view of the importance of PH in wheat yield, it is imperative to identify more candidate genes responsible for PH from wheat germplasm resources. In the past two decades, numerous major and minor QTL influencing PH have been identified on 21 chromosomes in wheat, and some of them have been applied in wheat breeding (Peng et al., 1999; Liu et al., 2002; Griffiths et al., 2012; Würschum et al., 2015; Tian et al., 2017; Hassan et al., 2019). Additionally, several PH genes have been cloned, such as *Rht-B1* and *Rht-D1* (located on chromosome 4B and 4D, respectively), and highly adopted in breeding practices during the green revolution; they encode the DELLA proteins, participating in gibberellin signaling, and thereby affecting PH (Peng et al., 1999; Pearce et al., 2011). *Rht18* encodes a gibberellic acid (GA) 2-oxidase protein, which regulates the balance of GA intermediates and inactive GA, leading to a semi-dwarf phenotype in wheat (Ford et al., 2018). In *Arabidopsis*, extensively studied dwarf mutants such as the *yda* and *pat10* mutants, which are defective in growth and development, have been shown to significantly differ from the wild-type plants in terms of PH (Lukowitz et al., 2004; Zhou et al., 2013).

In this study, a recombinant inbred lines (RILs) population was used for QTL mapping of PH with a genetic map using the 90K SNP array and phenotyping in six environments to identify major QTL for PH. The effects of the major QTL for PH were further assessed in different genetic backgrounds.

MATERIALS AND METHODS

Plant Materials

Three populations of RILs were generated by single-seed descent in the field in Sichuan Agricultural University, Wenjiang (103°51′E, 30°43′N), with H461 as a common parent. These three populations were as follows: H461/CN16 (HCN; 249 F_8 lines), H461/CM107 (HCM; 200 F_7 lines), and H461/MM37 (HMM; 142 F_6 lines).

The HCN population was used for QTL mapping, whereas the other populations (HCM and HMM) were used for validating the major QTL identified in the HCN population.

Phenotypic Evaluation

The three populations were planted in six different environments for phenotypic evaluation: Wenjiang in 2015 and 2019 (2015WJ and 2019WJ); Chongzhou (103°38′E, 30°32′N) in 2015, 2017, and 2019 (2015CZ, 2017CZ, and 2019CZ); and Ya'an (103°0′E, 29°58′N) in 2015 (2015YA). Each plot consisted of three rows, with a length of 1.5 m and an inter-row spacing of 30 cm; the sowing density was 15 seeds per row. For each plot, five plants were randomly chosen to measure PH, from the plant base to the tip of the spike, and calculate the mean PH. The main spike of five plants were selected to measure the spikelet number per spike (SN) and spike length (SL). The TKW was measured using an electronic balance with three replications. Flowering time (FT) was recorded as the date when half of the plants in each plot flowered after sowing. The SC was calculated by dividing the SL by the SN.

Analysis of variance (ANOVA) and calculation of Pearson's correlation coefficients among different environments were performed using SPSS 22 (IBM SPSS, Armonk, NY, United States). Frequency distribution was processed using MS Excel, and the best linear unbiased prediction (BLUP) for target traits was calculated using R version 3.5.2 (Team, 2013). Broad-sense heritability (h^2) was calculated across environments as described by Smith et al. (1998). The correlations between PH and the factors SN, SL, SC, TKW, and FT were calculated based on the BLUP values, and Student's *t*-test was performed to determine significant differences between two groups using SPSS 22.

QTL Mapping

The HCN population was used for constructing a whole-genome genetic linkage map using the 90K SNP array (Wang et al., 2016) for QTL mapping, consisting of 7808 SNP polymorphic markers in parents distributed in 50 linkage groups and covers a total genetic distance of 3486.44 cM, with an average distance of 0.45 cM between the adjacent markers.

MapQTL 6.0 (Van Ooijen and Kyazma, 2009) was used for the QTL analysis. Kruskal–Wallis test was used to evaluate the degree of association between markers and PH. Interval mapping (IM) was then used to identify major QTL and markers significantly associated with PH. For each trial, a test of 1000 permutations was performed to identify the LOD threshold corresponding to a genome-wide false discovery rate of 1%. Based on the permutation test, threshold LOD values between 2.4 and 3.3 were used to confirm the presence of a QTL. The QTL were named based on the International Rules of Genetic Nomenclature[2]. "Ph" and "sicau" stand for "plant height" and "Sichuan Agricultural University," respectively.

Validation of the Major QTL

The flanking markers of the major QTL were mapped to the physical map of the wheat cultivar Chinese Spring (IWGSC RefSeq v1.0), and the sequence information in the QTL interval was obtained. To develop Kompetitive allele specific PCR (KASP) markers closely linked to the QTL, the partial sequence information of the QTL interval was amplified in CN16 and H461 by PCR to search for polymorphic sites. The newly developed KASP markers were remapped into the genetic map.

The markers closely linked to the QTL were used for identifying alleles in different genetic backgrounds (populations HCM and CMM). The lines were classified into two groups: genotypes with homozygous alleles from H461 (designated AA) and those with homozygous alleles from alternative parents (designated BB). The mean PH from homozygotes was used for measuring the QTL effects, and Student's *t*-test was used to determine the significance of differences between the two groups in each population.

Predicted Candidate Genes

The gene information of the QTL interval was obtained from IWGSC RefSeq v1.1 annotation. Expression values as transcripts per million (TPM) were obtained from the expVIP Wheat

[2]http://wheat.pw.usda.gov/ggpages/wgc/98/Intro.htm

TABLE 1 | Phenotypic variation of the mapping population H461 × CN16 and parental lines in different environments.

Trait	Environment	Parents		Population			
		H461	CN16	Range	Mean	SD	h^2
PH	2015WJ	79.00**	61.60	64.32–115.84	84.72	8.08	
	2015YA	82.80**	74.02	62.47–114.44	83.25	8.87	
	2015CZ	85.76**	75.20	64.42–124.20	87.45	8.89	
	2017CZ	85.13**	76.75	55.33–111.17	79.73	9.14	
	2019WJ	86.94**	70.61	51.50–109.00	78.60	8.36	
	2019CZ	88.22**	76.17	58.10–110.78	81.10	7.80	
	BLUP	82.33	73.75	63.01–105.93	80.53	5.71	0.83
SN	BLUP	22.50	19.71	19.75–23.55	21.30	0.55	0.65
SL	BLUP	14.06	10.92	10.41–13.95	12.13	0.56	0.72
SC	BLUP	1.60	1.80	1.51–1.99	1.77	0.09	0.80
TKW	BLUP	52.49	44.66	38.99–58.31	49.41	3.35	0.81
FT	BLUP	142.06	141.10	137.78–149.16	141.75	2.12	0.84

*SD, standard deviation; h^2, broad-sense heritability; BLUP, phenotype values based on BLUP. **indicates significant differences at $P < 0.01$.*

Expression Browser[3] (Borrill et al., 2016), genes with a low expression (TPM < 0.5) in various tissues were excluded and the mean expression values were visualized by TBtools (Chen et al., 2020). The remaining genes were annotated by KOBAS v3.0 (Ai and Kong, 2018) BLAST against the corresponding protein sequences in rice and *Arabidopsis thaliana*. The genomic DNA of parents was extracted from the leaf samples using the Plant Genomic DNA kit (Biotechnologies, CA) and used to amplify candidate genes for sequence analysis.

RESULTS

Phenotyping of the HCN Population

In different environments, the PH of H461 ranged from 79.00 to 88.22 cm, and that of CN16 ranged from 61.60 to 76.75 cm. Moreover, significant differences in PH were observed between H461 and CN16 (**Table 1**). The frequency of PH in the HCN population showed continuous distribution, ranging from 51.50 to 124.2 cm (**Table 1** and **Supplementary Figure S1**); this implied that PH was affected by multiple loci. The h^2 of PH was 0.83, and Pearson's correlation coefficients between the different environments ranged from 0.232 to 0.872 ($P < 0.01$; **Table 2**).

The BLUP values of PH, SN, SL, SC, TKW, and FT are shown in **Table 1**. Phenotypic correlation coefficients between PH and other spike-related traits were obtained based on the BLUP values (**Table 3**). PH was highly significantly correlated with the TKW ($P < 0.01$) and significantly correlated with the SN and SL ($P < 0.05$). No significant correlation was observed between PH and SC or FT (**Table 3**).

Identification of QTL for PH

Three QTL for PH were identified using the IM analysis (**Table 4**). The first QTL (*QPh.sicau-4B*) was located on the short arm of chromosome 4B, between the markers Tdurum_contig64772_417 and Excalibur_rep_c113261_400. *QPh.sicau-4B* was a stable major QTL with the additive effects from H461, and it explained 10.1–21.3% of the phenotypic variance, with LOD values ranging from 4.15 to 9.39. It was identified in five environments and the combined analysis (BLUP). The second QTL (*QPh.sicau-6D*) was located on the short arm of chromosome 6D, between the markers IACX10982 and BS00063175_51. *QPh.sicau-6D* was a stable major QTL with the additive effects from H461; it explained 6.6%–13.6% of the phenotypic variance, with LOD values ranging from 2.67 to 5.80, identified in all environments and using BLUP. The third QTL

[3]http://www.wheat-expression.com/

TABLE 2 | Correlation coefficients for plant height (PH) in the HCN population evaluated in different environments.

	2015WJ	2015YA	2015CZ	2017CZ	2019WJ
2015YA	0.334**				
2015CZ	0.549**	0.661**			
2017CZ	0.613**	0.242**	0.444**		
2019WJ	0.570**	0.232**	0.405**	0.633**	
2019CZ	0.632**	0.314**	0.477**	0.644**	0.872**

***indicates significant differences at $P < 0.01$.*

TABLE 3 | Correlation coefficients among the BLUP value for plant height (PH) with spikelet number per spike (SN), spike length (SL), spikelet compactness (SC), thousand kernel weight (TKW) and flowering time (FT).

Trait	PH
SN	0.180*
SL	0.182*
SC	−0.075
TKW	0.249**
FT	0.073

***and *indicate significant correlations at $P < 0.05$ and $P < 0.01$, respectively.*

TABLE 4 | Quantitative trait loci (QTL) for plant height identified in the H461 × CN16 recombinant inbred line population evaluated in different environments.

QTL	Environment	Interval (cM)	Flanking markers	LOD	PVE (%)[a]	Add[b]
QPh.sicau-4B	2015WJ	94.00~102.26	Tdurum_contig64772_417 and BS00023766_51	9.39	21.3	3.79
	2015CZ	94.00~101.67	Tdurum_contig64772_417 and Excalibur_rep_c113261_400	7.76	17.9	3.97
	2015YA	94.00~101.67	Tdurum_contig64772_417 and Excalibur_rep_c113261_400	5.19	12.7	3.35
	2019WJ	94.00~102.26	Tdurum_contig64772_417 and BS00023766_51	4.15	10.1	2.70
	2019CZ	94.00~102.26	Tdurum_contig64772_417 and BS00023766_51	4.84	11.6	2.70
	BLUP	94.00~101.67	Tdurum_contig64772_417 and Excalibur_rep_c113261_400	6.19	14.4	2.28
QPh.sicau-6D	2015WJ	25.02~31.43	Kukri_c34967_226 and BS00063175_51	3.97	9.6	2.54
	2015CZ	25.02~31.43	Kukri_c34967_226 and BS00063175_51	3.49	8.5	2.64
	2015YA	28.53~31.43	IACX10982 and BS00063175_51	3.30	8.3	2.58
	2017CZ	25.02~31.43	Kukri_c34967_226 and BS00063175_51	5.71	13.4	3.37
	2019WJ	25.02~31.43	Kukri_c34967_226 and BS00063175_51	2.84	7.0	2.24
	2019CZ	25.02~31.43	Kukri_c34967_226 and BS00063175_51	2.67	6.6	2.03
	BLUP	28.53~31.43	IACX10982 and BS00063175_51	5.80	13.6	2.13
QPh.sicau-3B	2017CZ	96.74~98.46	BS00099633_51 and Kukri_c6907_80	3.27	7.9	2.65

[a] *Percentage of the phenotypic variation explained.* [b] *Additive effect.*

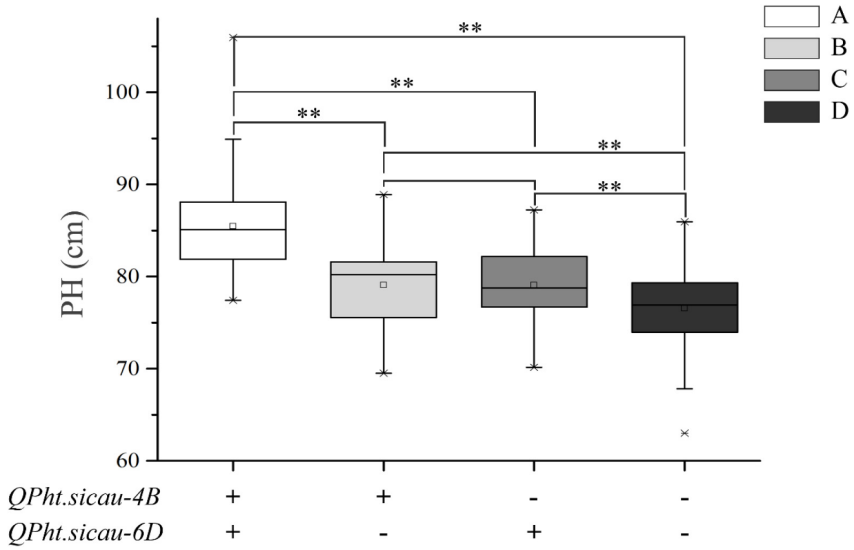

FIGURE 1 | Effects of the *QPh.sicau-4B* and *QPh.sicau-6D* on PH in HCN population (A: carrying both the additive alleles of the two major QTL; B: only carrying the additive allele of *QPh.sicau-4B*; C: only carrying the additive allele of *QPh.sicau-6D*; and D: not carrying the additive alleles of the two major QTL). **indicates significant differences at $P < 0.01$.

(*QPh.sicau-3B*) was located on 3B, identified only in 2017CZ, and it explained 7.9% of the phenotypic variance.

Effects of the Two Major PH QTL on PH and Other Panicle Traits

To identify the effect of the two major PH QTL (*QPh.sicau-4B* and *QPh.sicau-6D*) for other panicle traits, the BLUP values across six environments were used. For *QPh.sicau-4B*, lines with homozygous alleles from H461 and lines with homozygous alleles from CN16, classified into two groups, showed a significant difference ($P < 0.05$) for FT (**Supplementary Figure S2**). For *QPh.sicau-6D*, lines with homozygous alleles from H461 and those with homozygous alleles from CN16, classified into two groups, showed significant differences ($P < 0.05$) for SL, SC, and TKW (**Supplementary Figure S3**).

For PH, the HCN population could be divided into the following four groups based on markers: (A) carrying both the additive alleles of two major QTL, (B) only carrying the additive allele of *QPh.sicau-4B*, (C) only carrying the additive allele of *QPh.sicau-6D*, and (D) not carrying the additive alleles of *QPh.sicau-4B* and *QPh.sicau-6D*. Comparative analyses among the four groups showed that the group A had the highest effect on PH, which was significantly higher than that of the groups B, C, and D. Furthermore, the groups B and C had significantly higher effects than that of the group D. Thus, *QPh.sicau-4B* and *QPh.sicau-6D* might significantly affect PH, with both having a significant effects on PH (**Figure 1**).

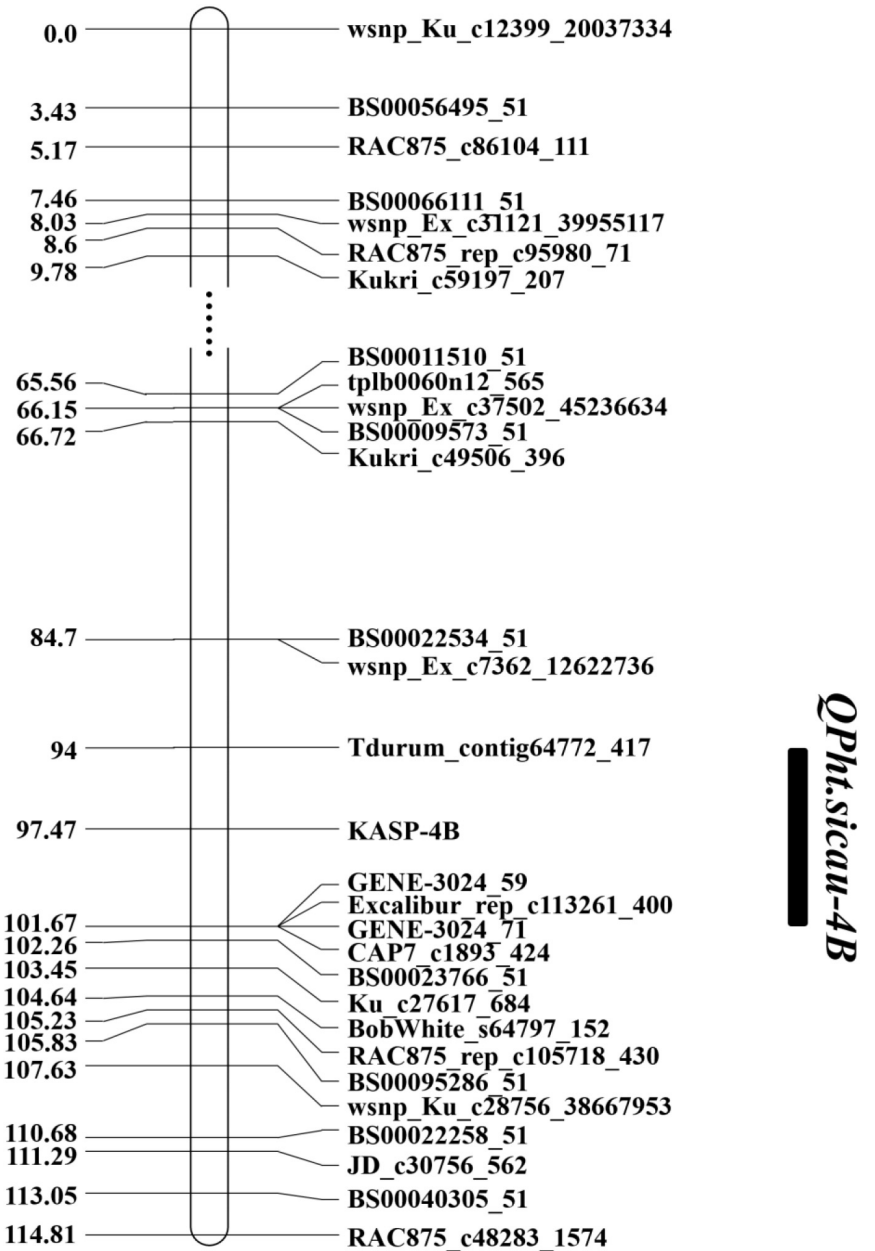

FIGURE 2 | Genetic map of the major QTL *QPh.sicau-4B* with the marker KASP-4B.

Validation of QTL in Different Genetic Backgrounds

Based on the QTL mapping results and Sanger sequencing of the PCR products of H461 and CN16, two KASP markers (KASP-4B and KASP-6D, **Supplementary Table S1**) were developed and used to reconstruct the genetic map. The KASP-4B marker was found to be closely linked to *QPh.sicau-4B*, whereas the KASP-6D marker was closely linked to *QPh.sicau-6D* (**Figures 2**, **3**).

Two populations (HCM and HMM) were used for evaluating the effects of the two major QTL in different genetic backgrounds, and the KASP markers were used to identify the genotype. For *QPh.sicau-4B*, KASP-4B was used to identify the alleles in the HCM and HMM populations and were classified into two groups. Significant differences ($P < 0.05$) were detected between "AA" and "BB" genotypes in three environments for HCM and four environments for HMM (**Table 5**). The differences in PH ranged from 2.56 to 9.24% in the HCM and HMM populations. For *QPh.sicau-6D*, KASP-6D was used to identify the alleles in the HCM and HMM populations and were classified into two groups. Significant differences ($P < 0.05$) were detected between "AA" and "BB" genotypes in four environments for HCM and four environments for HMM (**Table 6**). The differences in PH ranged

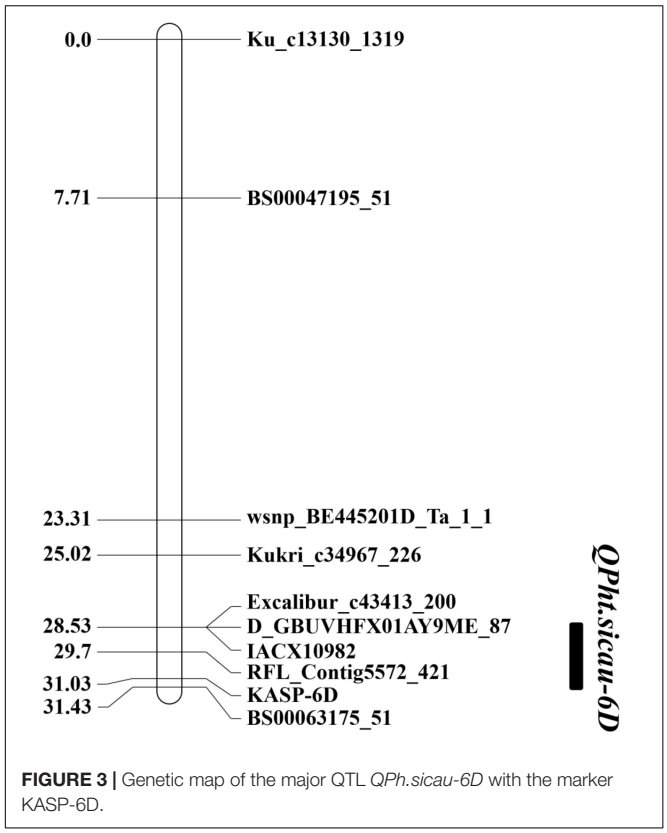

FIGURE 3 | Genetic map of the major QTL *QPh.sicau-6D* with the marker KASP-6D.

from 3.77 to 12.41% in the HCM and HMM populations. And the effects of Qph.sicau-6D was higher than Qph.sicau-4B in the validation populations, which may be responsible by different genetic backgrounds.

Potential Candidate Genes

A total of 224 high-confidence (HC) genes were selected from the *QPh.sicau-4B* and *QPh.sicau-6D* intervals, and 62 HC genes with low expression in various tissues were excluded (**Figure 4**). Finally, 47 HC genes from *QPh.sicau-4B* were selected for gene annotation, including *Rht-B1*. And 115 HC genes from *QPh.sicau-6D* were selected for gene annotation, including *YDA*, *BUB1*, and *PAT10* (**Supplementary Table S2** and **Table 7**).

Sequence analysis of the four candidate genes, revealed 2 SNPs and 1 insertion/deletion (indel) between H461 and CN16 for *Rht-B1*, one T for C substitution (C(190)T) converts the codon (CGA) to a translational stop codon (TGA) in CN16 (**Figure 5A**). For *TraesCS6D02G227300*, 1 SNP in the intron region between H461 and CN16 was found (**Figure 5B**). Five SNPs in the coding sequence, 4 SNPs in the intron region, and 1 SNP in the promoter region for *TraesCS6D02G233000* were detected between H461 and CN16 (**Figure 5C**). However, no sequence variation between H461 and CN16 was identified for *TraesCS6D02G234900*.

DISCUSSION

Plant height is a critical trait that influences plant architecture and grain yield potential in wheat, and it is controlled by multiple genes functioning together (Spielmeyer et al., 2007; Singh et al., 2016). Exploring PH QTL and genes is essential for wheat breeding, and the identification of QTL associated with PH on different chromosomes has been widely reported (Zhang et al., 2010; Liu et al., 2014; Gao et al., 2015; Chai et al., 2018). In this study, two stable and major QTL for PH were identified in different environments and were validated in different genetic backgrounds. *QPh.sicau-4B* was located in a 7.67-cM interval and mapped between 26.49 and 31.88 Mb on the physical map of chromosome 4B (**Figures 2**, **4**). *QPh.sicau-6D* was located in a 2.9-cM interval and mapped between 315.06 and 339.69 Mb on the physical map of chromosome 6D (**Figures 3**, **4**).

QPh.sicau-6D Is a Novel QTL for PH

QPh.sicau-6D was physically mapped between 315.06 and 339.69 Mb of chromosome 6D. Several known QTL responsible for PH have been mapped on chromosome 6D, including *QPh.spa-6D*, and *QPh.cau-6D* (**Supplementary Table S3**). We

TABLE 5 | Effects of *QPh.sicau-4B* in two validation populations.

Population	Environment	Parent[a]		AA[b]	BB[c]	Difference	*P*-value
		Parent1	Parent2				
HCM	2015YA	82.80	76.25	92.50	89.43	3.43%*	<0.05
HCM	2015WJ	79.00	71.19	93.07	90.22	3.16%*	<0.05
HCM	2015CZ	85.76	73.22	94.63	90.61	4.44%**	<0.01
HCM	2017CZ	85.13	84.44	86.25	84.03	2.64%	0.09
HCM	BLUP	82.33	72.3	89.58	87.34	2.56%*	<0.05
HMM	2015WJ	79.00	58.36	84.29	77.42	8.87%**	<0.01
HMM	2017CZ	85.13	80.33	86.80	79.46	9.24%**	<0.01
HMM	2019WJ	86.94	70.83	75.82	70.79	7.11%*	<0.05
HMM	2019CZ	88.22	56.25	82.67	76.21	8.48%**	<0.01
HMM	BLUP	82.33	70.53	82.09	76.14	7.81%**	<0.01

***and *indicate extremely significant difference and significant difference, respectively. [a]"Parent1" represents H461, "Parent2" represents CM107 or MM37. [b]"AA" represents homozygous alleles from H461. [c]"BB" represents homozygous alleles from non-H461 parents.*

TABLE 6 | Effects of *QPh.sicau-6D* in two validation populations.

Population	Environment	Parent[a]		AA[b]	BB[c]	Difference	*P*-value
		Parent1	Parent2				
HCM	2015YA	82.80	76.25	93.92	89.27	5.21%**	<0.01
HCM	2015WJ	79.00	71.19	93.19	88.45	5.36%**	<0.01
HCM	2015CZ	85.76	73.22	94.76	89.61	5.75%**	<0.01
HCM	2017CZ	85.13	84.44	86.78	83.56	3.85%**	0.01
HCM	BLUP	82.33	72.3	90.01	86.74	3.77%**	<0.01
HMM	2015WJ	79.00	58.36	85.95	77.25	11.26%**	<0.01
HMM	2017CZ	85.13	80.33	89.50	79.62	12.41%**	<0.01
HMM	2019WJ	86.94	70.83	76.25	71.29	6.96%*	<0.05
HMM	2019CZ	88.22	56.25	83.02	76.70	8.24%**	<0.01
HMM	BLUP	82.33	70.53	82.96	76.36	8.64%**	<0.01

***and *indicate extremely significant difference and significant difference, respectively. [a]"Parent1" represents H461, "Parent2" represents CM107 or MM37. [b]"AA" represents homozygous alleles from H461. [c]"BB" represents homozygous alleles from non-H461 parents.*

tried to compare the physical positions of these QTL to assess their relationship with *QPh.sicau-6D*. Physical mapping showed that *QPh.spa-6D* (Singh et al., 2016) was mapped to intervals 3.6–5.8 Mb of the chromosome 6D, was far away from *QPh.sicau-6D*, indicating that they were different loci. The confidence interval of *QPh.cau-6D* (Guan et al., 2018) was mapped between 283.97 and 292.07 Mb on the physical map. No physical interval of *QPh.cau-6D* overlapped with *QPh.sicau-6D*, verifying that they were different loci. These comparisons implied that *QPh.sicau-6D* was likely a novel QTL for PH.

QPh.cau-4B.2 (Guan et al., 2018) and *QPH.caas-4BS.2* (Gao et al., 2015) were mapped to intervals 29.0–35.5 Mb and 21.4–46.6 Mb, respectively, in the 4B chromosome physical map (**Supplementary Table S3**). Furthermore, the two QTL overlapped with *QPh.sicau-4B*, implying that *QPh.sicau-4B* was likely the same locus as *QPh.cau-4B.2* and *QPH.caas-4BS.2*.

Correlations Between the Major PH QTL and Other Spike-Related Traits

Gao et al. (2015) and Guan et al. (2018) reported that PH was significantly positively correlated with the TKW but not with the other spike-related traits. Zhai et al. (2016) reported that PH was positively correlated with the SL but negatively correlated with the SN and SC. In this study, Pearson's correlation analysis showed that PH was positively correlated with the SL, SN, and TKW (**Table 3**). This might be because of lines carrying different PH QTL that affect the correlation of PH with other traits. Further analysis of QTL responsible for significant differences in the SL, SC, and TKW between lines with different alleles at *QPh.sicau-6D* (**Supplementary Figure S3**) suggested that *QPh.sicau-6D* confers pleiotropic effects on the SL, TKW, and SC. This interesting perspective warrants further investigation. Additionally, *QPh.sicau-4B* and *QPh.sicau-6D* demonstrated superimposed effects on PH (**Figure 1**), which will allow these two QTL to be simultaneously applied for modifying plant morphological architecture.

TABLE 7 | The information of the candidate genes.

Gene ID	Gene name
TraesCS4B02G043100	*Rht1b*
TraesCS6D02G227300	*YDA*
TraesCS6D02G233000	*HUB1*
TraesCS6D02G234900	*PAT10*

Candidate Genes for *QPh.sicau-4B* and *QPh.sicau-6D*

In wheat, several reduced height (*Rht*) genes have been cloned, e.g., *Rht-b1*, *Rht-d1* (Peng et al., 1999), and *Rht18* (Ford et al., 2018). *Rht-B1* encodes a DELLA transcription factor protein, which participates in gibberellin signaling and thus confers the dwarfing trait to the plant (Peng et al., 1999; Sun, 2010). In the *QPh.sicau-4B* interval, 24 genes with low expression were removed (**Figure 4**), and 47 genes were further annotated using KOBAS 3.0 (**Supplementary Table S2**), which included *Rht-B1*. The sequence analysis of the *Rht-B1* region revealed that a T for C substitution (C(190)T) converts the codon (CGA) to a translational stop codon (TGA) in CN16 (**Figure 5**), which corresponded to dwarfing gene *Rht-B1b* (Pearce et al., 2011). Thus, *QPh.sicau-4B* possibly corresponded to dwarfing gene *Rht-B1*.

In the *QPh.sicau-6D* interval, 38 genes with low expression were removed (**Figure 4**), and 115 genes were further annotated using KOBAS 3.0 (**Supplementary Table S2**). Among these genes, three have been reported to be involved in plant growth and development and to affect PH in *Arabidopsis* and rice. *YDA* encodes a ubiquitously expressed MAPKK kinase and is sensitive to the hormone signal transduction pathway in dwarf phenotype mutants (Lukowitz et al., 2004). *PAT10* encodes an S-acyltransferase protein, which is critical for development, and the *pat10* mutant demonstrates characteristics such as slow cell expansion and cell division and dwarfism (Zhou et al., 2013). *HUB1* is an important regulatory gene for normal plant development as it is involved in histone

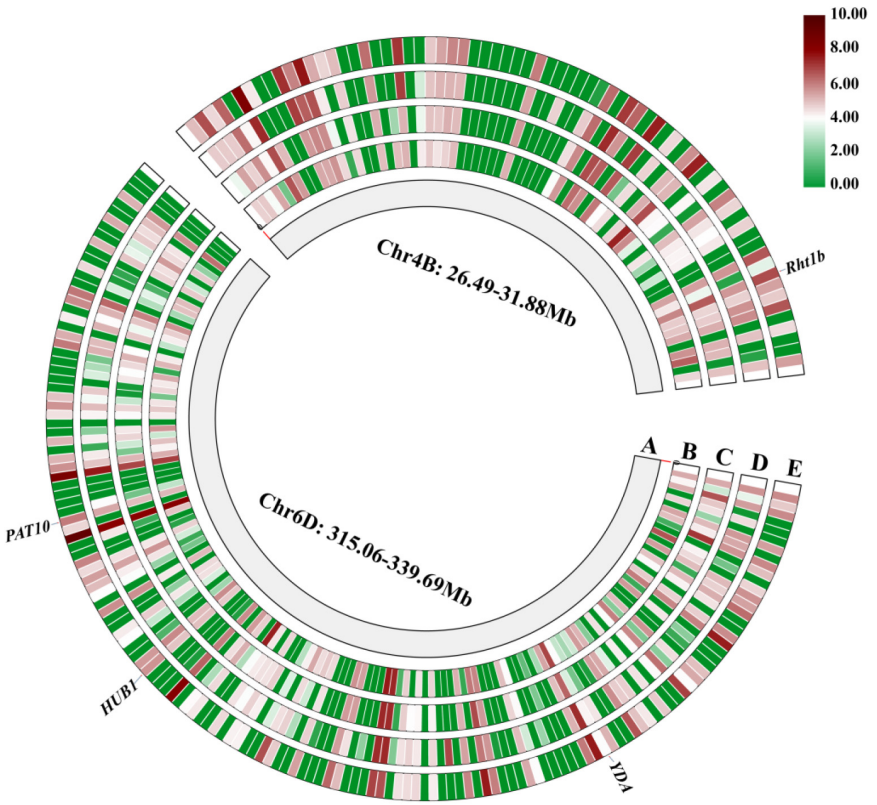

FIGURE 4 | The genes expression of in various tissues in the *QPh.sicau-4B* interval and *QPh.sicau-6D* interval from the expVIP Wheat Expression Brower (A: physical map of chromosome 4B and chromosome 6D, B: root, C: leaf, D: spike, E: grain).

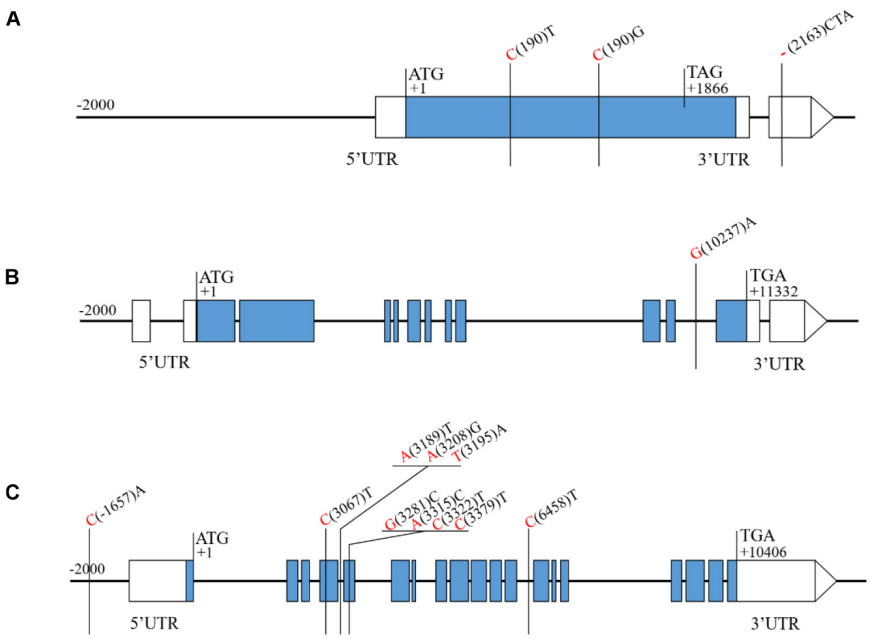

FIGURE 5 | Sequence analysis of the candidate genes showing the SNPs and Indels between H461 and CN16. The nucleotide of H461 and CN16 are shown in red and black, respectively (**A:** *Rht-1B*, **B:** *TraesCS6D02G227300*, **C:** *TraesCS6D02G233000*).

H2B monoubiquitination; the *hub1* mutants showed a dwarf phenotype compared with the wild type in *Arabidopsis* and rice (Fleury et al., 2007; Cao et al., 2015). Of these three candidate genes (*TraesCS6D02G227300*, *TraesCS6D02G233000*, and *TraesCS6D02G234900*) for *QPh.sicau-6D*, *TraesCS6D02G233000* has five SNPs in the coding sequence and five SNPs in the non-coding sequence between H461 and CN16 (**Figure 5**), which led to the substitution of three amino acids (V/A, R/G, T/N). Thus, the gene *TraesCS6D02G233000* might be the candidate gene for further research on *QPh.sicau-6D*.

CONCLUSION

In conclusion, two major stable QTL controlling PH were identified in the HCN population across different environments and were validated in the HCM and HMM populations. *QPh.sicau-4B* possibly corresponded to dwarfing gene *Rht-B1*. *QPh.sicau-6D* appears to be a novel QTL for PH, with pleiotropic effects on the SL, TKW, and SC, and thus, *QPh.sicau-6D* is a potential locus worth exploring further for genetic improvement in wheat breeding programs.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

ZW and HH drafted and revised the manuscript. XJ, YT, YL, FW, and SH conducted phenotype data analysis and contributed to QTL analysis. CL performed the phenotypic evaluation and helped with data analysis. SL and GC helped to draft the manuscript. YXL designed and coordinated this study and revised the manuscript. All authors have read and approved the final manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (31771794), the National Key Research and Development Program of China (2016YFD0101004 and 2017YFD0100900), the outstanding Youth Foundation of the Department of Science and Technology of Sichuan Province (2016JQ0040), the Key Technology Research and Development Program of the Department of Science and Technology of Sichuan Province (2016NZ0057), and the International Science & Technology Cooperation Program of the Bureau of Science and Technology of Chengdu China (No. 2015DFA306002015-GH03-00008-HZ).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.602495/full#supplementary-material

Supplementary Figure 1 | Frequency distribution of PH in the HCN population.

Supplementary Figure 2 | Student's test the groups carry the alleles from H461 or CN16 at *QPh.sicau-4B* for spikelet number per spike (SN), spike length (SL), spikelet compactness (SC), thousand kernel weight (TKW), anthesis date FT.

Supplementary Figure 3 | Student's test the groups carry the alleles from H461 or CN16 at *QPh.sicau-6D* for spikelet number per spike (SN), spike length (SL), spikelet compactness (SC), thousand kernel weight (TKW), anthesis date FT.

Supplementary Table 1 | The primers used in validation population.

Supplementary Table 2 | The HC-genes in the *QPh.sicau-4B* interval and *QPh.sicau-6D* interval.

Supplementary Table 3 | Comparison of the locations of reported PH quantitative trait loci on chromosome 4B with *QPh.sicau-4B* and chromosome 6D with *QPh.sicau-6D*.

REFERENCES

Ai, C., and Kong, L. (2018). CGPS: a machine learning-based approach integrating multiple gene set analysis tools for better prioritization of biologically relevant pathways. *J. Genet. Genomics* 45, 489–504. doi: 10.1016/j.jgg.2018.08.002

Borrill, P., Ramirez-Gonzalez, R., and Uauy, C. (2016). expVIP: a customizable RNA-seq data analysis and visualization platform. *Plant Physiol.* 170, 2172–2186. doi: 10.1104/pp.15.01667

Cao, H., Li, X., Wang, Z., Ding, M., Sun, Y. Z., Dong, F. Q., et al. (2015). Histone H2B monoubiquitination mediated by HISTONE MONOUBIQUITINATION1 and HISTONE MONOUBIQUITINATION2 is involved in anther development by regulating tapetum degradation-related genes in rice. *Plant Physiol.* 168, 1389–1405. doi: 10.1104/pp.114.256578

Chai, L. L., Chen, Z. Y., Bian, R., Zhai, H. J., Cheng, X. J., Peng, H. R., et al. (2018). Dissection of two quantitative trait loci with pleiotropic effects on plant height and spike length linked in coupling phase on the short arm of chromosome 2D of common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 131, 2621–2637. doi: 10.1007/s00122-018-3177-4

Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., et al. (2020). TBtools - an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Donald, C. T. (1968). The breeding of crop ideotypes. *Euphytica* 17, 385–403.

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. doi: 10.1007/s00122-013-2257-8

Fleury, D., Himanen, K., Cnops, G., Nelissen, H., Boccardi, T. M., Maere, S., et al. (2007). The *Arabidopsis* thaliana homolog of yeast BRE1 has a function in cell cycle regulation during early leaf and root growth. *Plant Cell* 19, 417–432. doi: 10.1105/tpc.106.041319

Ford, B., Foo, E., Sharwood, R., Karafiatova, M., Vrána, J., MacMillan, C., et al. (2018). Rht18 semi-dwarfism in wheat is due to increased expression of GA 2-oxidaseA9 and lower GA content. *Plant Physiol.* 177, 168–180. doi: 10.1104/pp.18.00023

Gao, F. M., Wen, W., Liu, J. D., Rasheed, A., Yin, G. H., Xia, X. C., et al. (2015). Genome-wide linkage mapping of QTL for yield components, plant height and

yield-related physiological traits in the chinese wheat cross zhou 8425B/chinese spring. *Front. Plant Sci.* 6:1099. doi: 10.3389/fpls.2015.01099

Griffiths, S., Simmonds, J., Leverington, M., Wang, Y. K., Fish, L., Sayers, L., et al. (2012). Meta-QTL analysis of the genetic control of crop height in elite European winter wheat germplasm. *Mol. Breeding* 29, 159–171. doi: 10.1007/s11032-010-9534-x

Guan, P. F., Lu, L. H., Jia, L. J., Kabir, M. R., Zhang, J. B., Lan, T. 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. doi: 10.3389/fpls.2018.00529

Hassan, M. A., Yang, M. J., Fu, L. P., Rasheed, A., Zheng, B. Y., Xia, X. C., et al. (2019). Accuracy assessment of plant height using an unmanned aerial vehicle for quantitative genomic analysis in bread wheat. *Plant Methods* 15:37. doi: 10.1186/s13007-019-0419-7

Kowalsk, A. M., Gooding, M., Ferrante, A., Slafer, G. A., Orford, S., Gasperini, D., et al. (2016). Agronomic assessment of the wheat semi-dwarfing gene Rht8 in contrasting nitrogen treatments and water regimes. *Field Crops Res.* 191, 150–160. doi: 10.1016/j.fcr.2016.02.026

Liu, D. C., Gao, M. Q., Guan, R. X., Li, R. Z., Cao, S. H., Guo, X. L., et al. (2002). Mapping quantitative trait loci for plant height in wheat (*Triticum aestivum* L.) using a F2: 3 population. *Acta Genet. Sin.* 29, 706–711.

Liu, G., Jia, L. J., Lu, L. H., Qin, D. D., Zhang, J. P., Guan, P. F., et al. (2014). Mapping QTL of yield-related traits using RIL population derived from common wheat and Tibetan semi-wild wheat. *Theor. Appl. Genet.* 127, 2415–2432. doi: 10.1007/s00122-014-2387-7

Liu, Y. X., Tao, Y., Wang, Z. Q., Guo, Q. L., Wu, F. K., Yang, X. L., et al. (2018). Identification of QTL for flag leaf length in common wheat and their pleiotropic effects. *Mol. Breeding* 38:11. doi: 10.1007/s11032-017-0766-x

Lukowitz, W., Roeder, A., Parmenter, D., and Somerville, C. (2004). A MAPKK kinase gene regulates extra-embryonic cell fate in *Arabidopsis*. *Cell* 116, 109–119. doi: 10.1016/S0092-8674(03)01067-5

Pearce, S., Saville, R., Vaughan, S. P., Chandler, P. M., Wilhelm, E. P., Sparks, C. A., et al. (2011). Molecular characterization of Rht-1 dwarfing genes in hexaploid wheat. *Plant Physiol.* 157, 1820–1831. doi: 10.1104/pp.111.183657

Peng, J. R., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., et al. (1999). 'Green revolution'genes encode mutant gibberellin response modulators. *Nature* 400:256. doi: 10.1038/22307

Sakamoto, T., and Matsuoka, M. (2004). Generating high-yielding varieties by genetic manipulation of plant architecture. *Curr. Opin. Biotechnol.* 15, 144–147. doi: 10.1016/j.copbio.2004.02.003

Singh, A., Knox, R. E., DePauw, R. M., Singh, A. K., Cuthbert, R. D., Kumar, S., et al. (2016). Genetic mapping of common bunt resistance and plant height QTL in wheat. *Theor. Appl. Genet.* 129, 243–256. doi: 10.1007/s00122-015-2624-8

Smith, S. E., Kuehl, R. O., Ray, I. M., Hui, R., and Soleri, D. (1998). Evaluation of simple methods for estimating broad-sense heritability in stands of randomly planted genotypes. *Crop Sci.* 38, 1125–1129. doi: 10.2135/cropsci1998.0011183X003800050003x

Spielmeyer, W., Hyles, J., Joaquim, P., Azanza, F., Bonnett, D., Ellis, M. E., et al. (2007). A QTL on chromosome 6A in bread wheat (*Triticum aestivum*) is associated with longer coleoptiles, greater seedling vigour and final plant height. *Theor. Appl. Genet.* 115, 59–66. doi: 10.1007/s00122-007-0540-2

Sun, T. P. (2010). Gibberellin-GID1-DELLA: a pivotal regulatory module for plant growth and development. *Plant Physiol.* 154, 567–570. doi: 10.1104/pp.110.161554

Team, R. C. (2013). *R: A language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Tian, X. L., Wen, W. E., Xie, L., Fu, L. P., Xu, D. G., Fu, C., et al. (2017). Molecular mapping of reduced plant height gene Rht24 in bread wheat. *Front. Plant Sci.* 8:1379. doi: 10.3389/fpls.2017.01379

Van Ooijen, J. W., and Kyazma, B. V. (2009). *MapQTL 6. Software for the Mapping of Quantitative Trait Loci in Experimental Populations of Diploid Species Kyazma BV*. Wageningen: Wageningen University.

Wang, Z. Q., Liu, Y. X., Shi, H. R., Mo, H. J., Wu, F. K., Lin, Y., et al. (2016). Identification and validation of novel low-tiller number QTL in common wheat. *Theor. Appl. Genet.* 129, 603–612. doi: 10.1007/s00122-015-2652-4

Würschum, T., Langer, S. M., and Longin, C. F. H. (2015). Genetic control of plant height in European winter wheat cultivars. *Theor. Appl. Genet.* 128, 865–874. doi: 10.1007/s00122-015-2476-2

Zhai, H. J., Feng, Z. Y., Li, J., Liu, X. Y., Xiao, S. H., Ni, Z. F., et al. (2016). QTL analysis of spike morphological traits and plant height in winter wheat (*Triticum aestivum* L.) using a high-density SNP and SSR-based linkage map. *Front. Plant Sci.* 7:1617. doi: 10.3389/fpls.2016.01617

Zhang, L. Y., Liu, D. C., Guo, X. L., Yang, W. L., Sun, J. Z., Wang, D. W., et al. (2010). Genomic distribution of quantitative trait loci for yield and yield-related traits in common wheat. *J. Integr. Plant Biol.* 52, 996–1007. doi: 10.1111/j.1744-7909.2010.00967.x

Zhou, L. Z., Li, S., Feng, Q. N., Zhang, Y. L., Zhao, X. Y., Zeng, Y. L., et al. (2013). PROTEIN S-ACYL TRANSFERASE10 is critical for development and salt tolerance in *Arabidopsis*. *Plant Cell* 25, 1093–1107. doi: 10.1105/tpc.112.108829

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 06 November 2020
doi: 10.3389/fgene.2020.609184

Genome-Wide Characterization of the *HSP20* Gene Family Identifies Potential Members Involved in Temperature Stress Response in Apple

Fuwen Yao†, Chunhui Song†, Hongtao Wang, Shangwei Song, Jian Jiao, Miaomiao Wang, Xianbo Zheng and Tuanhui Bai**

College of Horticulture, Henan Agricultural University, Zhengzhou, China

OPEN ACCESS

Edited by:
Jian Ma,
Sichuan Agricultural University, China

Reviewed by:
Yanling Ma,
Chinese Academy of Agricultural Sciences (CAAS), China
Quan Xie,
Nanjing Agricultural University, China

***Correspondence:**
Tuanhui Bai
tuanhuibai88@163.com
Xianbo Zheng
Xianboz@163.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *22 September 2020*
Accepted: *20 October 2020*
Published: *06 November 2020*

Citation:
Yao F, Song C, Wang H, Song S, Jiao J, Wang M, Zheng X and Bai T (2020) Genome-Wide Characterization of the HSP20 Gene Family Identifies Potential Members Involved in Temperature Stress Response in Apple. Front. Genet. 11:609184. doi: 10.3389/fgene.2020.609184

Apple (*Malus domestica* Borkh.), an economically important tree fruit worldwide, frequently suffers from temperature stress during growth and development, which strongly affects the yield and quality. Heat shock protein 20 (*HSP20*) genes play crucial roles in protecting plants against abiotic stresses. However, they have not been systematically investigated in apple. In this study, we identified 41 *HSP20* genes in the apple 'Golden Delicious' genome. These genes were unequally distributed on 15 different chromosomes and were classified into 10 subfamilies based on phylogenetic analysis and predicted subcellular localization. Chromosome mapping and synteny analysis indicated that three pairs of apple *HSP20* genes were tandemly duplicated. Sequence analysis revealed that all apple *HSP20* proteins reflected high structure conservation and most apple *HSP20* genes (92.6%) possessed no introns, or only one intron. Numerous apple *HSP20* gene promoter sequences contained stress and hormone response *cis*-elements. Transcriptome analysis revealed that 35 of 41 apple *HSP20* genes were nearly unchanged or downregulated under normal temperature and cold stress, whereas these genes exhibited high-expression levels under heat stress. Subsequent qRT-PCR results showed that 12 of 29 selected apple *HSP20* genes were extremely up-regulated (more than 1,000-fold) after 4 h of heat stress. However, the heat-upregulated genes were barely expressed or downregulated in response to cold stress, which indicated their potential function in mediating the response of apple to heat stress. Taken together, these findings lay the foundation to functionally characterize *HSP20* genes to unravel their exact role in heat defense response in apple.

Keywords: apple, *HSP20* family, heat stress, genome-wide analysis, gene expression

INTRODUCTION

Temperature is an important factor affecting plant growth and geographical distribution (Wang et al., 2016). Most plants undergo optimal growth and development within a narrow temperature range and can only tolerate minor fluctuations. Fluctuations beyond optimal range result in temperature stress, which is one of the most severe environmental stresses affecting plant growth, development and survival worldwide (Peleg and Blumwald, 2011). High and low temperature

stresses have rapid and severe effects on plant cell physiology, altering gene expression, protein levels, and energy consumption (Wang et al., 2017; Shen et al., 2019; Suzuki, 2019). Plants have developed a series of physiological and molecular strategies to overcome temperature stress over evolutionary time (Asea et al., 2016; Huo et al., 2020). Heat shock proteins (HSPs) are one of the strategies, and HSPs are essential in regulating growth, development and stress response in plants (Waters et al., 1996; Wang et al., 2004; Asea et al., 2016; He et al., 2019).

HSPs can be divided into five categories according to their molecular weight: *HSP100s*, *HSP90s*, *HSP70s*, *HSP60s,* and *HSP20s* (Waters, 2013; Zhao et al., 2018). Of these groups, *HSP20* is commonly associated with temperature stress in plants (Waters, 2013). As genomes for more species are sequenced, the *HSP20* gene family has been identified in various plants. Nineteen *HSP20* genes have been identified in *Arabidopsis* (Scharf et al., 2001), 39 in rice (Ouyang et al., 2009), 42 in tomato (Yu et al., 2016), 44 in watermelon (He et al., 2019), and 48 in grape (Ji et al., 2019). Previous studies have suggested that *HSP20* genes are involved in regulating a diverse array of developmental processes and responses to abiotic stresses, especially in heat stress (Guo et al., 2015; He et al., 2019; Ji et al., 2019). Yu et al. (2016) identified tomato *HSP20* family genes and analyzed their functions in abiotic-stress responses. Most pepper *HSP20* genes were highly induced by heat stress (Guo et al., 2015). Among the *GmHSP20* genes, five were shown to be involved in the soybean response to cold stress (Lopes-Caitar et al., 2013). Interestingly, the same *HSP20* genes exhibited a different expression pattern in the heat tolerant and sensitive plants. These differences in expression pattern indicate the roles of *HSP20* in heat tolerance. In addition, some studies have further verified the role of *HSP20s i*n stress tolerance using transgenic methods. For example, overexpressing with *WsHSP26* in *Arabidopsis* showed improved heat tolerance (Mu et al., 2013). Similarly, transgenic rice over-expressing *OsHSP17.7* conferred enhanced tolerance to heat stress (Murakami et al., 2004). Together, these studies reveal the crucial role of *HSP20* genes in mediating temperature stress tolerance.

Apple (*Malus domestica* Borkh.), an economically important fruit crop, is widely planted in temperate zones (Dobránszki and Teixeira da Silva, 2010). However, apple trees frequently suffer from both high and low temperature stresses during their life cycle, which strongly affect apple quality and yield. After suffering continuous heat stress in summer, the leaf and fruit of apple can be severely damaged; resulting in tissue discoloration and sunburn of the fruit surface (Torres et al., 2017). It is reported that fruit sunburn causes 10–40% yield losses in all major apple growing regions around the world (Wang et al., 2020). After suffering cold stress in early spring, the pollination, new leaves and shoots of apple can be severely damaged, thereby greatly reducing the yield and quality of apple. The entire genome of apple has been sequenced, providing powerful resource for the mining and identification of *HSP20* gene family members at the whole genome level.

In the present study, we identified *HSP20* genes from the apple genome using bioinformatics methods, and determined their chromosomal locations, gene duplication, phylogenetic relationships, gene structures, and conserved domains, as well as *cis*-elements. Furthermore, we analyzed the expression patterns of the apple *HSP20* genes using qRT-PCR in order to determine their roles in response to heat and cold stresses. Our findings provide valuable information for subsequent research on the functions and regulatory mechanisms of potentially important *HSP20* genes that are crucial in modulating heat stress tolerance in apple.

MATERIALS AND METHODS

Plant Materials and Stress Treatments

Seeds of *M. hupehensis* were stratified using sand at 4°C for 40 days, as previously described (Bai et al., 2013). Then, one germinated seed was planted in one plastic pot (6 cm × 6 cm) filled with soil/organic substrate (1:5, v: v) in a greenhouse under natural light, 22–28°C (day) and 5–10°C (night), and a relative humidity of 60–70%. After growing for 70 days, uniform seedlings were randomly divided into three groups and transferred to different chambers for temperature stress: (1) control (CK), growth chamber were maintained at 25 ± 1°C, 16-h photoperiod (160 μmol m^{-2} s^{-1}) and relative humidity of 60–70%; (2) heat stress (HS), growth chamber maintained at 40 ± 1°C; (3) cold stress (CS), growth chamber maintained at 4 ± 1°C, 16-h photoperiod (160 μmol m^{-2} s^{-1}) and relative humidity of 60–70%. Each treatment contained three biological replicates and 30 plants for each replicate. At 0, 4, 8, 12, and 24 h after treatments, the samples were rapidly frozen in liquid nitrogen and stored at −80°C until RNA extraction.

Genome-Wide Identification of the *HSP20* Genes in Apple

The reference apple genome and protein sequences were downloaded from the Genome Database for Rosaceae (GDR[1]). The apple *HSP20* candidates with an e-value ≤ 0.001 were identified based on the Hidden Markov Model (HMM) profile (PF00011) downloaded from Pfam protein family database[2]. The SMART database[3] was used to further confirm the conserved *HSP20* gene domain. ProtParam[4] was used to predict the potential chemical characteristics of the *HSP20* genes. ProtComp[5] was used to predict the subcellular localization.

Phylogenetic Analysis and Classification of Apple *HSP20* Genes

The full-length amino acid sequences of *HSP20* genes derived from *Arabidopsis* and rice downloaded from the Ensembl Plants Database[6] were combined with newly identified *HSP20* genes in

[1]https://www.rosaceae.org/
[2]http://pfam.xfam.org
[3]http://smart.embl-heidelberg.de/
[4]https://web.expasy.org/protparam/
[5]http://linux1.softberry.com/
[6]http://plants.ensembl.org/index.html

apple and used for phylogenetic analysis. The phylogenetic tree was constructed using MEGA 6.0[7].

Structure and Domain Analysis of Apple *HSP20* Genes

The structures of *HSP20* genes were identified using TBtools software (Chen et al., 2018). The conserved motifs of *HSP20* genes were identified using MEME Suite 5.1.1[8], and the parameters were as follows: optimum motif width ranges from 6 to 200 amino acid residues and maximum of 10 misfits. The upstream 2.0 kb promoter sequence of the apple *HSP20* genes was downloaded from the GDR and submitted to PlantCARE[9] to identify the *cis*-elements (Lescot et al., 2002).

[7]http://www.megasoftware.net/

[8]http://meme-suite.org/tools/meme

[9]http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

TABLE 1 | Characteristics of *HSP20* genes identified in apple.

Gene name	Sequence ID	Chr	Genomic location	ORF (bp)	AA	MW (kDa)	pI	Subcellular localization
HSP20-1	MD00G1181700	0	42820094-42820435	342	113	12.93	9.49	Cytoplasm
HSP20-2	MD01G1144400	1	25475100-25479690	810	269	30.82	9.38	Chloroplast
HSP20-3	MD01G1227600	1	31728955-31730100	480	159	18.05	5.94	Chloroplast
HSP20-4	MD02G1259500	2	31287370-31288929	1092	363	39.47	9.53	Chloroplast
HSP20-5	MD03G1081800	3	6580874-6581326	453	150	16.75	4.60	Chloroplast
HSP20-6	MD04G1140600	4	22893411-22893881	471	156	18.17	6.87	Cytoplasm
HSP20-7	MD05G1183400	5	31196823-31197443	621	206	23.14	6.46	Chloroplast
HSP20-8	MD05G1240300	5	37103604-37104083	480	159	18.32	5.72	Cytoplasm
HSP20-9	MD05G1343700	5	46361886-46362915	414	137	16.11	5.86	Chloroplast
HSP20-10	MD06G1060300	6	9340096-9341581	789	262	29.46	9.85	Nucleus
HSP20-11	MD06G1188500	6	32557693-32558803	717	238	26.43	5.82	Chloroplast
HSP20-12	MD07G1210400	7	28942672-28943139	468	155	17.53	5.99	Cytoplasm
HSP20-13	MD07G1210700	7	28969132-28969626	468	155	17.55	5.39	Cytoplasm
HSP20-14	MD07G1210800	7	28970024-28970494	471	156	17.86	7.08	Cytoplasm
HSP20-15	MD07G1253800	7	32097833-32098492	660	219	24.43	6.99	Chloroplast
HSP20-16	MD07G1298000	7	35766162-35768343	672	223	24.68	7.87	Chloroplast
HSP20-17	MD08G1068000	8	5407860-5408476	465	154	17.10	6.61	Cytoplasm
HSP20-18	MD08G1068200	8	5414896-5415389	471	156	17.41	5.93	Cytoplasm
HSP20-19	MD08G1068300	8	5433838-5434323	486	161	17.84	4.77	Cytoplasm
HSP20-20	MD08G1068500	8	5449001-5449663	663	220	24.25	5.66	Chloroplast
HSP20-21	MD08G1068700	8	5470111-5470842	732	243	26.17	6.24	Chloroplast
HSP20-22	MD08G1068800	8	5474013-5474495	483	160	17.90	6.41	Cytoplasm
HSP20-23	MD08G1249100	8	31349994-31350611	618	205	22.94	7.08	Chloroplast
HSP20-24	MD09G1271100	9	34531591-34533232	438	145	16.36	6.92	Cytoplasm
HSP20-25	MD10G1171200	10	26409043-26409660	618	205	23.37	6.93	Endoplasmic reticulum
HSP20-26	MD10G1218300	10	31639809-31640195	387	128	14.87	5.94	Chloroplast
HSP20-27	MD10G1319400	10	40169759-40170934	411	136	15.88	5.39	Chloroplast
HSP20-28	MD11G1087100	11	7308309-7308791	483	160	18.21	5.81	Cytoplasm
HSP20-29	MD11G1088300	11	7365514-7366036	483	160	17.98	5.34	Cytoplasm
HSP20-30	MD11G1089300	11	7421465-7421947	483	160	18.22	5.39	Cytoplasm
HSP20-31	MD13G1108500	13	7793385-7800082	723	240	26.81	9.16	Chloroplast
HSP20-32	MD13G1120200	13	8835673-8835939	267	88	9.98	4.86	Cytoplasm
HSP20-33	MD15G1053500	15	3658892-3659347	456	151	16.79	9.09	Cytoplasm
HSP20-34	MD15G1053600	15	3662890-3663356	402	133	15.14	6.15	Chloroplast
HSP20-35	MD15G1053800	15	3675367-3676018	471	156	17.44	5.57	Cytoplasm
HSP20-36	MD15G1443700	15	54408605-54409228	624	207	23.17	6.31	Chloroplast
HSP20-37	MD16G1108600	16	7581626-7584321	711	236	26.71	9.36	Nucleus
HSP20-38	MD17G1020300	17	1535600-1536025	426	141	16.40	7.83	Cytoplasm
HSP20-39	MD17G1151000	17	13922992-13924497	471	156	17.36	7.72	Chloroplast
HSP20-40	MD17G1209800	17	25460073-25461508	732	243	27.26	8.45	Chloroplast
HSP20-41	MD17G1269200	17	32927931-32928852	339	112	12.53	6.30	Cytoplasm

ORF, open reading frame; AA, amino acid; MW, molecular weight; pI, isoelectric point.

Chromosomal Location and Synteny Analysis

All identified *HSP20* genes were mapped to apple chromosomes using TBtools software based on the information available at GDR. Synteny analysis of *HSP20* genes was conducted using Circos v. 0.63[10].

RNA Extraction, cDNA Library Construction, and Sequencing

Total RNA was extracted from the leaf tissues according to the CTAB method (Chang et al., 1993). Each sample was 0.5 g and three biological replicates were performed. RNA concentrations were determined using a NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA, United States) and quality was assayed on a 1% agarose gel. The sample libraries were prepared according to the RNA-Seq library constructed flow path and sequenced on an Illumina HiSeq 4000 system. The raw sequence data from the sequence was used for analysis. After filtering the low quality reads and contaminant sequences, the clean reads were aligned to the apple genome (GDDH13 Version 1.1[11]) (Daccord et al., 2017) using HISAT2 (Kim et al., 2015). Stringtie software was used to assemble the transcript (Pertea et al., 2016). Gene expression was calculated using the fragments per kilobase of transcript per million (FPKM) fragments mapped Reads

[10]http://circos.ca/

[11]https://iris.angers.inra.fr/gddh13/the-apple-genome-downloads.html

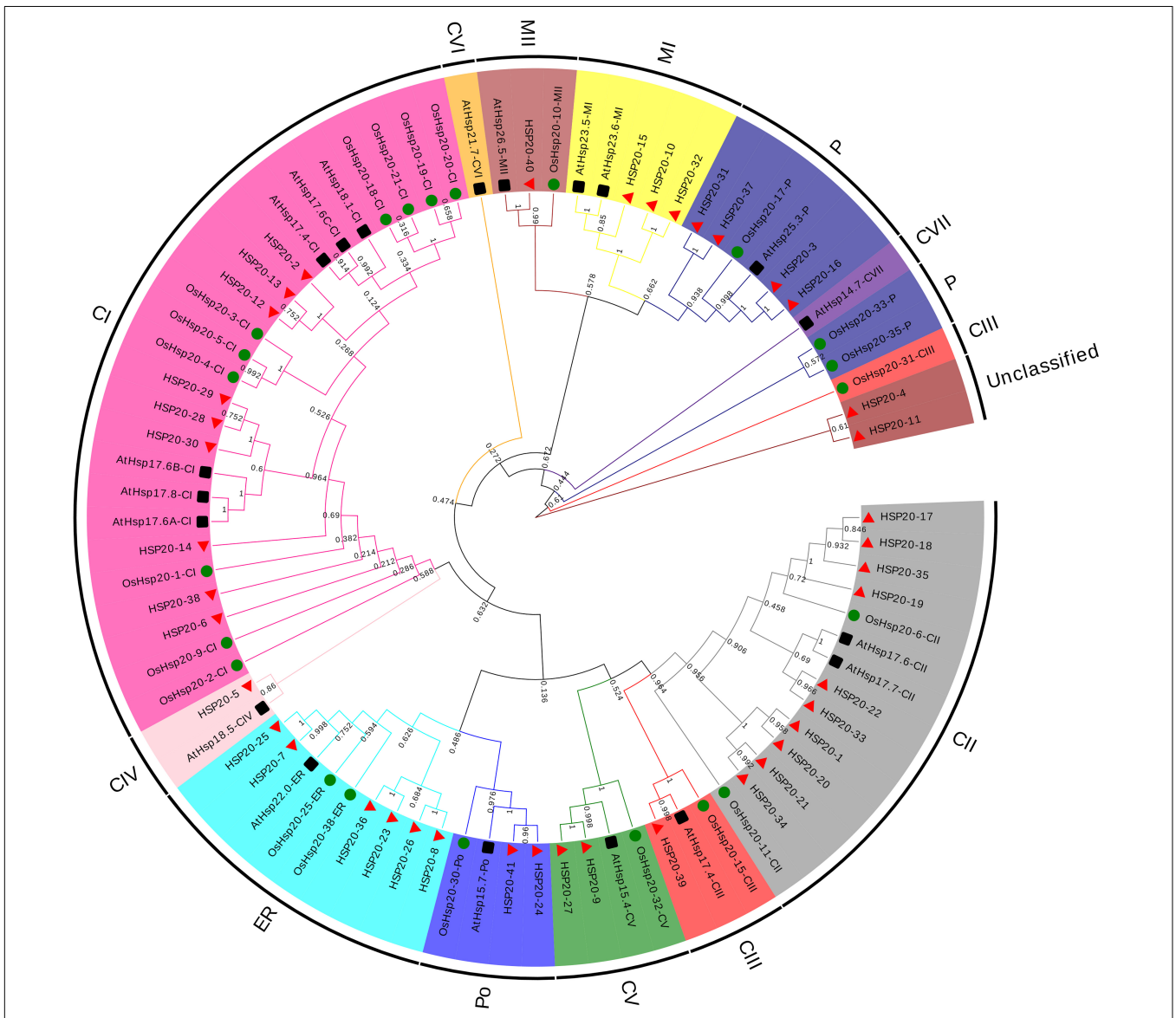

FIGURE 1 | Phylogenetic analysis of HSP20 proteins from *Arabidopsis*, rice and apple. The phylogenetic tree was constructed using MEGA 6.0 based on the neighbor joining method with 1000 bootstrap replicates. The 12 subfamilies were distinguished with different colored arcs.

method (Mortazavi et al., 2008). TBtools was used to generate the heatmap. The RNA-seq data were available at NCBI[12].

qRT-PCR Analysis

Quantitative real-time PCR (qRT-PCR) was used to analyze the gene expression. Primers (**Supplementary Table S1**) were designed to amplify products of 150–250 bp using Primer 5.0 software. qRT-PCR was performed using ABI-7500 Connect Real-Time PCR Detection System. cDNAs were diluted to 200 ng and run in three technical replicates, with 1 μL template in a reaction volume of 20 μL. PCR amplification conditions were as follows: 95°C for 5 min for initial denaturation, then 45 cycles of 94°C for 20 s, 55°C for 20 s, and 72°C for 10 s. Fluorescence was measured at the end of each cycle. A melting curve analysis was performed to determine whether a single product was amplified. The apple Actin gene was used as an internal standard in the analysis. The relative expression level of each gene was calculated according to the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). Values for mean expression and standard error (SE) were calculated from the results of three independent biological replicates.

[12]https://www.ncbi.nlm.nih.gov/sra/PRJNA665791

RESULTS

Genome-Wide Identification of *HSP20* Genes in Apple

A total of 45 *HSP20* protein sequences were found in the apple cultivar "Golden Delicious" genome. Among 45 *HSP20* sequences, four sequences lacked the conserved domain. Ultimately, 41 sequences were identified as genes in the apple *HSP20* family and named *HSP20-1* to *HSP20-41* based on the position of the genes on the chromosomes (**Supplementary Table S2**). Gene name, gene ID, chromosomal location, open reading frame (ORF), amino acid (AA), molecular weight (MW) and isoelectric point (pI) for each gene are in **Table 1**. Sequence

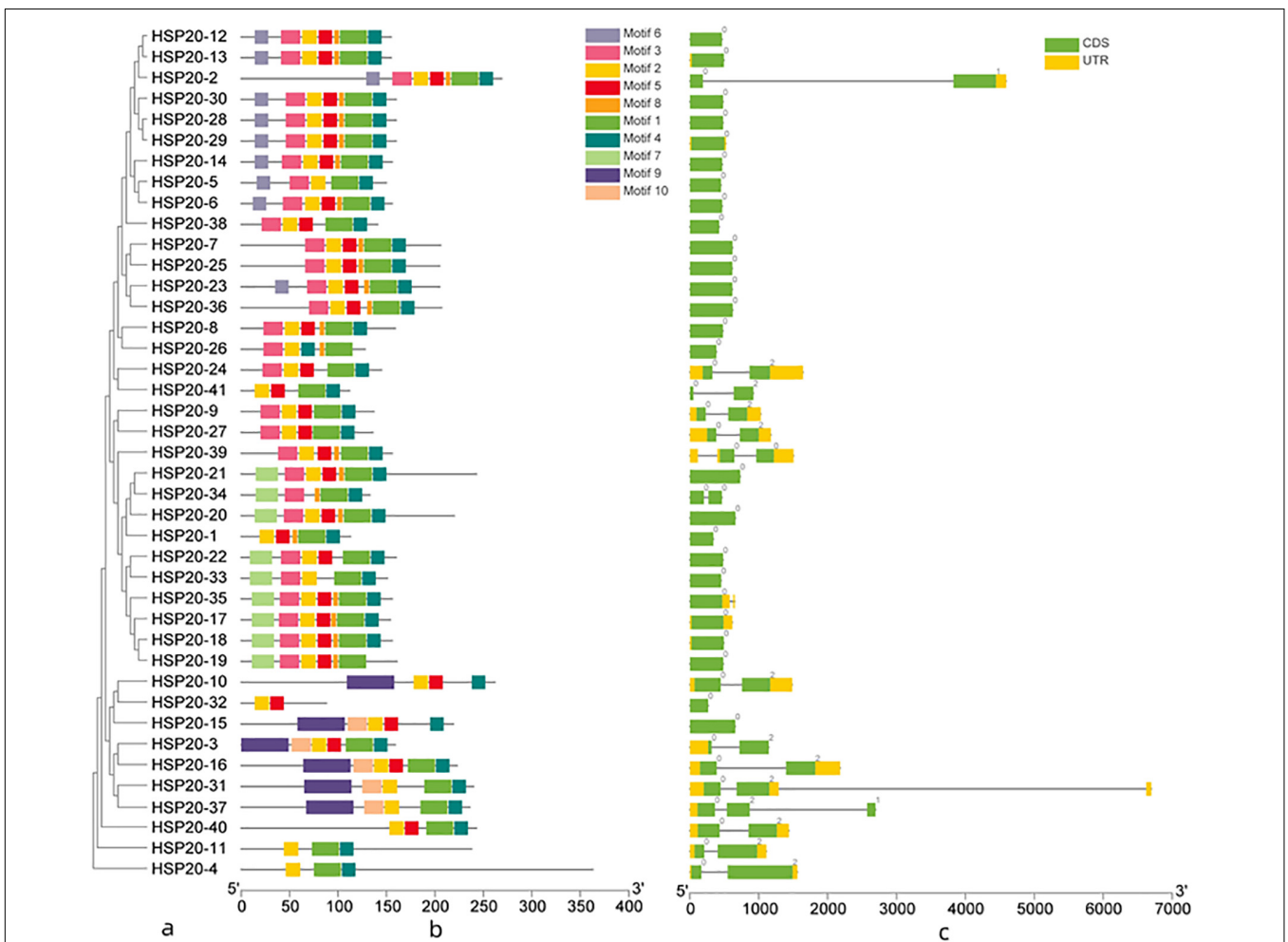

FIGURE 2 | Phylogenetic tree, gene structure and domain analyses of HSP20s. **(a)** Phylogenetic tree of *HSP20s*. **(b)** Domain analyses of *HSP20* proteins. **(c)** Gene structure of *HSP20* genes. CDS sequences are represented with yellow round-corner rectangles and introns with gray lines, UTRs are shown with green boxes. Different colored boxes represent the different types of motifs.

analysis showed that these *HSP20* proteins vary widely in length, from 88 (*HSP20-32*) to 363 (*HSP20-4*) AAs, and predicted MWs from 9.98 kDa (*HSP20-32*) to 39.47 kDa (*HSP20-4*). The ORF lengths of the *HSP20* genes ranged from 267 bp (*HSP20-32*) to 1,092 bp (*HSP20-4*), and the predicted pI-values of *HSP20* proteins ranged from 4.60 (*HSP20-5*) to 9.85 (*HSP20-10*).

Phylogenetic Relationships of Plant *HSP20* Genes

A phylogenetic tree was constructed based on the amino acid sequences of *HSP20* genes (**Figure 1**), 19 from *Arabidopsis*, 22 from rice (*Oryza sativa*) and 41 from apple (*M. domestica*) (**Supplementary Table S3**). The 82 *HSP20s* were divided into 12 subfamilies, 25 cytosol Is (CIs), 14 CIIs, 4 CIIIs, 2 CIVs, 4 CVs, 1 CVI, 1 CVII, five mitochondria Is (MIs), three MIIs, four peroxisomes (Pos), 9 from the endoplasmic reticulum (ER), and 8 plastids (Ps) based on the phylogeny and the subcellular localization. However, two apple *HSP20* genes (*HSP20-4* and *HSP20-11*) failed to cluster into any subfamily and were thus unclassified. Of the 12 subfamilies, 10 (CIs, CIIs, CIIIs, CIVs, CVs, MIs, MIIs, Pos, ER, and Ps) contained apple *HSP20* proteins. Except for the two unclassified *HSP20s,* 23 (59%) *HSP20s* were classified into CI-CVI, indicating that cytoplasm could be the primary functional area of the *HSP20* family in apple.

Conserved Motifs and Gene Structure of *HSP20* Genes

The conserved motifs of apple *HSP20* gene family were identified and divided into 10 motifs (**Figure 2b**). The lengths of the 10 motifs ranged from 6 to 50 AAs, with the longest motif (9) containing 50 AAs and the shortest motif (8) containing six AAs; motifs 4, motifs 5, and motifs 6 have 15 AAs (**Supplementary Table S4**). The number of the conserved motifs for each *HSP20* gene ranged from 2 to 7. Most apple *HSP20* genes had two to seven conserved motifs, however, *HSP20-32* only contained two conserved motifs. The results suggested that the *HSP20* genes exhibited extreme divergence during the evolutionary process.

To gain insight into the evolutionary relationships of apple *HSP20* genes, the exon–intron structure of the *HSP20* genes was analyzed (**Figure 2c**). Among the *HSP20s*, 25 (60.9%) were intronless, 13 (31.7%) possessed one intron, and three genes (7.3%) – *HSP20-31*, *HSP20-37,* and *HSP20-38* – possessed two introns. Most *HSP20* genes thus have no introns or only one intron, suggesting relatively simple gene structures. Gene structure analysis showed that the genes with similar exon–intron patterns were grouped in the same cluster (**Figure 2a**).

Chromosomal Location, Gene Duplication, and Synteny Analysis

A total of 41 apple *HSP20* genes were mapped on 15 chromosomes (Chr), except Chr 12 and 14, with an obviously non-uniform distribution (**Figure 3**). One *HSP20* gene (*HSP20-1*) could not be mapped on any of the apple chromosomes, so we mapped it on a pseudo-chromosome, named Chr00. In addition, most of the apple *HSP20* genes were located on the distal ends of the chromosomes. The biggest cluster was seven *HSP20* genes together on Chr 8, whereas the fewest *HSP20s* were found on Chrs 0, 2, 3, 4, 9, and 12 (one per Chr). Moreover, we analyzed the duplication events of apple *HSP20* genes (**Figure 4**). In total, 37 (90.2%) *HSP20* genes in apple exhibited segmental

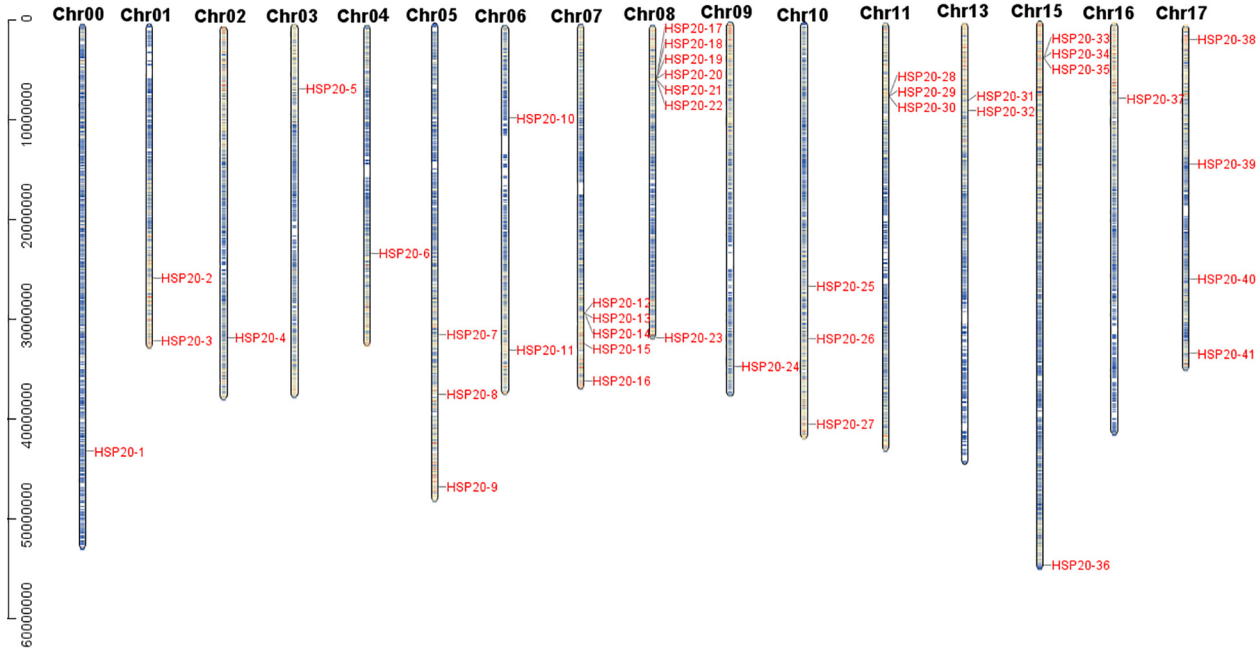

FIGURE 3 | Distribution of *HSP20* genes in apple chromosomes. Forty-one *HSP20* genes was mapped to the 16 linkage groups (Chr 01 through Chr 17, except Chr 12 and Chr 14), whereas one *HSP20* gene were mapped on apseudo-chromosome, designated as Chr00.

or tandem duplication. Twenty were segmentally duplicated and three pairs of genes (*HSP20-13* and *HSP20-14*, *HSP20-18* and *HSP20-19*, *HSP20-21*, and *HSP20-22*) were regarded as tandemly-duplicated genes (**Supplementary Table S5**). Chr 7 had the most duplication events, which could partly explain the larger numbers of *HSP20* genes on Chr 7, while Chr 0 and 12 did not contain any duplicated genes.

Analysis of *Cis*-Element in Apple *HSP20* Gene Promoters

To further investigate the potential regulatory mechanisms of the apple *HSP20* genes in response to temperature stress, the promoter in the upstream 2 kb region of 41 *HSP20* genes was analyzed to detect the *cis*-regulatory element. The results showed that three categories of *cis*-elements, including stress-related (heat, defense and stress, low-temperature and light), hormone-related (abscisic acid, auxin, gibberellin, MeJa, and salicylic acid), and plant development-related *cis*-elements (meristem expression and circadian control), were identified (**Figure 5a**). Among the stress-related *cis*-elements, 13 apple *HSP20* genes had the heat response elements (HRE) in their promoter regions, 26 apple *HSP20* genes had the low temperature response elements (LTR), which suggests a potential stress response under temperature conditions (**Figure 5b**). Among the hormone-related *cis*-elements, abscisic acid responsive (ABRE), salicylic acid responsiveness (TCA-element), auxin responsive (AUXRR-core), and MeJA-responsiveness (CGTCA-motif) were identified in the promoters of apple *HSP20* genes. All HSP20 genes contained light signal response elements, which indicate that *HSP20s* are essential in plant growth. The results indicate that the *HSP20* gene family is not only involved in the stress response, but is also involved in other physiological response processes.

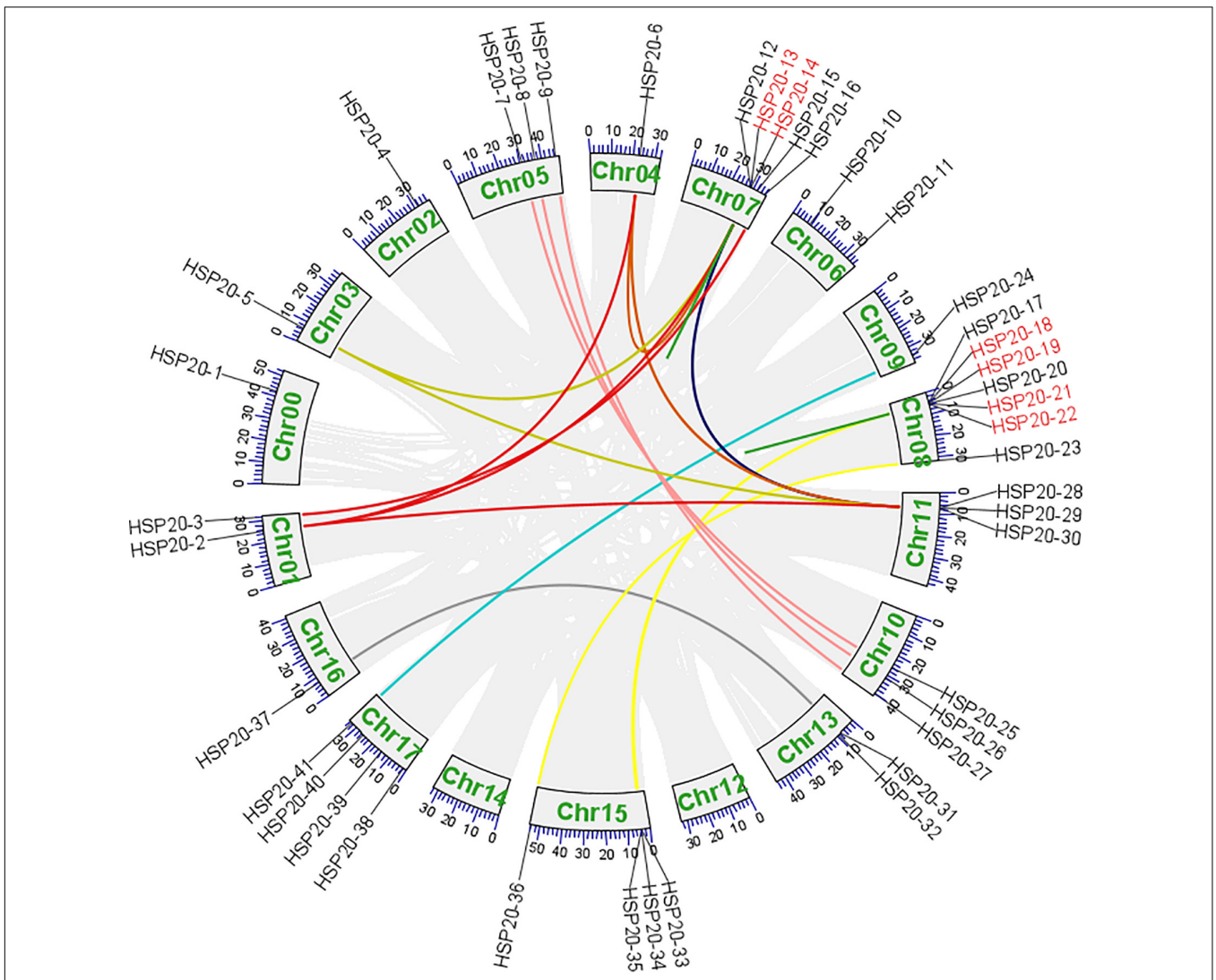

FIGURE 4 | Syntenic relationships and gene duplications of the apple *HSP20* genes. The segmental duplicated genes are indicated by differently colored lines and tandem duplicated genes are indicated by red gene names.

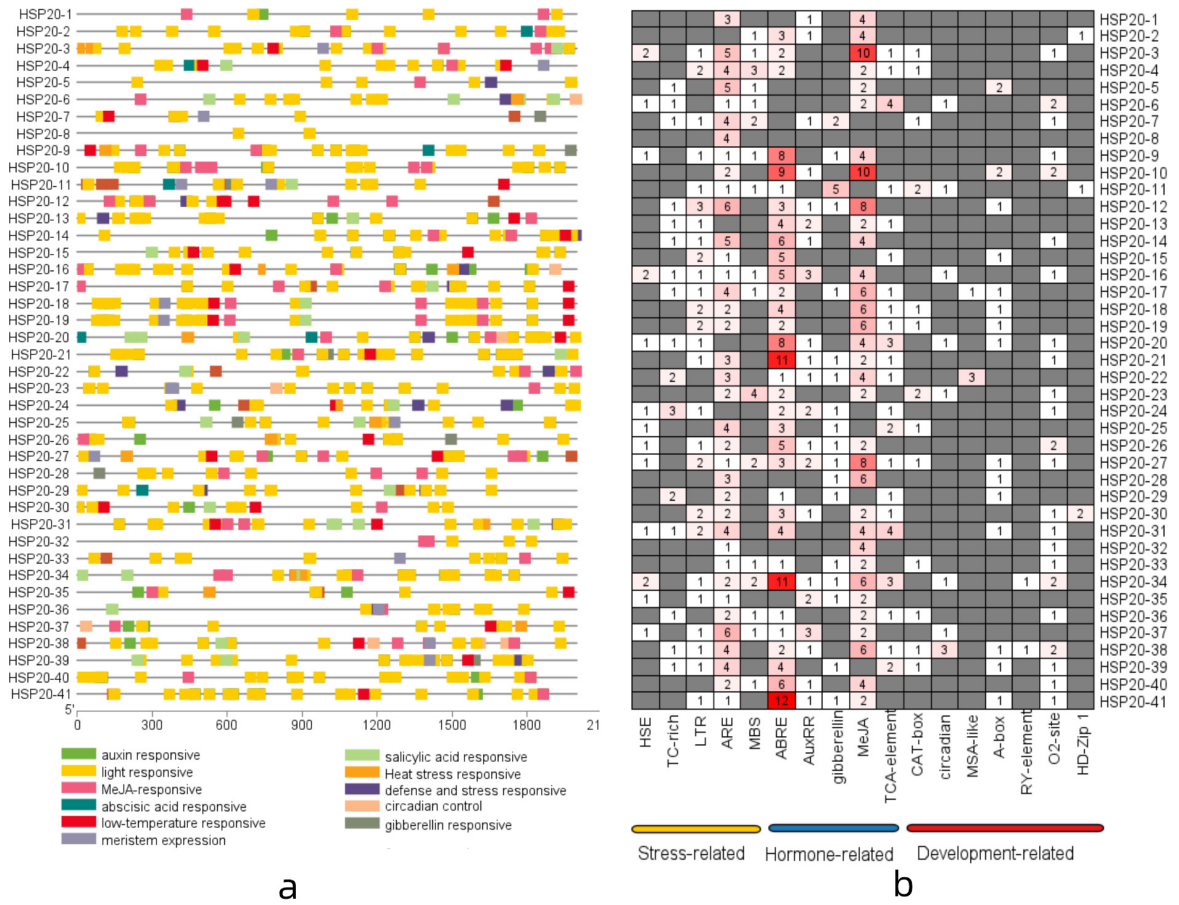

FIGURE 5 | *Cis*-Element analysis of apple *HSP20* gene promoters. **(a)** The different colored blocks represent the different types of *cis*-acting elements and their locations in each *HSP20* gene. **(b)** The number of each *cis*-acting element in the promoter region of each apple *HSP20* gene.

Expression Patterns of *HSP20s* in Response to Cold and Heat Stress

For a preliminary investigation of the functions of apple *HSP20* genes in response to heat and cold stress, nine RNA-seq libraries, including three independent biological replicates for the control, cold-treated and heat-treated, were constructed and sequenced. A heatmap of 41 apple *HSP20* genes was constructed using FPKM values from RNA-Seq data to estimate the expression levels of these genes (**Figure 6**). The heat map showed that the 41 *HSP20* genes clustered in three groups. Cluster A contains one member (*HSP20-33*) of 41 detectable *HSP20* genes, which was barely expressed after heat and cold stress treatment compared with the control. We found *HSP20-33* has no HREs in promoter region. This may be why the gene does not respond to heat stress. All 35 members from cluster B were mainly upregulated after 4 h of heat stress. However, these genes were nearly unchanged or downregulated under cold treatment. Cluster C contains five members, which had similar expression with cluster A, which was barely expressed after heat and cold stress treatment compared with the control.

To further identify which of these *HSP20* genes are most important in mediating heat and cold stress tolerances, 29 differentially expressed *HSP20* genes were selected to be further validated by qRT-PCR based RNA-Seq analysis (**Figure 7** and **Supplementary Table S6**). Consistent with the RNA-Seq data, all 29 selected *HSP20* genes were up-regulated under heat stress. The expression levels of 12 *HSP20* genes (*HSP20-8, 13, 16, 17, 18, 19, 23, 29, 35, 36, 37,* and *38*) were extremely up-regulated (more than 1,000-fold) after 4 h of heat stress. Under heat stress, 29 *HSP20* genes were similarly expressed over time, with peak expression levels at 4 h, except for *HSP20-40*, with peak expression levels at 24 h, while most *HSP20* (*HSP20-2, 5, 7, 13, 17, 18, 19, 22, 23, 25, 28, 29, 30, 31, 35, 36, 37,* and *38*) genes were barely expressed in response to cold stress. However, eight apple HSP20 genes (*HSP20-3, 10, 12, 14, 15, 16, 27,* and *40*) were up-regulated under cold stress.

DISCUSSION

HSP20s are considered to be the most abundant plant stress responsive class among HSPs (Waters, 2013). They have been identified in potato, pepper, tomato, and soybean in responding to temperature stress (Lopes-Caitar et al., 2013; Guo et al., 2015;

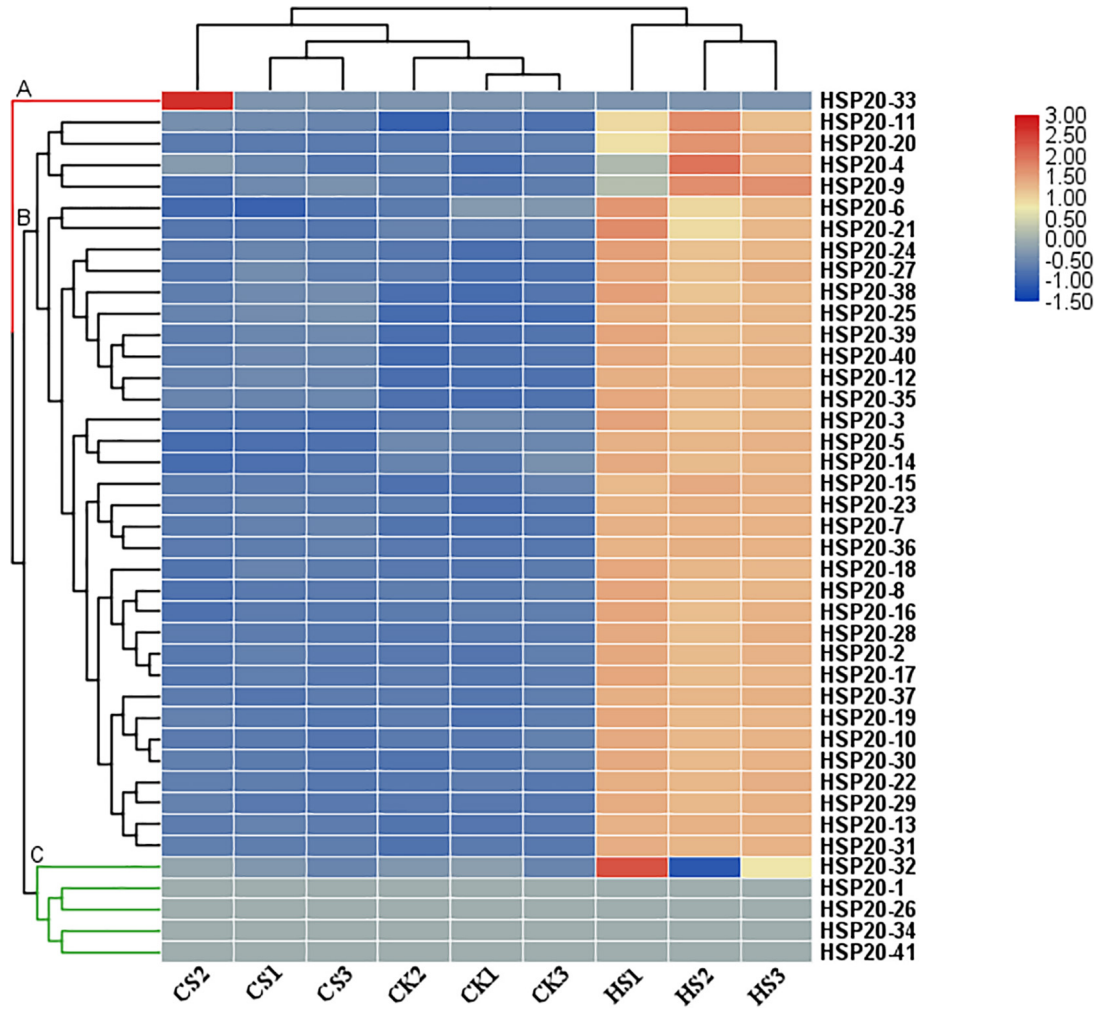

FIGURE 6 | Heat map of the expression profiles of 41 apple *HSP20* genes in response to heat and cold stresses based on RNA-Seq data. Log2 based FPKM value was used to create the heat map with clustering. The color scale representing the relative expression values is shown on the left. CK, control, plants were maintained at 25 ± 1°C; HS, heat stress, plants were maintained at 40 ± 1°C; CS, cold stress, plants were maintained at 4 ± 1°C.

Yu et al., 2016; Zhao et al., 2018). But not studies have conducted an overall identification and characterization of the apple *HSP20* genes. Completion of high-quality apple genome sequencing has provided an opportunity to identify and characterize *HSP20* genes at the whole-genome level.

In the present study, we identified 41 *HSP20* genes and investigated their characteristics from the apple genome database. The number of apple *HSP20* genes was higher than that of *Arabidopsis* (19) (Scharf et al., 2001), slightly higher than rice (39) (Ouyang et al., 2009) but lower than that of watermelon (44) (He et al., 2019), potato (48) (Zhao et al., 2018), and grape (48) (Ji et al., 2019). This difference is most likely due to the fact that apple had gene duplications during evolution (Velasco et al., 2010; Ma et al., 2018). Gene duplication was reported to play an important role in the expansion of the number of gene families in plants (Blanc and Wolfe, 2004; Han et al., 2011). In the current study, 41 apple *HSP20* genes were unevenly mapped on 15 Chrs and most of the *HSP20* genes were located on the distal ends of the Chrs, which might contribute to the occurrence of duplication events in the apple *HSP20* gene family. We confirmed many tandem and segmental duplications in apple *HSP20* genes: 37 of 41 *HSP20* genes were affected by gene duplication, 20 of which were segmental duplication and 17 gene clusters were from tandem duplication. Ma et al. (2018) and Zuo et al. (2018) also found many tandem and segmental duplications in apple receptor-like kinase1-like kinase (CrRLK1L) genes and malate dehydrogenase (MDH) genes, respectively.

To determine the evolutionary relationships of *HSP20* genes, a phylogenetic tree was constructed based on the amino acid sequences of apple, *Arabidopsis* and rice *HSP20* genes. The phylogenetic analysis indicated that the apple *HSP20* family could be divided into 10 subfamilies (CIs, CIIs, CIIIs, CIVs, CVs, MIs, MIIs, Pos, ER, and Ps), which is in line with previous evolutionary classification of *HSP20* genes in *Arabidopsis* and rice (Scharf et al., 2001; Ouyang et al., 2009), indicating a close

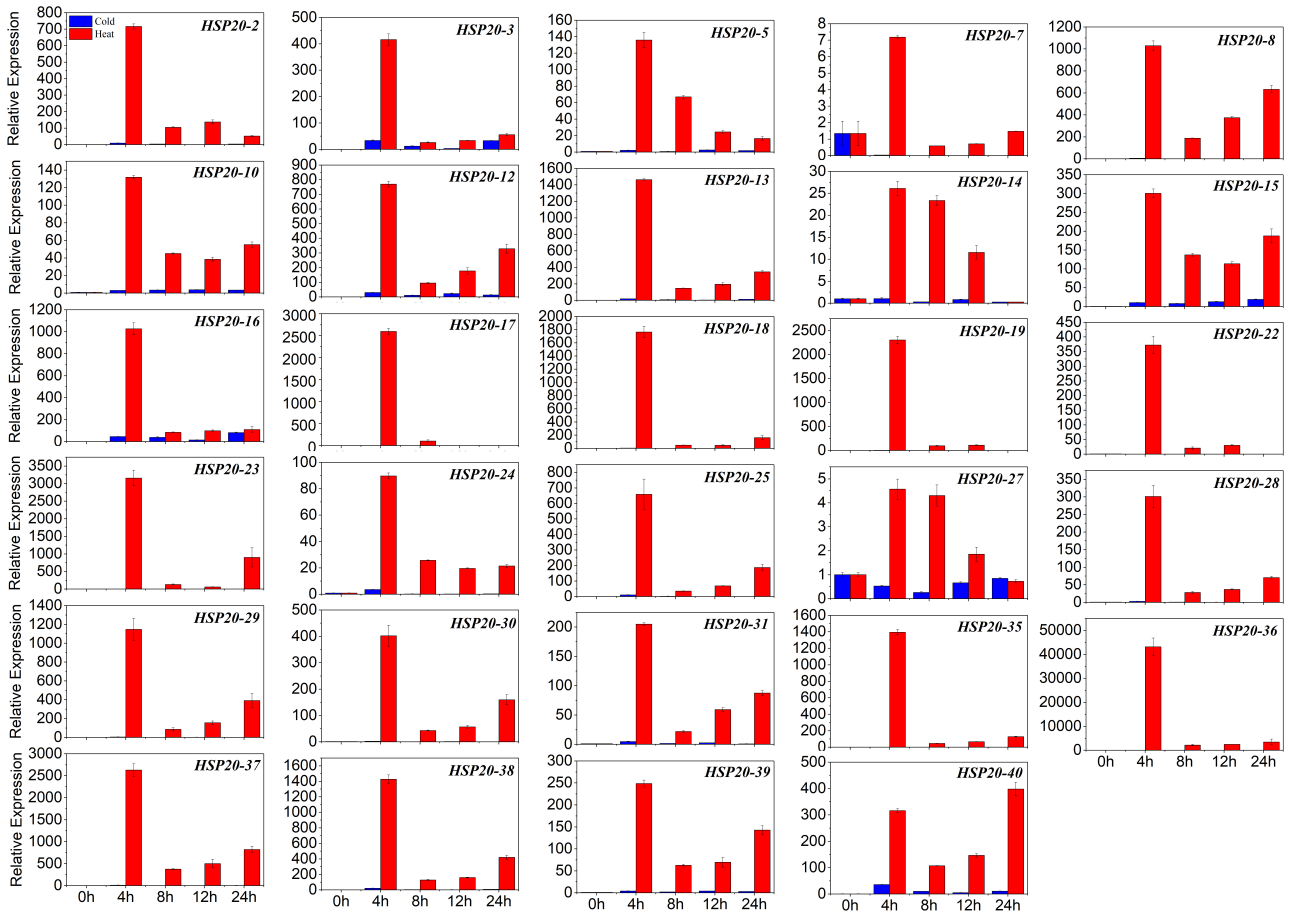

FIGURE 7 | The relative expression levels of apple *HSP20* genes under heat and cold stresses. Mean expression value was calculated from three independent replicates. Vertical bars indicate standard deviation from three independent technical replicates.

relationship among *HSP20* genes from apple, *Arabidopsis* and rice. In addition, most apple *HSP20* genes were classified into CI-CVI, indicating that cytoplasm could be the primary functional area of the *HSP20* family in apple. Gene structure has been documented to function directly in the evolution of multiple gene families (Ji et al., 2019). Gene structure analysis indicated that most apple *HSP20* genes have no introns (60.9%) or one intron (31.7%), suggesting relatively simple gene structures. Similarly, most (93.8%) grape *HSP20* genes have no introns or one short intron (Ji et al., 2019).

Genes with few or no introns are considered to be rapidly activated in response to various stresses (Jeffares et al., 2008). In our study, most apple *HSP20* genes were rapidly induced after 4 h of heat stress, which may support the rapid response. To more comprehensively investigate the evolution of *HSP20* genes, the encoded conserved motifs were also studied. Our results showed that most of the apple *HSP20* genes had five to seven conserved motifs and almost all the *HSP20* genes contained motif 1. This indicates a slow evolutionary rate. Furthermore, we found that most *HSP20* genes in the same subfamilies showed conserved motifs and similar exon/intron structures, supporting their close evolutionary relationship and the classification of subfamilies. Genes in the same subfamily tends to share similar motif and exon–intron organization, which was also reported in tomato (Yu et al., 2016).

Cis-elements in the promoters of genes have been documented as essential in plant physiological response and environmental stress (Yamaguchi-Shinozaki and Shinozaki, 2005). We identified *cis*-elements in the putative promoter regions of apple *HSP20* genes. Numerous hormone responsive, stress-responsive and plant development-related *cis*-elements were found. Among these *cis*-elements, the hormone responsive elements accounted for the highest proportion. Most *HSP20* genes contained stress-related response elements. The present results suggest that most apple *HSP20* genes might be significantly related to stress response. Similar regulatory patterns of *HSP20* genes were also found in pepper and grape (Guo et al., 2015; Ji et al., 2019). In addition, all apple *HSP20* genes contain light signal response elements, which indicate that *HSP20s* were essential in plant growth and development.

Previous studies have revealed that *HSP20s* function directly in plant responses to various stresses (Waters et al., 1996; Guo et al., 2015; Zhao et al., 2018; He et al., 2019; Ji et al., 2019). In this study, the expression profiles of apple *HSP20* genes under

heat stress revealed that the apple *HSP20* genes are involved in heat response. qRT-PCR analysis indicated that most apple *HSP20* genes were up-regulated under heat stress. It is interesting to note that the relative expression levels of 12 *HSP20* genes (*HSP20-8, 13, 16, 17, 18, 19, 23, 29, 35, 36, 37,* and *38*) were extremely up-regulated after 4 h of heat stress. The results from this study suggest that these genes might be mainly involved in the heat stress biological pathway. Similarly, most of the pepper and potato *HSP20s* were also up-regulated in response to heat stress (Guo et al., 2015; Zhao et al., 2018).

Transgenic research has demonstrated the positive role of *HSP20* genes in responding to heat, such as *WsHSP26* in *Arabidopsis* (Mu et al., 2013), and *OsHsp17.7* and *OsHSP20* in rice (Murakami et al., 2004; Guo et al., 2020). In addition, some *HSP20s* showed the same expression profile in response to heat stress, being upregulated with peak expression levels at 4 h, suggesting that these *HSP20s* were co-expressed in response to heat stress. Furthermore, Guo et al. (2015) found that the inducibilities of *HSP20* genes in response to heat stress were obviously different in pepper with different tolerance. Collectively, these results indicate that *HSP20* genes may be positively involved in heat stress responses in plants. Induction of *HSP20* genes by heat stress is well-known (Waters et al., 1996). However, some heat-regulated genes were barely expressed or downregulated in response to cold stress, which indicated that *HSP20* genes were negatively or only slightly involved in the response to cold stress. These results imply that the signaling pathways in plant response to heat and cold stress might be different.

CONCLUSION

This study identified 41 *HSP20* genes in apple. These genes are unequally distributed on 15 chromosomes and were classified into 10 subfamilies based on the phylogenetic tree and subcellular localization. The basic classification, genome distribution, gene structures, conserved motifs, and *cis* acting elements of these genes were analyzed, which will be helpful for a better understanding of the evolutionary relationships of the *HSP20* gene family. Transcriptome analysis revealed that most apple *HSP20* genes were highly induced by heat stress, whereas these genes were nearly unchanged or downregulated under cold stress, indicating that *HSP20* genes were positively involved in heat stress responses in apple. Additionally, we identified several *HSP20* genes that may be utilized as candidates for improving heat stress tolerance. The results presented here will lay a solid foundation for functional characterization of *HSP20* genes through gene-transfer techniques to improve heat tolerance of apple.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

TB designed the experiment, analyzed the data, and drafted the manuscript. FY, CS, and HW collected the public dataset and performed bioinformatics analysis. XZ and SS analyzed the data. JJ and MW assisted with revisions to the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the National Key Research and Development Program of China (2018YFD1000300 and 2019YFD1000100), the National Natural Science Foundation of China (31872058 and 31801844), Program of Young-Backbone Teacher of University in Henan Province (2018GGJS029), and the Special Fund for Henan Agriculture Research System (S2014-11-G02).

ACKNOWLEDGMENTS

The authors thank Anita K. Snyder for critical reading of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.609184/full#supplementary-material

Supplementary Table 1 | qRT-PCR primers used in this study.

Supplementary Table 2 | The CDS sequences, deduced amino acid and promoter sequences of apple *HSP20* genes.

Supplementary Table 3 | Information of *HSP20* genes from apple, rice and *Arabidopsis*.

Supplementary Table 4 | List of the putative motifs of HSP20 proteins.

Supplementary Table 5 | Tandem and segmental duplicated *HSP20* gene pairs within apple genome.

Supplementary Table 6 | Expression levels for selected apple *HSP20* genes under heat and cold stress.

REFERENCES

Asea, A. A. A., Kaur, P., and Calderwood, S. K. (2016). *Heat Shock Proteins and Plants*. Berlin: Springer International Publishing.

Bai, T. H., Li, C. Y., Li, C., Liang, D., Ma, F. W., and Shu, H. R. (2013). Contrasting hypoxia tolerance and adaptation in *Malus* species linked to differences in the stomatal behavior and photosynthesis. *Physiol. Planta* 147, 514–523. doi: 10.1111/j.1399-3054.2012.01683.x

Blanc, G., and Wolfe, K. H. (2004). Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 16, 1679–1691. doi: 10.1105/tpc.021410

Chang, S. J., Puryear, J., and Cairney, J. (1993). A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* 11, 113–116. doi: 10.1007/bf02670468

Chen, C. J., Chen, H., He, Y. H., and Xia, R. (2018). TBtools, a Toolkit for biologists integrating various biological data handling tools with a user-friendly interface. *bioRxiv* [Preprint], doi: 10.1101/289660

Daccord, N., Celton, J. M., Linsmith, G., Becke, C., Choisne, N., Schijlen, E., et al. (2017). High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat. Genet.* 49, 1099–1106. doi: 10.1038/ng.3886

Dobránszki, J., and Teixeira da Silva, J. A. (2010). Micropropagation of apple-A review. *Biotechnol. Adv.* 28, 462–488. doi: 10.1016/j.biotechadv.2010.02.008

Guo, L. M., Li, J., He, J., Liu, H., and Zhang, H. M. (2020). A class I cytosolic *HSP20* of rice enhances heat and salt tolerance in different organisms. *Sci. Rep.* 10:1383.

Guo, M., Liu, J. H., Lu, J. P., Zhai, Y. F., Wang, H., Gong, Z. H., et al. (2015). Genome-wide analysis of the *CaHsp20* gene family in pepper: comprehensive sequence and expression profile analysis under heat stress. *Front. Plant Sci.* 6:806. doi: 10.3389/fpls.2015.00806

Han, Y. P., Zheng, D. M., Vimolmangkang, S., Khan, M. A., Beever, J. E., and Korban, S. S. (2011). Integration of physical and genetic maps in apple confirms whole-genome and segmental duplications in the apple genome. *J. Exp. Bot.* 62, 5117–5130. doi: 10.1093/jxb/err215

He, Y. J., Fan, M., Sun, Y. Y., and Li, L. L. (2019). Genome-wide analysis of watermelon *HSP20s* and their expression profiles and subcellular locations under stresses. *Int. J. Mol. Sci.* 20:12. doi: 10.3390/ijms20010012

Huo, L. Q., Sun, X., Guo, Z. J., Jia, X., Che, R. M., Sun, Y. M., et al. (2020). MdATG18a overexpression improves basal thermotolerance in transgenic apple by decreasing damage to chloroplasts. *Hortic. Res.* 7:21.

Jeffares, D. C., Penkett, C. J., and Bähler, J. (2008). Rapidly regulated genes are intron poor. *Trends Genet.* 24, 375–378. doi: 10.1016/j.tig.2008.05.006

Ji, X. R., Yu, Y. H., Ni, P. Y., Zhang, G. H., and Guo, D. L. (2019). Genome-wide identification of small heatshock protein (*HSP20*) gene family in grape and expression profile during berry development. *BMC Plant Biol.* 19:433. doi: 10.1186/s12870-019-2031-4

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C (T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lopes-Caitar, V. S., De Carvalho, M. C., Darben, L. M., Kuwahara, M. K., Nepomuceno, A. L., Dias, W. P., et al. (2013). Genome-wide analysis of the *Hsp20* gene family in soybean: comprehensive sequence, genomic organization and expression profile analysis under abiotic and biotic stresses. *BMC Genom.* 14:577. doi: 10.1186/1471-2164-14-577

Ma, B. Q., Yuan, Y. Y., Gao, M., Qi, T. H., Li, M. J., and Ma, F. W. (2018). Genome-wide identification, molecular evolution, and expression divergence of aluminum-activated malate transporters in apples. *Int. J. Mol. Sci.* 19:2807. doi: 10.3390/ijms19092807

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Mu, C. J., Zhang, S. J., Yu, G. Z., Chen, N., and Li, X. F. (2013). Overexpression of small heat shock protein LimHSP16.*45* in *Arabidopsis* enhances tolerance to abiotic stresses. *PLoS One* 8:e82264. doi: 10.1371/journal.pone.0082264

Murakami, T., Matsuba, S., Funatsuki, H., Kawaguchi, K., Saruyama, H., Tanida, M., et al. (2004). Overexpression of a small heat shock protein, sHSP17.7, confers both heat tolerance and UV-B resistance to rice plants. *Mol. Breed.* 13, 165–175. doi: 10.1023/b:molb.0000018764.30795.c1

Ouyang, Y., Chen, J. J., Xie, W. B., Wang, L., and Zhang, Q. F. (2009). Comprehensive sequence and expression profile analysis of *hsp20* gene family in rice. *Plant Mol. Biol.* 70, 341–357. doi: 10.1007/s11103-009-9477-y

Peleg, Z., and Blumwald, E. (2011). Hormone balance and abiotic stress tolerance in crop plants. *Curr. Opin. Plant Biol.* 14, 290–295. doi: 10.1016/j.pbi.2011.02.001

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016.095

Scharf, K. D., Siddique, M., and Vierling, E. (2001). The expanding family of *Arabidopsis thaliana* small heat stress proteins and a new family of proteins containing alpha-crystallin domains (acd proteins). *Cell Stress Chaperon.* 6:225. doi: 10.1379/1466-1268(2001)006<0225:tefoat>2.0.co;2

Shen, J. Z., Zhang, D. Y., Zhou, L., Zhang, X. Z., Liao, J. R., Duan, Y., et al. (2019). Transcriptomic and metabolomic profiling of Camellia sinensis L.cv.'Suchazao' exposed to temperature stresses reveals modification in protein synthesis and photosynthetic and anthocyanin biosynthetic pathways. *Tree Physiol.* 39, 1583–1599. doi: 10.1093/treephys/tpz059

Suzuki, N. (2019). Temperature stress and responses in plants. *Int. J. Mol. Sci.* 20:2001. doi: 10.3390/ijms20082001

Torres, C. A., Sepúlveda, G., and Kahlaoui, B. (2017). Phytohormone interaction modulating fruit responses to photooxidative and heat stress on apple (*Malus domestica* Borkh.). *Front. Plant Sci.* 8:2129. doi: 10.3389/fpls.2017.02129

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., et al. (2010). The genome of the domesticated apple (Malus_domestica Borkh.). *Nat. Genet.* 42, 833–839.

Wang, B., Ranjan, R., Khot, L. R., and Peters, R. T. (2020). Smartphone application-enabled apple fruit surface temperature monitoring tool for in-field and real-time sunburn susceptibility prediction. *Sensors* 20:608. doi: 10.3390/s20030608

Wang, N., Guo, T. L., Sun, X., Jia, X., Wang, P., Shao, Y., et al. (2017). Functions of two *Malus hupehensis* (Pamp.) Rehd. YTPs (MhYTP1 and MhYTP2) in biotic- and abiotic-stress responses. *Plant Sci.* 261, 18–27. doi: 10.1016/j.plantsci.2017.05.002

Wang, Q. S., Liu, N. A., Yang, X. Y., Tu, L. L., and Zhang, X. L. (2016). Small RNA-mediated responses to low-and high-temperature stresses in cotton. *Sci. Rep.* 6:35558.

Wang, W. X., Vinocur, B., Shoseyov, O., and Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends Plant Sci.* 9, 244–252. doi: 10.1016/j.tplants.2004.03.006

Waters, E. R. (2013). The evolution, function, structure, and expression of the plant sHSPs. *J. Exp. Bot.* 64, 391–403. doi: 10.1093/jxb/ers355

Waters, E. R., Lee, G. J., and Vierling, E. (1996). Evolution, structure and function of the small heat shock proteins in plants. *J. Exp. Bot.* 47, 325–338. doi: 10.1093/jxb/47.3.325

Yamaguchi-Shinozaki, K., and Shinozaki, K. (2005). Organization of cis-acting regulatory elements in osmotic-and cold-stress-responsive promoters. *Trends Plant Sci.* 10, 88–94. doi: 10.1016/j.tplants.2004.12.012

Yu, J. H., Cheng, Y., Feng, K., Ruan, M. Y., Ye, Q. J., Wang, R. Q., et al. (2016). Genome-wide identification and expression profiling of tomato *Hsp20* gene family in response to biotic and abiotic stresses. *Front. Plant Sci.* 7:1215. doi: 10.3389/fpls.2016.01215

Zhao, P., Wang, D. D., Wang, R. Q., Kong, N. N., Zhang, C., Yang, C. H., et al. (2018). Genome-wide analysis of the potato *Hsp20* gene family: identification, genomic organization and expression profiles in response to heat stress. *BMC Genom.* 19:61. doi: 10.1186/s12864-018-4443-1

Zuo, C. W., Zhang, W. N., Ma, Z. H., Chu, M. Y., Mao, J., An, Z. S., et al. (2018). Genome-wide identification and expression analysis of the CrRLK1L gene family in apple (*Malus domestica*). *Plant Mol. Biol. Rep.* 36, 844–857. doi: 10.1007/s11105-018-1125-8

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 27 November 2020
doi: 10.3389/fgene.2020.603920

QTL Analysis and Fine Mapping of a Major QTL Conferring Kernel Size in Maize (*Zea mays*)

Guiying Wang[1†], Yanming Zhao[1,2†], Wenbo Mao[1], Xiaojie Ma[1] and Chengfu Su[1,2]*

[1] College of Agronomy, Qingdao Agricultural University, Qingdao, China, [2] Shandong Provincial Key Laboratory of Dryland Farming Technology, Qingdao Agricultural University, Qingdao, China

Kernel size is an important agronomic trait for grain yield in maize. The purpose of this study is to map QTLs and predict candidate genes for kernel size in maize. A total of 199 F_2 and its $F_{2:3}$ lines from the cross between SG5/SG7 were developed. A composite interval mapping (CIM) method was used to detect QTLs in three environments of F_2 and $F_{2:3}$ populations. The result showed that a total of 10 QTLs for kernel size were detected, among which were five QTLs for kernel length (KL) and five QTLs for kernel width (KW). Two stable QTLs, *qKW-1,* and *qKL-2,* were mapped in all three environments. Three QTLs, *qKL-1, qKW-1,* and *qKW-2,* were overlapped with the QTLs identified from previous studies. In order to validate and fine map *qKL-2*, near-isogenic lines (NILs) were developed by continuous backcrossing between SG5 as the donor parent and SG7 as the recurrent parent. Marker-assisted selection was conducted from BC_2F_1 generation with molecular markers near *qKL-2*. A secondary linkage map with six markers around the *qKL-2* region was developed and used for fine mapping of *qKL-2.* Finally, *qKL-2* was confirmed in a 1.95 Mb physical interval with selected overlapping recombinant chromosomes on maize chromosome 9 by blasting with the Zea_Mays_B73 v4 genome. Transcriptome analysis showed that a total of 11 out of 40 protein-coding genes differently expressed between the two parents were detected in the identified *qKL-2* interval. GRMZM2G006080 encoding a receptor-like protein kinase FERONIA, was predicted as a candidate gene to control kernel size. The work will not only help to understand the genetic mechanisms of kernel size of maize but also lay a foundation for further fine mapping and even cloning of the promising loci.

Keywords: candidate gene, maize, fine mapping, kernel size, NILs

OPEN ACCESS

Edited by:
Jian Ma,
Sichuan Agricultural University, China

Reviewed by:
Zhi Zheng,
Plant Industry (CSIRO), Australia
Baohua Wang,
Nantong University, China
Yanru Cui,
University of California, Riverside, United States

****Correspondence:***
Chengfu Su
chfsu2008@163.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *08 September 2020*
Accepted: *05 November 2020*
Published: *27 November 2020*

Citation:
Wang G, Zhao Y, Mao W, Ma X and Su C (2020) QTL Analysis and Fine Mapping of a Major QTL Conferring Kernel Size in Maize (Zea mays). Front. Genet. 11:603920. doi: 10.3389/fgene.2020.603920

INTRODUCTION

Maize is an important agricultural crop. It can be served as food, animal feed, and industrial materials (Li et al., 2017) and plays a special role in food security (Liu et al., 2014). High grain yield has always been the most important goal of maize breeding. But most yield traits are quantitative traits controlled by multiple genes (Lynch and Walsh, 1998; Xu, 2010). KL and KW are both considered to be important yield traits (Doebley et al., 2006). Kernel size traits, especially KW, has been revealed to be significantly correlated with grain yield of maize (Li et al., 2013). The improvement of kernel size is therefore of great significance in maize breeding.

To date, numerous studies on maize grain yield traits have been reported at phenotypic levels (Rafiq et al., 2010; Nzuve et al., 2014). However, the genetic architecture and molecular mechanisms underlying natural quantitative variation in kernel yield have not been completely elucidated (Chen et al., 2016). Along with the first genetic linkage map of maize, published in 1986 (Helentjaris et al., 1986), molecular markers based on polymerase chain reaction (PCR) technology have greatly developed and applied for constructing genetic maps. Then, increasing QTLs controlling important agronomic traits in maize were detected by analyzing phenotypic value based on constructed genetic maps. These identified QTLs were distributed on all 10 maize chromosomes (Qiu et al., 2011). Many QTL mapping or fine mapping works for kernel size or weight have been carried out in recent years (Liu et al., 2014; Zhang et al., 2014; Chen et al., 2016). Till now, more than 150 QTLs for kernel size or weight have been identified by using different maize populations (Gramene QTL database). Liu et al. (2020) detected 50 QTLs for kernel size traits in the intermated B73 × Mo17 (IBM) Syn10 doubled haploid (DH) population, of which eight were repetitively identified in at least three environments. A total 55 and 28 QTL for kernel traits were identified by using composite interval mapping (CIM) for single-environment analysis and mixed linear model-based CIM for joint analysis, respectively, with 270 $F_{2:3}$ families derived from the cross between V671 (large kernel) × Mc (small kernel)in five environments (Liu et al., 2014). It is critically important that QTLs should be validated and fine mapped for applying in further marker assisted breeding process. The near-isogenic line (NIL) is one of the most widely accepted populations commonly used for QTL fine mapping. NILs have been successfully used in confirming and fine-mapping QTLs in many species, such as rice (Lin et al., 2003; Li et al., 2004; Xie et al., 2006) and wheat (Xue et al., 2013; Zheng et al., 2015). In maize, a major QTL *qkrnw4* associated with kernel row number was mapped by using a NIL population (Nie et al., 2019). Gao et al. (2019) mapped *qLRI4,* which conferred leaf rolling index by using NIL populations. Yang et al. (2018) mapped a major QTL *qkc7.03* to a 416.27 kb physical interval for kernel cracking with NILs developed.

Great achievements in QTL mapping or isolating underlying genes for kernel size have been made in many species such as rice (Wan et al., 2006; Song et al., 2007; Li et al., 2011; Qiu et al., 2012; Kang et al., 2018), *Arabidopsis thaliana* (Xia et al., 2013; Du et al., 2014), soybean (Xu et al., 2011; Han et al., 2012), and wheat (Sun et al., 2009; Ramya et al., 2010). In particular, genes controlling rice kernel size or weight, such as *GS3* (Fan et al., 2006), *GS5* (Li et al., 2011), *qGL3* (Zhang et al., 2012), *GW2* (Song et al., 2007), *GW8* (Wang et al., 2012), *GS2* (Hu et al., 2015), *qGW7/GL7* (Wang et al., 2015), have been successfully cloned. The study of identifying and cloning kernel-size-related genes has lagged in maize. To a certain extent, the reason for this might be due to the genome of maize is large and complicated for many transposable elements and repetitive sequences exist (Gaut et al., 2000; Feuillet and Eversole, 2009). In addition, most complex traits, such as kernel yield and kernel size, are controlled by many genes with small effects (Edwards et al., 1987; Tian et al., 2011). QTLs identified in different genetic backgrounds across multiple environments have a higher chance of being positionally cloned (Chen et al., 2016).

Based on previous studies, the purposes of this study were as follows: (1) to map QTLs for kernel size in three environments by using F_2 and $F_{2:3}$ populations from the same cross SG5/SG7; (2) to validate and fine map the identified major QTL *qKL-2* by using BC_3F_1NILs; and (3) to reveal differently expressed genes (DEGs) between SG5 and SG7 by RNA-seq technology and predict candidate genes responsible for KL. In the study, we constructed an F_2 and an $F_{2:3}$ populations using two maize inbred lines SG5 and SG7 and evaluated them in three environments for mapping QTLs for kernel size. Furthermore, we finely mapped a major QTL by using the NILs from the cross of SG5 and SG7 and used RNA-seq technology to reveal the DEGs between parental lines SG5 and SG7. Finally, the candidate genes for *qKL-2* were predicted.

RESULTS

Phenotype Evaluation for Segregation Populations

Two kernel size traits, i.e., KL and KW were estimated. The trait values of F_2 population were investigated in 2016, while the phenotypic values of $F_{2:3}$ populations were collected in 2018 and 2019, and these were recorded as $F_{2:3}$-2018 and $F_{2:3}$-2019, respectively. **Table 1** presents the mean values of KL and KW investigated from F_2 and $F_{2:3}$ populations. The two inbred lines SG5 and SG7 were significantly different in both KL and KW traits. KL showed extremely significantly different between SG5 and SG7 ($P < 0.01$, **Figures 3A,B**). The data of two kernel size traits both emerged on normal distribution (**Supplementary Figure 1**). The two traits correlated positively with each other, with Pearson's correlation coefficient being 0.20, 0.25, and 0.24 in F_2-2016, $F_{2:3}$-2018, and $F_{2:3}$-2019, respectively.

QTL Mapping

CIM procedure was applied to map QTLs conferring KL and KW. Manhattan plots were shown in **Figure 1**. A total of 10 QTLs were mapped in total for KL and KW from F_2 and $F_{2:3}$ populations. The information is summarized in **Table 2**. For KL, two major QTLs were mapped on maize chromosome 9 in F_2 population. A total four QTLs were mapped on chromosomes 3, 7, and 9 in $F_{2:3}$-2018 population while three QTLs were mapped

TABLE 1 | Descriptive statistics of KL and KW traits in the F_2 and $F_{2:3}$ mapping populations of maize derived from the cross between SG5 and SG7.

Generations	Trait	SG5 (mm)	SG7 (mm)	Min (mm)	Max (mm)	Mean (mm)	SD (mm)
F_2	KL	9.93	8.99	8.07	12.87	10.42	0.85
	KW	8.07	11.17	8.03	11.80	9.98	0.78
$F_{2:3}$-2018	KL	9.93	8.99	8.49	13.21	10.34	0.74
	KW	8.07	11.17	8.24	12.12	10.07	0.74
$F_{2:3}$-2019	KL	9.93	8.99	8.40	13.35	10.34	0.75
	KW	8.07	11.17	8.28	12.22	10.08	0.73

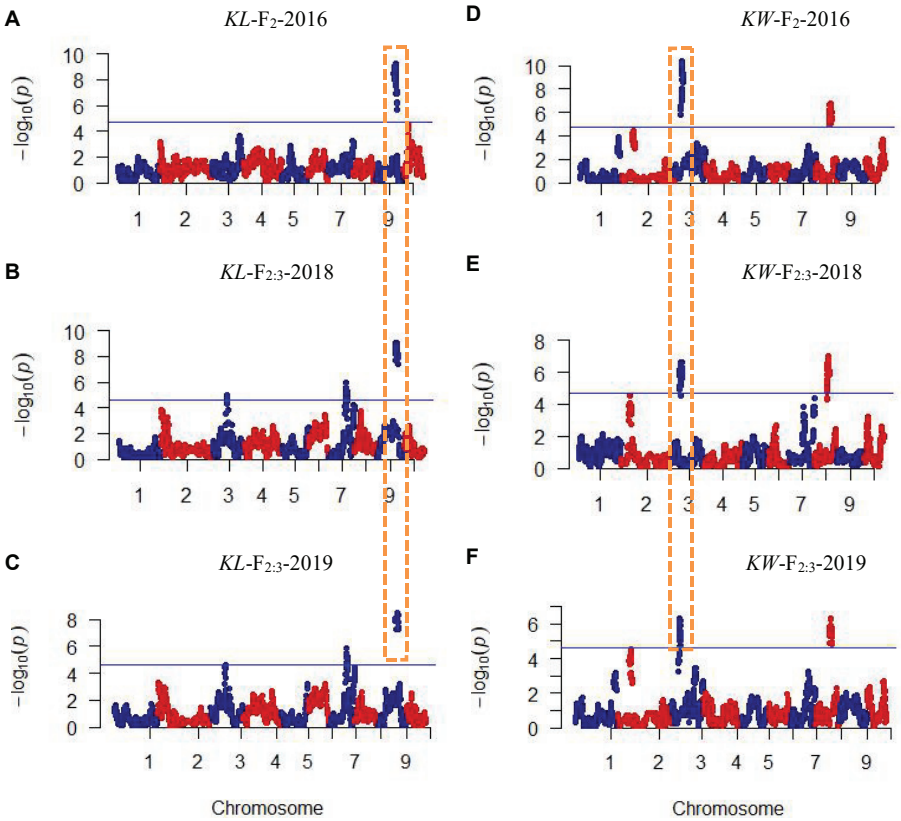

FIGURE 1 | Plots of test statistic−Log10(p) against genome location for KL and KW traits in maize using the CIM method. The horizontal blue line of each panel is the critical value of the test statistic generated from 1,000 permuted samples. **(A–C)** Indicate KL mapping results in three environments while **(D–F)** indicate KW mapping results in three environments. Dotted rectangles with orange color indicate these QTLs were mapped repeatedly in all three environments.

on chromosomes 7 and 9 in $F_{2:3}$-2019 population. For KW, three QTLs were mapped on maize chromosomes 3 and 8. A total three QTLs were mapped on maize chromosomes 3 and 8 in both $F_{2:3}$-2018 and $F_{2:3}$-2019 populations, respectively. The phenotypic variation explained by these QTLs ranged from 8.4 to 23.0%, with a mean value of 14.25 and 14.46%, 14.03 and 12.97%, and 10.83 and 13.67% for KL and KW in F_2-2016, $F_{2:3}$-2018, and $F_{2:3}$-2019, respectively. The LOD score ranges from 4.0 for *qKL-7* to 9.5 for *qKW-1*. Among the 10 QTLs, *qKL-2* for KL, and *qKW-1* for KW were detected in all the three environments (**Figure 2**, highlighted in green color circle). That is, they were stable QTLs in the study. Four QTLs (*qKW-2*, *qKL-7*, *qKW-3,* and *qKL-10*) were detected in two environments, highlighted in blue color circle in **Figure 2**. In addition, three QTLs, *qKL-1*, *qKW-1*, and *qKW-2,* overlapped with the QTLs identified from the metaQTL analysis (Chen et al., 2017).

Fine Mapping qKL-2 With NILs

From 2017 to 2019, a NIL population, consisting of 998 BC_3F_1 lines, was developed by introgressing the *qKL-2* genomic region of SG5 into the SG7 genetic background. A secondary linkage map with six markers (**Supplementary Table 1**) around *qKL-2* was generated. The six markers were located at 115.23, 130.51, 133.34, 135.29, 139.75, and 153.88 Mb on chromosome 9 by blasting maize B73 RefGen_v4 (**Figure 3C**). The secondary linkage map was 43.35 cM in length, and the genetic distances between every two adjacent markers were 16.75, 8.39, 0.80, 5.67, and 11.74 cM.

Then the major QTL *qKL-2* was detected with the secondary linkage map of NILs by CIM method in QTL Cartographer v2.5. The *qKL-2* had an additive effect of 0.97 mm and explained 16% of phenotypic variation. The LOD peak indicated that *qKL-2* was most likely located between SSR3 and SSR5, the LOD peak position was located between SSR3 and SSR4 (**Figure 3C**). To confirm the narrowed *qKL-2* interval, five recombinant types, namely, Class 1–Class 5, were selected from 998 NILs. Class 1 indicates that 28 recombinants with SSR1 and SSR2 homozygous and SSR3–SSR6 heterozygous. Class 2 indicates 33 recombinants with SSR1 and SSR2 heterozygous and SSR3–SSR6 homozygous. Class 3 indicates three recombinants with SSR1–SSR3 heterozygous and SSR4–SSR6 homozygous. Class 4 indicates 20 recombinants with SSR1–SSR4 heterozygous and SSR5–SSR6 homozygous. Class 5 indicates 47 recombinants with SSR1–SSR5 heterozygous and SSR6 homozygous. At SSR3 and SSR4 loci, Classes 2 and 3 were homozygous while Classes 1, 4, and 5 were heterozygous. There was significantly difference in phenotypic values between the two set of recombinants Classes 2 and 3 and Classes 1, 4, and 5 (**Figure 3D**). The progeny test

TABLE 2 | QTL identified for KL and KW traits of maize using high-density SNP bin-map from composite interval mapping (CIM).

Environments	QTL	Chr	Flanking markers	Positions (Mb)	Interval (Mb)	Physical length (Mb)	LOD[c]	ADD[a]	DOM[b]	R^2 (%)
					Kernel length (KL) trait					
F_2-2016	***qKL-1***	**9**	**mk3093–mk3100**	**126.05**	**124.90–127.55**	**2.65**	**8.4**	**0.34**	**0.18**	**14.8**
	qKL-2	**9**	**mk3106–mk3114**	**134.65**	**133.20–135.75**	**2.55**	**7.7**	**0.26**	**0.24**	**13.7**
	qKL-3	3	mk1343–mk1350	217.85	215.90–218.1	2.20	2.9	−0.13	−0.17	4.7
	qKL-4	2	mk624–mk638	2.75	0.10–3.55	3.45	2.5	0.06	0.254	4.2
	qKL-5	10	mk3228–mk3236	105.05	96.45–109.2	12.75	3.8	−0.19	−0.14	6.4
$F_{2:3}$-2018	***qKL-2***	**9**	**mk3105–mk3110**	**134.65**	**131.55–134.75**	**3.20**	**8.1**	**0.23**	**0.25**	**15.3**
	qKL-4	2	Mk624–mk634	2.85	0.10–3.00	2.90	3.1	−0.02	0.31	5.4
	qKL-6	**3**	**mk1145–mk1167**	**115.20**	**103.65–117.85**	**14.2**	**4.2**	**−0.38**	**0.17**	**12.1**
	qKL-7	**7**	**mk2510–mk2525**	**143.00**	**137.30–145.70**	**8.40**	**5.1**	**−0.40**	**0.32**	**13.4**
	qKL-8	7	mk2618–mk2622	174.6.	174.35–174.90	0.55	3.5	−0.28	0.04	6.5
	qKL-9	8	mk2689–mk2701	21.85	21.60–33.80	12.20	3.1	−0.30	0.20	8.0
	qKL-10	**9**	**mk3077–mk3084**	**115.15**	**113.65–120.70**	**7.05**	**8.1**	**0.30**	**0.19**	**15.3**
$F_{2:3}$-2019	***qKL-2***	**9**	**mk3100–mk3104**	**128.0.**	**127.55–130.05**	**2.50**	**6.9**	**0.27**	**0.17**	**13.4**
	qKL-4	2	mk624–mk637	2.85	0.10–3.45	3.35	2.6	0.00	0.27	4.5
	qKL-6	3	mk1158–mk1167	115.2.	114.80–117.85	3.05	3.9	−0.37	0.19	11.3
	qKL-7	**7**	**mk2517–mk2525**	**143.00**	**141.25–145.70**	**4.45**	**4.0**	**0.40**	**0.34**	**13.0**
	qKL-8	7	mk2617–mk2622	174.60	174.05–174.90	0.85	3.7	−0.26	-0.01	5.9
	qKL-10	**9**	**mk3077–mk3084**	**115.15**	**113.65–120.70**	**7.05**	**7.6**	**0.28**	**0.20**	**14.6**
	qKL-11	5	mk2183–mk2188	215.30	214.55–215.55	1.00	2.6	0.25	−0.26	5.9
	qKL-12	6	mk2298–mk2303	149.80	143.05–157.50	14.45	2.5	0.09	0.24	4.4
					Kernel width (KW) trait					
F_2-2016	***qKW-1***	**3**	**mk1042–mk1060**	**30.85**	**30.20–44.55**	**14.35**	**9.5**	**−0.54**	**0.17**	**23.0**
	qKW-2	**8**	**mk2806–mk2814**	**148.95**	**144.55–151.00**	**6.45**	**5.8**	**−0.34**	**−0.02**	**10.0**
	qKW-3	**8**	**mk2814–mk2820**	**152.25**	**151.00–157.95**	**6.95**	**5.9**	**−0.35**	**0.00**	**10.9**
	qKW-4	1	mk577–mk603	292.20	288.45–295.55	7.10	3.1	−0.29	0.09	7.2
	qKW-5	2	mk667–mk673	17.05	13.60–17.55	3.95	3.4	−0.23	−0.07	4.5
	qKW-6	2	mk676–mk686	24.05	19.55–24.35	4.80	3.7	−0.19	−0.13	6.1
	qKW-7	10	mk3284–mk3294	145.50	144.70–146.95	2.25	3.0	−0.15	0.33	4.9
$F_{2:3}$-2018	***qKW-1***	**3**	**mk1040–mk1047**	**30.85**	**29.30–32.75**	**3.45**	**5.8**	**−0.37**	**−0.04**	**9.9**
	qKW-2	**8**	**mk2811–mk2814**	**148.95**	**147.80–151.00**	**3.20**	**5.8**	**−0.43**	**0.13**	**14.5**
	qKW-5	2	mk667–mk676	16.70	13.60–19.55	5.95	3.2	−0.12	−0.20	5.3
	qKW-8	**8**	**mk2799–2808**	**141.80**	**140.30–145.60**	**5.30**	**6.2**	**−0.43**	**0.11**	**14.5**
	qKW-9	2	mk657–mk662	9.70	9.30–11.30	2.00	3.8	−0.06	−0.29	6.2
$F_{2:3}$-2019	***qKW-1***	**3**	**mk1035–mk1042**	**28.95**	**27.10–30.20**	**3.10**	**5.5**	**−0.47**	**0.19**	**14.5**
	qKW-3	**8**	**mk2814–mk2817**	**152.25**	**151.00–153.30**	**2.30**	**5.5**	**−0.33**	**−0.01**	**9.6**
	qKW-4	1	mk579–mk591	292.55	289.45–293.25	3.80	2.5	−0.23	0.05	5.0
	qKW-6	2	mk676–mk686	23.75	19.55–24.35	4.80	3.8	−0.17	−0.14	6.2
	qKW-10	**8**	**mk2826–mk2835**	**163.25**	**160.70–164.95**	**4.25**	**5.0**	**−0.24**	**−0.14**	**8.4**

[a] *Estimated additive effect.*
[b] *Estimated dominance effect.*
[c] *The logarithm of odds (LOD) value 3.86 for detected traits was set up by 1,000 times permutations results (in bold). QTL statistics were also reported for those in which the LOD score exceeded 2.5 but was less than 3.86 (in no bold).*

of homozygous segregants indicated that *qKL-2* was located in an interval of 1.95 Mb (133.34–135.29) and flanked by SSR3 and SSR4 physical interval (**Table 3**). The selected overlapping recombinant chromosomes also supported the location of *qKL-2*.

Candidate Genes for qKL-2 Prediction

RNA-seq procedure was conducted for 18 RNA grain samples at different developmental stages. Results showed that the 1.95 Mb physical intervals of *qKL-2* encompassed 40 protein coding genes (**Figure 3E**). After DEGs analysis, a total of 11 protein coding genes differently expressed and left in the *qKL-2* physical intervals (**Table 4**). Previous studies indicated that FERONIA receptor kinase controls seed size in *Arabidopsis thaliana* (Yu et al., 2014). GRMZM2G006080 encodes receptor-like protein kinase FERONIA and was predicted as a candidate gene of *qKL-2,* which is most likely responsible for KL.

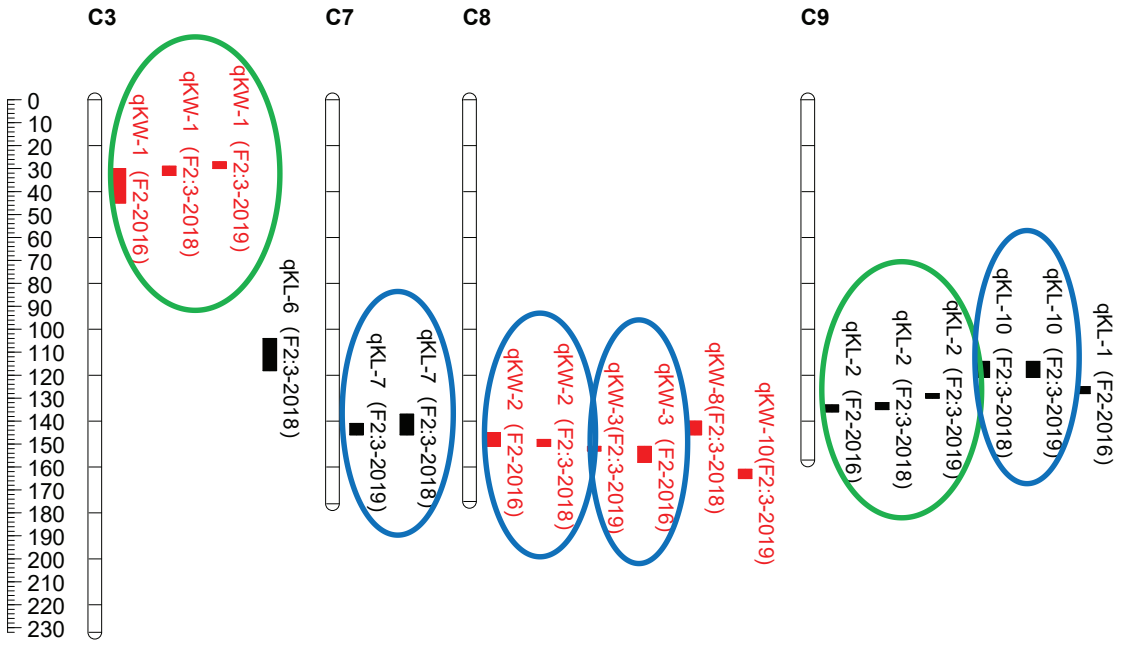

FIGURE 2 | QTL locations for KL and KW traits studied in the F_2 and $F_{2:3}$ populations from cross SG5/SG7. QTLs were represented in different colors for kernel size traits including red for KW (kernel width, mm), black for KL (kernel length, mm) on maize chromosomes 3, 7, 8, and 9. QTLs represented by bars are shown on the right of the linkage groups close to their corresponding markers. Supported intervals for each QTL are indicated by the length of vertical bars. The QTLs circled in green were stably detected in three environments while QTLs circled in blue were repeatedly detected in two environments.

DISCUSSION

Kernel size controlled by multiple genes is an important component of grain yield in maize. Grain yield was influenced significantly by kernel size, especially KL (Li et al., 2009, 2013). Stable QTLs are of great significance for marker-assisted breeding, while false positive QTLs are of no use. Normally, two steps, i.e., primary mapping and fine mapping, are needed for QTL analysis unless experiments were conducted in multiple environments with as many as sample size and marker numbers. In this study, primary mapping was carried out in three environments, and two kernel-size QTLs, *qKL-2,* and *qKW-1,* detected in all three environments were stable. The two QTLs could be benefit for further marker assisted breeding. Chen et al. (2017) conducted metaQTL analysis based on collecting information on QTLs conferring maize yield-related traits from 33 published studies. A total of 76 MQTLs for maize yield and its related traits were identified across the whole genome, with the number per chromosome ranging from four on chromosome 4–10 on chromosome 5 (Chen et al., 2017). After comparing with the metaQTL analysis results, *qKL-1, qKW-1,* and *qKW-2* detected in this study all overlapped with those MQTLs for kernel-related traits but with more decreased physical intervals (**Table 2**).

For *qKL-2* locus, primary mapping results showed that the physical intervals were 133.20–135.75, 131.55–134.75, and 127.55–130.05 Mb on chromosome 9, respectively, in three environments. In order to confirm and fine map *qKL-2*, a NIL population was developed by continuous backcross with markers assisted selection for confirming and fine mapping *qKL-2*. Finally, *qKL-2* was mapped in a 1.95 Mb (133.34–135.29 Mb) interval on maize chromosome 9. Compared with metaQTL analysis results from Chen et al. (2017), MQTL-66, which includes 16 QTLs related to grain yield, ear-related traits, and kernel-related traits located in 120.2–133.6 physical interval on chromosome 9. There was only 0.26 Mb physical distance overlap for *qKL-2* (133.34–135.29 Mb) and MQTL-66 (120.2–133.6 Mb). It is very likely that *qKL-2* was a new locus to control KL.

It is of critical importance that the less genes the better in target QTL interval for map-based cloning. In this study, RNA-seq technology was applied for transcriptomic analyzing DEGs between SG5 and SG7 grains in different developmental stages. DEGs identified were located in the *qKL-2* interval. After DEGs analysis, only 11 protein coding genes were left in the QTL *qKL-2* intervals (**Table 4**). The potential functional genes in QTLs physical intervals decreased significantly after DEGs analysis. According to gene annotation from Blast swiss prot, the function of 11 genes include endoglucanase, 17.0 kDa class II heat shock protein, phospholipid-transporting ATPase 1, receptor-like protein kinase FERONIA, calcium-binding protein, selenium-binding protein 2, NAC domain-containing protein, and thioredoxin-like 1–2, chloroplastic. Further comparative genomics analysis was applied for predicting candidate genes. The evidence on studies of rice or *Arabidopsis thaliana* showed that kernel size was regulated by multiple signaling pathways, including

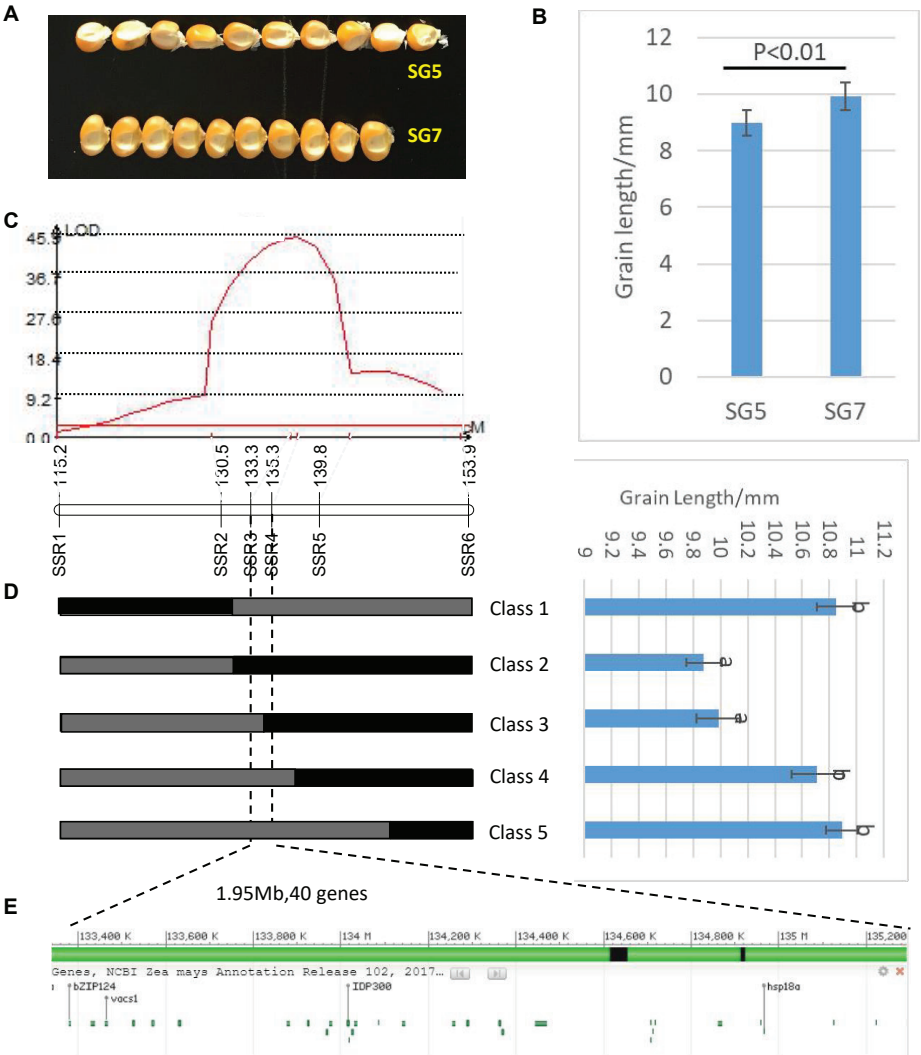

FIGURE 3 | (A) Grain length difference between SG5 and SG7. **(B)** Significance test of difference between SG5 and SG7, $P < 0.01$ means that difference was extremely significant between SG5 and SG7. **(C)** LOD profile of *qKL-2*, which was identified in the BC_3F_1 populations. **(D)** Major QTL*qKL-2* was mapped in SSR3 and SSR5interval by using 998 BC_3F_1 plants. Class 1–Class 5 indicates the recombinants with different recombination types. Graphical genotypes and their phenotypic values of recombinant lines separated in BC_3F_1. Black bars and gray bars represent the chromosomal segments for the homozygous SG7 and heterozygous alleles, respectively. The progeny test of homozygous segregants indicated that *qKL-2* was located at an interval of 1.95 Mb and flanked by SSR3 and SSR4. Different letters indicate significant difference at the 0.05 level. **(E)** Physical positions of the 40 protein coding genes in the mapped 1.95 Mb region (B73 RefGen_v4).

ubiquitin-proteasome degradation (Verma et al., 2004), the transcription factor pathway, the phytohormone signaling pathway, and the G protein independent pathway. Yu et al. (2014) concluded that receptor kinase FERONIA involved in a signaling pathway negatively regulated the elongation of integument cells and then controlled the seed size in *A. thaliana*. Based on the above function analysis of 11 protein coding genes, GRMZM2G006080, which encodes receptor-like protein kinase FERONIA, was predicted as a candidate gene of kernel size. The predicted candidate gene will not only be helpful for underlying genetic mechanism for kernel size but also provides a basis for improving kernel size traits in maize.

MATERIALS AND METHODS

Segregation Population Development and Phenotypic Evaluation

Two maize inbred lines, SG5 and SG7, were used in the study. The seeds were provided by the Institute of Grain and Oil, Liupanshui Academy of Agricultural Sciences, Liupanshui, China. We developed an F_2 population by crossing SG5 and SG7 in Liupanshui, Guizhou province of China in the summer of 2013 and 2014. A total of 199 F_2 individuals were planted at the Panxian Maize Breeding Station in Sanya, China, in the winter of 2014. Then, an $F_{2:3}$ segregation population containing

TABLE 3 | Statistical analysis of phenotypic values from different types of recombinants in NILs around the *qKL-2* region.

	No. of recombinants	Grain length/mm
Class[a] 1	28	10.85 ± 0.14
Class 2	33	9.88 ± 0.13
Class 3	3	9.99 ± 0.16
Class 4	20	10.71 ± 0.19
Class 5	47	10.89 ± 0.11

[a]*Class 1–Class 5 indicates different types of recombinants with six SSR markers. Class 1 indicates that 28 recombinants with SSR1 and SSR2 homozygous and SSR3–SSR6 heterozygous. Class 2 indicates 33 recombinants with SSR1 and SSR2 heterozygous and SSR3–SSR6 homozygous. Class 3 indicates three recombinants with SSR1–SSR3 heterozygous and SSR4–SSR6 homozygous. Class 4 indicates 20 recombinants with SSR1–SSR4 heterozygous and SSR5–SSR6 homozygous. Class 5 indicates 47 recombinants with SSR1–SSR5 heterozygous and SSR6 homozygous.*

199 lines was developed by selfing each F_2 individuals. The $F_{2:3}$ population was planted at the Panxian Maize Breeding Station in Sanya for kernel size evaluation in the summer of 2018 and 2019. Field experiment was performed in a randomized block design with three replications. Single-row plots with row spacing of 50 cm were adopted, and each plot grew 15 plants with plant spacing of 35 cm. Kernel size traits, including KL and KW, were investigated in both F_2 and $F_{2:3}$ populations after corns were harvested and dried naturally. For F_2 generation, the traits were estimated by the mean value of three repeats including 10 kernels of an ear. For $F_{2:3}$, kernel size evaluation was based on eight ears from the middle part of each plot. KL and KW were estimated by mean value of three repeats including 10 kernels randomly selected from bulked kernels of eight ears. The measured kernels were all sampled from the middle part of an ear.

Young leaves were collected from each F_2 individual for DNA extraction. The methods of genomic DNA extraction, genotype sequencing, and grouping, single nucleotide polymorphisms (SNPs) identification, and high-density linkage map construction were presented in our previous study (Su et al., 2017). The forward regression model of CIM method in QTL Cartographer v2.5 was applied for QTL mapping with walking speed of 1 cM. The likelihood of odds (LOD) value 3.86 was used to declare a QTL, which was based upon 1,000 times permutations analysis. QTL statistics were also reported for those in which the LOD score exceeded 2.5. LOD peaks were used for determining the position of a significant QTL on chromosomes. The positive additive effect value of a QTL indicates that the increase in phenotypic value is provided by SG5 alleles while negative value indicates the decrease in phenotypic value is provided by SG7 alleles. MapChart 2.32 software (Voorrips, 2002) was used for the graphical presentation of QTLs. The QTLs that are mapped in F_2 and $F_{2:3}$ populations were compared, and the consistent one will be regarded as stable QTL.

NILs Development and qKL-2 Fine Mapping

NILs for the *qKL-2* locus were developed by using continuous backcrossing combined with marker-assisted selection methods.

TABLE 4 | Differentially expressed genes out of 40 protein coding genes in 1.95 Mb physical interval on chromosome 9 and candidate gene predicted for *qKL-2*.

GeneID (B73 RefGen_v3)	Start (bp)	End (bp)	Length	Annotation from blast swiss prot	LogFC[a] or RCP1/RCP2[b]		
					Day 5[c]	Day 10	Day 15
GRMZM2G099101	132,128,943	132,133,712	2,349	Endoglucanase 9	0.36	0.42	1.66
GRMZM5G899188	132,973,437	132,974,250	814	17.0 kDa class II heat shock protein	2.03	0.98	3.23
GRMZM2G398288	131,627,588	131,635,485	4,521	Phospholipid-transporting ATPase 1	−0.40[d]	0.36	−0.22
GRMZM2G027437	131,778,230	131,786,437	1,277		−0.55	−0.45	−0.79
GRMZM2G006080	131,829,772	131,832,943	3,172	Receptor-like protein kinase FERONIA	0.41	−0.09	−0.04
GRMZM2G309327	131,924,112	131,925,141	1,030	Probable calcium-binding protein	−1.87	−2.20	−1.41
GRMZM2G159500	133,066,105	133,068,297	2,048	NAC domain-containing protein	−0.76	-0.22	−0.21
GRMZM2G102382	133,137,426	133,140,041	1,413	Thioredoxin-like 1–2, chloroplastic	0.85	0.95	1.01
GRMZM2G102657	13,279,7892	132,800,350	2,053		0/9.54	0/18.55	0/7.55
GRMZM2G404249	132,803,660	132,804,501	842	17.0 kDa class II heat shock protein	1.70	2.29	1.27
GRMZM5G875954	132,964,847	132,972,815	1,206	Selenium-binding protein 2	−5.07	−3.20	−1.20

[a]*Log 2 ratio, number of folds the gene is differentially expressed in RNA-seq.*
[b]*Different of read counts between P_1 and P_2.*
[c]*Day 5, day 10, day 15 indicate grain samples collected after selfing 5, 10, and 15 days between the two parents SG5 and SG7.*
[d]*Positive sign indicates gene transcript expressed high in SG5 while negative sign indicates gene transcript expressed high in SG7.*

The SSR molecular markers that are near *qKL-2* and are polymorphic between donor parent SG5 and recurrent parent SG7 were used for marker-assisted selection of the BC_2F_1 generation. These SSR markers, based on resequencing maize genome results, were all developed by Xu et al. (2013). We choose SSRs that are near *qKL-2* position with high polymorphism information content (PIC) values. These SSRs were used for screening polymorphism between our parental lines SG5 and SG7. SSRs with clearly bands of polyacrylamide gel electrophoresis (PAGE) and polymorphism between SG5 and SG7 were selected for developing secondary linkage map and further fine mapping works. Phenotypic value for BC_3F_1 lines was investigated in the same way as for the F_2 population. Young healthy leaves were collected from each of the 998 BC_3F_1 line for genomic DNA extraction. Plant Genomic DNA Kit (TIANGEN, Beijing, China) were used and the manufacturer's protocols were followed. DNA purity was checked by 1% agarose gel and NanoPhotometer R spectrophotometer (IMPLEN, CA, United States). DNA concentration was then measured using an Qubit R DNA Assay Kit in Qubit R 2.0 Flurometer (Life Technologies, CA, United States).

The secondary linkage map around *qKL-2* was generated by JoinMap 3.0 software (Van Ooijen and Voorrips, 2001). QTL Cartographer v2.5 was applied for QTL mapping with the CIM method, walking speed 1 cM, and a LOD threshold of 10.0.

Candidate Gene for qKL-2 Prediction

Grains of SG5 and SG7 were sampled on the 5th, 10th, and 15th days after selfing three biological replicates. All collected samples were immediately frozen in liquid nitrogen and then transferred to a −80°C environment before RNA extraction. We finally got 18 grain samples in total. All the samples were sequenced at the Illumina NovaSeq platform. Raw reads with fastq format were firstly handled by in-house perl scripts. Clean reads were then obtained after deleting reads containing adapters and ploy-N and removing reads of a low quality in raw data. In the meantime, the GC content and Q20 and Q30 of the clean reads were calculated. High-quality clean data were then carried out for further downstream analyzing. Reference genome was downloaded directly from genome website[1], and correlated files of gene annotation were also downloaded from the same website. Bowtie v2.2.3 was used for building reference genome index and TopHat v2.0.12 (Trapnell et al., 2013) was used for aligning paired-end clean reads to the reference genome. The number of reads mapped to each gene was counted by HTSeq v0.6.1. For each gene, the expected number of fragments per kilobase of transcript sequence per millions base pairs (FPKM) was calculated by analyzing the gene length and reads mapped to the gene. FPKM is a widely accepted method currently to evaluate levels of gene expression based on considering sequencing depth effect and gene length of the read count simultaneously (Trapnell et al., 2010). The DEGSeq R package (1.20.0) was applied for analyzing differential expression between two conditions. The *P*-values adjusted by using the Benjamini and Hochberg method were used. The threshold of corrected *P*-value 0.005 and log 2 (Fold change) of 1 (absolute value) were considered as significantly differential expression. More information about the methods for reference genome index construction, paired-end clean reads alignment and count, FPKM calculation and DEGs analysis referred to our previous study (Zhao and Su, 2019).

Through analyzing DEGs between SG5 and SG7, the DEGs that were overlaid on to a physical interval of *qKL-2* were considered as candidate genes for kernel size in maize. The detected DEGs were further annotated from Blast Swiss Prot database.

[1] ftp://ftp.ensemblgenomes.org/pub/plants/release-29/fasta/zea_mays/dna/Zea_mays.AGPv3.29.dna.toplevel.fa.gz

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in NCBI BioProject PRJNA673992, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA673992/. We have uploaded our SNP sequencing data to FigShare repository https://doi.org/10.6084/m9.figshare.13120322.v1.

AUTHOR CONTRIBUTIONS

GW and YZ developed the F_2, $F_{2:3}$, and BC_3F_1 population. WM and XM performed the phenotype investigation. CS developed the genotyping of $F_{2:3}$ progeny, analyzed the data, and drafted the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant #31460359), Research Foundation for Advanced Talents of Qingdao Agricultural University (Grant #6631118035), the Natural Science Foundation of Guizhou Province of China [Grant #Qian Kehe J word (2014)2155 and Grant #Qian Kehe LH word (2015)7605].

ACKNOWLEDGMENTS

This manuscript has been released as a pre-print at ResearchSquare (Wang et al., 2020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.603920/full#supplementary-material

REFERENCES

Chen, L., An, Y., Li, Y., Li, C., Shi, Y., Song, Y., et al. (2017). Candidate loci for yield-related traits in maize revealed by a combination of metaQTL analysis and regional association mapping. *Front. Plant Sci.* 8:2190. doi: 10.3389/fpls.2017.02190

Chen, L., Li, Y-X, Li, C., Wu, X., Qin, W., Li, X., et al. (2016). Fine-mapping of qGW4. 05, a major QTL for kernel weight and size in maize. *BMC Plant Biol.* 16:81. doi: 10.1186/s12870-016-0768-6

Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321.

Du, L., Li, N., Chen, L., Xu, Y., Li, Y., Zhang, Y., et al. (2014). The ubiquitin receptor DA1 regulates seed and organ size by modulating the stability of the ubiquitin-specific protease UBP15/SOD2 in Arabidopsis. *Plant Cell* 26, 665–677. doi: 10.1105/tpc.114.122663

Edwards, M., Stuber, C., and Wendel, J. (1987). Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics* 116, 113–125.

Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., et al. (2006). GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theoret. Appl. Genet.* 112, 1164–1171. doi: 10.1007/s00122-006-0218-1

Feuillet, C., and Eversole, K. (2009). Solving the maze. *Science* 326, 1071–1072. doi: 10.1126/science.1183463

Gao, L., Yang, G., Li, Y., Fan, N., Li, H., Zhang, M., et al. (2019). Fine mapping and candidate gene analysis of a QTL associated with leaf rolling index on chromosome 4 of maize (Zea mays L.). TAG. *Theor. Appl. Genet.* 132, 3047–3062. doi: 10.1007/s00122-019-03405-1

Gaut, B. S., D'ennequin, M. L. T., Peek, A. S., and Sawkins, M. C. (2000). Maize as a model for the evolution of plant nuclear genomes. *Proc. Natl. Acad. Sci.* 97, 7008–7015. doi: 10.1073/pnas.97.13.7008

Han, Y., Li, D., Zhu, D., Li, H., Li, X., Teng, W., et al. (2012). QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. *Theoret. Appl. Genet.* 125, 671–683. doi: 10.1007/s00122-012-1859-x

Helentjaris, T., Slocum, M., Wright, S., Schaefer, A., and Nienhuis, J. (1986). Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. *Theoret. Appl. Genet.* 72, 761–769. doi: 10.1007/bf00266542

Hu, J., Wang, Y., Fang, Y., Zeng, L., Xu, J., Yu, H., et al. (2015). A rare allele of GS2 enhances grain size and grain yield in rice. *Mol. Plant* 8, 1455–1465. doi: 10.1016/j.molp.2015.07.002

Kang, Y.-J., Shim, K.-C., Lee, H.-S., Jeon, Y.-A., Kim, S.-H., Kang, J.-W., et al. (2018). Fine mapping and candidate gene analysis of the quantitative trait locus gw8. 1 associated with grain length in rice. *Genes Genom.* 40, 389–397. doi: 10.1007/s13258-017-0640-6

Li, C., Li, Y., Sun, B., Peng, B., Liu, C., Liu, Z., et al. (2013). Quantitative trait loci mapping for yield components and kernel-related traits in multiple connected RIL populations in maize. *Euphytica* 193, 303–316. doi: 10.1007/s10681-013-0901-7

Li, H., Yang, Q., Fan, N., Zhang, M., Zhai, H., Ni, Z., et al. (2017). Quantitative trait locus analysis of heterosis for plant height and ear height in an elite maize hybrid zhengdan 958 by design III. *BMC Genet* 18:36. doi: 10.1186/s12863-017-0503-9

Li, J., Thomson, M., and McCouch, S. R. (2004). Fine mapping of a grain-weight quantitative trait locus in the pericentromeric region of rice chromosome 3. *Genetics* 168, 2187–2195. doi: 10.1534/genetics.104.034165

Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., et al. (2011). Natural variation in GS5 plays an important role in regulating grain size and yield in rice. *Nat. Genet.* 43:1266. doi: 10.1038/ng.977

Li, Y., Wang, Y., Shi, Y., Song, Y., Wang, T., and Li, Y. (2009). Correlation analysis and QTL mapping for traits of kernel structure and yield components in maize. *Sci. Agricult. Sin.* 42, 408–418.

Lin, H., Liang, Z.-W., Sasaki, T., and Yano, M. (2003). Fine mapping and characterization of quantitative trait loci Hd4 and Hd5 controlling heading date in rice. *Breed. Sci.* 53, 51–59. doi: 10.1270/jsbbs.53.51

Liu, M., Tan, X., Yang, Y., Liu, P., Zhang, X., Zhang, Y., et al. (2020). Analysis of the genetic architecture of maize kernel size traits by combined linkage and association mapping. *Plant Biotechnol. J.* 18, 207–221. doi: 10.1111/pbi.13188

Liu, Y., Wang, L., Sun, C., Zhang, Z., Zheng, Y., and Qiu, F. (2014). Genetic analysis and major QTL detection for maize kernel size and weight in multi-environments. *Theoret. Appl. Genet.* 127, 1019–1037. doi: 10.1007/s00122-014-2276-0

Lynch, M., and Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Abington: Elsevier.

Nie, N., Ding, X., Chen, L., Wu, X., An, Y., Li, C., et al. (2019). Characterization and fine mapping of qkrnw4, a major QTL controlling kernel row number in maize. TAG. *Theor. Appl. Genet.* 132, 3321–3331. doi: 10.1007/s00122-019-03427-9

Nzuve, F., Githiri, S., Mukunya, D., and Gethi, J. (2014). Genetic variability and correlation studies of grain yield and related agronomic traits in maize. *J. Agricult. Sci.* 6, 166–176.

Qiu, L., Guo, Y., Li, Y., Wang, X., Zhou, G., Liu, Z., et al. (2011). Novel gene discovery of crops in China: status, challenging, and perspective. *Acta Agronom. Sin.* 37, 1–17. doi: 10.3724/sp.j.1006.2011.00001

Qiu, X., Gong, R., Tan, Y., and Yu, S. (2012). Mapping and characterization of the major quantitative trait locus qSS7 associated with increased length and decreased width of rice seeds. *Theoret. Appl. Genet.* 125, 1717–1726. doi: 10.1007/s00122-012-1948-x

Rafiq, C. M., Rafique, M., Hussain, A., and Altaf, M. (2010). Studies on heritability, correlation and path analysis in maize (Zea mays L.). *J. Agricult. Res.* 48, 35–38.

Ramya, P., Chaubal, A., Kulkarni, K., Gupta, L., Kadoo, N., Dhaliwal, H., et al. (2010). QTL mapping of 1000-kernel weight, kernel length, and kernel width in bread wheat (Triticum aestivum L.). *J. Appl. Genet.* 51, 421–429. doi: 10.1007/bf03208872

Song, X.-J., Huang, W., Shi, M., Zhu, M.-Z., and Lin, H.-X. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* 39, 623. doi: 10.1038/ng2014

Su, C., Wang, W., Gong, S., Zuo, J., Li, S., and Xu, S. (2017). High density linkage map construction and mapping of yield trait QTLs in maize (Zea mays) using the genotyping-by-sequencing (GBS) technology. *Front. Plant Sci.* 8:706. doi: 10.3389/fpls.2017.00706

Sun, X.-Y., Wu, K., Zhao, Y., Kong, F.-M., Han, G.-Z., Jiang, H.-M., et al. (2009). QTL analysis of kernel shape and weight using recombinant inbred lines in wheat. *Euphytica* 165:615. doi: 10.1007/s10681-008-9794-2

Tian, F., Bradbury, P. J., Brown, P. J., Hung, H., Sun, Q., Flint-Garcia, S., et al. (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.* 43, 159–162. doi: 10.1038/ng.746

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31, 46–53. doi: 10.1038/nbt.2450

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Van Ooijen, J., and Voorrips, R. (2001). *Software for the calculation of genetic linkage maps*. Netherlands: Plant Research International Wageningen.

Verma, R., Oania, R., Graumann, J., and Deshaies, R. J. (2004). Multiubiquitin chain receptors define a layer of substrate selectivity in the ubiquitin-proteasome system. *Cell* 118, 99–110. doi: 10.1016/j.cell.2004.06.014

Voorrips, R. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Heredity* 93, 77–78. doi: 10.1093/jhered/93.1.77

Wan, X., Wan, J., Jiang, L., Wang, J., Zhai, H., Weng, J., et al. (2006). QTL analysis for rice grain length and fine mapping of an identified QTL with stable and major effects. *Theoret. Appl. Genet.* 112, 1258–1270. doi: 10.1007/s00122-006-0227-0

Wang, G., Zhao, Y., Zhu, Z., Zhang, X., Jiang, M., Cheng, M., et al. (2020). QTL analysis and fine mapping of a major QTL conferring kernel size in maize (*Zea mays*). ResearchSquare doi: 10.21203/rs.3.rs-20357/v1

Wang, S., Li, S., Liu, Q., Wu, K., Zhang, J., Wang, S., et al. (2015). The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat. Genet.* 47:949. doi: 10.1038/ng.3352

Wang, S., Wu, K., Yuan, Q., Liu, X., Liu, Z., Lin, X., et al. (2012). Control of grain size, shape and quality by OsSPL16 in rice. *Nat. Genet.* 44:950. doi: 10.1038/ng.2327

Xia, T., Li, N., Dumenil, J., Li, J., Kamenski, A., Bevan, M. W., et al. (2013). The ubiquitin receptor DA1 interacts with the E3 ubiquitin ligase DA2 to regulate

seed and organ size in Arabidopsis. *Plant Cell* 25, 3347–3359. doi: 10.1105/tpc.113.115063

Xie, X., Song, M.-H., Jin, F., Ahn, S.-N., Suh, J.-P., Hwang, H.-G., et al. (2006). Fine mapping of a grain weight quantitative trait locus on rice chromosome 8 using near-isogenic lines derived from a cross between Oryza sativa and Oryza rufipogon. *Theoret. Appl. Genet.* 113, 885–894. doi: 10.1007/s00122-006-0348-5

Xu, Y. (2010). *Molecular plant breeding*. United Kingdom: CABI.

Xu, J., Liu, L., Xu, Y., Chen, C., Rong, T., Ali, F., et al. (2013). Development and characterization of simple sequence repeat markers providing genome-wide coverage and high resolution in maize. *DNA Res.* 20, 497–509. doi: 10.1093/dnares/dst026

Xu, Y., Li, H.-N., Li, G.-J., Wang, X., Cheng, L.-G., and Zhang, Y.-M. (2011). Mapping quantitative trait loci for seed size traits in soybean (Glycine max L. Merr.). *Theoret. Appl. Genet.* 122, 581–594. doi: 10.1007/s00122-010-1471-x

Xue, S., Xu, F., Li, G., Zhou, Y., Lin, M., Gao, Z., et al. (2013). Fine mapping TaFLW1, a major QTL controlling flag leaf width in bread wheat (Triticum aestivum L.). *Theoret. Appl. Genet.* 126, 1941–1949. doi: 10.1007/s00122-013-2108-7

Yang, M., Chen, L., Wu, X., Gao, X., Li, C., Song, Y., et al. (2018). Characterization and fine mapping of qkc7.03: a major locus for kernel cracking in maize. TAG. *Theor. Appl. Genet.* 131, 437–448. doi: 10.1007/s00122-017-3012-3

Yu, F., Li, J., Huang, Y., Liu, L., Li, D., Chen, L., et al. (2014). FERONIA receptor kinase controls seed size in Arabidopsis thaliana. *Mol. Plant* 7, 920–922. doi: 10.1093/mp/ssu010

Zhang, X., Wang, J., Huang, J., Lan, H., Wang, C., Yin, C., et al. (2012). Rare allele of OsPPKL1 associated with grain length causes extra-large grain and a significant yield increase in rice. *Proc. Natl. Acad. Sci.* 109, 21534–21539. doi: 10.1073/pnas.1219776110

Zhang, Z., Liu, Z., Hu, Y., Li, W., Fu, Z., Ding, D., et al. (2014). QTL analysis of kernel-related traits in maize using an immortalized F2 population. *PLoS One* 9:e89645. doi: 10.1371/journal.pone.0089645

Zhao, Y., and Su, C. (2019). Mapping quantitative trait loci for yield-related traits and predicting candidate genes for grain weight in maize. *Sci. Rep.* 9, 1–10.

Zheng, Z., Ma, J., Stiller, J., Zhao, Q., Feng, Q., Choulet, F., et al. (2015). Fine mapping of a large-effect QTL conferring Fusarium crown rot resistance on the long arm of chromosome 3B in hexaploid wheat. *BMC Genom.* 16:850. doi: 10.1186/s12864-015-2105-0

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 12 January 2021
doi: 10.3389/fgene.2020.611414

Genome-Wide Investigation of the Phospholipase C Gene Family in *Zea mays*

*Jiantang Zhu†, Yuanyuan Zhou†, Jiale Li and Hui Li**

School of Biological Science and Technology, University of Jinan, Jinan, China

Phospholipase C (PLC) is one of the main hydrolytic enzymes in the metabolism of phosphoinositide and plays an important role in a variety of signal transduction processes responding to plant growth, development, and stress. Although the characteristics of many plant PLCs have been studied, *PLC* genes of maize have not been comprehensively identified. According to the study, five phosphatidylinositol-specific PLC (PI-PLC) and six non-specific PLC (NPC) genes were identified in maize. The PI-PLC and NPC genes of maize are conserved compared with homologous genes in other plants, especially in evolutionary relationship, protein sequences, conserved motifs, and gene structures. Transient expression of ZmPLC-GFP fusion protein in *Arabidopsis* protoplast cells showed that *ZmPLCs* are multi-localization. Analyses of transcription levels showed that *ZmPLCs* were significantly different under various different tissues and abiotic stresses. Association analysis shown that some *ZmPLCs* significantly associated with agronomic traits in 508 maize inbred lines. These results contribute to study the function of ZmPLCs and to provide good candidate targets for the yield and quality of superior maize cultivars.

Keywords: maize, PLC, genome-wide, expression pattern, stress, association analysis

OPEN ACCESS

Edited by:
Jian Ma,
Sichuan Agricultural University, China

Reviewed by:
Jun Ma,
China Agricultural University, China
Xiao-Wei Zhang,
Chinese Academy of Agricultural Sciences, China

***Correspondence:**
Hui Li
bio_lih@ujn.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *29 September 2020*
Accepted: *26 November 2020*
Published: *12 January 2021*

Citation:
Zhu J, Zhou Y, Li J and Li H (2021) Genome-Wide Investigation of the Phospholipase C Gene Family in Zea mays. Front. Genet. 11:611414. doi: 10.3389/fgene.2020.611414

INTRODUCTION

Phospholipids are important basic structural components of biological membranes and also as key signaling components responding to the plant development and various environmental stresses (Pokotylo et al., 2013). Phospholipids could be degraded into various products, such as diacylglycerol (DAG), phosphatidic acid (PA), free fatty acids (FFAs), and lysophospholipids (LPLs) by phospholipases, which include phospholipase C (PLC), phospholipase D (PLD), and phospholipase A (PLA) (Tuteja and Sopory, 2008; Hong et al., 2016). Among them, PLC is recognized as an important lipid hydrolase in animals and plants and has a profound effect on membrane lipid remodeling and intracellular signaling (Meldrum et al., 1991).

Based on different substrate affinities and cellular functions, plant PLCs have two different types: phosphatidylinositol-specific PLCs (PI-PLCs) and phosphatidylcholine-PLC (PC-PLC) (Kocourková et al., 2011; Pokotylo et al., 2014). PI-PLC hydrolyzes phosphoinositides to produce inositol 1,4,5-trisphosphate (IP_3) and DAG, which may function as the second messengers (Berridge, 1987; Meldrum et al., 1991). IP_3 could be quickly synthesized into hexakisphosphate (IP_6) and trigger Ca^{2+} influx, while DAG could be phosphorylated by DAG kinase (DGK) and transformed into PAs (Wang et al., 2006). Unlike PI-PLC, PC-PLC, also known as non-specific PLC (NPC), preferentially hydrolyzes the common membrane phospholipids, for example, PC, phosphatidylethanolamine (PE), and phosphatidylserine (PS) (Kocourková et al., 2011).

In plants, two types of PLCs were composed of many gene members. For example, *Arabidopsis* contains nine PI-PLCs and six NPCs, whereas there are four PI-PLCs and five NPCs in rice, respectively (Nakamura et al., 2005; Zheng et al., 2012; Singh et al., 2013). In general, a typical plant PI-PLC enzyme structurally contains two domains at least, catalytic PI-PLC-X domain and PI-PLC-Y domain, which are necessary for PI-PLC to function as phosphoesterase (Hicks et al., 2008). Furthermore, the X and Y domains could form together a distorted triose phosphate isomerase (TIM) barrel structure, which contains the active-site residues (Chen et al., 2011). In addition, PI-PLCs have a C-terminal Ca^{2+}/phospholipid binding C2 domain and an N-terminal EF hand domain involved in calcium binding (Chen et al., 2011). Some plant NPCs contain a putative signal peptide at the N-terminus, while all NPCs have a phosphoesterase domain, which is necessary for the function of esterase, such as NPCs and acid phosphatases (Wimalasekera et al., 2010). Generally, the phosphoesterase domain contains two highly conserved motifs, ENRSFDxxxG and TxPNR, and two other invariable motifs, DExxGxxDHV, and GxRVPxxxxxP (Pokotylo et al., 2013).

Members of the plant PLC family play important roles in various biological processes, for example, plant growth, development, and stress response. *AtPI-PLC2* is required for female gametogenesis and embryo development in *Arabidopsis*, and loss of *AtPI-PLC2* resulted in defective male and female gametophyte development (Li et al., 2015; Di Fino et al., 2017). Similarly, *AtNPC2* and *AtNPC6* are involved in gametophyte and embryo development and glycerolipid metabolism in the flower buds (Ngo et al., 2018), while *AtNPC3* and *AtNPC4* have an important regulatory role in root development (Wimalasekera et al., 2010). In addition, plant *PLCs* have also been confirmed to participate in a variety of tolerances to abiotic or biotic stresses. *Arabidopsis AtPI-PLC* members, except *AtPI-PLC2*, could be induced under various abiotic stresses such as salinity, drought, and cold (Tasma et al., 2008). Overexpression of *Brassica napus PI-PLC2* in canola induces significant changes in the expression of stress-related genes and enhances drought tolerance (Georges et al., 2009). Overexpression of *ZmPI-PLC1* enhanced the grain yield of maize under drought conditions, while suppression of *ZmPI-PLC1* had an opposite effect (Wang et al., 2008). Knockout of *AtNPC4* in *Arabidopsis* could increase sensitivity to salt stress in root elongation, seedling biomass, and seed germination, while *AtNPC5* expression could be significantly upregulated under salt stress and positively regulate the development of lateral root under salt stress (Kocourková et al., 2011).

In this research, the PLC-encoding genes including PI-PLC and NPC were identified in the maize genome. The *ZmPLCs* have been analyzed in detail, including phylogenetic relationships, gene structures, conserved motifs, and subcellular localization. The transcription levels of the *ZmPLCs* were determined by qRT-PCR in different tissues and various abiotic stresses. In addition, the *ZmPLC* genes with quantitative trait loci (QTLs) associated with agronomic traits. Furthermore, the results presented herein provide valuable clues for studying the functions of *ZmPLCs* in response to the growth, development, and stress responses of maize.

MATERIALS AND METHODS

Genome-Wide Identification of the *PLC* Genes in Maize

To identify putative PLC proteins, the hidden Markov models (HMMs) of the two characteristic domains of a PLC protein from PFam (http://pfam.sanger.ac.uk/), PI-PLC-X (PF00388), and PI-PLC-X (PF00387) were used as query sequences in local HMM-based searches, setting *E*-values < 0.01. In addition, to identify ZmPLC proteins that might have been missed through HMM searching, ZmPLC sequences were further identified using the previously reported *Arabidopsis* PLC protein sequences from the Maize Genome Database (https://www.maizegdb.org) and Phytozomev12.0 (http://www.phytozome.net) under the *E*-value cutoff 0.1. The matched sequences were subjected to SMART (http://smart.embl.de/) analyses to detect the presence and number of the PLC domain. Therefore, 11 independent *ZmPLC* genes were identified in maize. The chromosomal location image was mapped by MapInspect software. The ExPASy (https://web.expasy.org/protparam/) was performed to calculate the molecular weight (MW) and the theoretical isoelectric point (PI).

Phylogenetic Analysis and Synteny Analysis

The multi-species PLC sequences, including 11 ZmPLCs from *Zea mays*, 15 AtPLCs from *Arabidopsis thaliana*, nine OsPLCs from *Oryza sativa*, 22 GhPLCs from *Gossypium hirsutum*, eight BdPLCs from *Brachypodium distachyon*, 10 SbPLCs from *Sorghum bicolor*, and 19 GmPLCs from *Glycine max* (**Supplementary File 1**), were constructed into a phylogenetic tree using the neighbor-joining (NJ) method in MEGA7.1 (Tamura et al., 2011). *ZmPLCs* were named basing on the phylogenetic relationship with *AtPLCs* and *OsPLCs*. Syntenic gene pairs among *Zea mays* and between *Arabidopsis thaliana* and *Oryza sativa* were identified using the TBtools (Chen et al., 2020).

Analyses of Gene Structures, *cis*-Acting Elements, and Motifs

The exon/intron structures of *PLCs* were analyzed by Gene Structure Display Server (GSDS) (http://gsds.cbi.pku.edu.cn/). The genomic sequences 2,000 bp upstream of *ZmPLCs* predicted the *cis*-acting elements using PlantCARE software (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/?tdsourcetag=s_pcqq_aiomsg). The conserved motifs were analyzed with MEME (http://meme.sdsc.edu/meme4_3_0/intro.html).

Subcellular Localization of ZmPLCs

In order to confirm the subcellular localization, the coding sequences of selected *ZmPLC* genes were amplified using gene-specific primers. Then, ZmPLCs minus the stop codons were cloned and inserted into pBI221:eGFP, and the corresponding expression vectors were introduced into *Arabidopsis* protoplasts. The green fluorescent protein (GFP) fluorescence was excited with a confocal laser scanning microscope LSM 800 (Zeiss).

Plant Growth Conditions and Treatments

The maize inbred line W22 (from Huazhong Agricultural University) was used for all the experimental treatments. The seeds were sterilized with 70% ethanol for 5 min and then washed three times with sterile water. The seedlings were grown in a Hoagland solution in a greenhouse with a regime of 16 h light/8 h dark and at 28°C. When the maize seedlings were raised to the three-leaf stage, seedlings were selected for stress treatments according to Lin et al. (2014), including 20% PEG 6000 for drought stress, 200 mM NaCl for salt stress, 4°C for cold stress, and 20 μmol/L Cu^{2+} for heavy metal stress. The leaves were collected at different points in time after treatment. Each treatment consisted of three replicates. Adult plants were grown in the field, and then all organs and tissues were harvested. The root was harvested at the three-leaf stage; the stem and leaf were collected at five fully extended leaves; the silk, cob, and anther were harvested at 13 extended leaves; the kernel was harvested at 10 days after pollination (DAP). All materials were immediately frozen in liquid nitrogen after harvesting and stored at −80°C prior to RNA isolation.

Expression Analysis and Quantitative Real-Time PCR

For quantitative real-time PCR, total RNAs were extracted from various maize tissues with the RNAprep Pure Plant Kit (Tiangen, Beijing, China). According to supplier instructions, total cDNA was synthesized using PrimeScript™ RT Reagent Kit with gDNA Eraser (Tiangen, Beijing, China). The Primer Premier 5.0 was used to design the primers for qRT-PCR (**Supplementary Table 1**), and the maize TUB-ribosylation factor was selected as an internal control. The reaction was performed on Bio-Rad CFX Connect™ using SYBR-Green to detect gene expression levels. For all qRT-PCR analyses, triplicate biological samples were collected. Data were analyzed using Bio-Rad CFX Manager software.

Candidate Gene-Based Association Mapping of *ZmPLC* Family Members in Maize

Regional association tests between the single-nucleotide polymorphisms (SNPs) of candidate genes and the 17 traits agronomic, including plant height, ear height, tassel branch number, ear diameter, 100 grain weight, silking time, heading date, leaf number above ear, tassel main axis length, ear length, kernel width, cob weight, pollen shed, kernel weight, kernel number per row, ear leaf width, ear leaf length, were conducted in the 508 inbred lines (Fu et al., 2013). The genotype and phenotypes of the association panel were detected by two genotyping platforms, resulting in 550,000 high-quality SNPs (Yang et al., 2011; Li et al., 2013), and only SNPs within the range of 100 kb upstream and downstream of candidate genes were used. The association analysis was estimated using a mixed linear model (MLM) incorporated in TASSEL V5.0 (Bradbury et al., 2007). $P \leq 0.05$ was considered the significance threshold.

TABLE 1 | General information of *PLC* genes in maize.

Gene Name	Gene ID (AGPv4)	Gene ID (AGPv3)	Chromosome Location	ORF (bp)	Exon No.	Protein (aa)	PI	MW (KD)	Type	Sub-Localization Prediction
ZmPI-PLC1	Zm00001d007229	GRMZM2G129238	chr2:225211077-225213907 (+)	1,764	7	587	6.46	66.97	PI-PLC	Chloroplast
ZmPI-PLC2	Zm00001d028746	GRMZM2G114354	chr1:45171565-45179033(+)	1,770	9	589	5.98	66.52	PI-PLC	Cytoplasm
ZmPI-PLC3a	Zm00001d014903	GRMZM5G889467	chr5:67438945-67442847(+)	1,761	9	586	6.40	65.68	PI-PLC	Chloroplast
ZmPI-PLC3b	Zm00001d014906	GRMZM2G137435	chr5:67681040-67684923(+)	1,761	9	586	6.40	65.85	PI-PLC	Chloroplast
ZmPI-PLC4	Zm00001d047447	GRMZM2G157760	chr9:130364956-130370673(+)	1,821	6	606	5.94	68.23	PI-PLC	Mitochondrion
ZmNPC1a	Zm00001d034714	GRMZM2G139041	chr1:300523648-300526448(+)	1,629	3	542	7.37	60.60	NPC	Chloroplast
ZmNPC1b	Zm00001d012924	GRMZM2G116876	chr5:2004892-2007704(+)	780	3	259	6.07	28.96	NPC	Nucleus
ZmNPC2	Zm00001d042057	GRMZM2G479112	chr3:148970593-148972670(+)	1,548	2	515	6.25	57.49	NPC	Mitochondrion
ZmNPC3	Zm00001d034875	GRMZM2G422670	chr1:304506320-304508029(+)	1,626	2	541	5.82	59.02	NPC	Cytoplasm
ZmNPC4	Zm00001d032907	–	chr1:242704706-242711363(+)	1,470	2	489	5.74	53.14	NPC	Cytoplasm
ZmNPC5	Zm00001d040205	GRMZM2G081719	chr3:32011584-32013797(+)	1,593	2	530	9.09	58.41	NPC	Chloroplast

ID, identification; ORF, open reading frame; PI, isoelectric point; MW, molecular weight; aa, amino acid.

RESULTS

Identification and Characterization of *PLCs* in Maize

Eleven ZmPLCs, including five PI-PLC and six NPC sequences, were identified in the maize genome according to their domain structures (**Supplementary Figure 1A**). The ZmPI-PLC group contains five members: ZmPI-PLC1, ZmPI-PLC2, ZmPI-PLC3a, ZmPI-PLC3b, and ZmPI-PLC4. The ZmNPC group has six members: ZmNPC1a, ZmNPC1b, ZmNPC2, ZmNPC3, ZmNPC4, and ZmNPC5. Domain analysis showed that the five ZmPI-PLCs contained the catalytic PI-PLC-X, PI-PLC-Y domain, and Ca^{2+}/phospholipid binding C2 domain, whereas an EF hand-like motif was found only in ZmPI-PLC4 (**Supplementary Figure 1A**). Six ZmNPCs had a phosphoesterase domain, which contains two highly conserved motifs, ENRSFDxxxG and TxPNR, and two invariable motifs, DExxGxxDHV and GxRVPxxxxxP (**Supplementary Figure 1B**). The gene identification (ID), gene name, open reading frame size, exon number, length, molecular weight, and isoelectric point of *ZmPLCs* were shown in **Table 1**. Especially, the number of amino acids (aa) in ZmPLCs of maize is comparable, with ZmPI-PLCs ranging from 586 to 606 aa and ZmNPCs ranging from 489 to 542 aa (ZmNPC1b is the only exception with 259 aa) (**Table 1**).

Phylogenetic Relationship, Gene Structure, and Conserved Motifs of *ZmPLCs* in Maize

To investigate the phylogenetic relationship of *PLC* among different species, a phylogenetic tree consisting of 11 ZmPLCs, 15 AtPLCs, nine OsPLCs, 19 GmPLCs, eight BdPLCs, 10 SbPLCs, and 22 GhPLCs was constructed using the NJ method (**Figure 1**). A total of 94 PLC protein sequences were classified into two subfamilies, named PI-PLC and NPC, based on the differences in the domain and phylogenetic relationships (**Figure 1**). The phylogenetic analysis showed that ZmPLCs shared high homology with those from other plants (**Figure 1**), especially ZmPI-PLC1, ZmPI-PLC2, ZmPI-PLC3a, and ZmPI-PLC3b made a separate small clade, while ZmPLC4 fell apart (**Figure 1**). While ZmNPC1a and ZmNPC1b; ZmNPC2; ZmNPC5; ZmNPC3 and ZmNPC4 made a separate small clade (**Figure 1**).

Corresponding to the evolutionary relationship, analysis of the exon/intron structures of the *PLC* genes revealed that these genes are also divided into two different types: PI-PLC with an exon-rich clade ($\geq$5 exons per gene) generally containing 5–10 exons, while NPC with an exon-poor clade ($\leq$5 exons per gene) containing 1–5 exons (**Figure 2**). It is worth noting that a similar exon/intron pattern exists in each clade, for instance, most NPC1 genes contain three exons, and the vast majority of PI-PLC1 genes have seven exons (**Figure 2**).

To further study the characteristic regions of the PLC proteins, a total of 10 conserved motifs were identified in PLCs using the online MEME tool (**Figure 2**). Interestingly, the members of the same clade usually had similar structures and lengths in terms of domain. The group PI-PLC members had the conserved motifs 1, 2, 4, 5, 7, and 9, which are PI-PLC-X, PI-PLC-Y, and Ca^{2+}/phospholipid binding C2 domain. While the conserved motifs 3, 6, 8, 9, and 10, which were annotated as the phosphoesterase domain, were found in group NPC family (**Figure 2**).

Chromosomal Location of the *ZmPLCs* and Synteny Analysis of *PLCs* Among Several Different Species

According to the data of the gene locus, 11 *ZmPLC* genes were found in five different chromosomes (**Table 1**, **Figure 3A**). *ZmPI-PLC2*, *ZmNPC1a*, *ZmNPC3*, and *ZmNPC4* are localized on chromosome 1, *ZmPI-PLC1* and *ZmPI-PLC4* are distributed on chromosomes 2 or 9, and *ZmNPC2* and *ZmNPC5* are located on chromosomes 3, while chromosomes 5 contained *ZmNPC1b*, *ZmPI-PLC3a*, and *ZmPI-PLC3b* (**Figure 3A**).

To understand more about the phylogeny of *ZmPLC* genes family, syntenic analysis was performed between maize and two other plant species, including *Arabidopsis thaliana* and *Oryza sativa* (**Figure 3B**). There were three *ZmPLC* genes that were synchronized with those in *Arabidopsis thaliana* (**Figure 3B**, **Supplementary Table 2**). The comparative syntenic maps of maize associated with rice were analyzed further, and six out of 11 *ZmPLC* genes had collinear genes in rice (**Figure 3B**, **Supplementary Table 2**), indicating that these genes may be derived from a common ancestor.

Analysis of *cis*-Elements in the Promoters of *ZmPLCs*

To study the expression regulation patterns and the potential function of *ZmPLCs*, putative *cis*-elements on promoter regions were searched in the Plant CARE database. *Cis*-elements related to developmental processes, such as the meristem expression (CAT-box), were found in the promoter regions of *ZmPI-PLC4*, *ZmNPC1a*, *ZmNPC1b*, *ZmNPC2*, *ZmNPC3*, and *ZmNPC5*, suggesting that *ZmPLC* genes play important roles in differentiation (**Figure 4**). The hormone-responsive elements, including abscisic acid (ABA) responsive elements (ABREs), MeJA-responsive elements (TGACG-motif and CGTCA-motif), salicylic acid-responsive element (TCA-element and SARE), auxin-responsive element (TGA-element and AuxRR-core), and gibberellin-responsive element (TATC-box, P-box, and GARE-motif), were also found in some *ZmPLC* gene promoters (**Figure 4**), showing that *ZmPLC* genes may play a role in growth and development. Furthermore, stress-responsive elements, such as low-temperature responsive (LTR) element, drought-responsive element, anaerobic induction responsive element (ARE), and defense and stress responsive element (AT-rich and TC-rich repeats), were observed in some promoters of *ZmPLC* genes. For example, LTR elements were detected in *ZmPI-PLC1/2/3a/3b* and *ZmNPC3/4/5* gene promoters (**Figure 4**), while drought-responsive elements were found in the promoters of *ZmPI-PLC1/2* and *ZmNPC1a/1b/5* (**Figure 4**).

Expression Profiles of *ZmPLCs* in Different Tissues and Developmental Stages of Maize

To study the roles of *ZmPLCs* in growth and development of maize, qRT-PCR was used to determine the expression profiles

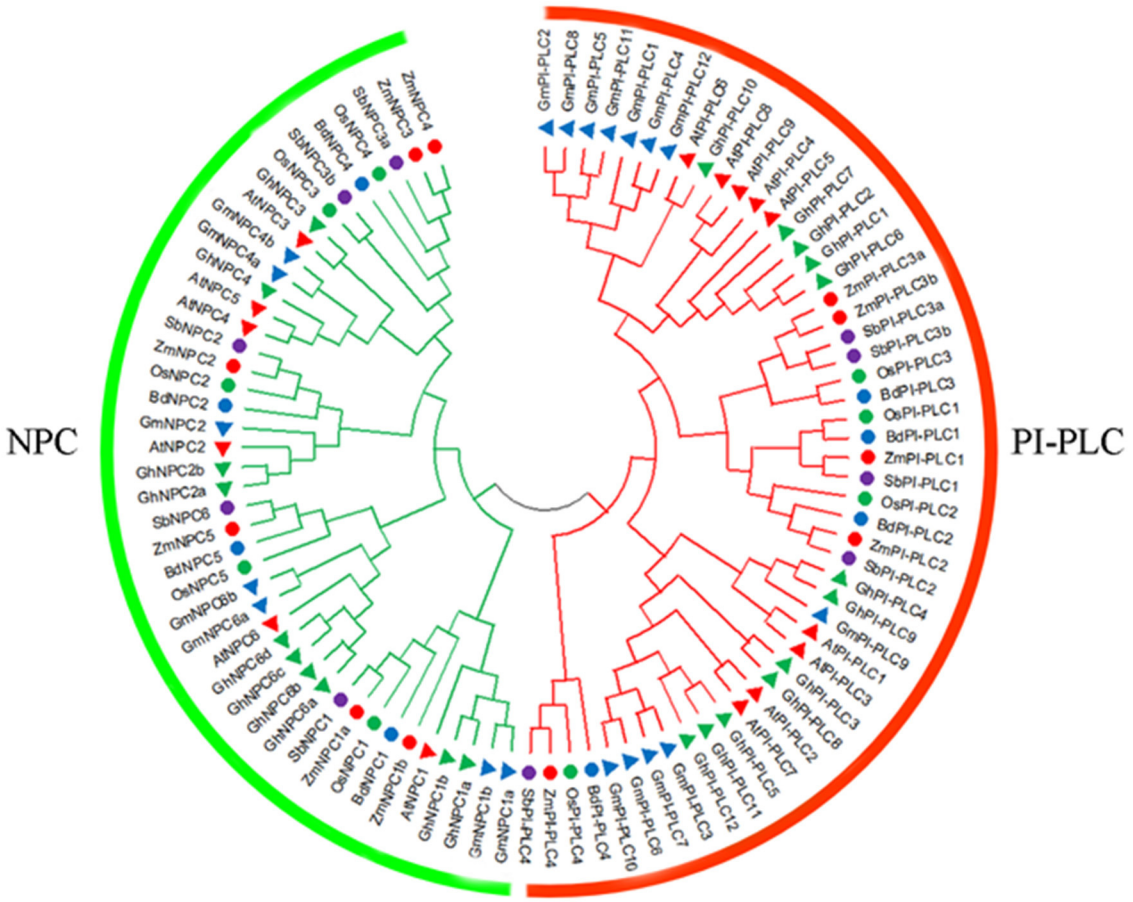

FIGURE 1 | Phylogenetic analysis of the *PLC* family. The phylogenetic tree was made by MEGA 7.1 software using the neighbor-joining (NJ) method with bootstrapping analysis. The analysis involved 94 amino acid sequences from various plants, including *Arabidopsis thaliana* (At), *Glycine max* (Gm), *Oryza sativa* (Os), *Sorghum bicolor* (Sb), *Gossypium hirsutum* (Gh), *Brachypodium distachyon* (Bd).

of the 11 *ZmPLC* mRNAs in different tissues, including root, stem, leaf, silk, immature cob, anther, tassel, kernel (10 DAP) (**Figure 5**). The results showed that *ZmPLCs* were significantly different under various different tissues. For example, *ZmPI-PLC1* was more highly expressed in silk; *ZmPI-PLC2* and *ZmNPC1a* were more highly expressed in the leaf than in other tissues; *ZmPI-PLC3a* and *ZmPI-PLC3b* had similar expression patterns with relatively lower expressions in anther, while *ZmNPC2*, *ZmNPC3*, and *ZmNPC4* had relatively high expressions in anther (**Figure 5**). Except for the relatively high expression in root, *ZmPI-PLC4* and *ZmNPC1b* showed low expression in others organs (**Figure 5**), and *ZmNPC5* was relatively highly expressed in 10 DAP of kernel (**Figure 5**). These results indicated that *ZmPLCs* may have important functions in different developmental stages and tissues of maize.

Expression Profiling of *ZmPLCs* in Response to Various Stresses

To verify the expression changes of *ZmPLCs* under stresses, the expression profiles were examined by qRT-PCR under different abiotic stress treatments, including 20% PEG6000, 200 mM NaCl, 4°C, and 20 μmol/L Cu^{2+} (**Figure 6**). The results show that *ZmPLC* genes, except *ZmNPC4* and *ZmNPC5*, were induced by 20% PEG 6000, peaking at 6 or 12 h after treatments, and then reduced to the lowest level at 48 h (**Figure 6A**). *ZmPI-PLC1*, *ZmPI-PLC3a*, *ZmNPC1a*, *ZmNPC1b*, *ZmNPC4*, and *ZmNPC5* were rapidly induced by high salinity, and *ZmPI-PLC3b*, *ZmPI-PLC4*, and *ZmNPC3* mRNAs were almost unaffected; however, the expressions of *ZmPI-PLC2* and *ZmNPC2* were suppressed (**Figure 6B**). The expressions of *ZmPI-PLC1*, *ZmPI-PLC2*, *ZmNPC1b*, *ZmNPC2*, *ZmNPC3*, *ZmNPC4*, and *ZmNPC5* were decreased after cold treatment and then reached the lower level at 12 h, while *ZmPI-PLC3a*, *ZmPI-PLC3b*, and *ZmNPC1a* mRNAs were induced with cold treatment, then still relatively high level at 12 h (**Figure 6C**). Under 20-μmol/L Cu^{2+} treatments, *ZmPI-PLC3a*, *ZmPI-PLC3b*, *ZmPI-PLC4*, *ZmNPC2*, and *ZmNPC3* mRNAs were suppressed within 48 h, while *ZmPI-PLC1*, *ZmPI-PLC2*, *ZmNPC1a*, *ZmNPC1b*, *ZmNPC2*, and *ZmNPC5* were induced within 48 h and peaked at 3 or 6 h after treatment (**Figure 6D**). These results show that

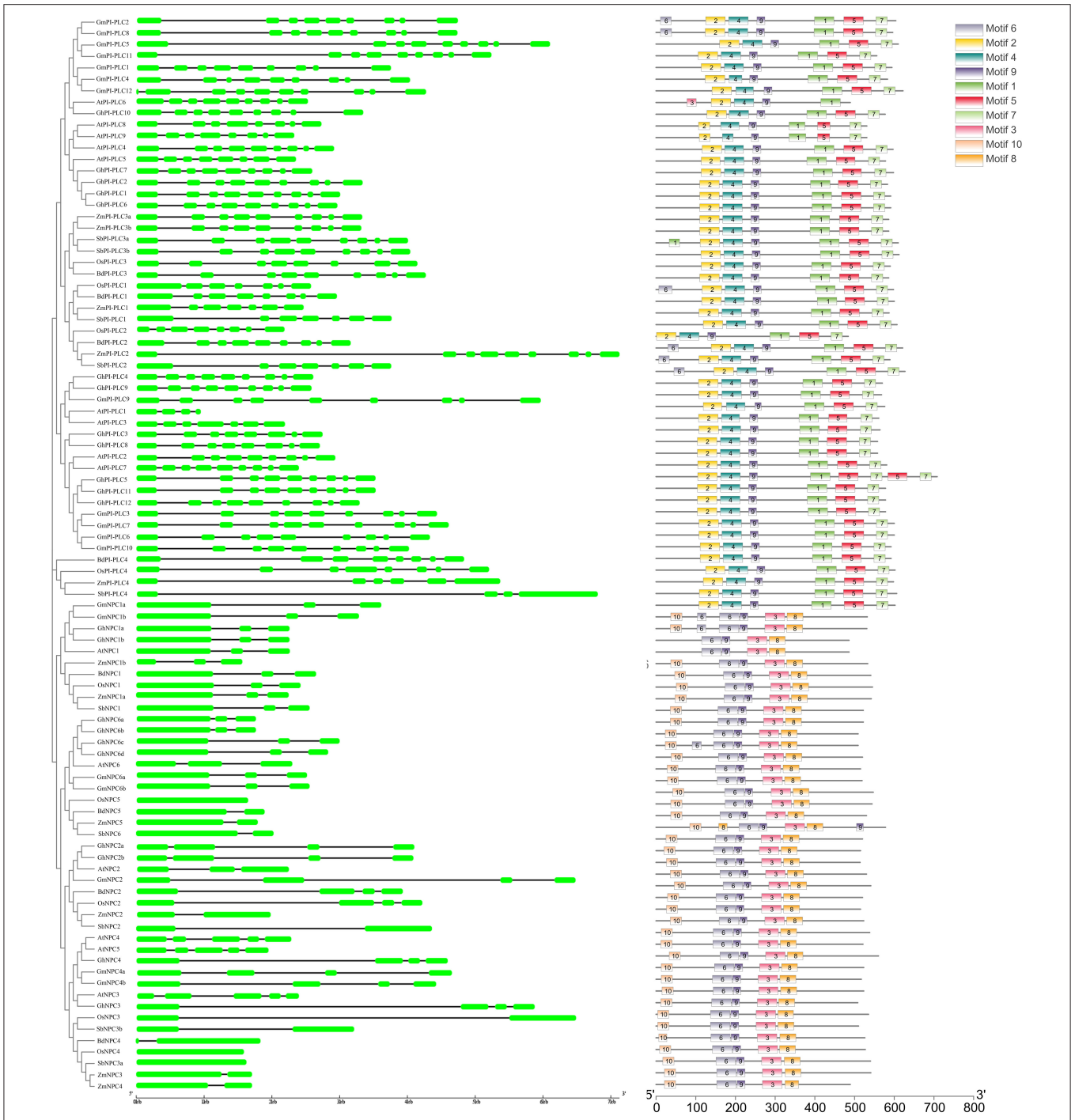

FIGURE 2 | Phylogenetic relationships, gene structures, and conserved protein motifs of maize *PLC* family. Green boxes indicate exons, and black lines indicate introns. The motifs, numbered 1–10, are displayed in different colored boxes. The length of the protein can be estimated using the scale at the bottom.

ZmPLCs genes might play important roles in response to various abiotic stresses.

Subcellular Localization of ZmPLC Proteins

Previously, it was reported that the substrates of PLCs, PI4P and PI(4,5)P_2, are located in the plasma membrane (Munnik et al., 1998), so it is speculated that the subcellular localization of PLCs should also be in the plasma membrane; however, the sub-localization prediction shows that ZmPLCs may be multi-localized (**Table 1**). In order to confirm the subcellular localization, two randomly selected ZmPLCs, ZmPI-PLC2 and ZmNPC3, were transiently expressed in *A. thaliana* mesophyll protoplasts to analyze subcellular

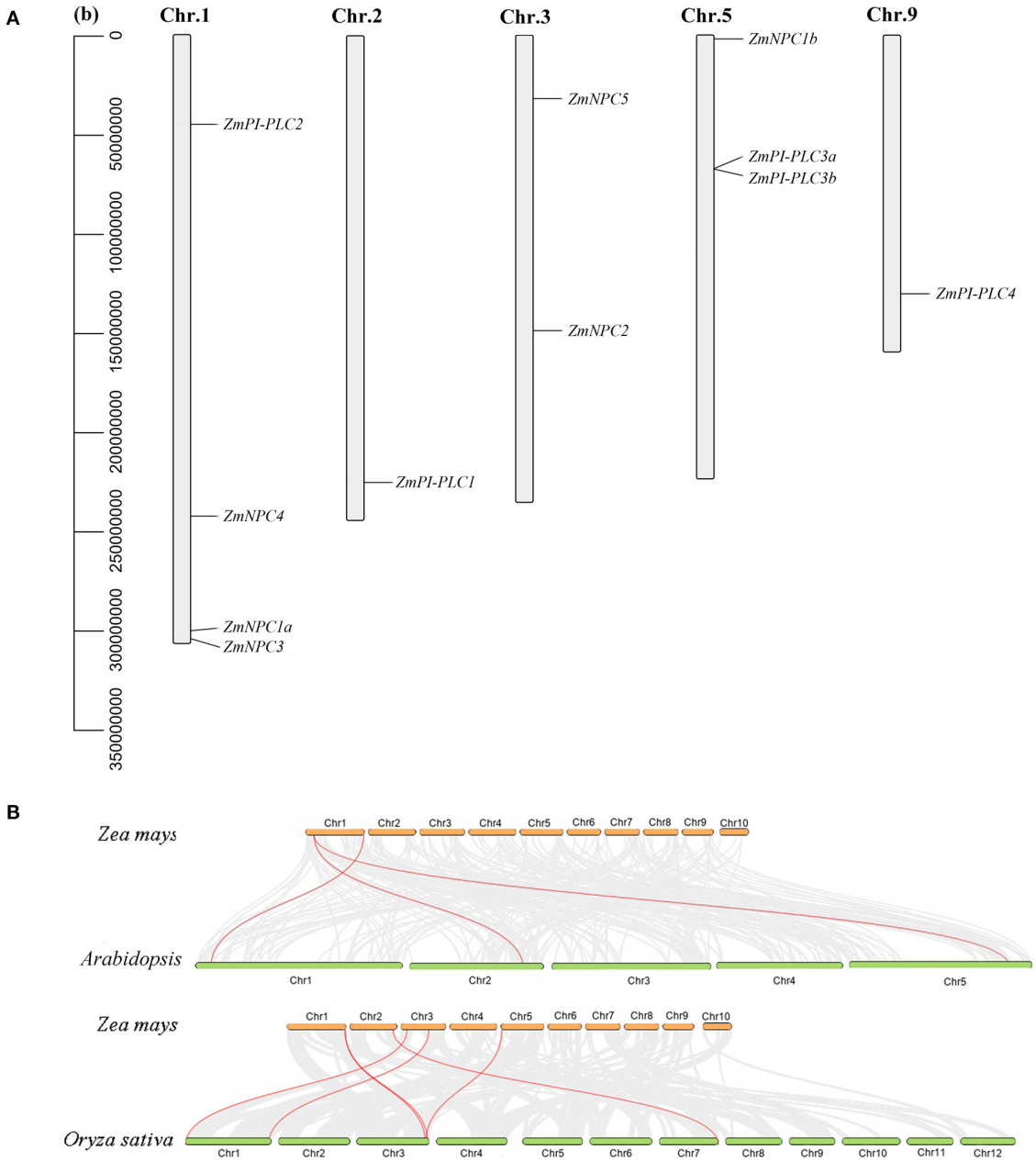

FIGURE 3 | The chromosomal distribution of *ZmPLC* genes **(A)** and synteny analysis between *ZmPLCs* and *PLCs* from *Arabidopsis* or rice **(B)**. The number of the chromosome is shown on each chromosome. Gray lines in the background indicate the collinear blocks within maize and other plant genomes, while the red lines highlight the syntenic *PLC* gene pairs.

localization by fusing to the N-terminus of GFP. In *Arabidopsis* leaf protoplasts, the fluorescence of the vector control was distributed throughout the nucleus and cytoplasm (**Figure 7**); however, the signals of ZmPI-PLC2 and ZmNPC3 were only observed in the cytoplasm (**Figure 7**), showing that ZmPI-PLC2 and ZmNPC3 are located in the cytosol.

Regional Association Mapping of *ZmPLCs* for Agronomic Traits in Maize

To understand the possible functions of the *ZmPLCs*, an analysis of agronomy -related trait QTLs from 508 maize inbred lines was performed. There were significant correlations between all 11 *ZmPLC* genes and more than one agronomic trait at the $P \leq 0.05$ level, such as kerner number per row,

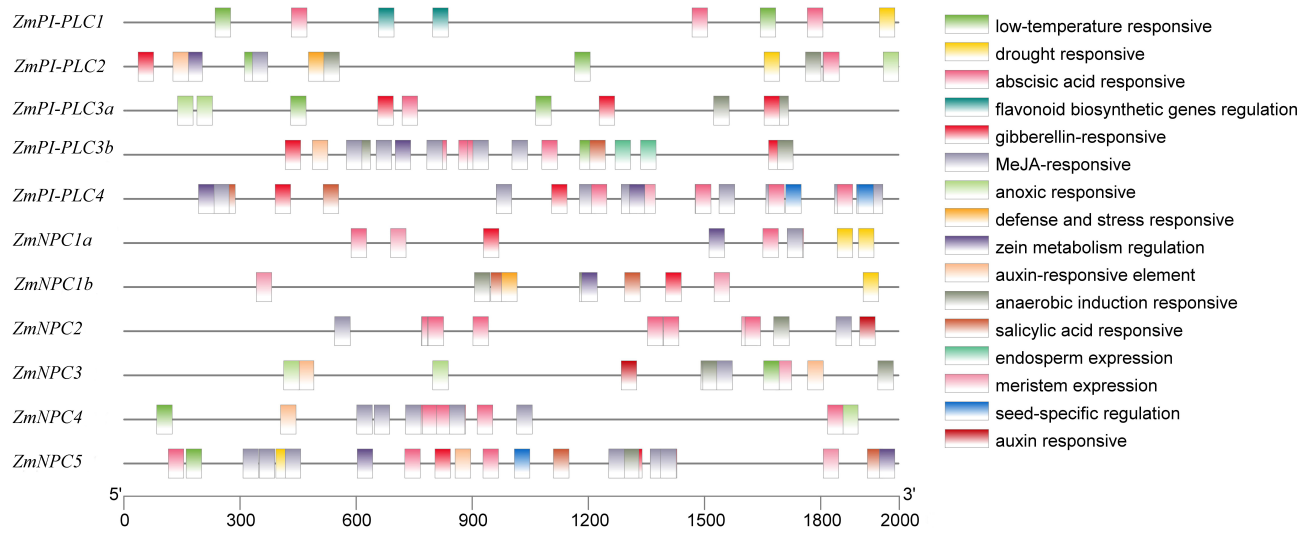

FIGURE 4 | Putative regulatory *cis*-elements in *ZmPLC* family gene promoters. Different *cis*-elements are represented with different color box.

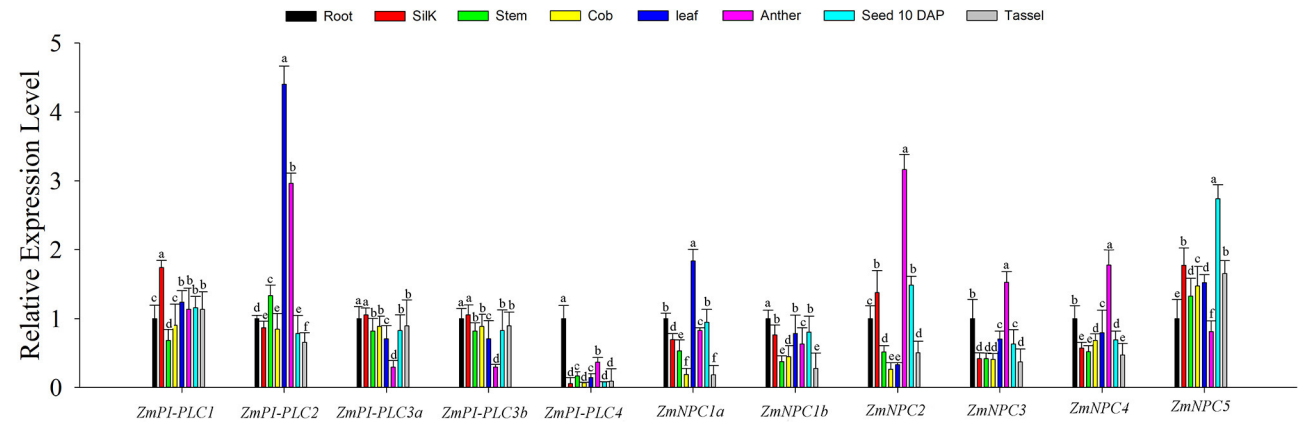

FIGURE 5 | Tissue-specific expression patterns of the 11 *ZmPLC* genes in maize root, stem, leaf, silk, immature cob, anther, and kernel [10 days after pollination (DAP)] by using qRT-PCR. The expression levels were normalized to that of TUB. Bars represent SE ($n = 3$). The entire experiment was repeated three times. Different letters indicate values that departed significantly from those of root.

100 grain weight, cob weight, and so on (**Figure 8**). At $P \leq 0.01$, 10 *ZmPLC* genes were identified to be related to some important agronomic traits (**Supplementary Table 3**). For example, *ZmPI-PLC1* was significantly correlated with ear height, ear leaf length, ear length, kerner number per row, plant height, pollen shed, and tassel branch number (**Supplementary Table 3**, **Figure 8A**); *ZmPI-PLC4* significantly affected heading date, pollen shed, 100 grain weight, silking time, ear leaf width, cob weight, kerner number per row, and so on (**Supplementary Table 3**, **Figure 8E**); *ZmNPC1a* significantly affected ear leaf length, kerner number per row, pollen shed, and plant height (**Supplementary Table 3**, **Figure 8F**).

DISCUSSION

ZmPLC Gene Identification and Evolutionary Analysis in Maize

At present, *PLC* genes and their functions in many plants have been studied and reported, such as *Arabidopsis*, rice, and soybean (Wang et al., 2015; Zhang et al., 2018). It was reported that nine, 15, and 19 *PLC* genes were found in rice (Singh et al., 2013), *Arabidopsis* (Zheng et al., 2012), and cotton (Zhang et al., 2018), respectively. In this study, 11 ZmPLCs, five PI-PLCs, and six NPC members were identified according to the domain structure and phylogenetic analysis (**Supplementary Figure 1**, **Figure 2**). Based on the phylogenetic analysis, although PLCs

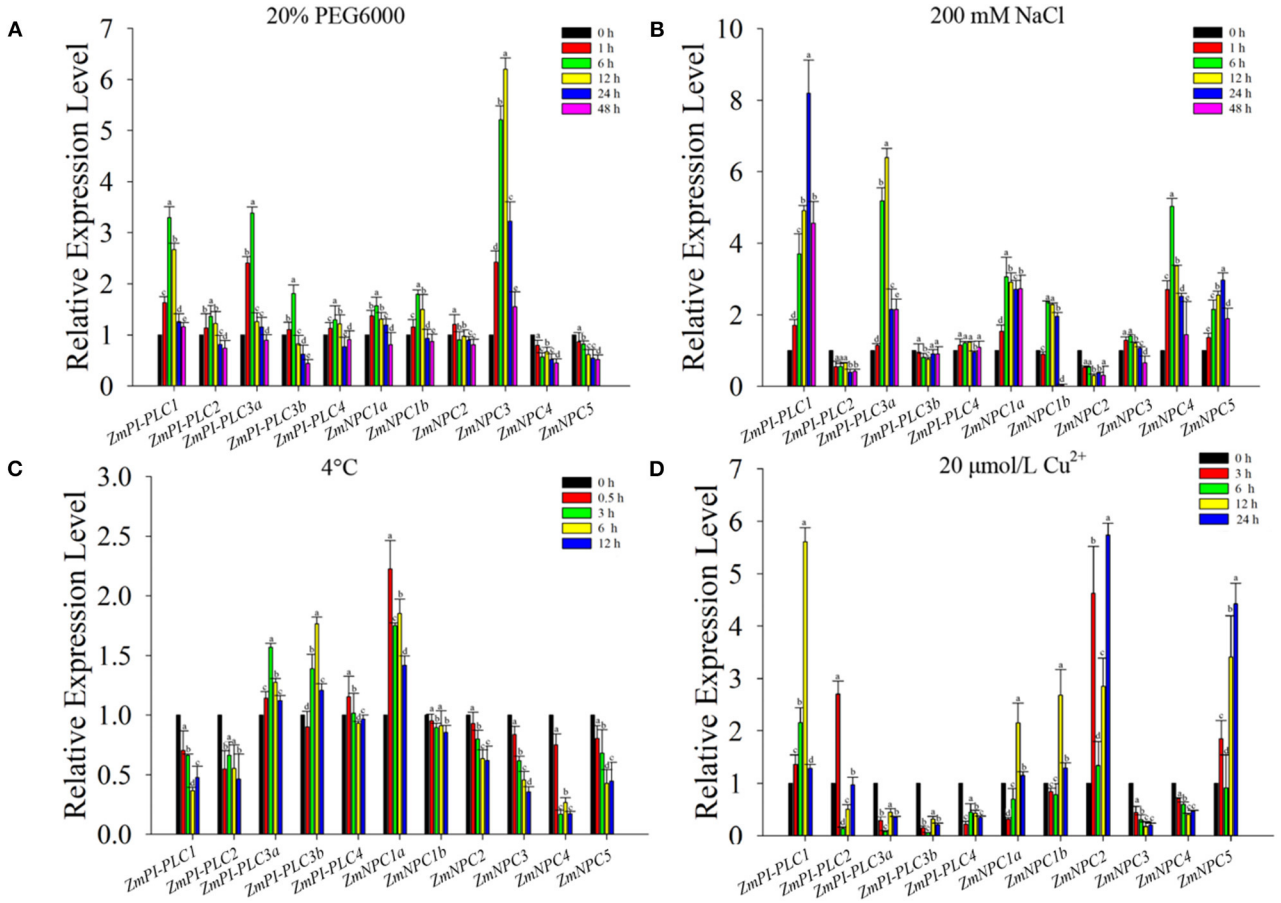

FIGURE 6 | Quantitative real-time PCR analysis of *ZmPLCs* expression under stress treatments: **(A)** 20% PEG 6000, **(B)** 200 mM NaCl, **(C)** 4°C, **(D)** 20 μmol/L Cu^{2+}. The expression levels were normalized to that of TUB. Bars represent SE ($n = 3$). The entire experiment was repeated three times. Different letters indicate values that departed significantly from those of without stress.

from closely related species made a phylogenetic clade due to different genetic relationships (**Figure 1**), maize PI-PLCs made a phylogenetic clade with dicots PI-PLCs with very low bootstrap; however, maize NPCs fall in the same clade with dicot NPCs with a relative high bootstrap (**Figure 1**), suggesting that PLCs from different species diversify during the course of evolution even after being originated from a common ancestor. In addition, there are generally more PLCs in dicots than monocots (Zheng et al., 2012; Zhang et al., 2018), especially PI-PLC, which indicates that the natural selection of the *PLC* genes is a variant for different plant species. Domain and motif analysis of the PI-PLC protein sequences in maize with homologs from other plants contained the conserved and typical PI-PLC-X and PI-PLC-Y catalytic domains and C2 domain (**Supplementary Figure 1B**, **Figure 2**). It is worth noting that an EF hand-like motif was only found in ZmPI-PLC4 (**Supplementary Figure 1A**). Similarly, all the NPC members harbored a highly conserved phosphoesterase domain (**Supplementary Figure 1A**). The gene structure of *PLCs* from different species could be distinguished into two types: exon-rich (PI-PLC) and exon-poor (NPC) (**Figure 2**). Genes in the same branch usually have similar exon–intron structures, whereas the gene structures differed markedly among the different branches (**Figure 2**). Furthermore, there are similar results in the motif compositions of the PLC protein (**Figure 2**). Synteny analysis showed that the number of *PLC* homologous genes between maize and rice were more than those between maize and *Arabidopsis*, which indicating that PLC has a higher homology among related species (**Figure 3**). These findings indicated that highly conserved PLC from different plants might have a similar function in the evolution of plants.

Functional Analysis of the *ZmPLCs* Response to the Development and Abiotic Stress in Maize

The *cis*-regulatory elements in the promoter play an important role in regulating the expression patterns of the genes (Vedel and Scotti, 2011). According to analysis of the *cis*-acting elements in *ZmPLC* promoter region, the *ZmPLC* promoters contain various element responses to plant development and stresses,

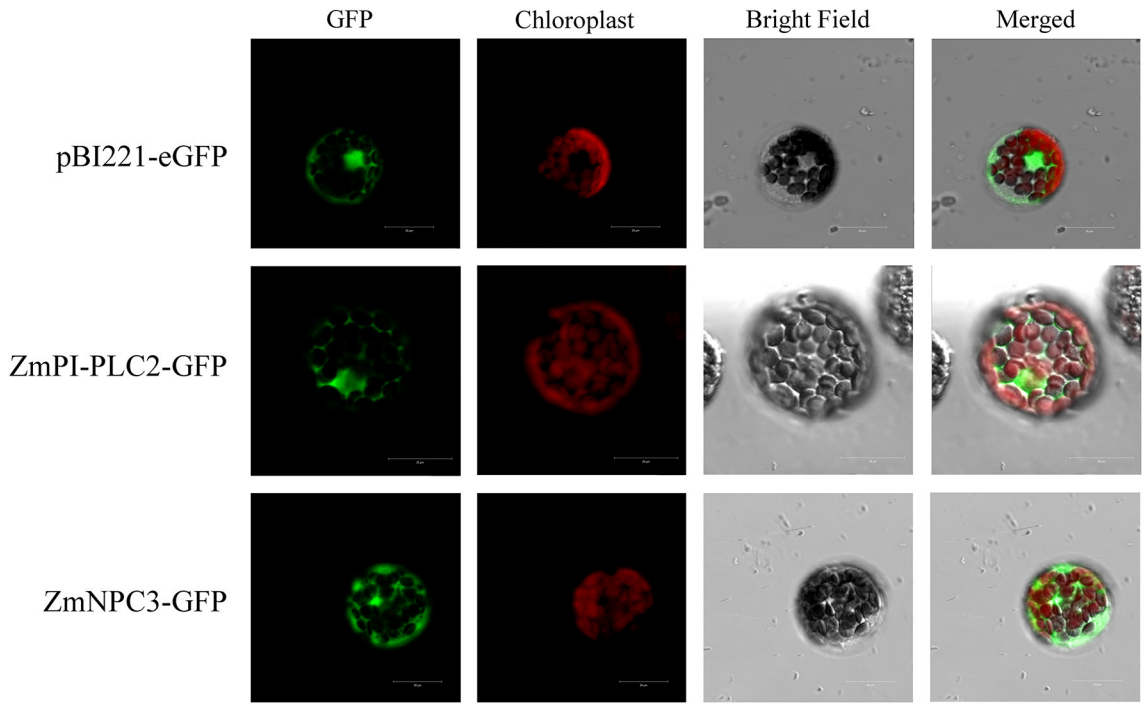

FIGURE 7 | Subcellular localization of ZmPLC proteins in *Arabidopsis* leaf protoplasts.

for example, light-responsive elements, ABA-response elements, MeJA-responsive elements, LTR element, and so on (**Figure 4**). These results mean that *ZmPLC* genes could be involved not only in the regulation of the growth and development of maize but also in various stress responses.

The expression patterns of *PLC* genes have already been studied in different tissues and developments of many plants, such as *Arabidopsis*, rice, soybean, and so on. In this study, the expression profiles of *ZmPLC* genes were validated using qRT-PCR under different tissues. All *ZmPLC* genes were expressed in our tested tissues, while the expression abundance of different *ZmPLC* genes is significantly different in tissues (**Figure 5**). For example, the transcript levels of *ZmPI-PLC2* were relatively high in all organs, especially leaf and anther (**Figure 5**). Interestingly, *AtPI-PLC2*, involved in reproductive organ development, was also highly expressed in leaves, stems, roots, and flowers (Tasma et al., 2008; Li et al., 2015). *OsPLC* genes from rice are also expressed differentially during reproductive developmental phases including stages of panicle and seed development (Singh et al., 2013). Furthermore, the regional association analysis between 11 *ZmPLCs* and 17 agronomic traits in maize indicated that some *ZmPLC* genes are significantly correlated with many development-related traits (**Figure 8**). These results suggested that *ZmPLC* genes have important functions in the regulation of growth and development in maize and could act as important candidate genes to improve maize agronomic traits for breeding.

Many previous studies have reported that *PLC* members are involved in various abiotic stress-triggered signaling transductions in many plant species. For example, *AtPI-PLC3*, *AtPI-PLC9*, and *AtNPC1* genes play an important function in regulating heat tolerance (Gao et al., 2014; Krčková et al., 2015), whereas *AtNPC4* and *AtNPC5* genes could respond to salt stress (Kocourková et al., 2011; Peters et al., 2014). *OsPI-PLC4* plays a positive role in osmotic stress response (Deng et al., 2019). Overexpression of maize *PI-PLC1* could enhance drought tolerance of transgenic plants (Wang et al., 2008). In this study, the expression of *ZmPLC* genes showed different changes under different stresses; however, they presented their own characteristics (**Figure 6**). For example, all maize *PI-PLC* genes were induced by osmotic treatments, while only *ZmPI-PLC1* and *ZmPI-PLC3a* were rapidly induced by high salinity (**Figures 6A,B**). *ZmPI-PLC3a* and *ZmPI-PLC3b* were upregulated by cold treatment and suppressed by copper ion treatment (**Figures 6C,D**). *ZmNPC4* and *ZmNPC5* were downregulated by osmotic treatments; however, they were rapidly induced by high salinity (**Figure 6B**). The different expression changes of *ZmPLCs* under different stresses may be related to the *cis*-acting elements in their promoters; for example, the *cis*-regulatory elements including MBS (MYB binding site), ABRE, and defense and stress responsive element (AT-rich and TC-rich repeats) are known to regulate various stress responses (Abe et al., 2003; Narusaka et al., 2003). In this study, the LTR could be detected in *ZmPI-PLC1/2/3a/3b* and *ZmNPC3/4/5* gene promoters (**Figure 4**), and *ZmPI-LC3a* and *ZmPI-PLC3b* mRNAs were induced by cold treatments.

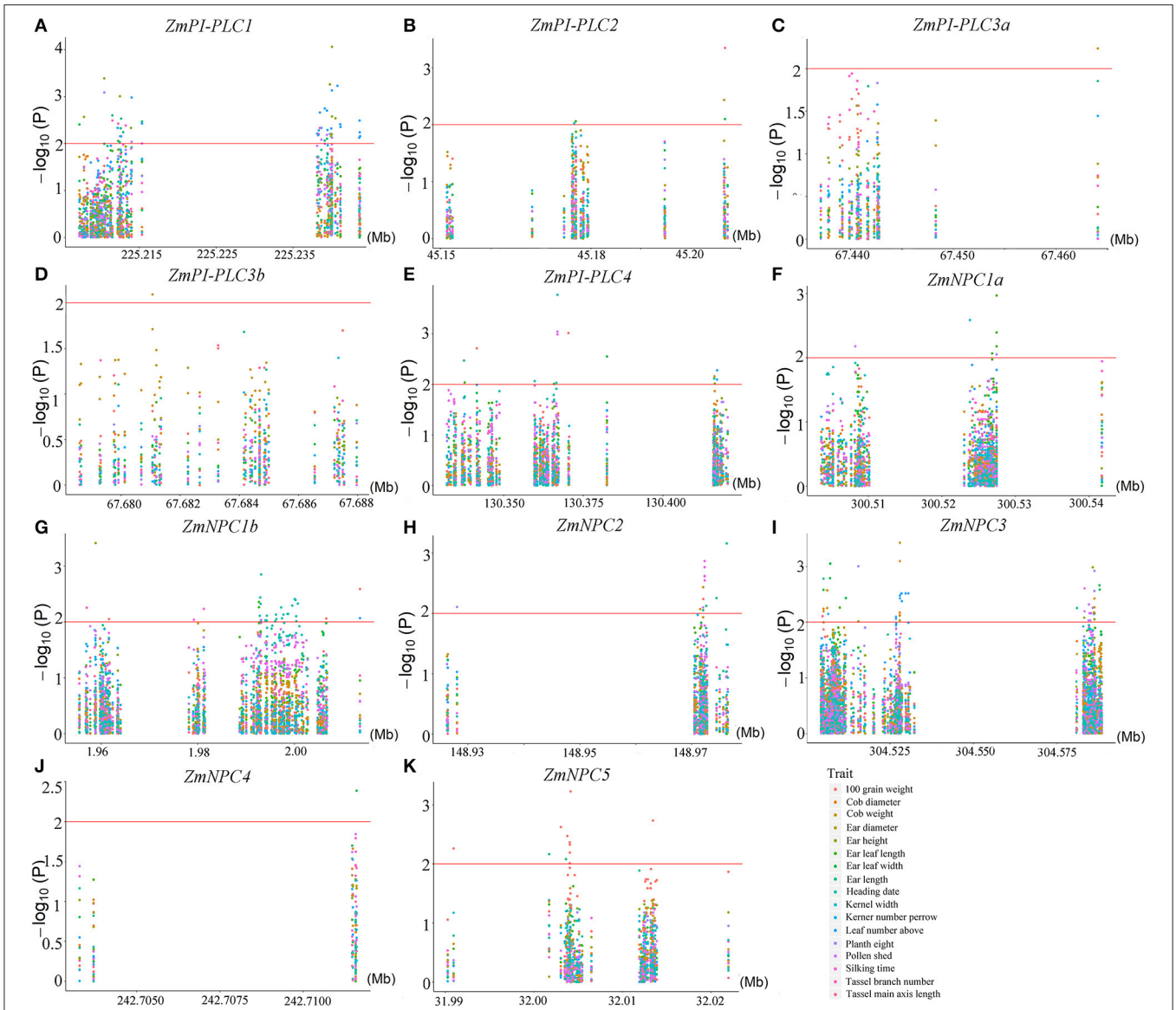

FIGURE 8 | Regional correlation analysis between *ZmPLC* genes with the agronomic traits in maize. **(A–K)** show the correlation between *ZmPLC* genes with the agronomic traits, respectively. Local Manhattan plot (top) and linkage disequilibrium heatmap (bottom) surrounding the SNPs for agronomic traits on the range of 100 kb upstream/downstream of *ZmPLC* genes. An logarithm of the odds (LOD) > 2 indicates that a SNP is significantly related to kernel-related traits.

The substrates of PLCs, phosphoinositides, PC, or PE, are mainly distributed in the plasma membrane (Munnik et al., 1998), so it is speculated that the subcellular localization of PLCs should also be in the plasma membrane. However, the sub-localization from many plants showed that PLCs maybe multi-localized. For example, *Arabidopsis* AtPLC9 and soybean GmPLC10-GFP fluorescence were located in the plasma membrane (Zheng et al., 2012), OsPLC1 and OsPLC4 were distributed throughout the cytoplasm and nucleus, while OsNPC3 protein might be localized in the chloroplast/plastids (Singh et al., 2013). DAG derived from PI-PLC or NPC activities can be phosphorylated to PA by DGK, while DAG could act as a substrate to produce various lipid species and also significantly affect properties of cell membranes as sites of crucial cell activity (Darwish et al., 2009; Hong et al., 2016). On the other hand, PI-PLC hydrolyzes PtdIns(4,5)P_2 into Ins(1,4,5)P_3, and Ins(1,4,5)P_3 is synthesized into InsP_6, which could bind to IP_6 receptors and lead to release of Ca^{2+} (Hong et al., 2016). Ca^{2+} and PA could act as a signaling molecule to play vital roles in various signaling pathways such as plant development, hormone signaling, and abiotic or biotic stresses to produce favorable response to plants (Xue et al., 2007). There is a possibility that the diverse localization of plant PLCs implies that they might take part in different cellular processes during development and abiotic stress.

CONCLUSIONS

In this study, 11 *ZmPLCs* were identified by genome-wide analysis. The results of evolution, gene structures, and motifs of PLCs indicated that ZmPLCs were highly conserved compared with their homologous genes from other plants. Additionally, the *cis*-elements, expression profiles, and association analysis between *ZmPLC* genes and agronomic traits for *ZmPLC* genes were also analyzed, which showed that *ZmPLC* genes may have important functions in regulating development and various stresses. Taken together, these results provide useful information for further study of the roles of *ZmPLCs* in plant development and environmental stress conditions.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

JZ and HL wrote the paper. JZ, YZ, and JL performed the experiments. JZ contributed to the data analysis. All authors read and approved the manuscript.

FUNDING

This study was funded by grants from the National Natural Science Foundation of China (31901497) and the scientific research leaders studio of Jinan (2019GXRC052).

ACKNOWLEDGMENTS

We are very grateful to Dr. Jianbing Yan (Huazhong Agricultural University, China) for providing maize inbred line W22.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.611414/full#supplementary-material

Supplementary File 1 | List of the PLC sequences from multi-species in this study.

Supplementary Figure 1 | Schematic domain structures of ZmPLC **(A)** and Multiple sequence alignments of ZmPLCs **(B)**. Protein sequences of ZmPLC were aligned by Bioedit. The conserved and characteristic domains and sequences of PI-PLCs and NPCs have been marked with colored box and red underline, respectively.

Supplementary Table 1 | Primers and their sequences used in this study.

Supplementary Table 2 | Synteny blocks of *PLC* genes between maize and two representative plant species.

Supplementary Table 3 | The *ZmPLC* genes list associated with agronomic traits in maize. The red mark indicates the significant value after Bonferroni correction ($p < 0.05/n$).

REFERENCES

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* 15, 63–78. doi: 10.1105/tpc.006130

Berridge, M. J. (1987). Inositol trisphosphate and diacylglycerol: two interacting second messengers. *Annu. Rev. Biochem.* 56, 159–193. doi: 10.1146/annurev.bi.56.070187.001111

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. doi: 10.1093/bioinformatics/btm308

Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant.* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, G., Snyder, C. L., Greer, M. S., and Weselake, R. J. (2011). Biology and biochemistry of plant phospholipases. *Crit. Rev. Plant Sci.* 30, 239–258. doi: 10.1080/07352689.2011.572033

Darwish, E., Testerink, C., Khalil, M., El-Shihy, O., and Munnik, T. (2009). Phospholipid signaling responses in salt-stressed rice leaves. *Plant Cell Physiol.* 50, 986–997. doi: 10.1093/pcp/pcp051

Deng, X. J., Yuan, S., Cao, H. S., Lam, S. M., Shui, G. H., Hong, Y. Y., et al. (2019). Phosphatidylinositol-hydrolyzing phospholipase C4 modulates rice response to salt and drought. *Plant Cell Environ.* 42, 536–548. doi: 10.1111/pce.13437

Di Fino, L. M., D'Ambrosio, J. M., Tejos, R., Wijk, R., Lamattina, L., Munnik, T., et al. (2017). Arabidopsis phosphatidylinositol-phospholipase C2 (PLC2) is required for female gametogenesis and embryo development. *Planta* 245, 717–728. doi: 10.1007/s00425-016-2634-z

Fu, J., Cheng, Y., Linghu, J., Yang, X., Kang, L., Zhang, Z., et al. (2013). RNA sequencing reveals the complex regulatory network in the maize kernel. *Nat Commun.* 4:2832. doi: 10.1038/ncomms3832

Gao, K., Liu, Y. L., Li, B., Zhou, R. G., Sun, D. Y., and Zheng, S. Z. (2014). *Arabidopsis thaliana* phosphoinositide-specific phospholipase C isoform 3 (AtPLC3) and AtPLC9 have an additive effect on thermotolerance. *Plant Cell Physiol.* 55, 1873–1883. doi: 10.1093/pcp/pcu116

Georges, F., Das, S., Ray, H., Bock, C., Nokhrina, K., Kolla, V. A., et al. (2009). Over-expression of *Brassica napus* phosphatidylinositol-phospholipase C2 in canola induces significant changes in gene expression and phytohormone distribution patterns, enhances drought tolerance and promotes early flowering and maturation. *Plant Cell Environ.* 32, 1664–1681. doi: 10.1111/j.1365-3040.2009.02027.x

Hicks, S. N., Jezyk, M. R., Gershburg, S., Seifert, J. P., Harden, T. K., and Sondek, J. (2008). General and versatile autoinhibition of PLC isozymes. *Mol. Cell* 31, 383–394. doi: 10.1016/j.molcel.2008.06.018

Hong, Y., Zhao, J., Guo, L., Kim, S. C., Deng, X., Wang, G., et al. (2016). Plant phospholipases D and C and their diverse functions in stress responses. *Prog. Lipid Res.* 62, 55–74. doi: 10.1016/j.plipres.2016.01.002

Kocourková, D., Krcková, Z., Pejchar, P., Veselková, S., Valentová, O., Wimalasekera, R., et al. (2011). The phosphatidylcholine-hydrolysing phospholipase C NPC4 plays a role in response of Arabidopsis roots to salt stress. *J. Exp. Bot.* 62, 3753–3763. doi: 10.1093/jxb/err039

Krčková, Z., Brouzdová, J., Daněk, M., Kocourková, D., Rainteau, D., Ruelland, E., et al. (2015). Arabidopsis non-specific phospholipase C1: characterization and its involvement in response to heat stress. *Front. Plant Sci.* 6:928. doi: 10.3389/fpls.2015.00928

Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., et al. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* 45, 43–50. doi: 10.1038/ng.2484

Li, L., He, Y., Wang, Y., Zhao, S., Chen, X., Ye, T., et al. (2015). Arabidopsis PLC2 is involved in auxin modulated reproductive development. *Plant J.* 84, 504–515. doi: 10.1111/tpj.13016

Lin, Y., Zhang, C., Lan, H., Gao, S., Liu, H., Liu, J., et al. (2014). Validation of potential reference genes for qpcr in maize across abiotic stresses, hormone treatments, and tissue types. *PLoS ONE* 9:e95445. doi: 10.1371/journal.pone.0095445

Meldrum, E., Parker, P. J., and Carozzi, A. (1991). The PtdIns-PLC superfamily and signal transduction. *Biochim. Biophys. Acta.* 1092, 49–71. doi: 10.1016/0167-4889(91)90177-Y

Munnik, T., Irvine, R. F., and Musgrave, A. (1998). Phospholipid signaling in plants. *Biochim. Biophys. Acta* 1389, 222–272. doi: 10.1016/S0005-2760(97)00158-6

Nakamura, Y., Awai, K., Masuda, T., Yoshioka, Y., Takamiya, K. I., and Ohta, H. (2005). A novel phosphatidylcholine-hydrolyzing phospholipase C induced by phosphate starvation in *Arabidopsis*. *J. Biol. Chem.* 280, 7469–7476. doi: 10.1074/jbc.M408799200

Narusaka, Y., Nakashima, K., Shinwari, Z. K., Sakuma, Y., Furihata, T., Abe, H., et al. (2003). Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis rd29A* gene in response to dehydration and high-salinity stresses. *Plant J.* 34, 137–148. doi: 10.1046/j.1365-313X.2003.01708.x

Ngo, A. H., Lin, Y. C., Liu, Y. C., Gutbrod, K., Peisker, H., Dörmann, P., et al. (2018). A pair of nonspecific phospholipases C, NPC2 and NPC6, are involved in gametophyte development and glycerol lipid metabolism in Arabidopsis. *New Phytol.* 219, 163–175. doi: 10.1111/nph.15147

Peters, C., Kim, S. C., Devaiah, S., Li, M., and Wang, X. (2014). Non-specific phospholipase C5 and diacylglycerol promote lateral root development under mild salt stress in *Arabidopsis*. *Plant Cell Environ.* 37, 2002–2013. doi: 10.1111/pce.12334

Pokotylo, I., Kolesnikov, Y., Kravets, V., Zachowski, A., and Ruelland, E. (2014). Plant phosphoinositide-dependent phospholipases C: variations around a canonical theme. *Biochimie* 96, 144–157. doi: 10.1016/j.biochi.2013.07.004

Pokotylo, I., Pejchar, P., Potocký, M., Kocourková, D., Krčková, Z., Ruelland, E., et al. (2013). The plant non-specific phospholipase C gene family. Novel competitors in lipid signalling. *Prog. Lipid. Res.* 52, 62–79. doi: 10.1016/j.plipres.2012.09.001

Singh, A., Kanwar, P., Pandey, A., Tyagi, A. K., Sopory, S. K., Kapoor, S., et al. (2013). Comprehensive genomic analysis and expression profiling of phospholipase C gene family during abiotic stresses and development in rice. *PLoS ONE* 8:e62494. doi: 10.1371/journal.pone.0062494

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121

Tasma, I. M., Brendel, V., Whitham, S. A., and Bhattacharyya, M. K. (2008). Expression and evolution of the phosphoinositide-specific phospholipase C gene family in *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 46, 627–637. doi: 10.1016/j.plaphy.2008.04.015

Tuteja, N., and Sopory, S. K. (2008). Plant signaling in stress: G-protein coupled receptors,heterotrimeric G-proteins and signal coupling via phospholipases. *Plant Signal. Behav.* 3, 79–86. doi: 10.4161/psb.3.2.5303

Vedel, V., and Scotti, I. (2011). Promoting the promoter. *Plant Sci.* 180, 182–189. doi: 10.1016/j.plantsci.2010.09.009

Wang, C. R., Yang, A. F., Yue, G. D., Gao, Q., Yin, H. Y., and Zhang, J. R. (2008). Enhanced expression of phospholipase C 1 (*ZmPLC1*) improves drought tolerance in transgenic maize. *Planta* 227, 1127–1140. doi: 10.1007/s00425-007-0686-9

Wang, F. W., Deng, Y., Zhou, Y. G., Dong, J. Y., Chen, H., Dong, Y. Y., et al. (2015). Genome-wide analysis and expression profiling of the phospholipase C gene family in soybean (*Glycine max*). *PLoS ONE* 10:e0138467. doi: 10.1371/journal.pone.0138467

Wang, X. M., Devaiah, S. P., Zhang, W. H., and Welti, R. (2006). Signaling functions of phosphatidic acid. *Prog. Lipid. Res.* 45, 250–278. doi: 10.1016/j.plipres.2006.01.005

Wimalasekera, R., Pejchar, P., Holk, A., Martinec, J., and Scherer, G. F. (2010). Plant phosphatidylcholine-hydrolyzing phospholipases C NPC3 and NPC4 with roles in root development and brassinolide signaling in *Arabidopsis thaliana*. *Mol. Plant* 3, 610–625. doi: 10.1093/mp/ssq005

Xue, H. W., Chen, X., and Li, G. (2007). Involvement of phospholipid signaling in plant growth and hormone effects. *Curr. Opin. Plant Biol.* 10, 483–489. doi: 10.1016/j.pbi.2007.07.003

Yang, X. H., Gao, S. B., Xu, S. T., Zhang, Z. X., Prasanna, B. M., Li, L., et al. (2011). Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breed.* 28, 511–526. doi: 10.1007/s11032-010-9500-7

Zhang, B., Wang, Y. M., and Liu, J. Y. (2018). Genome-wide identification and characterization of phospholipase C gene family in cotton (*Gossypium* spp.). *Sci. China Life Sci.* 61, 88–99. doi: 10.1007/s11427-017-9053-y

Zheng, S. Z., Liu,. Y. L., Li, B., Shang, Z. L., Zhou, R. G., and Sun, D. Y. (2012). Phosphoinositide-specific phospholipase C9 is involved in the thermotolerance of Arabidopsis. *Plant J.* 69, 689–700. doi: 10.1111/j.1365-313X.2011.04823.x

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 26 January 2021
doi: 10.3389/fgene.2021.645443

Global Transcriptome and Weighted Gene Co-expression Network Analyses of Growth-Stage-Specific Drought Stress Responses in Maize

Songtao Liu[1,2,3], *Tinashe Zenda*[1,2,3], *Anyi Dong*[1,2,3], *Yatong Yang*[1,2,3], *Nan Wang*[1,2,3]* *and Huijun Duan*[1,2,3]*

[1] State Key Laboratory of North China Crop Improvement and Regulation, Hebei Agricultural University, Baoding, China, [2] North China Key Laboratory for Crop Germplasm Resources of the Education Ministry, Hebei Agricultural University, Baoding, China, [3] Department of Crop Genetics and Breeding, College of Agronomy, Hebei Agricultural University, Baoding, China

Drought is the major abiotic stress threatening maize (*Zea mays* L.) production globally. Despite recent scientific headway in deciphering maize drought stress responses, the overall picture of key genes, pathways, and co-expression networks regulating maize drought tolerance is still fragmented. Therefore, deciphering the molecular basis of maize drought tolerance remains pertinent. Here, through a comprehensive comparative leaf transcriptome analysis of drought-tolerant hybrid ND476 plants subjected to water-sufficient and water-deficit treatment conditions at flared (V12), tasseling (VT), the prophase of grain filling (R2), and the anaphase of grain filling (R4) crop growth stages, we report growth-stage-specific molecular mechanisms regulating maize drought stress responses. Based on the transcriptome analysis, a total of 3,451 differentially expressed genes (DEGs) were identified from the four experimental comparisons, with 2,403, 650, 397, and 313 DEGs observed at the V12, VT, R1, and R4 stages, respectively. Subsequently, 3,451 DEGs were divided into 12 modules by weighted gene co-expression network analysis (WGCNA), comprising 277 hub genes. Interestingly, the co-expressed genes that clustered into similar modules exhibited diverse expression tendencies and got annotated to different GO terms at different stages. MapMan analysis revealed that DEGs related to stress signal transduction, detoxification, transcription factor regulation, hormone signaling, and secondary metabolites biosynthesis were universal across the four growth stages. However, DEGs associated with photosynthesis and amino acid metabolism; protein degradation; transport; and RNA transcriptional regulation were uniquely enriched at the V12, VT, R2, and R4 stages, respectively. Our results affirmed that maize drought stress adaptation is a growth-stage-specific response process, and aid in clarifying the fundamental growth-stage-specific mechanisms regulating drought stress responses in maize. Moreover, genes and metabolic pathways identified here can serve as valuable genetic resources or selection targets for further functional validation experiments.

Keywords: hub gene, drought stress, RNA-seq, weighted gene co-expression network analyses, *Zea mays* L

OPEN ACCESS

Edited by:
Jian Ma,
Sichuan Agricultural University, China

Reviewed by:
Quan Xie,
Nanjing Agricultural University, China
Linyi Qiao,
Shanxi Agricultural University, China

***Correspondence:**
Nan Wang
wangnan@hebau.edu.cn
Huijun Duan
hjduan@hebau.edu.cn

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *23 December 2020*
Accepted: *04 January 2021*
Published: *26 January 2021*

Citation:
Liu S, Zenda T, Dong A, Yang Y, Wang N and Duan H (2021) Global Transcriptome and Weighted Gene Co-expression Network Analyses of Growth-Stage-Specific Drought Stress Responses in Maize. Front. Genet. 12:645443. doi: 10.3389/fgene.2021.645443

INTRODUCTION

Among all the abiotic stress factors that present threats to agricultural production, drought has the largest dramatic effect on crop growth and productivity, in both natural and manmade agricultural systems (Wheeler and von Braun, 2013). With the current evidence suggesting a continued increase in global warming, water shortage, and climate change, against a rising human population, crop breeders are faced with the biggest food security challenge in history (Hu and Xiong, 2014). It is estimated that the demand for agricultural products, including cereals, will increase by ∼50% by the year 2030, driven by population and income growth. This will require unprecedented sustained increases in the production of annual food crops (Farooq et al., 2009). Therefore, it is of top priority for crop breeders to develop drought-tolerant crop cultivars in order to sustain higher yields and global food security under the prevailing climate change scenario.

Globally, maize (*Zea mays* L.) ranks as the third staple crop after wheat (*Triticum aestivum* L.) and rice (*Oryza sativa* L.), contributing to both food security and industrial growth in some agro-based economies (Thirunavukkarasu et al., 2017). However, maize productivity and production expansion are negatively affected by drought stress, especially in the arid and semi-arid regions of South East Asia and Sub-Saharan Africa. More precisely, 60% of China's maize production region lies in such drier regions. Consequently, a 20–30% maize yield loss per year occurs owing to water-deficit stress (Gong et al., 2014). Therefore, the development of maize hybrids with enhanced drought tolerance, either through conventional or genetic engineering approaches, is a priority goal for most maize improvement programs.

Maize is susceptible to drought stress throughout its life span, with the most devastating effects being experienced at the reproductive stage (Hu and Xiong, 2014). Generally, drought stress results in stomatal closure and reduced transpiration rates, decreased cell turgor, diminished photosynthetic efficiency, and overall plant growth (Zhang et al., 2018). The photosynthetic and gas exchange responses are the most sensitive to drought and the survival of drought-tolerant plants hinges on the maintenance of relatively high photosynthetic activity levels (Aslam et al., 2015). To cope with drought stress, plants institute several developmental-stage-specific changes at physiological and molecular levels. Numerous genes are expressed and translated in response to drought and have been identified to interact with the environment, thus the networks associated with water deficit conditions are quite complex. When plants are exposed to stresses, stress receptors, and transporters on cell membranes coordinate stress perception and signal transmission to the target genes. Then, phytohormones such as abscisic acid (ABA), cytokinin, auxin, and ethylene, etc., regulate numerous drought-inducible genes (Khan et al., 2019). At the same time, transcription factors (TFs), including basic region/leucine zipper motif (bZIP), NAM/ATAF/CUC transcription factor (NAC), myeloblastosis (MYB), WRKY, and dehydration responsive element binding protein (DREB) interact with *cis*-regulatory sequences to execute transcriptional regulation of gene expression, thereby providing adaptive responses to water-deficit conditions (Joshi et al., 2016). Additionally, plants activate cellular redox homeostasis maintenance through metabolic adjustment; transduce stress signals for the synthesis of defense enzymes and other antioxidant systems to protect cells from reactive oxygen species (ROS) damages; and institute stress-responsive proteins (Mahajan and Tuteja, 2005). Previous studies have highlighted the role of late embryogenic abundant (LEA) and heat shock proteins (HSPs) in enhancing tolerance to dehydration by functioning as chaperons to combat cellular damage (Hanin et al., 2011).

As a result of the fast advancement and reduction in the cost of next-generation sequencing (NGS) technologies, RNA-sequencing (RNA-Seq) has become a powerful tool for whole genome-wide gene expression profiling and has been widely used to investigate complex gene regulatory networks. This has immensely contributed to our better understanding of the complex molecular networks involved in adaptation and tolerance to water-deficit stress (Miao et al., 2017). RNA-seq technology has also been used in *Sorghum bicolor* L. (Fracasso et al., 2016) and rice (Ma et al., 2016). Resultantly, several genes that respond to drought stress have been identified (Zenda et al., 2019). In our previous study, we highlighted the role protein ubiquitination play in coordinating cellular crosstalk between stress and hormone signaling in maize seedlings under drought stress conditions (Zenda et al., 2019). Further, the co-expression of genes associated with osmotic adjustments and transporter proteins-maintained cell water balance at the seedling stage (Thirunavukkarasu et al., 2017).

Although global gene expression profiles in response to drought stress have been monitored in different maize tissues by micro-arrays and RNA-Seq experiments (Zheng et al., 2010; Min et al., 2016; Wang et al., 2018; Zenda et al., 2019), most of these studies were conducted separately in different tissues and at various developmental stages. Consequently, little is known about the gene co-expression networks of different developmental stages. In other words, it is not yet clear how maize drought adaptation is regulated genetically and how stress signaling pathways crosstalk with the developmental signaling pathways.

Fortunately, co-expression network analysis has become an important tool in the identification of gene co-expression in relation to their functional associations, this method identifies gene subsets that are highly correlated with each other within the network (Thirunavukkarasu et al., 2017). Particularly, the weighted gene co-expression network analysis (WGCNA), an approach in systems biology used to describe gene-related patterns in microarray samples allows for the amalgamation of vast amounts of microarray data from different biological samples and multiple experiments to obtain insights into genes from various metabolic pathways that possess similar expression patterns. The method has been successfully used to screen biomarkers and detect hub genes involved in metabolic pathways in abiotic stress responses (Wang B. et al., 2019).

Meanwhile, plant responses to abiotic stresses and drought stress, in particular, are dependent on the specific developmental stage and tissue affected and the level and duration of

the stress. Therefore, the co-ordination of tissue-specific and developmental-stage-specific responses to the whole plant responses to drought stress needs to be considered (Wang B. et al., 2019). Thus, developing crops with higher resistance to water-deficit stress requires knowledge of the underlying physiological and molecular mechanisms of drought tolerance at various plant developmental stages.

Here, we utilized RNA-seq based approach to perform a comparative transcriptome analysis of growth-stage-specific drought stress responses between drought-tolerant maize hybrid Nongdan 476 (ND476) plants subjected to well-watered (control) and moisture-stressed (drought) conditions under short term conditions. Additionally, twenty-four RNA-Seq datasets were used to conduct WGCNA analysis in order to identify gene subsets possessing similar expression patterns and highly correlated with each other within different metabolic networks. Our results will be vital in clarifying the fundamental developmental-stage-specific molecular mechanisms regulating drought tolerance in the tolerant maize genotype.

MATERIALS AND METHODS

Plant Materials and Drought Stress Treatment

The maize germplasm of hybrid ND476 that was used in this experiment was bred and provided by the North China Key Laboratory for Crop Germplasm Resources of the Education Ministry (Hebei Agricultural University, China). Maize hybrid ND476 is a highly drought-tolerant cultivar (as screened/identified by the Dryland Research Institute of Hebei Academy of Agricultural and Forestry Sciences, China). The experiment was conducted between May and July 2018 in a fully automated rain-proof shelter at Qing Yuan Experimental Station, Baoding, China (115.56 E; 38.80 N; 118 m). Each experimental plot measured 25 m^2 (5 m × 5 m), with 60 cm × 30 cm plant spacings. The field arrangement was set up in a randomized complete block design, with the water-sufficient and water-deprived groups replicated three times. The soil water content was kept between 70 and 80% in the well-watered plots (control) and 15–20% in water-stressed plots (treatment). The soil water content across the water/drought treatments at all the four maize growth stages was consistently kept the same. This was made possible by the use of drip irrigation for agricultural water supply for the trials evaluated under both conditions (Wang N. et al., 2019). The relative soil water content of one meter underground was monitored by the TZS-1 soil moisture measurement instrument (Zhejiang Tuopu Technology Co., Ltd., Zhejiang, China). To prevent the transverse infiltration of soil moisture, building waterproof membranes of one-meter depth were put between control and treatment units.

Drought treatment was instituted at four different maize growth stages. Plants were water-deprived (a) from eight fully-expanded-leaf (FEL) to twelve FEL (V12) period (flared stage); (b) from twelve FEL until the tassel was visible (VT) (tasseling stage); (c) from self-pollination until 12 days post pollination (DPP), that is, the prophase of grain filling stage (R2); and (d) from 13 DPP until 24 DPP, that is, the anaphase of the grain filling stage (R4) (**Figure 1**). For each growth stage, leaf tissues were collected from the flag leaves of three replicates, of both control and drought treatment conditions. All the leaf samples were immediately frozen in liquid nitrogen and then stored at −80°C for further analysis.

Determination of Photosynthetic Rate of Maize Leaves

Physiological parameters were measured for the maize under well-watered and water-deprived conditions from V12 to R2 stages. Specifically, photosynthetic rate (Pn) was measured according to drought treatment time and weather conditions using Li-6400 portable photosynthesis system (LI-COR Biosciences Inc., Lincoln, NE, United States) from 9:00 to 11:00 in the morning, as well as the conditions for measurement, were set as follows: photosynthetic photon flux density, 1500 μmol m^{-2} s^{-1}; and chamber CO_2 concentration, 300 μmol s^{-1}. The student's *t*-test was used to detect any significant differences in the data measured between control and drought treatment at each time point.

Total RNA Extraction, cDNA Library Construction, and Transcriptome Sequencing

The total RNA was extracted from the leaf samples of the control and water-deprived plants using TRIzol reagent (Invitrogen, Carlsbad, CA, United States) following the manufacturer's protocols. Subsequently, RNA was treated with DNase I (QIAGEN, Pudong, Shanghai, China) to eliminate contaminating genomic DNA. RNA degradation and contamination (integrity) were monitored on 1% agarose gels, then RNA quality was determined by 2100 Bioanalyzer (Agilent) and quantified using the ND-2000 (NanoDrop Technologies Inc., Wilmington, DE, United States). A total amount of 1 μg RNA per sample was used as input material. Only a high-quality RNA sample (OD260/280 = 1.8∼2.2, OD260/230 ≥ 2.0, RIN ≥ 6.5, 28S:18S ≥ 1.0) was used to construct the sequencing library. RNA purification, reverse transcription, library construction, and sequencing were performed at Shanghai Majorbio Bio-pharm Biotechnology Co., Ltd (Shanghai, China) according to the manufacturer's instructions (Illumina, San Diego, CA, United States). RNA-seq transcriptome library was prepared following TruSeqTM RNA sample preparation Kit from Illumina (San Diego, CA, United States) using 1 μg of total RNA. Briefly, mRNA was isolated according to the polyA selection method by oligo (dT) beads and then fragmented by fragmentation buffer firstly. Secondly, double-stranded cDNA was synthesized using a SuperScript double-stranded cDNA synthesis kit (Invitrogen, Carlsbad, CA, United States) with random hexamer primers (Illumina). The short fragments (200 – 300 bp) were ligated with adapters and the suitable fragments were chose PCR amplified using Phusion DNA polymerase (NEB) for 15 PCR cycles. After quantified by TBS380 (Turner Biosystems, United States), the paired-end RNA-seq sequencing library was sequenced with the Illumina Novaseq 6000 (2 × 150 bp read length). Four cDNA libraries were prepared using mRNA isolated from the leaves of both water-deprived and well-watered maize plants

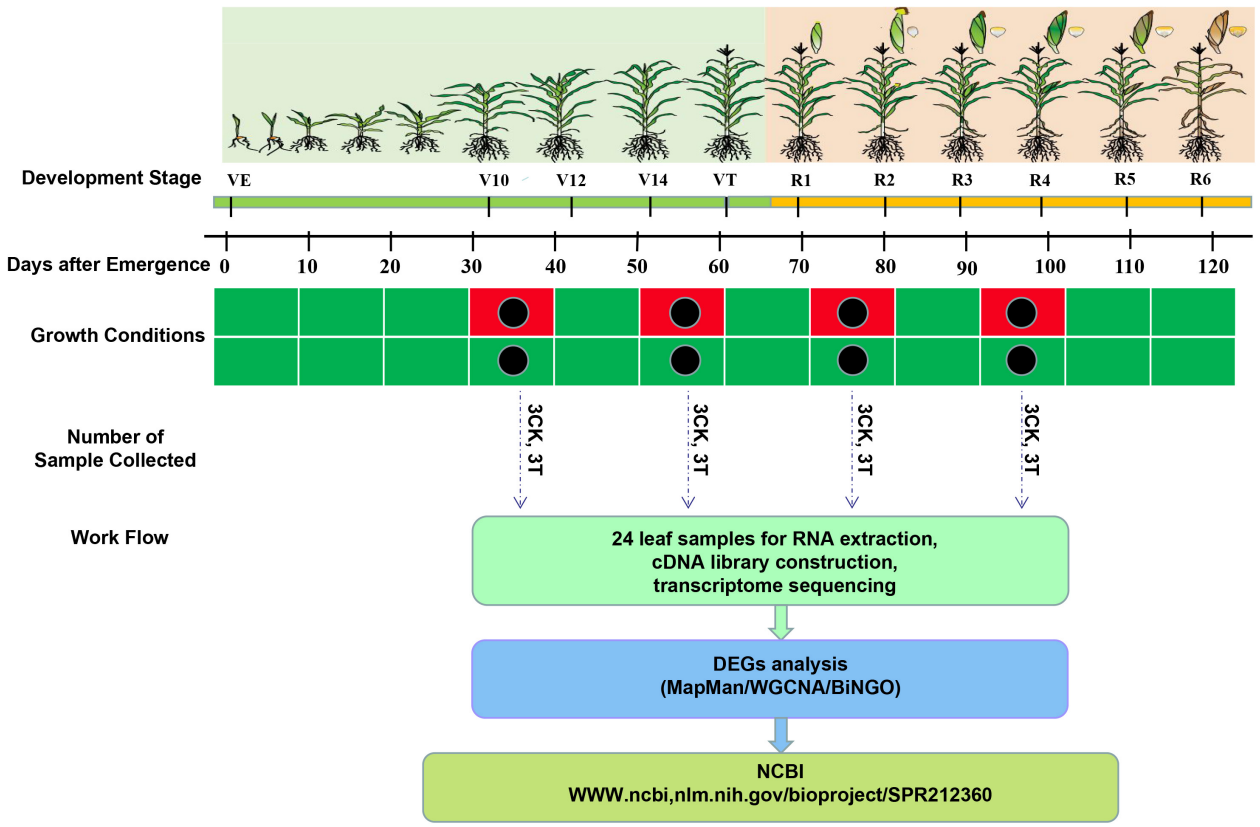

FIGURE 1 | Schematic representation of the leaf transcriptome sequencing experiment under control and drought stress treatment in drought-tolerant hybrid line ND476 from vegetative to reproductive development stages. On the crop development stage timeline, green and yellow colors represent vegetative and reproductive stages, respectively. On growth condition timelines, green color represents water-sufficient (control), whereas red color represents water-deficit (drought) treatment periods, respectively. Black dots show the time of leaf sample collection after 12 days of treatment exposure.

of drought-tolerant hybrid ND476 at four developmental stages. The libraries were denoted ND1_Con (NDCa, NDCb, NDCc) and ND1_Tre (NDDa, NDDb, NDDc) (the leaves of control and treatment maize at V12 stage), ND2_Con (NDCA, NDCB, NDCC) and ND2_Tre (NDDA, NDDB, NDDC) (the leaves of control and treatment maize at VT stage), ND3_Con (NDC1, NDC2, NDC3) and ND3_Tre (NDD1, NDD2, NDD3) (the leaves of control and treatment maize at R2 stage), ND4_Con (NDC4, NDC5, NDC6) and ND4_Tre (NDD4, NDD5, NDD6) (the leaves of control and treatment maize at R4 stage). To identify genes responsive to drought stress in maize leaves at various growth stages, global gene expression profiling was performed by Illumina RNA sequencing of these libraries.

Processing, Mapping of Sequencing Reads, and Gene Expression Quantification

Raw data (raw reads) generated by the Illumina Novaseq 6000 system were initially processed by SeqPrep[1] and Sickle[2] with default parameters. After trimming the adapter sequencing, removing low-quality bases, and filtering short reads, clean reads were separately aligned to the reference genome (ZmB73_RefGen_v4) with orientation mode using TopHat (version 2.1.1)[3] software. The mapping criterion was as follows: sequencing reads should be uniquely matched to the genome allowing up to 2 mismatches, without insertions or deletions. At the same time, Q20, Q30, GC-content, and sequence duplication level of the clean data (clean reads) were calculated. These high-quality reads were used in all the subsequent analyses. Subsequently, the gene expression level of each transcript was calculated according to the FPKM (fragments per-kilobase of the exon model per million mapped reads) based on the length of the reads count mapped to this transcript. For functional annotation, the quality reads were used for BLASTX alignment and annotation against non-redundant protein sequence database (Nr)[4], Swiss-port (a manually annotated and reviewed protein sequence database)[5], Clusters of

[1]https://github.com/jstjohn/SeqPrep

[2]https://github.com/najoshi/sickle

[3]http://tophat.cbcb.umd.edu/

[4]https://www.ncbi.nlm.nih.gov/

[5]https://web.expasy.org/docs/swiss-prot

Orthologous Groups (COG)[6], the Kyoto Encyclopedia of Genes and Genomes (KEGG)[7] and Gene Ontology (GO)[8] with the threshold *e*-value = 1E-5.

Differentially Expressed Genes (DEGs) Detection and Function Annotation of DEGs

Differential expression analysis of two samples was performed using the EdgeR package (Empirical Analysis of Digital Gene Expression in R)[9]. A differential expression analysis between stages was conducted using the ratio of FPKM values, and the *p*-value of each contrast corrected for multiplicity using the Benjamini and Hochberg method (Lei et al., 2015). In this study, genes with fold change (FC) $\geq$ 1.5 and *p*-value $<$ 0.05 found by EdgeR were assigned as DEGs.

To further characterize DEGs in response to drought stress, the DEGs were visualized using the Mercator web tool subsequently loaded into MapMan software for a functional and categories annotation. The mapping file was used for visualizing the functional classes and pathways belonging to hierarchical BINs and sub-BINs based on the putative function. Subsequently, WGCNA analysis was performed to establish the maize DEGs co-expression network using the free online platform – Majorbio Cloud Platform[10]. In a scale-free weighted gene network, a node corresponded to a DEG, and an edge was determined by the similarity expression profiles of paired genes calculated by Pearson correlation. We selected a soft threshold (β) 12 to construct the co-expression networks according to the adjacency matrix (**Supplementary Figure 1**). The other parameters were as follows: minModuleSize = 30, minKMEtoStay = 0.3, mergeCutHeight = 0.25. Clusters (Modules) were visualized using Cytoscape software (version 3.4.0). To further explore the modules' functions, BiNGO plugins of Cytoscape were used for GO enrichment analysis based on the hypergeometric test and Bonferroni correction method (FDR $<$ 0.05). Following WGCNA analysis, hub-genes were detected as the top 10% DEGs with the highest hub scores (Miao et al., 2017).

Quantitative Real Time-PCR (qRT-PCR) Analysis

To validate the Illumina sequencing data results, quantitative real-time PCR (qRT-PCR) was conducted on 24 RNA samples that were used in the preparation of sequencing libraries using a C1000 (CFX96 Real-Time System) Thermal Cycler (Bio-Rad). Twelve genes that co-expressed at two treatment stages were selected for qRT-PCR to verify the RNA-seq results. Specific primers for each DEG were designed according to individual gene sequences using Primer Premier 5 Designer (Premier Biosoft International, Palo Alto, CA, United States). The cDNA for qRT-PCR analyses was synthesized from 1 μg total RNA with HiFiscript cDNA Synthesis Kit (CWBIO, Beijing, China). QRT-PCR experiments were performed on a Bio-Rad iQ5 Thermo Cycler (Bio-Rad, Hercules, CA, United States) using 2 × Fast Super EvaGreen® qPCR Mastermix (US Everbright Inc., Daly City, CA, United States). Each PCR reaction mixture contained 10 μl of 2 × Fast Super EvaGreen® qPCR Mastermix, 1 μl of template cDNA, 1 μl of forwarding primer (50 pmol), 1 μl of reverse primer (50 pmol), and 7 μl ddH_2O to a final volume of 20 μl with three technical replicates of each gene. We used the maize gene *GAPDH* (accession no. X07156) as an internal control for data normalization. Additionally, a negative control was added. The relative mRNA abundance for each gene was determined for both the control and the drought-stressed samples by the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

RESULTS

Leaf Photosynthesis Rate Response of Maize Hybrid Cultivar ND476 to Drought Treatment

To determine whether the water-limited conditions could influence the physiological activities within the maize leaf tissues, in this research, we measured Pn of the drought-tolerant hybrid cultivar ND476 at different growth stages. Our analysis of the four stages showed that under well-watered conditions, at the V12 stage, Pn increased initially (from 1 to 9 days post-treatment exposure), and then decreased gradually thereafter. At the VT stage, Pn showed an increasing trend from day 1 onward. However, under well-watered conditions, at the R2 and R4 stages, Pn exhibited a slight gradual decrease throughout the treatment exposure period, starting from day 1 (**Supplementary Figure 2**). Meanwhile, all the four growth stages generally showed significantly reduced Pn under the water-deficit condition as compared to the well-watered condition (**Supplementary Figure 2**). This observation may indicate that with the increased drought exposure duration, leaf stomatal closure resulted in decreased leaf available CO_2, or there was increased photo-oxidative damage induced by an accumulation of ROS.

Illumina Paired-End Sequencing, Assembly and Annotation of Maize Leaf Transcriptomes

Resultantly, a total of 125.26 million raw reads were obtained. The raw sequencing data had been deposited in the NCBI Sequence Read Archive (SRA, Accession: SPR212360). After adaptors and low-quality reads were filtered out, 124.16 million clean reads were obtained, ranging from 23,029,648 to 72,636,578 for each sample. The clean reads were used for further analysis. Meanwhile, 20,432,633 (88.72%) to 64,399,057 (88.66%) clean reads were mapped onto unique positions on the maize reference genome (ZmB73_Ref-Gen_v4) (**Supplementary Table 1**). The Q30 base percentage and GC percentages exceeded 94.46% and 54.6%, respectively (**Supplementary Table 1**).

Subsequently, for functional annotation of the assembled transcriptome sequences, all the sequences were mapped onto

[6]https://www.ncbi.nlm.nih.gov/COG/

[7]http://www.genome.jp/kegg

[8]http://www.geneontology.org

[9]http://www.bioconductor.org/packages/2.12/bioc/html/edgeR.html

[10]www.majorbio.com

the public genome database with an *E*-value threshold of 1e-5. We annotated 179,093 (93.1%) and 137,849 (71.66%) genes in the NCBI Nr database and the Swiss-Port protein database, respectively (**Supplementary Table 2**). Based on KEGG analysis, only 82,758 genes were successfully annotated, accounting for 43.02% of the total number. In addition, 169,691 (88.21%) and 133,525 (69.41%) genes were annotated using COG and GO databases, respectively (**Supplementary Table 2**).

Additionally, the similarities or differences of the twenty-four samples were analyzed using principal component analysis (PCA). The PCA results of all the samples showed a clear separation between treatment and control samples at different stages (**Supplementary Figure 3**). To measure the gene expression levels for the three replicates for each sample, Pearson correlation coefficients between samples were calculated. The results shown by way of heatmap revealed that each R2 between the two samples was higher than 90% except for one comparison (NDC4_vs_NDDb) (**Supplementary Table 3**). These results indicated the overall reproducibility and quality of the assay, which met the demands for further analysis.

Identification of DEGs in Response to Drought

In order to reveal the transcriptional responses of maize leaves to water-stressed conditions, we compared the genes identified under water-sufficient and water-deficit conditions at four different growth stages. Gene expression levels were calculated and normalized to the RPKM values. Based on this analysis, a total of 3,451 DEGs were identified at four various maize growth stages. We obtained most numbers of DEGs at the V12 stage, including 1,203 upregulated and 1,200 downregulated. Meanwhile, 352 upregulated and 298 downregulated DEGs were identified at the VT stage. Similarly, we fished out 397 DEGs (95 upregulated and 302 downregulated) and 313 DEGs (112 upregulated and 201 downregulated) at the R2 and R4 stages, respectively (**Figure 2A**).

The number of DEGs showing overlaps and specific responses under drought stress in different growth stages is visualized in **Figure 2B**. A large number of DEGs were period-specific; there were 2,164, 483, 307, and 198 DEGs, respectively, at the four different stages. However, a limited number of common DEGs were detected. The area I represent 117 DEGs shared between V12 and VT stages after drought treatment, that is, the common DEGs identified in the two vegetative stages. Area II represents 15 DEGs shared between R2 and R4 stages after drought treatment, that is, the common DEGs identified in the two reproductive stages. There were 180 drought-responsive DEGs identified in the vegetative stages (V12 or VT) and also differentially expressed at the reproductive stages (R2 or R4) after drought treatment; that is, the DEGs identified at both vegetative and reproductive stages (Area III, **Figure 2B**).

To further understand the gene expressions between different stages, we performed the hierarchical clustering analysis of the identified DEGs (**Figures 2C,D** and **Supplementary Figure 4**). A total of 28 DEGs showed downregulated and 69 DEGs showed upregulated both of V12 and VT stages, but 20 DEGs showed the opposite trend of expression of the two stages (**Figure 2C**). Thirteen DEGs were downregulated both of R2 and R4 stages, 2 DEGs showed the opposite trend of expression of the two stages (**Figure 2D**). The others shared DEGs in different stages also showed different expression patterns (**Supplementary Figures 4A–D**). These results indicated that there were different mechanisms of maize drought stress responses at various growth phases.

Functional Annotation of DEGs Using MapMan

All DEGs of four growth stages were assigned to MapMan functional categories. The DEGs were grouped into 35 BINs with putative functions (**Supplementary Figure 5**). We found out that 720, 183, 147, and 72 DEGs of the V12, VT, R2, and R4 stages, respectively, were not assigned to any functional group (BIN 35) due to lack of annotation information (**Supplementary Figure 5**). The DEGs of the V12 stage were mainly annotated to the cell wall, lipid metabolism, photosynthesis (PS), protein synthesis, and degradation, abiotic stress, secondary metabolites biosynthesis and hormone metabolism (**Figure 3A**). The highly enriched categories of the VT stage DEGs included lipid metabolism, amino acid metabolism, and hormone metabolism (**Figure 3A**). Meanwhile, the enriched categories of the R2 stage DEGs included lipid metabolism, protein degradation, and secondary metabolites biosynthesis, whereas the R4 stage DEGs related to transport, PS, hormone metabolism, secondary metabolites biosynthesis, and RNA transcriptional regulation were highly enriched (**Figure 3A**).

An overview analysis of DEGs was generated with the MapMan tool and the drought-inducible regulated genes were classified into different regulatory processes (**Table 1**). The DEGs involved in PS showed upregulated expression under drought stress conditions in all the stages, except the VT stage. A total of 171 DEGs related to transport were altered in expression among the four growth stages. Moreover, 266 protein kinases including serine/threonine-protein kinases, leucine-rich repeat receptor-like proteins, receptor-like kinases, phospholipases, and protein phosphatases were observed to be mostly upregulated at the vegetative stages, whilst showing downregulation at the reproductive stages. Several plant hormones, which functions as regulatory compounds, were identified to be responsive to drought stress including abscisic acid (ABA), auxin, brassinosteroids (BR), gibberellic acid (GA), salicylic acid (SA), and ethylene. DEGs related to auxin and SA were downregulated, whilst the DEGs involved in other hormones showed an increased expression under drought stress. Additionally, 289 differentially expressed TFs were identified, such as C_2H_2, bHLH, HB, MYB domain, MYB-related, and WRKY domain (**Figure 3B**). In the current study, more increased abundance TFs were identified in the vegetative stages compared to the reproductive stages.

The DEGs were involved in maintaining redox homeostasis by a series of enzymatic compounds including thioredoxin (TRX), glutathione S-transferases (GST), and peroxidase (POD), which played major roles in protecting maize from oxidative

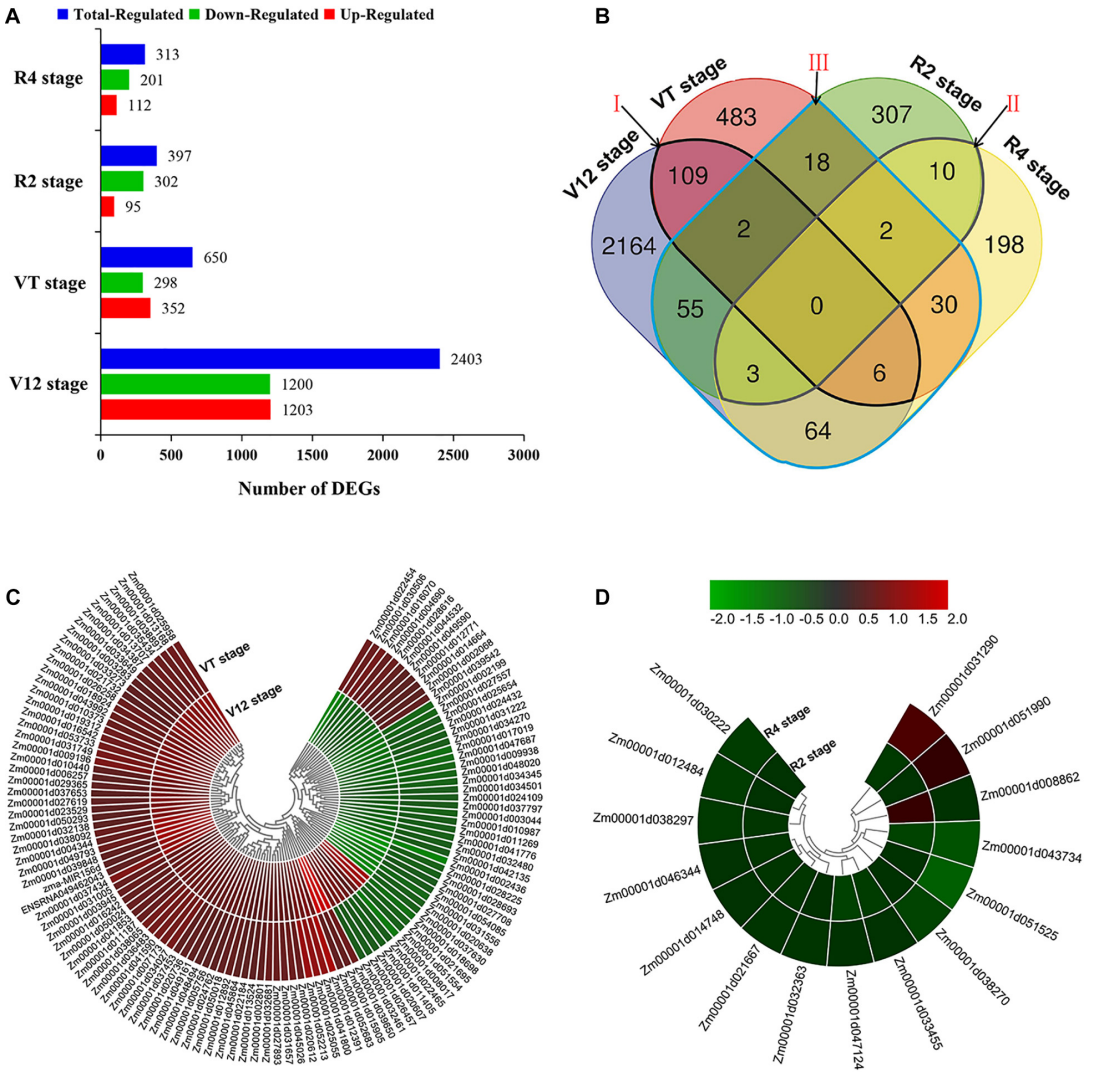

FIGURE 2 | Gene differential expression and clustering analysis. **(A)** Number of DEGs expressed in different experimental stages **(B)** Venn diagram analysis of DEGs. The regions labeled Area I–III identify genes described under section 'Identification of DEGs in Response to Drought' above **(C)** Clustering analysis of DEGs common in vegetative stages (Area I). **(D)** Clustering analysis of DEGs common in reproductive development stages (Area II).

damage. Additionally, 61 DEGs were annotated to stress defense. HSPs mainly showed upregulated, whist LEA and pathogenesis-related proteins (PRPs) had increased abundance under drought. The identification of such a great number of regulatory DEGs showed that there were multiple signaling mediators and intricate pathways in response to drought stress. Subsequently, we also obtained 10, 3, and 2 DEGs of the V12, VT, and R4 stages that were annotated to "response to drought/salt" (BinCode: 20.2.3) (**Table 1**). Among them, responsive to dehydration 22 (RD22, *AT5g25610*) mediated by ABA was identified in the V12 and VT stages and was involved in response to desiccation. Drought-responsive family protein *AT3g05700* and ERD (early-responsive to dehydration stress, *AT4g22120*) family protein may play a role in maize response to water-stress in the V12 stage. In addition, gene encoding AOC (allene oxide cyclase, *AT3g25780*), which is involved in jasmonic acid biosynthesis, is suggested to play a functional role in maize response to drought at the V12 stage. Taken together, these differentially expressed genes were speculated to be the vital cogs in maize drought stress tolerance, and hence aroused our keen interest for further discussion.

Co-expression Network Analysis of DEGs by WGCNA

To capture crucial shifts in gene networks in maize under water-stressed conditions, we further applied the WGCNA approach to perform a network-level analysis of co-expression relationships among 3,451 DEGs based on their expression patterns throughout the four growth stages. After filtering, a total of 2,771 DEGs were divided into 12 modules (clusters)

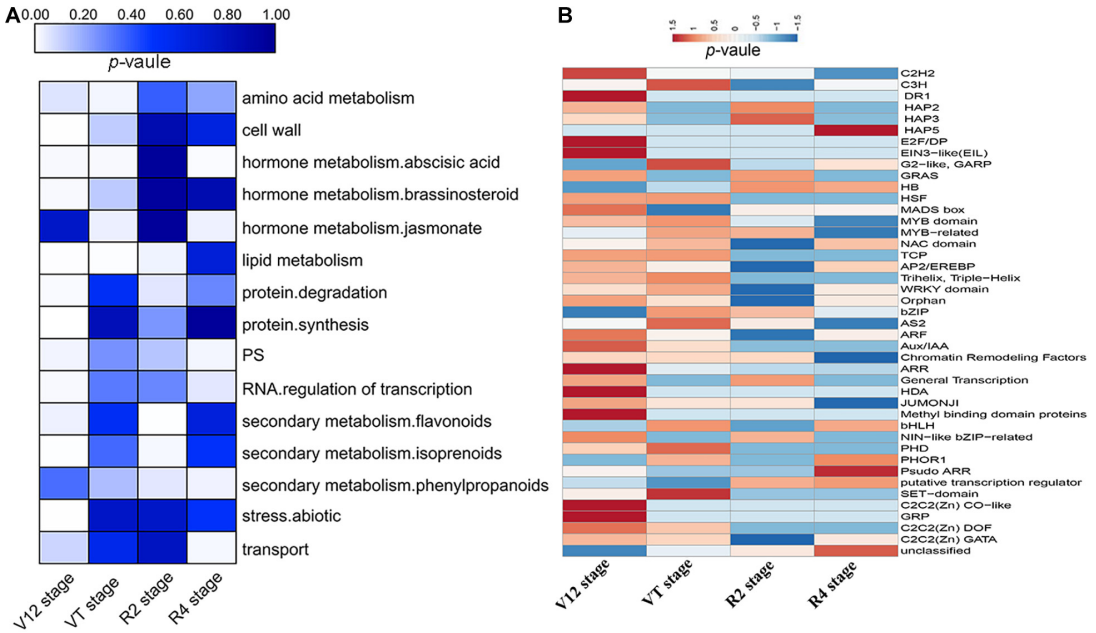

FIGURE 3 | Enrichment analysis of the functional categories of DEGs for the four maize growth stages by MapMan. **(A)** Enrichment of DEGs into functional categories for different stages. **(B)** Enrichment of transcription factor subcategories.

(designated M1-M12) comprising of 32 to 1,155 highly co-expressed genes (**Figure 4**, **Supplementary Figure 6**, and **Supplementary Table 4**). GO enrichment analysis of each module by BiNGO highlighted vital biological processes represented by a series of co-expressed genes. Module M1 formed the largest cluster of 1,155 DEGs enriched in functions related to metabolic processes (cellular amino acid, oxoacid, organic acid, and small molecule) and response to temperature stimulus (**Figure 5** and **Supplementary Table 5**). We also observed that cluster M5 had 172 DEGs enriched in functions related to metabolism process (peptide, cellular amide, and protein) and biosynthetic process (peptide, amide, and cellular macromolecule) (**Supplementary Table 5**). Meanwhile, module M11, comprising a cluster of 34 DEGs, had its DEGs annotated to biosynthetic process (cinnamic acid, phenylpropanoid, and carboxylic acid) and metabolism process (cinnamic acid, aromatic amino acid, and benzene-containing compound) (**Supplementary Table 5**). Modules M2, M10 and M12 showed enrichment of GO terms related to photosynthesis (**Supplementary Table 5**). Additionally, module M10 represents DEGs that showed high-expression specifically in the VT-stage and were enriched in GO terms associated with ion homeostasis and transport (**Supplementary Table 5**). The black module M7 included 52 DEGs involved in response to an abiotic or environmental stimulus (**Supplementary Table 5**). Further, the brown module M3 included DEGs related to ribosome and ribonucleoprotein complex biogenesis, gene expression, and RNA processing (**Supplementary Table 5**). Modules M4, M6, M8, and M9 did not get enriched in any GO term. By combining our gene expression pattern and GO enrichment analysis results, we concluded that the DEGs mainly participated in metabolic and biosynthetic processes, photosynthesis, ion homeostasis, transport, and response to abiotic stimulus under drought stress conditions.

Identification of Hub Genes Within Network Modules

There were some genes with extremely high connectivity with other genes, and these were designated as hub genes in each network module. Owing to their central location within the network clusters, the hub genes were considered to be vital components of the networks. Selecting only the top 10% of genes that showed high connectivity degree, a total of 277 DEGs were identified as hub genes (**Figure 6A**), including 17 TFs represented from distinct families including WRKY, MYB-related, C_2H_2, MYB, and NAC TFs (**Supplementary Table 6**). Seven hub genes were also identified as crucial enzymes playing a key role during maize drought stress response (**Supplementary Table 6**). Besides TFs and enzymes, two HSP90 genes that were also observed to respond to water-deficit stress conditions. Six hub-genes (*Zm00001d005410*, *Zm00001d025920*, *Zm00001d008462*, *Zm00001d019363*, *Zm00001d020272*, and *Zm00001d047235*) were observed to respond to drought stress by taking part in photosynthesis (**Supplementary Table 6**). Further, we conducted significant KEGG pathway enrichment analysis of these hub genes by using the hypergeometric test. Resultantly, by comparing the top ten pathways that were most enriched in the hub-genes, we discovered that starch and sucrose metabolism (6 genes), photosynthesis (4), linoleic acid metabolism (2), and photosynthesis - antenna proteins (2) were dominant under droughts stress conditions

TABLE 1 | Classification of drought-response regulated DEGs into different categories according to MapMan annotation.

Category	V12 stage		VT stage		R2 stage		R4 stage	
	Up-	Down-	Up-	Down-	Up-	Down-	Up-	Down-
Photosynthesis								
Photosystem II	3	1	0	1	1	0	3	0
Photosystem I	2	0	0	3	0	0	1	0
Electron carrier (ox/red)	3	0	1	2	1	0	0	0
Calvin cycle	0	0	0	2	1	1	0	0
Transport								
Transport sugars	8	3	1	2	0	1	0	3
Transport amino acids	11	7	1	4	2	0	1	4
Transport ammonium	0	2	0	1	1	0	0	1
Transport phosphate, nitrate, sulfate	2	4	1	2	0	0	2	4
Transport metal	3	2	3	0	0	3	0	0
Transport peptides and oligopeptides	11	6	1	1	0	1	0	5
Transport potassium	8	4	2	0	0	3	1	1
ABC transporters	6	4	2	1	0	2	1	3
Transport misc	9	6	2	5	0	4	1	2
Protein kinases and phosphatases								
Serine/threonine-protein kinase	11	4	2	1	0	5	0	1
LRR receptor-like protein	27	6	10	1	1	7	1	2
Receptor-like kinase	28	20	5	7	0	5	0	10
Phospholipase	8	2	5	0	0	1	2	0
Calmodulin	10	6	0	2	0	4	0	1
Mitogen-activated protein kinase kinase kinase	2	1	1	2	0	0	0	0
Protein phosphatase	2	6	3	0	0	0	4	0
Protein phosphatase 2C	4	0	2	0	0	0	2	2
Plant Hormones								
Abscisic acid	14	2	8	1	1	3	9	3
Auxin	4	7	3	3	1	2	1	2
Brassinosteroid	11	4	4	0	0	0	1	1
Jasmonic acid	3	3	4	2	1	1	0	3
Salicylic acid	1	4	1	2	0	0	0	2
Gibberellic acid	3	0	0	0	1	1	0	1
Ethylene	6	4	2	2	2	2	0	0
Transcription factors family								
Basic Helix-Loop-Helix	9	3	4	2	2	1	1	1
C2H2 zinc finger	6	3	0	1	0	2	0	1
Homeobox transcription factor	11	4	2	0	0	1	2	2
MYB domain transcription factor	4	3	3	3	0	2	0	2
MYB-related transcription factor	3	1	0	1	0	1	0	1
NAC domain transcription factor	2	4	3	2	0	0	0	1
WRKY domain transcription factor	3	3	1	1	0	0	0	1
bZIP transcription factor	3	0	1	2	0	3	1	1
G2-like transcription factor	2	0	1	3	0	1	0	2
Other transcription factor families	42	54	22	16	7	11	11	3
DEGs related to detoxification								
Thioredoxin	2	1	0	1	0	0	2	0
Glutathione S-transferases	6	3	1	4	0	1	0	2
Peroxidase	8	6	1	3	1	4	0	2
Ascorbate and glutathione	5	5	1	0	1	0	1	1
Glutaredoxins	5	1	0	1	0	0	1	0

(Continued)

TABLE 1 | Continued

Category	V12 stage		VT stage		R2 stage		R4 stage	
	Up-	Down-	Up-	Down-	Up-	Down-	Up-	Down-
DEGs involved in defense								
Heat shock proteins	4	20	8	7	0	1	0	0
Late embryogenesis proteins	6	0	2	1	0	0	0	0
Pathogenesis-related proteins	5	1	4	0	0	2	0	0
DEGs response to abiotic								
Response to heat	4	23	8	7	0	1	0	0
Response to drought/salt	6	4	3	0	0	0	2	0
Response to cold	3	1	0	2	0	2	0	2
Secondary metabolism								
Isoprenoids metabolism	13	2	5	0	3	0	0	5
Phenylpropanoids metabolism and biosynthesis	14	7	6	3	0	2	0	12
Flavonoids metabolism	9	15	1	2	7	1	2	4
Sulfur-containing metabolism	5	6	2	2	0	0	0	4

(**Figure 6B** and **Supplementary Table 7**). Moreover, the significantly enriched GO terms related to drought response were identified, including photosynthesis (GO: 0015979), response to stress (GO:0006950), and response to external stimulus (GO:0009605) (**Figure 6C**).

Validation of DEGs by Quantitative Real-Time PCR (qRT-PCR)

To confirm the accuracy of the RNA sequencing results, we conducted a validation experiment by qRT-PCR analysis for three biological replicates. The representative DEGs were chosen based on them being highly differentiated in response to drought and reported to be related to drought resistance. Precisely, the patterns of RNA-Seq expressions on all the 12 DEGs were consistent with the qRT-PCR data, suggesting that the patterns of the RNA-seq expression on all the sampled genes were replicated by the qRT-PCR approach (**Figure 7** and **Supplementary Table 8**). A correlation coefficient (R^2) (of the fold changes before and after drought treatment) of 93.01% was obtained (**Supplementary Figure 7**), endorsing our RNA-Seq data as reliable.

DISCUSSION

Drought stress during the transition from vegetative to reproductive development greatly affects grain production in maize (Zheng et al., 2010; Aslam et al., 2015). Thus, a full understanding of physiological, biochemical, and gene regulatory networks associated with water-deficit stress tolerance at these different growth stages in maize becomes imperative for breeding drought-tolerant cultivars. However, the complex adaptive mechanisms underpinning water-deficit stress tolerance from vegetative growth to reproductive development have remained elusive despite recent advances in molecular biology approaches (Bhanu et al., 2016). Therefore, in this report, we have employed RNA-seq based approach to perform a comprehensive comparative transcriptome analysis of drought-tolerant hybrid ND476 from the vegetative to reproductive growth stages to identify key regulatory genes and gene co-expression networks involved in maize drought stress response. We have further conducted photosynthetic parameter measurements to support the RNA-seq data. Additionally, functional validation by qRT-PCR analysis corroborated the differential expression of these identified genes. Our findings not only enrich our knowledge about maize drought stress tolerance mechanisms but also provide a valuable genetic resource or selection target for the genetic improvement of maize.

Photosynthesis Related Genes Were Differentially Altered in Response to Drought

Compared to other cereal food crops, maize is relatively sensitive to water-deficit stress (Pinheiro and Chaves, 2011). Photosynthesis is the most sensitive physiological process of plants subjected to abiotic stresses (Zhou et al., 2019). A stress-induced negative effect on any of the components of the photosynthesis systems may lead to a reduction in the overall photosynthetic performance (Lamaoui et al., 2018). To adapt to water-deprived conditions, plants will immediately close the stomata, thereby reducing the leaf gas exchange. This has negative influences on photosynthetic parameters. In the current investigation, our physiological analysis results showed that photosynthesis rate was significantly repressed by drought stress (**Supplementary Figure 2**). At the molecular level, MapMan annotation and analysis results of the common elements of modules M2, M10, and M12 by BiNGO found out that photosynthesis was significantly enriched mostly at the V12 and R4 stages in response to drought stress. Furthermore, we found several genes in the photosynthesis pathway with altered transcription abundance (**Table 1** and **Supplementary Figure 5**). The photosynthesis-related genes showed downregulated expression at the VT stage but showed

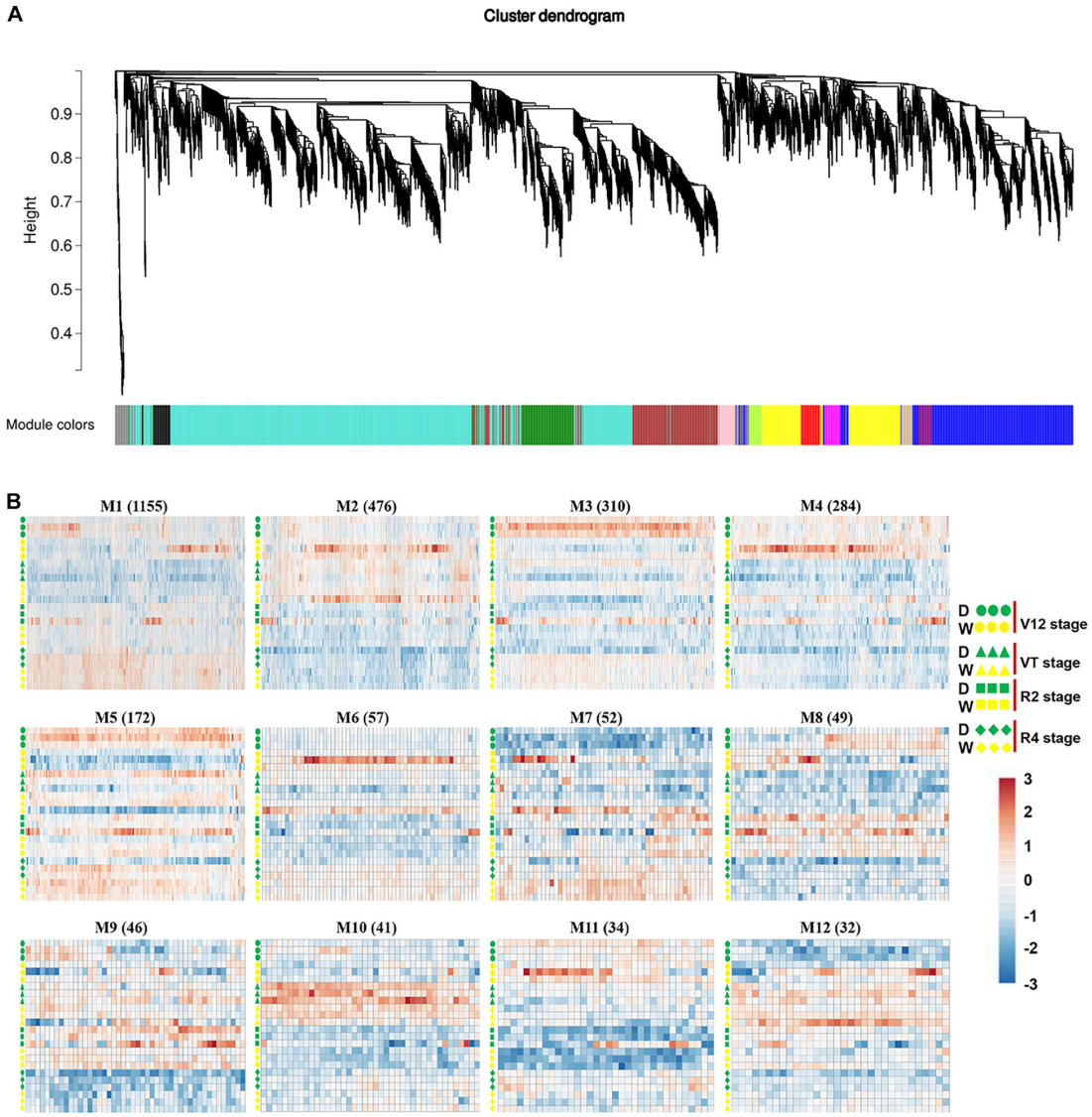

FIGURE 4 | Co-expression network analysis identifying gene modules underlying maize drought stress response at four different growth stages. **(A)** Hierarchical cluster trees showing co-expression modules identified using WGCNA of the differentially expressed genes. **(B)** Heatmap showing gene expression levels of the genes within the 12 modules across four growth stages. W, well-watered; D, drought treatment.

upregulation at the V12 and R4 stages (**Table 1**). This is consistent with the measured photosynthesis rate, and showing that drought-tolerant hybrid ND476 exhibited the well-maintained level of photosynthesis under drought stress conditions. Previously, Wang B. et al. (2019) observed that drought stress at the V9-V10 stages decreased net Pn in maize resulting in abnormal ear primordium development. Several other reports (Min et al., 2016; Lamaoui et al., 2018) have shown that drought stress represses photosynthesis in maize plants. However, Ma et al. (2016) found photosynthesis-related genes displaying upregulated expression in rice drought-tolerant line IAC1246, but the downregulated expression in rice drought-sensitive line IRAT109. Taken together, our results suggest that drought stress retards photosynthetic efficiency in plants, hence strategies aimed at improving photosynthesis under drought conditions can be vital for plants' survival.

Stress Signal Transduction and Protein Kinases Under Drought Stress Conditions

Drought-tolerance is typically a complicated trait since drought stress affects multiple aspects of plant physiology and metabolism, consequently, changing thousands of genes' expression (Min et al., 2016). At the initial stage of stress signal transduction, receptors, and transporters on cell membranes perceive stress bridging the gap between perception and transmission of the signals to the target genes

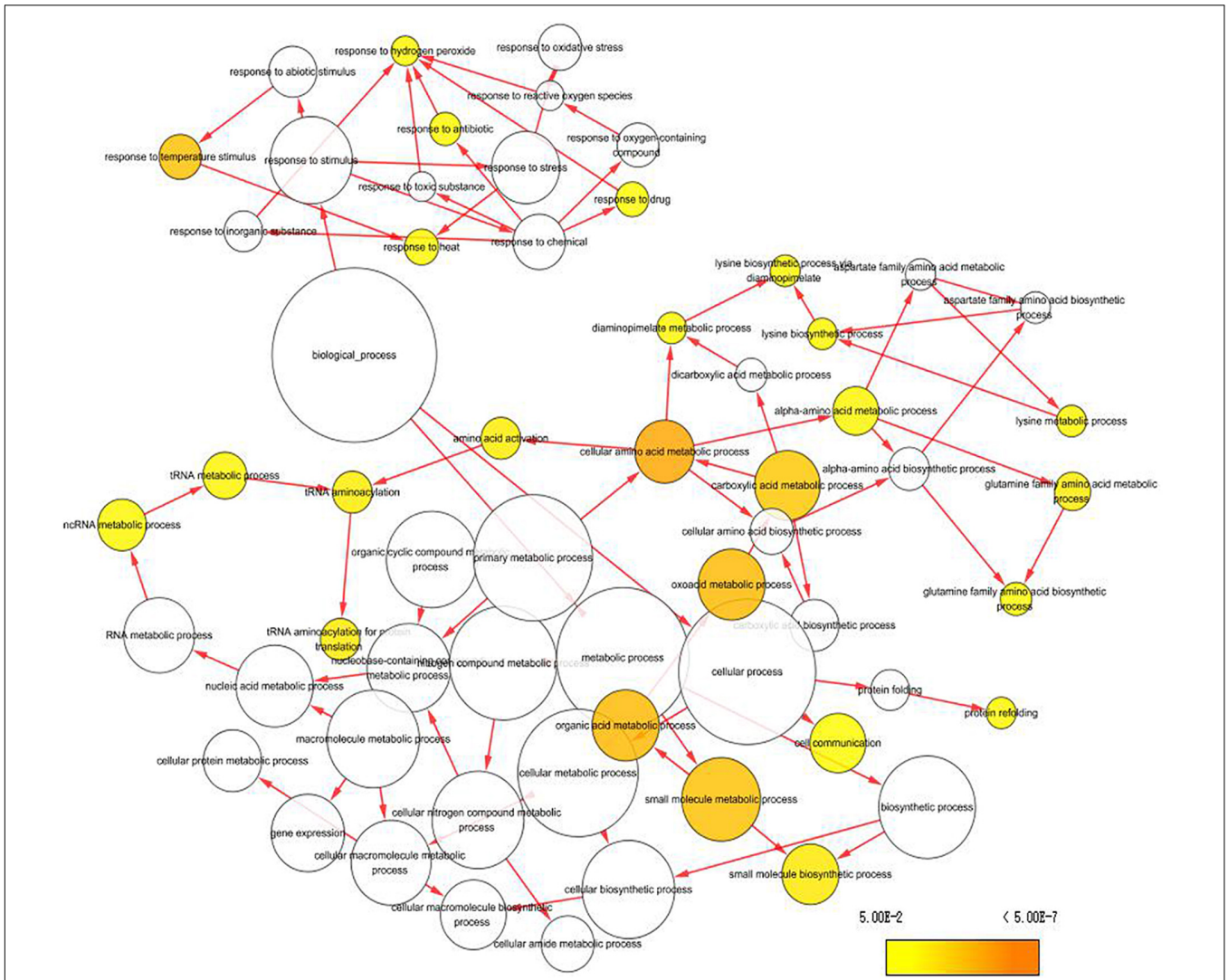

FIGURE 5 | Enrichment analysis of DEGs clustered in Module 1 (turquoise module) using GO terms from GO Slim. Significantly over-represented GO terms were visualized by BiNGO application in Cytoscape. The size of a node represents the proportion of the GO term to the number of targets in GO biological process category. The deeper the color, the higher the level of significance.

and contributing to plant survival. Water-limited conditions are often related to a significant increase in transport-proteins and channel proteins (Guo et al., 2006). In this study, several transporters were identified in maize response to drought stress conditions, including phosphate transporters, ion transporters and nitrate transporters (**Table 1**). Additionally, receptor kinases, another vital type of membrane protein, were found differentially expressed in response to drought stress, including serine/threonine-protein kinases, cysteine-rich receptor-like protein kinases, and proline-rich receptor-like protein kinases (**Table 1**). Among them, cysteine-rich receptor-like protein kinase (*Zm00001d008462*), proline-rich receptor-like protein kinase (*Zm00001d043480*) and receptor-like serine/threonine-protein kinase (*Zm00001d002199*) was identified as hub-genes (**Supplementary Table 6**). Receptor-like protein kinases (PLPKs) form the largest part of all plant kinases, and are well-known for playing an important role in abiotic stress responses. Several PLPKs detected in both leaves and roots of Bothriochloa ischaemum significantly changed their expression in response to drought stress (Li et al., 2019).

Protein kinases (PKs) are sensor responder genes which initiate phosphorylation cascades and play essential parts in water-deficit responses (Singh and Laxmi, 2015). In this current study, PKs were differentially expressed to regulate stress signaling transmitting in maize under water-limited conditions (**Table 1**). Moreover, serine/threonine protein phosphatase 2A (Zm00001d019363) was identified as a hub-gene (**Supplementary Table 6**). Similarly, the identical PKs were reported in faba bean (Vicia faba L.) drought-tolerant variety hassawi-2 response to drought stress (Khan et al., 2019). From this discussion, we can infer that a complex web of signaling was triggered under water-stressed conditions, which relayed

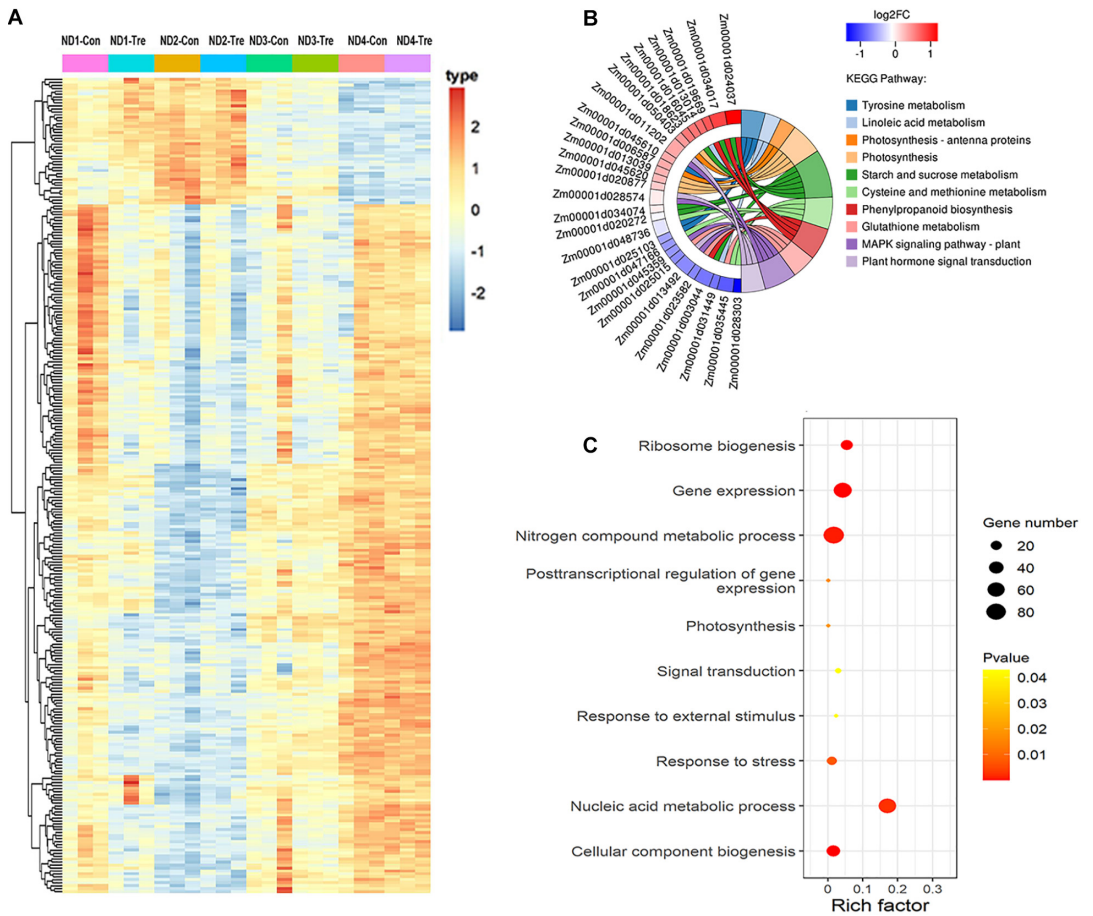

FIGURE 6 | Co-expression network analysis of hub-genes. **(A)** Heatmap showing the hub-genes that were expressed at four different maize growth stages. **(B)** KEGG pathway enrichment analysis of the hub-genes. **(C)** GO functional classification of the hub-genes.

messages through the plasma membrane to the cell, activating a signal transduction cascade. Then, TFs were modulated by the PKs, as a result affecting corresponding response to the downstream drought-responsive genes that enabled maize to regulate its growth and metabolism under drought stress (Wang et al., 2016).

Enhanced Cellular Redox Homeostasis May Contribute to Drought Tolerance in Maize

When plants are subjected to water-stressed conditions, there is a rapid and transient production of ROS which can damage cellular components and structures. In response, plants institute various mechanisms to re-establish the cellular redox balance and homeostasis, and avoid cellular components and structures damage caused by ROS. Cellular redox homeostasis transduced signals for the synthesis of defense and antioxidant enzymes contribute to the modification of the antioxidant system and cell turgor maintenance by osmotic adjustment (Mahajan and Tuteja, 2005). In the current study, twenty-five peroxidases, six TRXs, fourteen L-ascorbate peroxidase (APX), and seventeen GSTs were observed to be differentially altered in their expressions in response to drought stress at various maize growth stages (**Table 1**).

It is well known that peroxidases are central in neutralizing the damaging effects of toxic peroxides and other ROS that accumulate under oxidative stresses. It has been reported that the upregulated expression of peroxidases protected wheat plants from ROS-induced cell damage under drought stress (Sheoran et al., 2015). Similarly, Khan and Komatsu (2016) indicated peroxidases' essential role in soybean root ROS scavenging and cellular redox homeostasis. The chloroplast thioredoxin (TRX) systems compose an important component of the redox network, with thioredoxin reductase (TRs) functioning in re-establishing cellular redox homeostasis. Previous researchers have identified that TRX genes were upregulated or downregulated in response to drought stress, adjusting the cellular redox status in the process (Xie et al., 2016). Ascorbate peroxidases (APXs) are a vital cog of the complex stress response network. APXs play a role in detoxifying hydrogen peroxide (H_2O_2) in the chloroplasts and cytosol using ascorbate as a substrate (Xu and Huang, 2010).

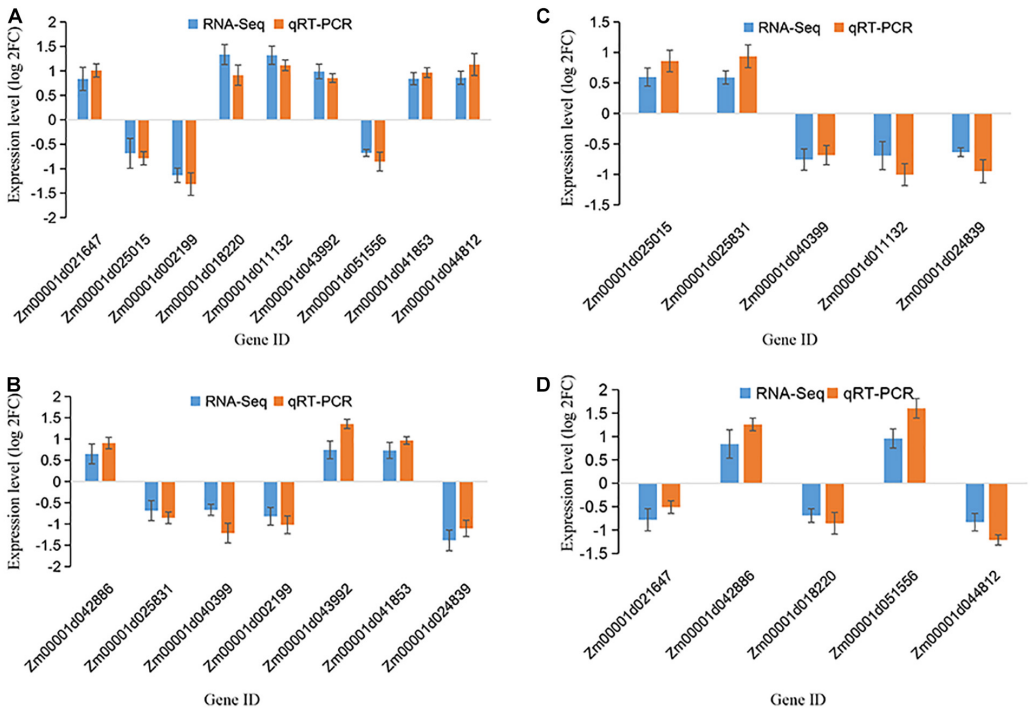

FIGURE 7 | qRT-PCR validation of selected genes among various maize growth stages. The *y*-axis represents the gene relative expression levels (fold changes) in real-time PCR analysis and RNA-seq data. **(A)** The DEGs identified at V12 stage; **(B)** The DEGs identified at VT stage; **(C)** The DEGs identified at R2 stage; **(D)** the DEGs identified at R4 stage. All genes with negative values means they were downregulated. Maize gene GAPDH (accession no. X07156) was used as the internal reference.

APXs expression has been reported to be significantly increased in winter rapeseed (*Brassica napus* L.) under drought stress conditions (Urban et al., 2017). Additionally, a previous study identified APXs being involved in ROS scavenging in maize in response to water-deficit stress (Miao et al., 2017). Glutathione-S-transferases (GSTs) are conjugating enzymes involved in the detoxification of a wide range of harmful substances such as ROS or xenobiotic compounds (Xu and Huang, 2010). Accumulation of GST in wheat has also been reported under drought stress conditions (Bazargani et al., 2011). However, GST showed downregulation in rapeseed subjected to drought stress (Mohammadi et al., 2007). Overall, our results here indicate that hybrid cultivar ND476 could endure water stress via increased activation of genes associated with ROS detoxification and oxidation-reduction processes, whereas the downregulated expression of some stress redox homeostasis genes may imply the complexity of the cell redox system in drought stress response.

Regulation of Drought Stress by Transcriptional Factors

As gene regulators, TFs play a key role in modulating gene expression and transmitting stress signals in plant cells. Therefore, TFs have been designated master regulators of abiotic stresses, including drought (Wang et al., 2016). In the current study, there were 42 classes of TFs that were annotated by MapMan among the four maize growth stages (**Figure 3B**). Meanwhile, among the total 277 hub-genes, 17 (6%) were annotated as TFs, belonging to 13 different families (**Supplementary Table 6**), indicating that differential transcription mechanisms function in the water stress signal transduction pathway in maize. A large number of identified TFs belong to MYB, NAC, WRKY, and bZIP families, which are well known for their roles in drought stress response (Thirunavukkarasu et al., 2017). Previously, Tran et al. (2004) observed that three *Arabidopsis* genes (ANAC019, ANAC055, and ANAC072) and two homologous maize NAC transcripts were abiotic-stress-inducible-expression genes. NACs were identified in foxtail millet (*Setaria italica* L.) (Shi et al., 2018) and maize (Song et al., 2017) responding to drought stress. WRKY TFs were identified as the key drought response elements by changing their differential expressions under drought stress (Yan et al., 2014). MYB factors were related to hormone signal transduction and abiotic stress response (Liu et al., 2015). Overexpression of R1R2R3-MYB TF in *Arabidopsis* significantly enhanced the tolerance of transgenic plants to drought stress (Zhang et al., 2019). More recently, Wu et al. (2019) reported that over-expressing ZmMYB3R enhanced drought and salt stress tolerance in transgenic maize. A previous study by Song et al. (2017), observed that C_2H_2 and bHLH TF factors were involved in drought stress response in maize leaves. Additionally, Zhang et al. (2014) observed that several NACs, MYBs, bZIPs, bHLHs, and other TFs expression was tightly coupled to plant water potential, indicating their involvement in *Medicago truncatula* L.

drought adaptation responses. Furthermore, 49 TFs from bHLH, bZIP, C_2H_2, MYB, and NAC families were found in maize spatio-temporal drought stress response (Miao et al., 2017). Taken together, the complex expression changes of these TFs crucially contribute to the drought stress tolerance of maize hybrid line ND476 as these TF genes interact with other molecular actors in complex networks.

Genes Related to Hormone Signaling Are Critical for Drought Stress Response

Plant hormones participate in numerous plant abiotic stress responses. For instance, a large number of genes related to biosynthesis or signaling of plant hormones such as auxin, ABA and ethylene were identified in a drought-tolerant faba bean variety under drought stress (Khan et al., 2019). Here, several auxin-responsive genes were altered in their expression levels in response to drought stress (**Table 1**). Previously, a series of auxin-responsive genes were induced by several abiotic stresses, revealing that auxin may function in abiotic stress signaling. Consistent with our findings, genes encoding auxin-responsive protein exhibited significantly increased abundance under partial desiccation treatment (Jing et al., 2017). It is well recorded that ABA is a critical messenger that mediates the adaptive response of plants to abiotic stresses. When plants are subjected to drought stress, a large number of ABA-regulated genes are accumulated. Previously, water deprivation resulted in high levels of ABA in maize, which stimulated the production of ROS and regulated the activity of the antioxidant defense system (Jiang and Zhang, 2002). Similarly, in our current study thirty-two (78%) genes related to ABA biosynthesis were upregulated. However, nine *cis*-epoxycarotenoid dioxygenase (NCED3) enzymes showed decreased expression under drought treatment (**Table 1**). By comparing our results with the previous observations, we realized that NCED3, a vital enzyme in ABA biosynthesis, was also downregulated in faba bean battling drought stress (Khan et al., 2019).

Meanwhile, we observed that sixteen (76%) genes encoding brassinosteroid (BRs) biosynthesis had an increased abundance in response to drought (**Table 1**). The BRs have critical functions in detoxifying oxidative damage by the expression of genes involved in ROS scavenging under drought stress conditions. Previously, BRs have also been identified in maize response to water-deficit conditions (Fracasso et al., 2016). The stress-tolerance ability of BR rests with its crosstalk with other plant hormones such as ABA (Divi et al., 2010). Additionally, we identified ethylene-, jasmonic acid- and salicylic acid-regulated genes that were responsive to drought treatment, suggesting that these plant hormones may play critical roles in drought stress signaling.

Stress Defense Genes Are Essential for Plant Response to Water-Stress Conditions

Expectedly, several genes involved in stress defense including HSPs, pathogenesis-related (PR) and LEA proteins showed significant changes in expression under drought stress conditions (**Table 1**). Two HSP90 genes (*Zm00001d041719* and *Zm00001d052809*) were identified as hub-genes in drought stress response. Meanwhile, twenty-eight HSPs exhibited decreased abundance at the four growth stages after drought treatment (**Table 1**). HSPs play important roles in helping proper folding or unfolding of proteins and preventing unwanted aggregation, as well as contributing to cellular homeostasis in cells under stress conditions (Wang et al., 2004). Here, the downregulation of the HSPs at the V12 stage may imply that short-term drought stress had a significant negative impact on the chaperon activities of these protective proteins at this stage. PRs are also thought to be involved in plants' developmental

FIGURE 8 | Model of the drought stress pathways occurring in maize from vegetative growth to reproductive development under drought stress. ABA, abscisic acid; JA, jasmonic acid; SA, salicylic acid; BR, brassinosteroid; ETH, ethylene; MAPK, mitogen-activated protein kinase; PP2C, protein phosphatase 2C; ERD4, early-response to dehydration stress; RD22, responsive to dehydration; POD, guaiacol peroxidase; TRX, thioredoxin; APX, ascorbate peroxidase; GST, glutathione S-transferases; HSP, heat shock proteins; PR, pathogenesis-related proteins; LEA, late embryogenesis proteins; PSI-D, photosystem I 20 kDa subunit II; PSI-N, photosystem I reaction center subunit N; PSI-G, photosystem I reaction center subunit V; PERK15, proline-rich receptor-like protein kinase 15; CP29, chlorophyll a-b binding protein CP29.1 chloroplastic.

processes and defense responses against abiotic stress (Kim et al., 2008). In our research, unlike HSPs, PR proteins showed significantly increased accumulation at the V12 and VT stages (**Table 1**). Previously, drought stress-induced response of PR proteins showed increased abundance in tobacco (Gharechahi et al., 2015) and grapevine (Król and Weidner, 2017).

Late embryogenic abundant proteins are commonly induced during water-deficit conditions, and their expressions are regulated by ABA and C_2H_2 TFs, helping to maintain the osmotic balance (Thirunavukkarasu et al., 2017; Khan et al., 2019). LEA proteins were induced in vegetative tissues by drought stress. Several LEA proteins showed increased abundance both at the V12 and VT stages in response to drought stress (**Table 1**). The relation of this protein with drought stress has been previously reported in wheat (Li et al., 2012) and faba bean (Khan et al., 2019). Moreover, the over-expression of the barley LEA HVA1 gene enhanced transgenic wheat biomass productivity and water use efficiency under water-limited conditions (Sivamani et al., 2000). Taken collectively, our results reveal that stress defense proteins, interacting with other proteins in complex networks, are essential for drought-tolerant ND476 maize plants' survival under water-deficit conditions.

The Gene Co-expression Network Analysis Offered an Essential Resource for Mining Novel Genes Related to Water-Deficit Stress Conditions

Given that the expression of a great number of DEGs was affected by drought treatment, WGCNA was used to construct a gene co-expression network to mine the major genes and dig out the key modules involved in the maize responses to drought stress from the vegetative to reproductive stages. In this study, a total of 12 modules were identified based on gene expression patterns, and several modules showed functional specificity in various stages, as genes were regulated dynamically under water-limited conditions (**Figure 4**). Then for investigating these DEGs' functional biological roles, BiNGO software was used (**Supplementary Table 5**). The functions of DEGs with known biological functions could be predicted according to their module, and this analysis found a series of biological processes that were affected by drought stress. By comparing our results with previous studies, similar biological processes were identified in plants' response to drought stress. Therefore, the gene co-expression network analysis provides an essential resource for mining novel genes related to water-deficit stress acclimation of maize. Particularly, the hub-genes are suggested to be the key players in maize drought stress response. Further downstream analysis studies will be essential in determining each of these hub genes' exact contribution to drought stress tolerance in maize.

Proposed Molecular Model for Maize Drought Stress Tolerance

According to our main findings of the drought-responsive DEGs and their related networks, in combination with the related relevant literature, we herein propose a molecular model for drought stress tolerance in maize at four different growth stages as shown in **Figure 8**.

CONCLUSION

In the present study, we performed a comprehensive comparative leaf transcriptome analysis of the drought-tolerant maize hybrid ND476 plants subjected to water-sufficient (control) and water-deficit treatment conditions at four different growth stages. Based on the transcriptome analysis, a total of 3,451 DEGs were identified from the four experimental comparisons, and changes in these genes affected corresponding metabolic pathway responses related to drought tolerance in maize. Subsequently, 3,451 DEGs were divided into 12 modules by the WGCNA analysis. Our results showed that maize drought stress adaptation is a stage-specific response process. Whereas DEGs related to stress signal transduction, detoxification, transcription factor regulation, hormone signaling and secondary metabolites biosynthesis were universal across the four crop growth stages, those associated with photosynthesis and amino acid metabolism, protein degradation, transport, and RNA transcriptional regulation were uniquely enriched at the V12, VT, R2, and R4 stages, respectively. Our findings may help in clarifying the important growth-stage-specific molecular mechanisms regulating maize drought stress responses. Further, the key genes and metabolic pathways identified here may serve as valuable genetic resources or selection targets for genetic engineering of drought-resistant maize cultivars.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

SL, TZ, NW, and HD conceived and designed the experiments. SL, TZ, HD, AD, and YY performed the experiments. SL, TZ, AD, and YY analyzed the data. SL and TZ wrote the manuscript. All authors have read and approved the final manuscript.

FUNDING

This research was supported by the National Key Research and Development Program of China (2018YFD0300501).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.645443/full#supplementary-material

REFERENCES

Aslam, M., Maqbool, M. A., and Cengiz, R. (2015). *Drought stress in maize (Zea mays L.): Effects, resistance mechanisms, global achievements and biological strategies for improvement. SpringerBriefs in Agriculture*. Cham: Springer.

Bazargani, M. M., Sarhadi, E., Bushehri, A. A., Matros, A., Mock, H. P., Naghavi, M. R., et al. (2011). A proteomics view on the role of drought-induced senescence and oxidative stress defense in enhanced stem reserves remobilization in wheat. *J. Proteomics.* 74, 1450–1462. doi: 101016/jjprot201105015

Bhanu, B. D., Ulaganathan, K., Shanker, A. R., and Desai, S. (2016). RNA-seq analysis of irrigated vs water stressed transcriptomes of Zea mays cultivar Z59. *Front. Plant Sci*. 7:239. doi: 103389/fpls201600239s

Divi, U. K., Rahman, T., and Krishna, P. (2010). Berseaarschs airnticolesteroid-mediated stress tolerance in Arabidopsis shows interactions with abscisic acid, ethylene and salicylic acid pathways. *BMC Plant Biol.* 10:151. doi: 101186/1471-2229-10-151

Farooq, M., Wahid, A., Kobayashi, N., Fujita, D., and Basra, S. M. A. (2009). Plant drought stress: effects, mechanisms and management. *Agron. Sustain. Dev*. 29, 185–212.

Fracasso, A., Trindade, L. M., and Amaducci, S. (2016). Drought stress tolerance strategies revealed by RNA-Seq in two sorghum genotypes with contrasting WUE. *BMC Plant Biol.* 16:115. doi: 10,1186/s12870-016-0800-x

Gharechahi, J., Hajirezaei, M. R., and Salekdeh, G. H. (2015). Comparative proteomic analysis of tobacco expressing cyanobacterial flavodoxin and its wild type under drought stress. *J. Plant Physiol*. 175, 48–58. doi: 101016/jjplph201411001

Gong, F., Yang, L., Tai, F., Hu, X., and Wang, W. (2014). "Omics" of maize stress response for sustainable food production: opportunities and challenges. *OMICS* 18, 714–732. doi: 10.1089/omi.2014.0125

Guo, L., Zi, Y. W., Lin, H., Wei, E. C., Chen, J., Liu, M., et al. (2006). Expression and functional analysis of the rice plasma-membrane intrinsic protein gene family. *Cell Res.* 16, 277–286. doi: 101038/sjcr7310035

Hanin, M., Brini, F., Ebel, C., Toda, Y., Takeda, S., and Masmoudi, K. (2011). Plant dehydrins and stress tolerance versatile proteins for complex mechanisms. *Plant Signal. Behav*. 6, 1503–1509. doi: 10.4161/psb.6.10.17088

Hu, H., and Xiong, L. (2014). Genetic engineering and breeding of drought-resistant crops. *Annu. Rev. Plant Biol.* 65, 715–741. doi: 10;1146/annurev-arplant-050213-040000

Jiang, M., and Zhang, J. (2002). Water stress-induced abscisic acid accumulation triggers the increased generation of reactive oxygen species and up-regulates the activities of antioxidant enzymes in maize leaves. *J. Exp. Bot*. 53, 2401–2410. doi: 101093/jxb/erf090

Jing, D., Zhang, J., Xia, Y., Kong, L., Ou, F., Zhang, S., et al. (2017). Proteomic analysis of stress-related proteins and metabolic pathways in Picea asperata somatic embryos during partial desiccation. *Plant Biotechnol. J.* 15, 27–38. doi: 101111/pbi12588

Joshi, R., Wani, S. H., Singh, B., Bohra, A., Dar, Z. A., Lone, A. A., et al. (2016). Transcription factors and plants response to drought stress: current understanding and future directions. *Front Plant Sci.* 7:1029. doi: 10.3389/fpls.2016.01029

Khan, M. A., Alghamdi, S. S., Ammar, M. H., Sun, Q., Teng, F., Migdadi, H. M., et al. (2019). Transcriptome profiling of faba bean (Vicia faba L.) drought-tolerant variety hassawi-2 under drought stress using RNA sequencing. *Electron. J. Biotechnol.* 39, 15–29. doi: 10.1016/j.ejbt.2019.02.004

Khan, M. N., and Komatsu, S. (2016). Proteomic analysis of soybean root including hypocotyl during recovery from drought stress. *J. Proteomics.* 144, 39–50. doi: 101016/jjprot201606006

Kim, S. T., Yu, S., Kang, Y. H., Kim, S. G., Kim, J. Y., Kim, S. H., et al. (2008). The rice pathogen-related protein 10 (JIOsPR10) is induced by abiotic and biotic stresses and exhibits ribonuclease activity. *Plant Cell Rep*. 27, 593–603. doi: 101007/s00299-007-0485-6

Król, A., and Weidner, S. (2017). Changes in the proteome of grapevine leaves (*Vitis vinifera* L) during long-term drought stress. *J. Plant Physiol*. 211, 114–126. doi: 101016/jjplph201611016

Lamaoui, M., Jemo, M., Datla, R., and Bekkaoui, F. (2018). Heat and drought stresses in crops and approaches for their mitigation. *Front. Chem.* 6:26. doi: 10.3389/fchem.2018.00026

Lei, L., Shi, J., Chen, J., Zhang, M., Sun, S., Xie, S., et al. (2015). Ribosome profiling reveals dynamic translational landscape in maize seedlings under drought stress. *Plant J.* 84, 1206–1218. doi: 10.1111/tpj.13073

Li, C., Dong, J., Zhang, X., Zhong, H., Jia, H., Fang, Z., et al. (2019). Gene expression profiling of Bothriochloa ischaemum leaves and roots under drought stress. *Gene* 691, 77–86. doi: 101016/jgene201812038

Li, Y. C., Meng, F. R., Zhang, C. Y., Zhang, N., Sun, M. S., Ren, J. P., et al. (2012). Comparative analysis of water stress-responsive transcriptomes in drought-susceptible and -tolerant wheat (*Triticum aestivum* L). *J. Plant Biol.* 55, 349–360. doi: 101007/s12374-011-0032-4

Liu, J. Y., Anne-Osbourn, A., and Ma, P. D. (2015). MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol. Plant.* 8, 689–708. doi: 101016/jmolp201503012

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2-\Delta\Delta CT$ method. *Methods* 2, 402–408. doi: 10.1006/meth.2001.1262

Ma, X., Xia, H., Liu, Y., Wei, H., Zheng, X., Song, C., et al. (2016). Transcriptomic and metabolomic studies disclose key metabolism pathways contributing to well-maintained photosynthesis under the drought and the consequent drought-tolerance in rice. *Front. Plant Sci*. 21:1886. doi: 10,3389/fpls,2016,01886, eCollection 2016

Mahajan, S., and Tuteja, N. (2005). Cold salinity and drought stresses: An overview. *Arch. Biochem. Biophys.* 444, 139–158. doi: 10,1016/j,abb,2005,10,018

Miao, Z. Y., Han, Z. X., Zhang, T., Chen, S. Y., and Ma, C. (2017). A systems approach to a spatiotemporal understanding of the drought stress response in maize. *Sci Rep*. 7:6590. doi: 10.1038/s41598-017-06929-y

Min, H. W., Chen, C. X., Wei, S. W., Shang, X. L., Sun, M. Y., Xia, R., et al. (2016). Identification of drought tolerant mechanisms in maize seedlings based on transcriptome analysis of recombination inbred lines. *Front. Plant Sci*. 7:1080. doi: 103389/fpls201601080

Mohammadi, M., Kav, N. N. V., and Deyholos, M. K. (2007). Transcriptional profiling of hexaploid wheat (*Triticum aestivum* L.) roots identifies novel, dehydration-responsive genes. *Plant Cell Environ.* 30, 630–645. doi: 10.1111/j1365 -3040.2007.01645.x

Pinheiro, C., and Chaves, M. M. (2011). Photosynthesis and drought: can we make metabolic connections from available data? *J. Exp. Bot*. 62, 869–882. doi: 10.1093/jxb/erq340

Sheoran, S., Thakur, V., Narwal, S., Turan, R., Mamrutha, H. M., Singh, V., et al. (2015). Differential activity and expression profile of antioxidant enzymes and physiological changes in wheat (*Triticum aestivum* L) under drought. *Appl. Biochem. Biotechnol*. 177, 1282–1298. doi: 101007/s12010-015-1813-x

Shi, W., Cheng, J., and Wen, X. (2018). Transcriptomic studies reveal a key metabolic pathway contributing to a well-maintained photosynthetic system under drought stress in foxtail millet (*Setaria italica* L). *Peer J.* 6:e4752. doi: 107717/peerj4752

Singh, D., and Laxmi, A. (2015). Transcriptional regulation of drought response: a tortuous network of transcriptional factors. *Front. Plant Sci.* 6:895. doi: 10.3389/fpls.2015.00895

Sivamani, E., Bahieldin, A., Wraith, J. M., Al-Niemi, T., Dyer, W. E., Ho, T. H. D., et al. (2000). Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley HVA1 gene. *Plant Sci.* 155, 1–9. doi: 101016/S0168-9452(99)00247-2

Song, K., Kim, H. C., Shin, S., Kim, K. H., Moon, J. C., Kim, J. Y., et al. (2017). Transcriptome analysis of flowering time genes under drought stress in maize leaves. *Front. Plant Sci*. 8:267. doi: 103389/fpls201700267

Thirunavukkarasu, N., Sharma, R., Singh, N., Shiriga, K., Mohan, S., Mittal, S., et al. (2017). Genomewide Expression and Functional Interactions of Genes under Drought Stress in Maize. *Int. J. Genomics.* 2017:2568706. doi: 10.1155/2017/2568706

Tran, L. S., Nakashima, K., Sakuma, Y., Simpson, S. D., Fujita, Y., Maruyama, K., et al. (2004). Isolation and functional analysis of arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress promoter. *Plant Cell.* 16, 2481–2498. doi: 101105/tpc104022699

Urban, M. O., Vasek, J., Klima, M., Krtkova, J., Kosova, K., Prasil, I. T., et al. (2017). Proteomic and physiological approach reveals drought-induced changes in rapeseeds: Water-saver and water-spender strategy. *J. Proteomics.* 152, 188–205. doi: 101016/jjprot201611004

Wang, B., Liu, C., Zhang, D., He, C., Zhang, J., and Li, Z. (2019). Effects of maize organ-specific drought stress response on yields from transcriptome analysis. *BMC Plant Biol.* 19:335. doi: 10.1186/s12870-019-1941-5

Wang, N., Liu, B., Liang, X., Zhou, Y., Song, J., Yang, J., et al. (2019). Genome-wide association study and genomic prediction analyses of drought stress tolerance in china in a collection of off-PVP maize inbred lines. *Mol. Breed.* 39:113.

Wang, H., Wang, H., Shao, H., and Tang, X. (2016). Recent advances in utilizing transcription factors to improve plant abiotic stress tolerance by transgenic technology. *Front. Plant Sci.* 7:67. doi: 103389/fpls201600067

Wang, N., Li, L., Gao, W. W., Wu, Y. B., Yong, H. J., and Weng, J. F. (2018). Transcriptomes of early developing tassels under drought stress reveal differential expression of genes related to drought tolerance in maize. *J. Integr. Agric.* 17, 1276–1288. doi: 10.1016/s2095-3119(17)61777-5

Wang, W., Vinocur, B., Shoseyov, O., and Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperons in the abiotic stress response. *Trends Plant Sci.* 9, 244–252. doi: 10.1016/j.tplants.2004.03.006

Wheeler, T., and von Braun, J. (2013). Climate change impacts on global food security. *Science* 341, 508–513. doi: 10,1126/science,1239402

Wu, J., Jiang, Y., and Liang, Y. (2019). Expression of the maize MYB transcription factor ZmMYB3R enhances drought and salt stress tolerance in transgenic plants. *Plant Physiol. Biochem.* 137, 179–188. doi: 101016/jplaphy201902010

Xie, H., Yang, D., Yao, H., Bai, G., Zhang, H., and Xiao, B. G. (2016). iTRAQ-based quantitative proteomic analysis reveals proteomic changes in leaves of cultivated tobacco (Nicotiana tabacum) in response to drought stress. *Biochem. Biophys. Res. Commun.* 469, 768–775. doi: 101016/jbbrc201511133

Xu, C. P., and Huang, B. (2010). Comparative analysis of drought responsive proteins in kentucky bluegrass cultivars contrasting in drought tolerance. *Crop Sci.* 50, 2543–2552. doi: 102135/cropsci2010030152

Yan, H., Jia, H., Chen, X., Hao, L., An, H., and Guo, X. (2014). The cotton WRKY transcription factor GhWRKY17 functions in drought and salt stress in transgenic Nicotiana benthamiana through ABA signalling and the modulation of reactive oxygen species production. *Plant Cell Physiol.* 55, 2060–2076. doi: 10.1093/pcp/pcu133

Zenda, T., Liu, S., Wang, X., Liu, G., Jin, H., Dong, A., et al. (2019). Key maize drought-responsive genes and pathways revealed by comparative transcriptome and physiological analyses of contrasting inbred lines. *Int. J. Mol. Sci.* 20:1268. doi: 10.3390/ijms20061268

Zhang, J. Y., Cruz de Carvalho, M. H., Torres-Jerez, I., Kang, Y., Allen, S. N., Huhman, D. V., et al. (2014). Global reprogramming of transcription and metabolism in Medicago truncatula during progressive drought and after rewatering. *Plant Cell Environ.* 7, 2553–2576. doi: 101111/pce12328

Zhang, X., Lei, L., Lai, J., Zhao, H., and Song, W. (2018). Effects of drought stress and water recovery on physiological responses and gene expression in maize seedlings. *BMC Plant Biol.* 18:68. doi: 10.1186/s12870-018-1281-x

Zhang, Y., Tang, W., Wang, L., Hu, Y., Liu, X., and Liu, Y. (2019). Kiwifruit (Actinidia chinensis) R1R2R3-MYB transcription factor AcMYB3R enhances drought and salinity tolerance in Arabidopsis thaliana. *J. Integr. Agric.* 18, 417–427. doi: 101016/S2095-3119(18)62127-6

Zheng, J., Fu, J., Gou, M., Huai, J., Liu, Y., Jian, M., et al. (2010). Genome-wide transcriptome analysis of two maize inbred lines under drought stress. *Plant Mol. Biol.* 72, 407–421. doi: 10.1007/s11103-009-9579-6

Zhou, R., Kan, X., Chen, J., Hua, H., Li, Y., Ren, J., et al. (2019). Drought-induced changes in photosynthetic electron transport in maize probed by prompt fluorescence, delayed fluorescence, P700 and cyclic electron flow signals. *Environ. Exp. Bot.* 158, 51–62. doi: 10.1016/j.envexpbot.2018.11.005

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 February 2021
doi: 10.3389/fgene.2021.625985

Characterizing the Leaf Transcriptome of *Chrysanthemum rhombifolium* (Ling et C. Shih), a Drought Resistant, Endemic Plant From China

Wenjie Zhang[1], *Hongyuan Xu*[1], *Xiaxia Duan*[1], *Jing Hu*[1], *Jingjing Li*[1], *Liang Zhao*[2] *and Yueping Ma*[1]*

[1] College of Life and Health Sciences, Northeastern University, Shenyang, China, [2] College of Life Sciences, Northwest A&F University, Yangling, China

OPEN ACCESS

Edited by:
Jian Ma,
Sichuan Agricultural University, China

Reviewed by:
Kaimei Zhang,
Nanjing Forestry University, China
Ahsan Habib,
Khulna University, Bangladesh
Zhumei Ren,
Shanxi University, China

***Correspondence:**
Yueping Ma
mypluna@sina.com

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *04 November 2020*
Accepted: *07 January 2021*
Published: *11 February 2021*

Citation:
Zhang W, Xu H, Duan X, Hu J, Li J, Zhao L and Ma Y (2021) Characterizing the Leaf Transcriptome of Chrysanthemum rhombifolium (Ling et C. Shih), a Drought Resistant, Endemic Plant From China. Front. Genet. 12:625985. doi: 10.3389/fgene.2021.625985

Chrysanthemum rhombifolium (Ling et C. Shih), an endemic plant that is extremely well-adapted to harsh environments. However, little is known about its molecular biology of the plant's resistant traits against stress, or even its molecular biology of overall plant. To investigate the molecular biology of *C. rhombifolium* and mechanism of stress adaptation, we performed transcriptome sequencing of its leaves using an Illumina platform. A total of 130,891 unigenes were obtained, and 97,496 (∼74.5%) unigenes were annotated in the public protein database. The similarity search indicated that 40,878 and 74,084 unigenes showed significant similarities to known proteins from NCBI non-redundant and Swissprot protein databases, respectively. Of these, 56,213 and 42,005 unigenes were assigned to the Gene Ontology (GO) database and Cluster of Orthologous Groups (COG), respectively, and 38,918 unigenes were mapped into five main categories, including 18 KEGG pathways. Metabolism was the largest category (23,128, 59.4%) among the main KEGG categories, suggesting active metabolic processes in *C. rhombifolium*. About 2,459 unigenes were annotated to have a role in defense mechanism or stress tolerance. Transcriptome analysis of *C. rhombifolium* revealed the presence of 12,925 microsatellites in 10,524 unigenes and mono, trip, and dinucleotides having higher polymorphism rates. The phylogenetic analysis based on *GME* gene among related species confirmed the reliability of the transcriptomic data. This work is the first genetic study of *C. rhombifolium* as a new plant resource of stress-tolerant genes. This large number of transcriptome sequences enabled us to comprehensively understand the basic genetics of *C. rhombifolium* and discover novel genes that will be helpful in the molecular improvement of chrysanthemums.

Keywords: Asteraceae, stress tolerance, ornamental plant, RNA-seq, SSR

INTRODUCTION

Chrysanthemum (*Chrysanthemum morifolium* (Ramat.)Tzvel.; Asteraceae) is among the most popular flowers in China, and the most important cut flowers in the world, having a great ornamental and economical value (Song et al., 2018; Su et al., 2019). However, the long-term artificial domestication of chrysanthemums often causes declines in their resistance to environmental stressors and adaptability (Da Silva, 2003; Chen et al., 2011, 2012; Song et al., 2014), thereby limiting their use in landscaping and industrial production. Therefore, the development of Chrysanthemum cultivars with increased environmental tolerance has always been a goal of breeders (Su et al., 2019). Many stress resistance traits, and corresponding stress resistance gene resources identified in the wild chrysanthemum species (Zhao et al., 2009; Lu et al., 2010; Li et al., 2013), have a great significance for the genetic improvement of chrysanthemum cultivars.

RNA sequencing (RNA-Seq) is a powerful tool for quantifying and analyzing different types of RNA molecules using deep-sequencing technologies (Wang et al., 2009). It provides us large-scale transcript data with high throughput, accuracy, sensitivity and reproducibility which enabled us to generate an unprecedented global view of the transcriptome of the species (Angeloni et al., 2011; Jain, 2011). RNA-seq has been widely used in plants, especially for some non-model species and some large and complex genomes, greatly accelerating the discovery of novel genes, understanding the complex tissue-specific expression patterns, and regulation networks in higher plants (Li and Dewey, 2011; Wang et al., 2014, 2017; Wu et al., 2016).

Chrysanthemum rhombifolium Ling et Shih is a perennial herb endemic to Wushan, Chongqing in China (Shih and Fu, 1983; Bremer and Humphries, 1993) and has a high ornamental value. It has diamond-shaped leaves with dense abaxial pubescence and semi-lignified stems and branches (**Figure 1**). The species is well-adapted to environments characterized by high temperatures, low soil fertility, and drought (Zhao et al., 2009, 2010). However, few studies performed on *C. rhombifolium* except using as a sample in molecular phylogeny of *Chrysanthemum* (Masuda et al., 2009; Zhao et al., 2010; Li et al., 2014; Ma et al., 2020) or in geographical distribution of *Chrysanthemum* (Zhao et al., 2009; Li et al., 2013). Here, little is known about its molecular biology of overall plant or the plant's resistant traits against stress. This prompted us to characterize its leaf transcriptome using high-throughput RNA sequencing and *de novo* assembly to provide a comprehensive resource for understanding the biology of *C. rhombifolium* in general, and gain insights in improving the breeding of chrysanthemums and other related crops.

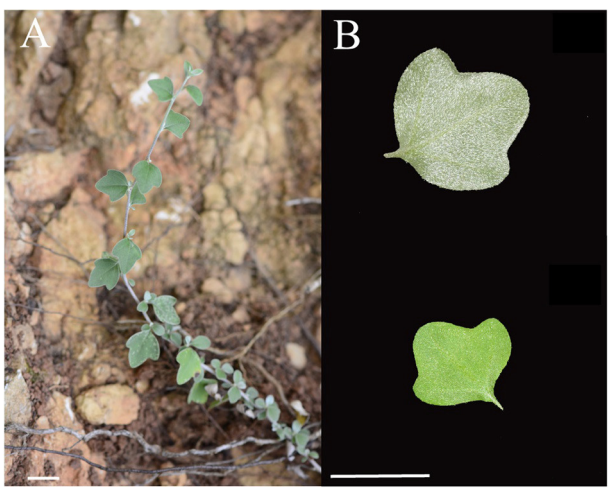

FIGURE 1 | The plant of *C. rhombifolium*. **(A)** The habitat; **(B)** leaf, up, abaxially, down, adaxially. Scale bars = 2 cm.

MATERIALS AND METHODS

Plant Materials

We collected *C. rhombifolium* plants from Wushan of Chongqing in China and planted them in the Nurse Garden of the Northeastern University, China. Fresh, mature leaves were washed with sterile water, immediately frozen in liquid nitrogen, and stored at −80°C.

RNA Isolation and cDNA Library Construction

Total RNA was isolated from the leaves using TRIzol reagent (Invitrogen™ Life Technologies, CA, USA) following the manufacturer's instructions. The RNA quality was assessed using formaldehyde denaturing gel electrophoresis (28S:18S>2), a NanoPhotometer®spectrophotometer (IMPLEN, CA, USA), and RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). For RNA-Seq analysis, three biological replicates were used. Sequencing libraries were generated with 1 μg RNA sample using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following the manufacturer's recommendations. The mRNA was purified from total RNA using beads with Oligo (dT), and cut into short fragments with fragmentation buffer. First-strand cDNA was synthesized using random hexamer primers and M-MuLV Reverse Transcriptase (NEB, USA), and second-strand cDNA was synthesized using buffer, dNTPs, RNase H, and DNA polymerase I. The remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of the 3′ ends of DNA fragments, NEBNext Adaptor with hairpin loop structure was ligated to prepare for hybridization. cDNA fragments, preferentially 250–300 bp in length, were selected by purifying the library fragments with AMPure XP system (Beckman Coulter, Beverly, USA). The size-selected, adaptor-ligated cDNA fragments were incubated with 3 μl USER Enzyme (NEB, USA) at 37°C for 15 min, followed by 5 min at 95°C before PCR. PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) Primer. The PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system.

Sequencing and *de novo* Assembly

We sequenced the transcriptome library using the Illumina HiSeq 2500 platform, and generated paired-end reads. We filtered the raw data using the Filterfq program (BGI, Shenzhen, China) to remove adaptor sequences, reads in which unknown nucleotides (N) were >5%, and low-quality sequences with >20% low-quality bases (quality value ≦10) to generate clean data. The raw data were deposited in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) with the Bioproject accession: PRJNA674029 and BioSample accessions:SAMN16633381- SAMN16633383.

We then used the Trinity software (v2.8.0; http://trinityrnaseq.sourceforge.net/) with default settings for *de novo* transcriptome assembly (Grabherr et al., 2011). Two contigs thus obtained were connected into a single scaffold to generate unigenes. These unigenes were further spliced to generate longer complete consensus sequences and to remove redundant sequences with TGICL (v 2.1; http://www.tigr.org/tdb/tgi/) (Pertea et al., 2003).

Functional Annotation and Classification of Unigenes

We annotated the obtained unigenes using the NCBI Nr (non-redundant protein database), NCBI Nt (non-redundant nucleotide sequences), Swiss-Prot, Gene ontology terms (GO), and Protein family (Pfam) using BLAST 2 with an E-value cut-off of 10^{-5} to obtain information on protein function annotation. We also performed functional annotation using Clusters of Orthologous Groups of proteins (KOG/COG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases to classify possible COG functions and KEGG pathways and predict possible functional classifications and molecular pathways, respectively (Conesa et al., 2005; Ye et al., 2006; Kanehisa et al., 2008).

Phylogenetic Analysis

GDP-D-mannose 3′, 5′- epimerase (GME), regulates cell wall biosynthesis and ascorbate accumulation, playing an important role in plant development and abiotic stress tolerance (Tao et al., 2018). We extracted annotated GME unigene sequences, aligned them with other GME homologs retrieved from Genbank (http://www.ncbi.nlm.nih.gov/entyez/query.fcgi). Multiple alignments were made using MUSCLE (Edgar, 2004) in Geneious v.8.1.2 (http://www.geneious.com/; Kearse et al., 2012) and adjusted manually. We contructed the phylogenetic tree by the neighbor-joining (NJ) method with 1,000 bootstrap replicates using MEGA 7 (Kumar et al., 2016).

Simple Sequence Repeats (SSRs) Prediction

We predicted SSR regions among all the assembled unigenes using MIcroSAtellite (MISA, http://pgrc.ipk-gatersleben.de/misa/; Zalapa et al., 2012). We detected the SSR motifs of mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides with a minimum of twelve, six, five, five, four, and four repeats, respectively. For other parameters, default settings were used.

TABLE 1 | Summary of sequence assembling of *C. rhombifolium* transcriptome.

Category		Transcripts	Unigene
Number	300–500 bp	102,216	59,296
	500-1 Kbp	74,883	41,353
	1 K-2 Kbp	53,959	21,862
	>2 Kbp	23,795	8,380
Total number		254,853	130,891
Mean length		921	807
N50		1,301	1,034
Total nucleotides		234,824,746	105,639,077

TABLE 2 | Summary of functional annotation of assembled unigenes in *C. rhombifolium*.

Database	Number of unigenes annotated	Percentage (%)
Annotated in NR	40,878	31.23
Annotated in NT	55,831	42.65
Annotated in KO	37,488	28.64
Annotated in SwissProt	74,084	56.59
Annotated in PFAM	56,213	42.94
Annotated in GO	56,213	42.94
Annotated in KOG	42,005	32.09
Annotated in all databases	7,286	5.56
Annotated in at least one database	97,496	74.48
Total unigenes	130,891	100

RESULTS AND DISCUSSION

Illumina Paired-End Sequencing and Assembly

The 151,037,024 raw sequencing reads obtained from the Illumina sequencing were cleaned by removing low-quality

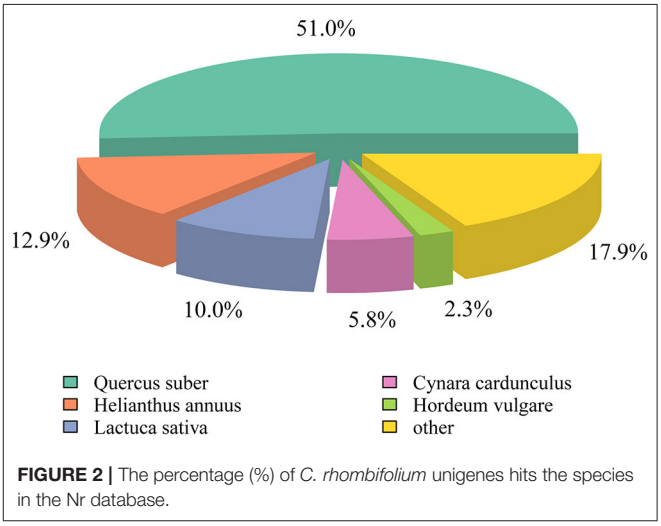

FIGURE 2 | The percentage (%) of *C. rhombifolium* unigenes hits the species in the Nr database.

data and adaptors, yielding 147,842,128 clean reads with Q20 bases at 97.9%, and a GC content of 44.44%. Using the overlapping information of high-quality reads from the Trinity software, 254,853 transcripts with an average length of 921 bp and N50 of 1,301 bp were generated. After that 130,891 unigenes with an average length of 807 bp and N50 of 1,034 bp were obtained (**Table 1**). The number and average length of the unigenes we obtained was larger and longer than the transcriptomes of the related species, *C. nankingense* (45,789) and *C. lavandulifolium* (108,737) (Wang et al., 2013, 2014), indicating the high quality of sequencing.

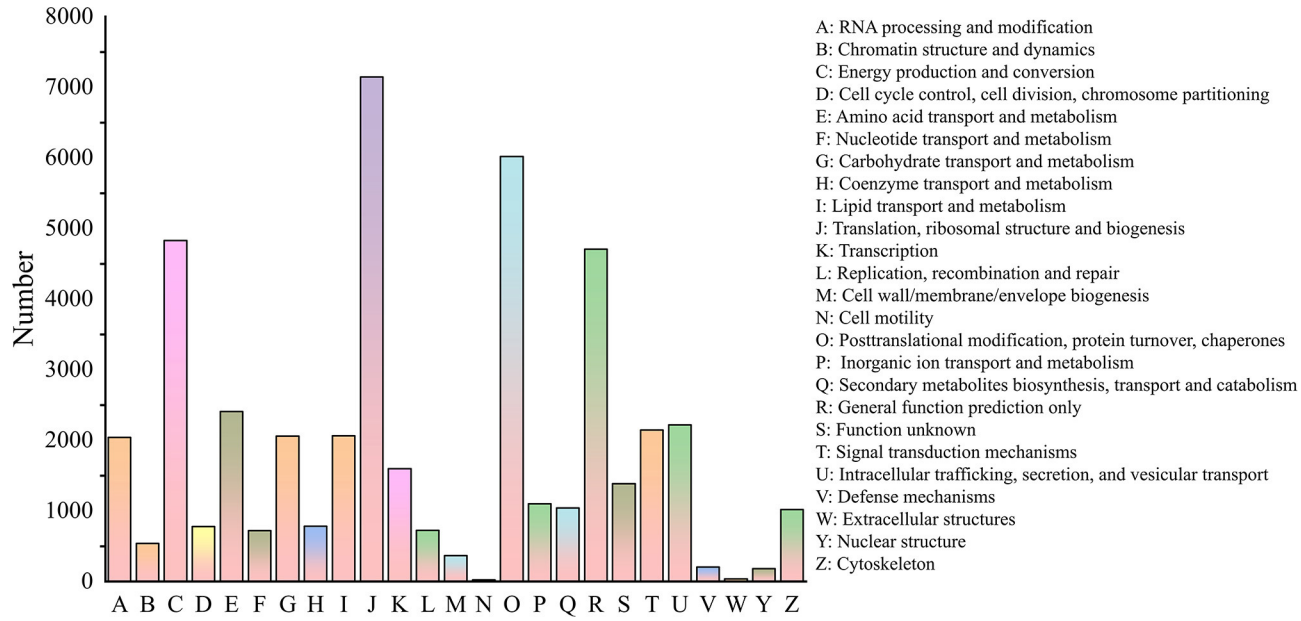

FIGURE 3 | COG classifications of unigenes in the transcriptome of *C. rhombifolium* unigenes.

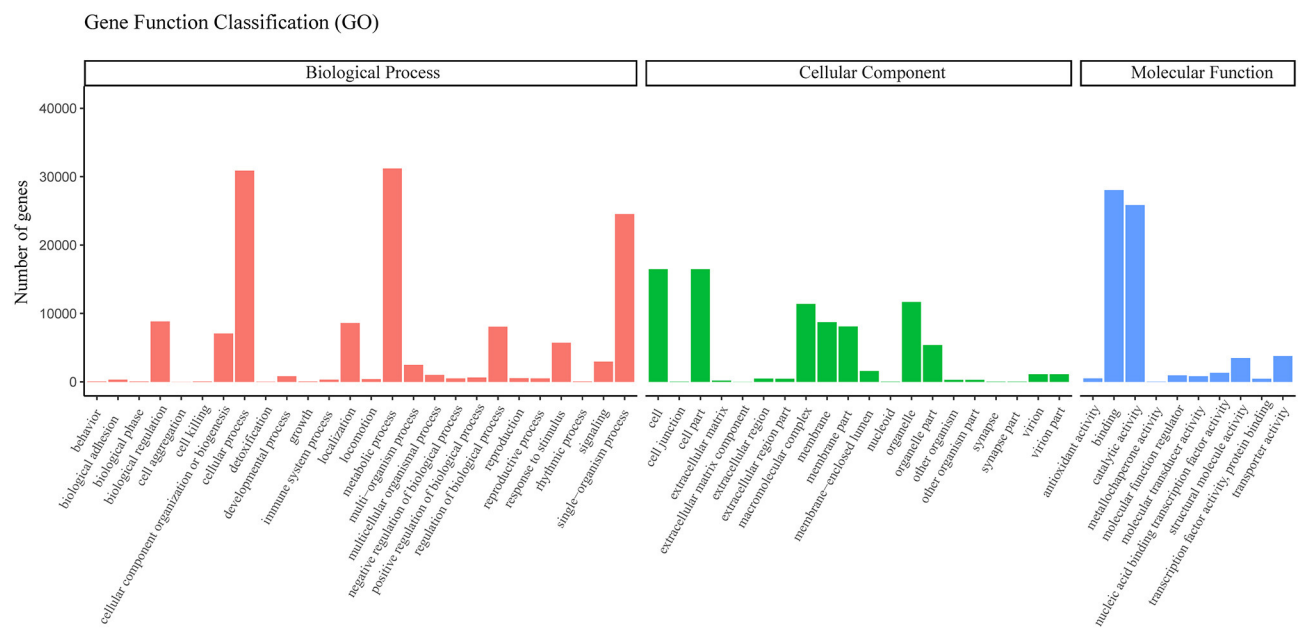

FIGURE 4 | Functional classification of *C. rhombifolium* unigenes based on Gene Ontology (GO) categorization. The results are summarized in three main GO categories: biological process (shown in red color), cellular component (shown in green color) and molecular function (shown in blue color). The x-axis indicates the subcategories, and the y-axis indicates the numbers related to the total number of GO terms present.

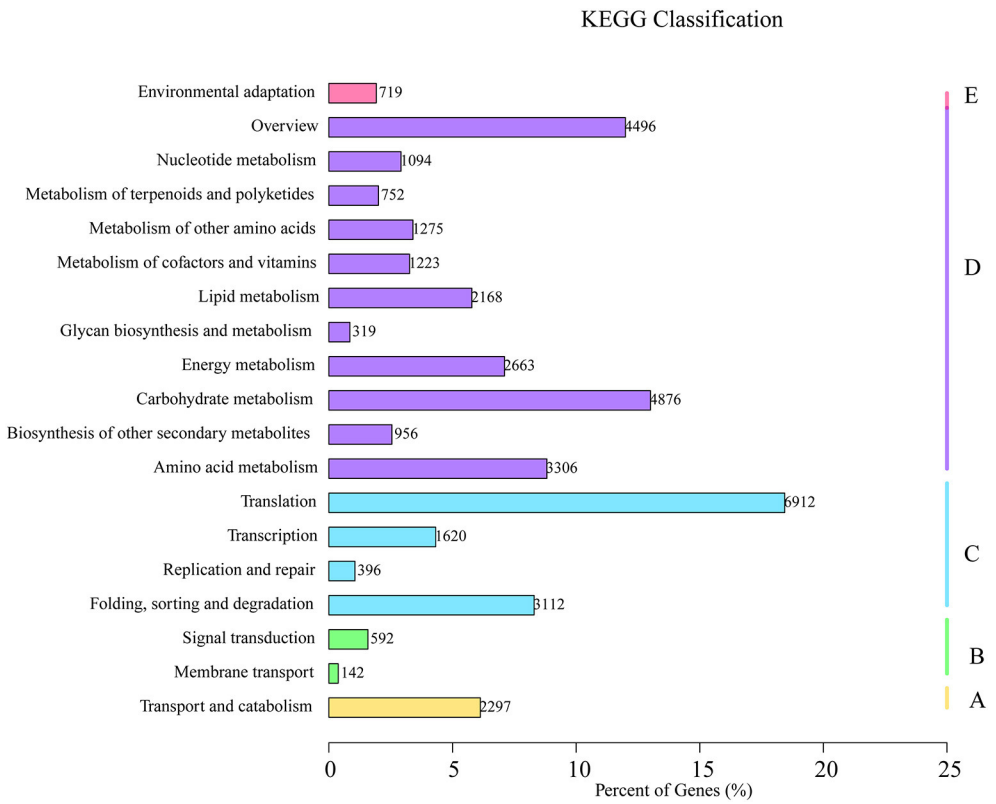

FIGURE 5 | Pathway assignment based on the Kyoto Encyclopedia of Genes and Genomes (KEGG). **(A)** Classification based on cellular process categories, **(B)** classification based on environmental information processing categories, **(C)** classification based on genetic information processing categories, **(D)** classification based on metabolism categories, and **(E)** classification based on organismal systems categories.

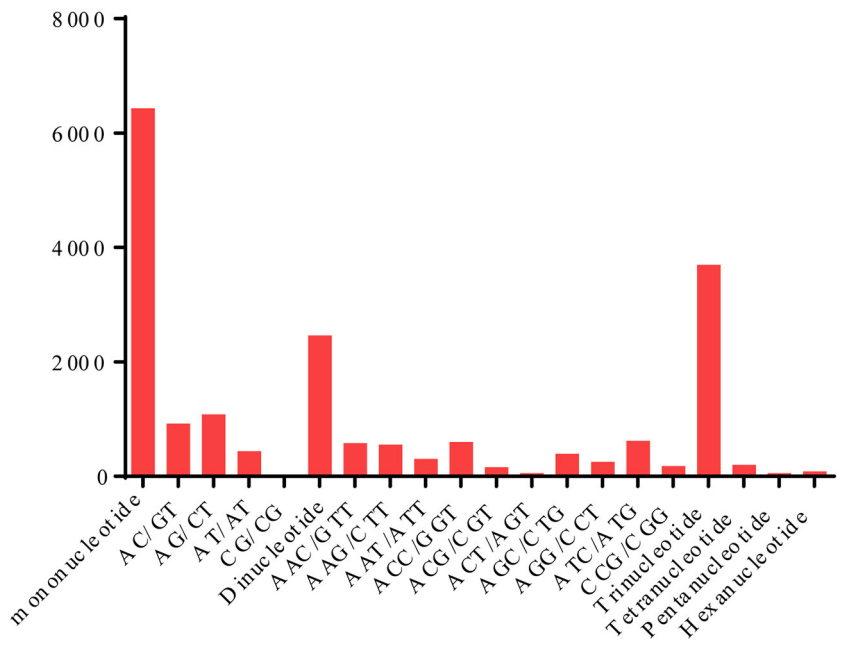

FIGURE 6 | Simple sequence repeat (SSR) types present in the *C. rhombifolium* transcriptome.

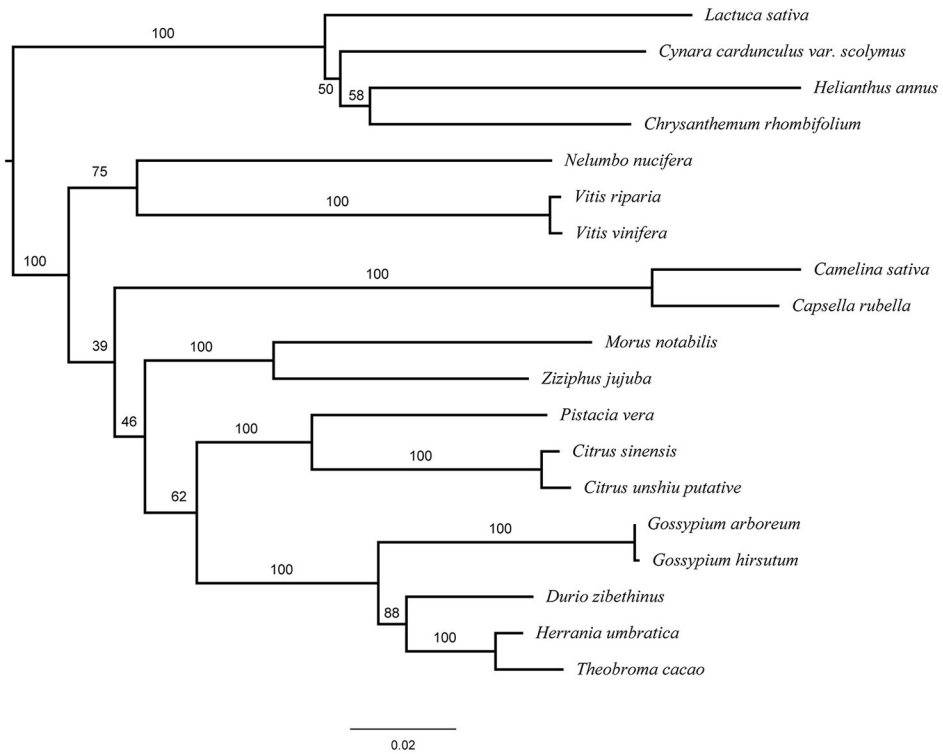

FIGURE 7 | Phylogenetic analysis of the GME in plants. *Helianthus annuus* XM_022174384.1; *Lactuca sativa* (XM_023912155.1); *Cynara cardunculus* var. scolymus (XM_025140401.1), *Chrysanthemum rhombifolium* (R|Cluster-20886.25690); *Camelina sativa* (XM_010456910.1); *Capsella rubella* (XM_006287892.2); *Nelumbo nucifera* (XM_010275940.2); *Morus notabilis* (XM_024169380.1); *Vitis riparia* (XM_034830226.1); *Vitis vinifera* (XM_002283862.4); *Ziziphus jujuba* (XM_016019918.2); *Gossypium arboretum* (XM_017764784.1); *Gossypium hirsutum* (XM_016869694.1); *Durio zibethinus* (XM_022891507.1); PREDICTED: *Herrania umbratica* (XM_021427500.1); *Theobroma cacao* (XM_018122066.1); *Pistacia vera* (XM_031412928.1); *Citrus sinensis* (XM_006471610.2); *Citrus unshiu* (HQ224947.1).

Functional Annotation and Classification of All Non-redundant Unigenes

We used the Nr, Nt, Pfam, KOG, Swiss-prot, GO, and KEGG databases to annotate the assembled unigenes. Among the 130,891 unigenes obtained, at least 97,496 unigenes (74.48%) could be annotated with the searched databases −40,878 (Nr), 55,831 (Nt), 37,488 (KO), 74,084 (Swiss-prot), 56,213 (Pfam), 56,213 (GO), and 42,005 (KOG/COG), suggesting that this project generated a substantial fraction of the expressed genes in this study (**Table 2**). The unigenes annotated with the Nr database mainly comprised *Quercus suber* L. (51%), *Helianthus annuus* L. (12.9%), *Lactuca sativa* Linn. (10%), *Cynara cardunculus* L., Sp. Pl. (5.8%), and *Hordeum vulgare* Linn. (2.3%) sequences (**Figure 2**). The highest similarity observed was with *Q. suber*, a species with resistance to wind, drought, and barren environments (Pereira-Leal et al., 2014).

The COG analysis enabled the functional classification of 42,005 unigenes (**Figure 3**). The most frequently identified classes were "Translation, ribosomal structure and biogenesis" (7,147; 17%), followed by "Posttranslational modification, protein turnover, chaperones" (6,016; 14.3%), "Energy production and conversion" (4,830; 11.5 %), "General function prediction only" (4,706; 11.2%), "Amino acid transport and dynamics" (2407; 5.7%), "Intracellular trafficking, secretion, and vesicular transport (2218; 5.3%), "signal transduction" (2,145; 5.1%), "Lipid transport and metabolism" (2065, 4.9%) and "Carbohydrate transport and metabolism (2057; 4.9%). The least frequently identified groups were "Nuclear structure" (183; 0.4%), "Extracellular structures" (36; 0.09%), and "Cell motility" (24; 0.06%). Similar patterns have been reported in some angiosperms, such as *Chrysanthemum nankingense* (Wang et al., 2013) and *Camelina sativa* (Liang et al., 2013). We found 205 unigenes belonging to "Defense mechanism," which indicated the existence of stress resistance genes in *C. rhombifolium*.

Based on the Nr annotation, 42,005 unigenes were assigned to three ontologies and classified into 48 functional GO categories using the Blast2GO software. Of these, 1050 (68.5%), 163 (10.6%), and 320 (20.9%) GO terms were related to cellular components, biological processes, and molecular functions, respectively (**Figure 4**). The assignment of GO terms in *C. rhombifolium* in this study focused on "cellular processes," "metabolic processes," "single-organism processes," "cell," "cell parts," "macromolecular complex," "membrane part," "organelles," and "binding and catalytic activity," which reflected the functional gene expression characteristics during its normal growth. This result was similar to those GO terms in some drought- resistance species, e.g., bread

wheat, oak, *Boea hygrometrica*, *Boea hygrometrica,* and so on (Gupta et al., 2003; Durand et al., 2010; Xiao et al., 2015; Zhu et al., 2015), which mainly due to the selective gene expression caused by various environments and physiological states.

Based on the sequence homology searches against the KEGG database, 56,213 unigenes were assigned to five ontologies and classified into 18 functional KEGG pathways (**Figure 5**). Among these pathways, the "translation pathway" (6,912; 12.3% of KEGG unigenes) was the largest category in "metabolism" followed by "carbohydrate metabolism" (4,876), "overview" (4,496), "amino acid metabolites" (3,306), "folding, sorting, and degradation" (3,112), "energy metabolism" (2,663), "transport and catabolism" (2,297), and "lipid metabolism" (2,168). In this study, we highlight the pathways enriched for the interaction between plants and their environment, including: "metabolism of terpenoids and polyketides" (752), "signal transduction" (592), "environmental adaptation" (719), and "replication and repair" (396). Our results are consistent with those from other studies identifying plant genes and gene products with important roles in drought-resistant plants (Gechev et al., 2012; Xiao et al., 2015). More than 74% of the unigenes from *C. rhombifolium* were mapped in the known databases, which is higher than that reported for *C. nankingense* (64%) (Wang et al., 2013).

Frequency and Distribution of SSRs

In total, 12,925 SSR regions were identified in 10,524 unigenes. Among the identified SSRs, 128 different motifs were identified, the distribution and frequencies of which are shown in **Figure 6**. Mononucleotide motifs were the most abundant, and A/T were the largest subset (6,328). Overall, 6,429 mononucleotide, 2,463 di-repeats, 3,694 tri-repeats, 199 tetra-repeats, 56 penta-repeats, and 84 hexa-repeats were found in the *C. rhombifolium* leaf transcriptome. Among the unigenes containing SSRs, 941 SSRs presented compound formation, and 1,874 contained more than one SSR. On average, one SSR was found every 8.17 kb. The observed number of SSR sequences in our study was higher than EST-SSR ever reports in *Chrysanthemum* (Wang et al., 2013). The SSR sequences may gain or loss of repeats at a locus in their rapid evolution for adaptation to various environments (King et al., 1997; Trifonov, 2004). The mass EST-SSR loci in *C. rhombifolium* may be caused by its harsh habitats. These ESTs will provide a valuable repository of abundant information for future functional SSR studies.

Phylogenetic Analysis of GME

Using the annotated sequence of *GME* in *C. rhombifolium* and other *GME* homologs, we constructed a phylogenetic tree among related species. All of the *GME* sequences from the same taxa were clustered together and *GME* in *C. rhombifolium* were grouped into a single clade with the sequences of *Helianthus annuus* and other Asteraceae species (**Figure 7**), this result revealed a close relationship of *C. rhombifolium* and other Asteraceas species, which consistent with the taxonomy based on morphology.

CONCLUSIONS

We obtained 130,891 unigenes from the leaf of *C. rhombifolium* by NGS transcriptomics, of which 97,496 (∼74.5%) unigenes were successfully annotated in the public protein database. A total of 12,925 SSRs were detected in 10,524 unigenes. This is the first genetic study of *C. rhombifolium* as a plant resource of stress-tolerant genes. These large numbers of transcriptome sequences have enabled us to comprehensively understand the basic genetics of *C. rhombifolium* and discover novel genes that will be helpful in the molecular improvement of chrysanthemums.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in GenBank. The accession numbers can be found in the article.

AUTHOR CONTRIBUTIONS

YM conceived and designed the experiments. LZ collected the plants. WZ, HX, XD, JL, and JH performed the experiments and analyzed the data. YM, WZ, and LZ wrote the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported by the National Nature Science Foundation of China (Nos. 31470699, 31872710, and 31770200).

REFERENCES

Angeloni, F., Wagemaker, C. A. M., Jetten, M. S. M., Camp, H. J. M. O. D., Janssen-Megens, E. M., Francoijs, K. J., et al. (2011). *De novo* transcriptome characterization and development of genomic tools for *Scabiosa columbaria* L. using next generation sequencing techniques. *Mol. Ecol. Resour*. 11, 662–674 doi: 10.1111/j.1755-0998.2011.02990.x

Bremer, K., and Humphries, C. J. (1993). Generic monograph of the Asteraceae-Anthemideae. *Bull. Nat. Hist. Mus. Lond.* 23, 71–177.

Chen, L., Chen, Y., Jiang, J., Chen, S., Chen, F., Guan, Z., et al. (2012). The constitutive expression of *Chrysanthemum dichrum ICE1* in *Chrysanthemum grandiflorum* improves the level of low temperature, salinity and drought tolerance. *Plant Cell Rep*. 31, 1747–1758. doi: 10.1007/s00299-012-1288-y

Chen, S., Cui, X., Chen, Y., Gu, C., Miao, H., Gao, H., et al. (2011). *CgDREBa* transgenic chrysanthemum confers drought and salinity tolerance. *Environ. Exp. Bot*. 74, 255–260 doi: 10.1016/j.envexpbot.2011.06.007

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676 doi: 10.1093/bioinformatics/bti610

Da Silva, J. A. T. (2003). Chrysanthemum: advances in tissue culture, cryopreservation, postharvest technology, genetics and transgenic biotechnology. *Biotechnol. Adv*. 21, 715–766. doi: 10.1016/S0734-9750(03)00117-4

Durand, J., Bodénès, C., Chancerel, E., Frigerio, J. M., Vendramin, G., Sebastiani, F., et al. (2010). A fast and cost-effective approach to develop

and map EST-SSR markers: oak as a case study. *BMC Genomics* 11:570. doi: 10.1186/1471-2164-11-570

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Gechev, T. S., Dinakar, C., Benina, M., Toneva, V., and Bartels, D. (2012). Molecular mechanisms of desiccation tolerance in resurrection plants. *Cell. Mol. Life Sci.* 69, 3175–3186. doi: 10.1007/s00018-012-1088-0

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Gupta, P. K., Rustgi, S., Sharma, S., Singh, R., Kumar, N., and Balyan, H. S. (2003). Transferable ESTSSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol. Genet. Genomics* 270, 315–323 doi: 10.1007/s00438-003-0921-4

Jain, M. (2011). Next-generation sequencing technologies for gene expression profiling in plants. *Brief. Funct. Genomics.* 11, 63–70. doi: 10.1093/bfgp/elr038

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480–D484. doi: 10.1093/nar/gkm882

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

King, D. G., Soller, M., and Kashi, Y. (1997). Evolutionary tuning knobs. *Endeavour* 21, 36–40. doi: 10.1016/S0160-9327(97)01005-3

Kumar, S., Stecher, G., and Tamura, K. (2016). Mega7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874 doi: 10.1093/molbev/msw054

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323

Li, J., Wan, Q., Guo, Y. P., Abbott, R. J., and Rao, G. Y. (2014). Should I stay or should I go: biogeographic and evolutionary history of a polyploid complex (Chrysanthemum indicum complex) in response to Pleistocene climate change in China. *New Phytol.* 201, 1031–1044. doi: 10.1111/nph.12585

Li, J., Wan, Q., Richard, J., Abbott, R. J., and Rao, G. Y. (2013). Geographical distribution of cytotypes in the *Chrysanthemum indicum* complex as evidenced by ploidy level and genome-size variation. *J. Syst. Evol.* 51, 196–204. doi: 10.1111/j.1674-4918.2012.00241.x

Liang, C., Liu, X., Yiu, S. M., and Lim, B. L. (2013). *De novo* assembly and characterization of *Camelina* sativa transcriptome by paired-end sequencing. *BMC Genomics* 14:146. doi: 10.1186/1471-2164-14-146

Lu, M., Zhang, Q. X., Mo, G. Z., and Chen, X. J. (2010). Collection and analysis of drought-resistant wild resources. *Heilongjiang Agric. Sci.* 7, 84–89. doi: 10.3969/j.issn.1002-2767.2010.07.028

Ma, Y. P., Zhao, L., Zhang, W. J., Zhang, Y. H., Xing, X., Duan, X.-X., et al. (2020). Origins of cultivars of Chrysanthemum - evidence from the chloroplast genome and nuclear *LFY* gene. *J. Syst. Evol.* 58, 925–944. doi: 10.1111/jse.12682

Masuda, Y., Yukawa, T., and Kondo, K. (2009). Molecular phylogenetic analysis of members of *Chrysanthemum* and its related genera in the tribe Anthemideae, the Asteraceae in East Asia on the basis of the internal transcribed spacer (ITS) region and the external transcribed spacer (ETS) region of nrDNA. *Chromosome Bot.* 4, 25–36. doi: 10.3199/iscb.4.25

Pereira-Leal, J. B., Abreu, I. A., Alabaça, C. S., Almeida, M. H., and Ricardo, C. P. P. (2014). A comprehensive assessment of the transcriptome of cork oak (*Quercus suber*) through EST sequencing. *BMC Genomics* 15:371. doi: 10.1186/1471-2164-15-371

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., et al. (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19, 651–652 doi: 10.1093/bioinformatics/btg034

Shih, C., and Fu, G. X. (1983). *Flora Reipublicae Popularis Sinicae 76*. Beijing: Science Press.

Song, A., An, J., Guan, Z., Jiang, J., Chen, F., Lou, W., et al. (2014). The constitutive expression of a two transgene construct enhances the abiotic stress tolerance of chrysanthemum. *Plant Physiol. Biochem.* 80, 114–120. doi: 10.1016/j.plaphy.2014.03.030

Song, C., Liu, Y., Song, A., Dong, G., Zhao, H., Sun, W., et al. (2018). The *Chrysanthemum nankingense* genome provides insights into the evolution and diversification of chrysanthemum flowers and medicinal traits. *Mol. Plant.* 11, 1482–1491. doi: 10.1016/j.molp.2018.10.003

Su, J., Jiang, J., Zhang, F., Liu, Y., and Chen, F. (2019). Current achievements and future prospects in the genetic breeding of chrysanthemum: a review. *Horticulture Res.* 6, 109–128 doi: 10.1038/s41438-019-0193-8

Tao, J. J., Wu, H., Li, Z. Y., Huang, C. H., and Xu, X. B. (2018). Molecular evolution of GDP-D-Mannose Epimerase (*GME*), a key gene in plant ascorbic acid biosynthesis. *Front. Plant Sci.* 9:1293. doi: 10.3389/fpls.2018.01293

Trifonov, E. N. (2004). "The tuning function of tandemly repeating sequences: a molecular device for fast adaptation," in *Evolutionary Theory and Processes: Modern Horizons* (Kluwer Academic Publishers; Springer), 115–138.

Wang, H., Jiang, J., Chen, S., Qi, X., Peng, H., Li, P., et al. (2013). Next-generation sequencing of the *Chrysanthemum nankingense* (Asteraceae) transcriptome permits large-scale unigene assembly and SSR marker discovery. *PLoS ONE* 8:e62293. doi: 10.1371/journal.pone.0062293

Wang, Y., Huang, H., Ma, Y. P., Fu, J. X., Wang, L. L., and Dai, S. L. (2014). Construction and *de novo* characterization of a transcriptome of *Chrysanthemum lavandulifolium*: analysis of gene expression patterns in floral bud emergence. *Plant Cell Tissue Organ Culture* 116, 297–309 doi: 10.1007/s11240-013-0404-1

Wang, Y., Liu, K., Bi, D., Zhou, S. B., and Shao, J. W. (2017). Characterization of the transcriptome and EST-SSR development in *Boea clarkeana*, a desiccation-tolerant plant endemic to China. *Peer J.* 5:e3422. doi: 10.7717/peer j.3422

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484

Wu, Q. J., Chen, Z. D., Sun, W. J., Deng, T. T., and Chen, M. J. (2016). *De novo* Sequencing of the leaf transcriptome reveals complex light-responsive regulatory Networks in *Camellia sinensis* cv. *Baijiguan. Front Plant Sci.* 7:332. doi: 10.3389/fpls.2016.00332

Xiao, L. H., Yang, G., Zhang, L. C., Yang, X. H., Zhao, S., et al. (2015). The resurrection genome of *Boea hygrometrica*: a blueprint for survival of dehydration. *Proc. Natl. Acad. Sci. U.S.A.* 112, 5833–5837 doi: 10.1073/pnas.1505811112

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., et al. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34(Suppl. 2), W293–W297. doi: 10.1093/nar/gkl031

Zalapa, J. E., Cuevas, H., Zhu, H. Y., Steffan, S., Senalik, D., Zeldin, E., et al. (2012). Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in plant sciences. *Am. J. Bot.* 99, 193–208. doi: 10.3732/ajb.11 00394

Zhao, H. B., Chen, F. D., Chen, S. M., Wu, G. S., and Guo, W. M. (2010). Molecular phylogeny of *Chrysanthemum*, *Ajania* and its allies (Anthemideae, Asteraceae) as inferred from nuclear ribosomal ITS and chloroplast trnL-F IGS sequences. *Plant Syst. Evol.* 284, 153–169. doi: 10.1007/s00606-009-0242-0

Zhao, H. E., Liu, Z. H., Hu, X., Yin, J. L., Li, W., Rao, G. Y., et al. (2009). *Chrysanthemum* genetic resources and related genera of Chrysanthemum collected in China. *Genet Resour Crop Ev.* 56, 937–946. doi: 10.1007/s10722-009-9412-8

Zhu, Y., Wang, B., Phillips, J., Zhang, Z. N., Du, H., Xu, T., et al. (2015). Global transcriptome analysis reveals acclimation-primed processes was involved in the acquisition of desiccation tolerance in *Boea hygrometrica*. *Plant Cell Physiol.* 56, 1429–1441. doi: 10.1093/pcp/pcv059

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 12 April 2021
doi: 10.3389/fgene.2021.663557

Genome-Wide Association Mapping of Seedling Biomass and Root Traits Under Different Water Conditions in Wheat

Iza Fatima[1†], Yutian Gao[1†], Xiangru Xu[1], Jingjing Jin[2], Shuonan Duan[1], Wenchao Zhen[2], Chaojie Xie[1] and Jun Ma[1]*

[1] College of Agronomy and Biotechnology, China Agricultural University, Beijing, China, [2] College of Agronomy, Hebei Agricultural University, Baoding, China

Drought is a major threat to global wheat production. In this study, an association panel containing 200 Chinese wheat germplasms was used for genome-wide association studies (GWASs) of genetic loci associated with eight root and seedling biomass traits under normal water and osmotic stress conditions. The following traits were investigated in wheat seedlings at the four-leaf stage: root length (RL), root number (RN), root fresh weight (RFW), root dry weight (RDW), shoot fresh weight (SFW), shoot dry weight (SDW), total fresh weight (TFW), and total dry weight (TDW). A total of 323 and 286 SNPs were detected under two water environments, respectively. Some of these SNPs were near known loci for root traits. Eleven SNPs on chromosomes 1B, 2B, 4B, and 2D had pleiotropic effects on multiple traits under different water conditions. Further analysis indicated that several genes located inside the 4 Mb LD block on each side of these 11 SNPs were known to be associated with plant growth and development and thus may be candidate genes for these loci. Results from this study increased our understanding of the genetic architecture of root and seedling biomass traits under different water conditions and will facilitate the development of varieties with better drought tolerance.

Keywords: wheat, seedling biomass, root traits, GWAS, osmotic stress

OPEN ACCESS

Edited by:
Jian Ma,
Sichuan Agricultural University, China

Reviewed by:
Zhi Zheng,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
Ahsan Habib,
Khulna University, Bangladesh

****Correspondence:***
Jun Ma
junma@cau.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *03 February 2021*
Accepted: *02 March 2021*
Published: *12 April 2021*

Citation:
Fatima I, Gao Y, Xu X, Jin J, Duan S, Zhen W, Xie C and Ma J (2021) Genome-Wide Association Mapping of Seedling Biomass and Root Traits Under Different Water Conditions in Wheat. Front. Genet. 12:663557. doi: 10.3389/fgene.2021.663557

INTRODUCTION

Wheat (*Triticum aestivum* L.) is a widely cultivated crop in the world that provides the main source of calories and protein in the human diet (Shewry, 2009). In many regions of the world, wheat production suffered significant losses due to drought stress (Trethowan and Pfeiffer, 2000). Drought stress can induce significant morphological and physiological changes in plants, including stomatal closure, reductions in photosynthesis and transpiration, shoot and root growth inhibition, antioxidant production, and changes in hormonal composition (Szegletes et al., 2000; Lawlor and Cornic, 2002; Zhu, 2002). The yield loss caused by drought can be up to 92% (Farooq et al., 2014). Due to climate change, the frequency and severity of drought stress will significantly increase in the future and pose a threat to the food security of the rapidly increasing world population (IPCC, 2014).

The root system is vital for plants to obtain water and nutrients from the soil. A positive correlation between root system architecture and agronomic traits was reported by Cane et al. (2014). During grain filling, every millimeter of water extracted from the soil increased wheat

yield by 55 kg/ha (Manschadi et al., 2006; Kirkegaard et al., 2007; Christopher et al., 2013). The root system is also a key structure to respond to water stress conditions and maintain yield under drought stress because plants with deep root systems and large root biomass could extract more water from deeper soil layers (Boyer, 1996; Sharma and Carena, 2016; Wasaya et al., 2018). A study on durum wheat showed that compared with shallow root genotypes, deep root genotypes had 16 to 35% larger grains and can increase grain yield by 35% and thousand-grain weight by 9% in environments with limited moisture (El Hassouni et al., 2018). Furthermore, unlike the consistent reduction of shoot biomass and grain yield under drought stress, the responses of wheat root biomass to drought might be negative, positive, or no response depending on the genotype and environmental factors, which make it an ideal target trait for the improvement of drought tolerance (Wasaya et al., 2018).

To date, a large number of quantitative trait loci (QTL) related to various root traits including root dry weight, seminal root number, total root length, root diameter, number of root tips, root number, etc., have been reported in wheat (Bai et al., 2013; Liu et al., 2013, 2019; Ayalew et al., 2017; Xie et al., 2017; Alahmad et al., 2019; Beyer et al., 2019). Of these QTL, some QTL were specific to water stress conditions. For example, Ayalew et al. (2017) identified four root length QTL specific to drought stress conditions on chromosomes 1A, 3A, and 7B. Similarly, Liu et al. (2013) reported several QTL for maximum root length and seminal root area on chromosomes 1A, 2A, 5A, and 5D. Several QTL controlling plant height and shoot dry weight also affected various root traits such as root length and root biomass, indicating the important roles of the root system on plant growth and development (Cao et al., 2014; Iannucci et al., 2017).

In this study, we performed a genome-wide association study (GWAS) to identify sets of markers associated with root and seedling biomass traits in a panel of 200 Chinese wheat germplasms under normal water and PEG-induced osmotic environments via hydroponic culture. The results will increase our understanding of the genetic architecture of root and seedling biomass traits under different water conditions and will facilitate the development of varieties with better drought tolerance.

MATERIALS AND METHODS

Plant Material

An association panel consisting of 200 Chinese varieties of wheat collected from different wheat production regions of China were used in this study (**Supplementary Table 1**). Most of them were from the Yellow and Huai River Valley, one of the major wheat-producing regions in China (Jin et al., 2020).

Phenotypic Evaluation

Seeds were first selected by removing small and shriveled kernels. Seeds from each variety were soaked in 70% sodium hypochlorite solution for 10 min to sterilize and then washed two to three times with distilled water. Following that, seeds were germinated in petri dishes at room temperature. The seedlings with 0.5 cm length of coleoptiles were rolled in germination paper after 1–2 days. Correspondingly, seedlings were transferred 1 day later to a container with 1/2 Hoagland nutrient solution in a greenhouse with 60% humidity, 25°C temperature, and 10/14 h (day/night)

TABLE 1 | Basic statistics of root traits for 200 accessions grown in controlled and drought conditions.

Environment	Trait	Min	Max	Ave	SD	Coef. of var (%)	H^2
Drought	RL (cm)	17.806	19.650	18.796	0.335	1.7	0.517
	RN	3.815	5.708	4.947	0.313	6.3	0.793
	RFW (g)	0.088	0.125	0.107	0.006	6.4	0.544
	RDW (g)	0.013	0.020	0.017	0.001	7.0	0.650
	SFW (g)	0.141	0.222	0.174	0.015	9.0	0.654
	SDW (g)	0.017	0.031	0.022	0.002	10.8	0.766
	TFW (g)	0.231	0.345	0.282	0.019	7.0	0.628
	TDW (g)	0.031	0.049	0.039	0.003	8.3	0.731
Control	RL (cm)	25.635	28.815	27.332	0.599	2.1	0.271
	RN	4.685	7.551	6.061	0.514	8.4	0.692
	RFW (g)	0.155	0.323	0.210	0.026	12.4	0.619
	RDW (g)	0.010	0.022	0.014	0.001	12.5	0.572
	SFW (g)	0.348	0.654	0.469	0.052	11.2	0.668
	SDW (g)	0.010	0.056	0.035	0.005	14.6	0.676
	TFW (g)	0.505	0.969	0.679	0.078	11.5	0.667
	TDW (g)	0.035	0.076	0.049	0.006	13.8	0.661

Trait abbreviations: RL, root length; RN, root number; RFW, root fresh weight; RDW, root dry weight; SFW, shoot fresh weight; SDW, shoot dry weight; TFW, total fresh weight; TDW, total dry weight; SD, standard deviation; H^2, broad-sense heritability.

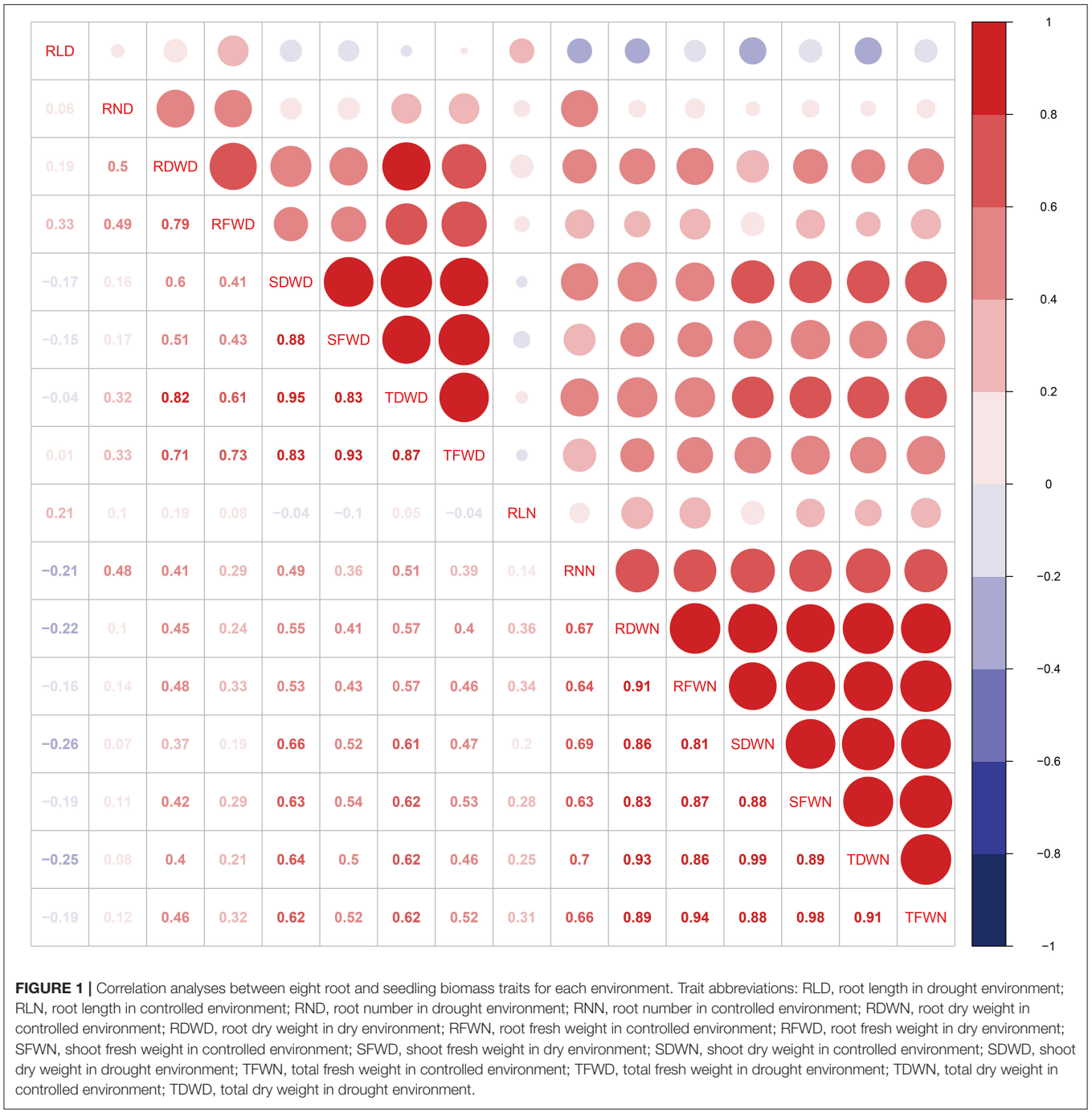

FIGURE 1 | Correlation analyses between eight root and seedling biomass traits for each environment. Trait abbreviations: RLD, root length in drought environment; RLN, root length in controlled environment; RND, root number in drought environment; RNN, root number in controlled environment; RDWN, root dry weight in controlled environment; RDWD, root dry weight in dry environment; RFWN, root fresh weight in controlled environment; RFWD, root fresh weight in dry environment; SFWN, shoot fresh weight in controlled environment; SFWD, shoot fresh weight in dry environment; SDWN, shoot dry weight in controlled environment; SDWD, shoot dry weight in drought environment; TFWN, total fresh weight in controlled environment; TFWD, total fresh weight in drought environment; TDWN, total dry weight in controlled environment; TDWD, total dry weight in drought environment.

timing using automatic timer. On the seventh day, eight seedlings of each variety were grown to 1/2 Hoagland solution with 20% polyethylene glycol (PEG) 6000 (Sinopharm Chemical Reagent Co. Ltd, China) for drought treatment, whereas the remaining eight seedlings were kept in 1/2 Hoagland solution for control. The solution containing PEG was changed every 3 days to keep the water potential stable. For both controlled and drought environments, the experiment was repeated three times.

When the seedlings were at the four-leaf stage, eight traits, including root length (RL), root number (RN), root fresh weight (RFW), root dry weight (RDW), shoot fresh weight (SFW), shoot dry weight (SDW), total fresh weight (TFW), and total dry weight (TDW), were evaluated under both normal and drought environments. The roots of the seedlings were first washed before measurement. The longest root among all the roots of a seedling was selected for the measurement of RL with the help of a measuring scale and was expressed in centimeters (cm). For

the measurement of RFW, the excessive water was removed by pressing the roots gently with a tissue paper sheet. For the measurement of RDW, samples were dried to constant weight by incubating them in small paper bags at 80°C for 48 h. The SFW and SDW were measured in a similar way.

Statistical Analysis

To analyze the variation among all eight traits, different statistical tools such as mean, median, sum, variance, range, standard error of the mean, confidence interval of the mean, standard deviation, and coefficient of variance were applied by using a statistical package "pastecs" in R software. The following formula was used for the calculation of broad sense heritability (H^2) of various traits:

$$H^2 = V_G/(V_G + V_E)$$

Here, V_G denotes genetic variance, and V_E represents environmental variance (Ehlers et al., 2009). R corrplot was employed to analyze the correlation of each trait. For each line, the lme4 package in R was used for the estimation of the best linear unbiased predictors (BLUP), which were later used for GWAS analysis. Each trait was analyzed within eight plants in each accession, and the mean of all eight traits was used for consecutive statistical analysis and GWAS.

Genome-Wide Association Analysis and Prediction of Candidate Gene

Molecular marker data of the 200 germplasms in the association panel was extracted from our previous publication (Jin et al., 2020). The Genomic Association and Prediction Integrated Tool (GAPIT) package in R (Version 4.0.3.) was used for GWAS analysis (Lipka et al., 2012). GWAS was performed using the mixed linear model (PCA + K), and the variance–covariance kinship matrix (K) was calculated by the VanRaden method in R (VanRaden, 2008; Zhang et al., 2010; Lipka et al., 2012; Jin et al., 2020). A suggestive threshold of P value equal to 1.0E−3 ($P = 1/n$, $n =$ effective SNP number) was used to estimate the significant SNPs (Sun et al., 2017; Jin et al., 2020). The CMplot package in R was used to draw Manhattan plots, which are showing the SNPs identified for root traits in GWAS using BLUP values of 200 wheat germplasm (LiLin-Yin, 2020).

The stable SNPs in the three experiments were selected for the favorable allele analysis. In the analysis of allele effects on each trait, alleles with positive effects leading to higher values of root traits were described as "favorable alleles," whereas those with lower values were "unfavorable alleles" (Liu et al., 2019). For the investigation of potential candidate genes, the EnsemPlants database (http://plants.ensembl.org) was used to download the genes within 4 Mb LD block on both sides of the significant SNPs in both controlled and drought environments (Jin et al., 2020). Gene annotation, the relative homologous rice, and the *Arabidopsis* gene of particular wheat genes were also investigated through the Triticeae-Gene Tribe (TGT) website (http://wheat.cau.edu.cn/TGT/index.html) (Chen et al., 2020).

RESULTS

Phenotypic Evaluation

All phenotypic traits in both controlled and drought environments showed continuous and significantly wide variations in 200 wheat germplasms. The basic statistics of these phenotypic traits is shown in **Table 1**. The mean values of RDW, SDW, TDW, RFW, SFW, TFW, RN, and RL were significantly higher in a controlled environment (0.014, 0.035, 0.049, 0.21, 0.469, 0.679, 6.061, and 27.332, respectively) compared with those in a drought environment (0.017, 0.022, 0.039, 0.107, 0.174, 0.282, 4.947, and 18.796, respectively) as shown in **Table 1**. The coefficient of variation for all eight traits in controlled and drought environments ranged from 2.1 to 14.6% and 1.7 to 10.8% respectively. The values of standard deviation (SD) ranged from 0.001 to 0.599 in the controlled environment and from 0.001 to 0.335 in the drought environment (**Table 1**).

Correlation coefficients (r^2) among different phenotypic traits were analyzed to quantify the relationship between all eight traits under both controlled and drought environments (**Figure 1**). A strong correlation was detected in various root traits under both environments. For example, TFWD is positively correlated with SFWD and TDWD with a value range of (0.83) and (0.87) in the drought environment, while RFWN was positively correlated with SDWN (0.81), SFWN (0.87), TDWN (0.86), and TFWN (0.94), respectively (**Figure 1**). Most of the investigated traits showed a strong broad sense heritability (H^2), ranging from 0.271 (RL) to 0.692 (RN) and 0.517 (RL) to 0.793 (RN) in the controlled and drought environments, respectively (**Table 1**). However, the H^2 of RL under controlled environments was relatively low (0.271), suggesting a significant environmental effect (**Table 1**).

Marker–Trait Associations

Based on the BLUP values in the three experiments, a total of 609 SNPs were significantly associated with eight root and seedling biomass traits in the current study (**Supplementary Table 2**). The B genome has the highest number of SNPs (440), followed by the D (93) and A (65) genomes. These significant SNPs were distributed on 21 chromosomes except 2A, 4A, 5A, and 7A. From these SNPs, a total of 323 SNPs related to root and seedling biomass traits were identified in the normal water condition. Chromosome 2B had the largest proportion of SNPs (19.50%), followed by chromosome 5B. Additionally, 115 SNPs were associated with multiple traits, including 30 SNPs on chromosomes 5B (22), 4D (6), and 5D (2) associated with RDW, RFW, SDW, SFW, TDW, and TFW (**Supplementary Table 2**). A total of 286 SNPs related to root and seedling biomass traits were observed under the drought environment. These SNPs were distributed on all the 21 chromosomes of wheat except 6A (**Supplementary Table 2**). The same as those in the controlled environment, most of these SNPs were on chromosome 2B (38.11%) and chromosome 5B (26.57%).

Moreover, statistical analysis identified 11 SNPs that appeared to be significantly associated ($P < 0.001$) with RN, RDW, SFW, SDW, TFW, and TDW under two water conditions (**Table 2**). These 11 loci were distributed on chromosomes 1B, 2B, 4B, and 2D, respectively. Two SNP loci (*Affx-111601113*

TABLE 2 | The SNP significantly associated with pleiotropic effect in different phenotyping traits in two environments.

Marker[a]	Chromosome	Position[b]	Pleiotropic effect	SNP[c]	*P* value
Affx-109979441	1B	533184310	RDWN, SFWN, SDWN, TFWN, TDWN, SDWD, TDWD	C	<0.001
Affx-109782056	2B	607064966	SDWN, SFWD, SDWD, TDWN, TFWD, TDWD	C	<0.001
Affx-111326878	2B	606037476	SDWN, TDWN, SFWD, SDWD, TDWD	T	<0.001
Affx-111434051	2B	606938708	RDWN, SDWN, SFWD, SDWD, TFWD, TDWD	T	<0.001
Affx-111601113	2B	448398290	RNN, SDWN, TDWN, SDWD	A	<0.001
Affx-92897136	2B	641102486	SDWN, SDWD	C	<0.001
Affx-109617080	2D	514348182	RDWN, SDWN, TDWN, SFWD, SDWD, TFWD, TDWD	C	<0.001
Affx-88425798	2D	375362652	RNN, SDWN, TDWN, SDWD	C	<0.001
Affx-109674658	4B	37700884	SDWN, TDWN, SDWD, TDWD	T	<0.001
Affx-88444969	4B	37845625	TDWN, SDWD	G	<0.001
Affx-88596529	4B	37735225	SDWN, SDWD, TDWN	C	<0.001

Trait abbreviations: RNN, root number in controlled environment; RDWN, root dry weight in controlled environment; SFWN, shoot fresh weight in controlled environment; SDWN, shoot dry weight in controlled environment; TFWN, total fresh weight in controlled environment; TDWN, total dry weight in controlled environment;, SFWD, shoot fresh weight in drought environment; SDWD, shoot dry weight in drought environment; TFWD, total fresh weight in drought environment; TDWD, total dry weight in drought environment.

[a] *Representative marker at a specific locus.*

[b] *Physical position of SNP marker in base pair (bp).*

[c] *Indication of favorable allele (SNP).*

on chromosome 2B and *Affx-88425798* on chromosome 2D) explained 14.5 to 30.6% of the phenotypic variance in RNN (**Table 2**, **Figure 2**). These two SNP loci also had significant effects on SDWN, TDWN, and SDWD. Similarly, one SNP (*Affx-109979441*) on chromosome 1B was associated with RDWN, SFWN, TFWN, TDWN, SDWD, and TDWD; five on chromosome 2B (*Affx-111601113*, *Affx-111326878*, *Affx-111434051*, *Affx-109782056*, and *Affx-92897136*) were associated with RNN, SFWN, SDWN, TDWN, SFWD, SDWD, TFWN, and TDWD; three on chromosome 4B (*Affx-109674658*, *Affx-88444969*, and *Affx-88596529*) showed pleiotropic effect for SDWN, TDWN, SDWD, and TDWD, respectively (**Table 2**).

Stability is an important parameter to evaluate a particular QTL. Of the SNPs identified by GWAS analyses based on BLUP value, four SNPs on chromosomes 2D, 4D, and 6D were stable in all of the three replicates (**Supplementary Table 2**). The favorable allele of SNP *Affx-111174209* on chromosome 4D led to an increase from 0.45 to 0.48 for SFWN, and from 0.65 to 0.71 for TFWN (**Figure 3**). The SNP (*Affx-109617080*) on chromosome 2D led to an increase from 0.0209 to 0.0254 for SDWD. Two SNPs on chromosome 6D including *Affx-109538680* and *Affx-108905447* led to an increase of 0.97 and 1.1 for RNN, respectively (**Figure 3**).

Candidate Genes for the SNPs Stable Under Two Water Conditions

Based on genes annotated in Chinese spring reference genome, a total of 442 genes were identified in the 4 Mb (LD block) region on each side of 11 significant SNPs (**Supplementary Table 3**). Some of these genes raised our interests due to their reported roles in plant growth and development. These genes included *TraesCS1B02G310200* near *Affx-109979441* on chromosome 1B, which encoded the transcription repressor OFP4, *TraesCS2B02G423500* on chromosome 2B and *TraesCS2D02G402400* on chromosome 2D encoding rolling and erect leaf 2 protein, two genes on chromosome 4B including *TraesCS4B02G048400* encoding G-type lectin S-receptor-like serine/threonine protein kinase, and *TraesCS4B02G049700* encoding 26S proteasome non-ATPase regulatory subunit 8, respectively.

DISCUSSION

Due to the difficulty in phenotyping root traits, direct selection for variation in root characteristics is impractical (Reynolds et al., 2007). Molecular screens are likely to have a considerable cost-benefit advantage over the phenotyping method (Tuberosa and Salvi, 2006). In this study, a total of 164 SNPs were associated with root traits under controlled environment, whereas 152 SNPs were identified under osmotic stress (**Supplementary Table 2**). These SNPs are useful genetic resources for the improvement of root traits and drought tolerance in wheat. Some of these SNP loci are closely located to the known loci associated with root traits based on the reference genome of IWGSC V1.1. For example, under drought stress condition, three SNP loci including *Affx-88733278*, *Affx-109672297*, and *Affx-110800753* for root length on chromosome 2B are 2.29, 1.79, and 3.39 Mb away from *AX_111251784*, *AX_94405934*, and *AX_108756976*, which were associated with total root length in a previous study (Liu et al., 2019). Three SNPs (*Affx-110656000*, *Affx-111718859*, and *Affx-88565514*) on chromosome 3A for root number was about 0.7 Mb away from an SNP (*S7_12487861*) for branched root length (Beyer et al., 2019). Under normal water conditions, *Affx-110668350* on chromosome 6A for RDW and *Affx-109434039* on chromosome 7D for RL were 0.48 and 1 Mb away from *S16_7327093* and *S21_99959518* for root diameter, respectively (Beyer et al., 2019).

Similar to the SNP specific to a single water condition, several SNPs under two water conditions were closely located to known QTL for root traits or grain yield-related traits

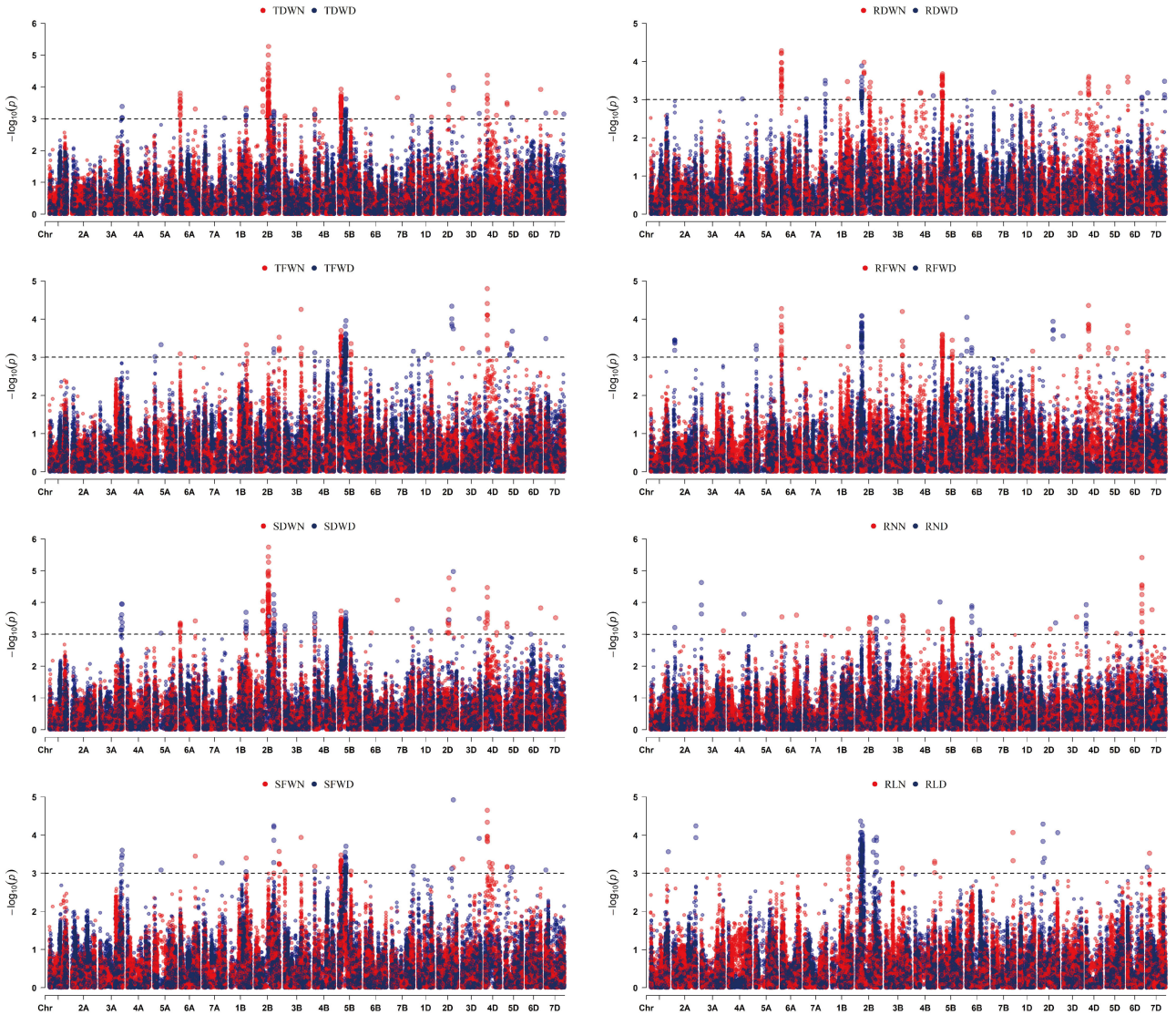

FIGURE 2 | Manhattan plots for the eight root and seedling biomass traits identified by genome-wide association study (GWAS) using BLUP values. The dashed line represents the significance threshold (–log10 $P = 3.0$). Trait abbreviations: RLD, root length in drought environment; RLN, root length in controlled environment; RND, root number in drought environment; RNN, root number in controlled environment; RDWN, root dry weight in controlled environment; RDWD, root dry weight in dry environment; RFWN, root fresh weight in controlled environment; RFWD, root fresh weight in dry environment; SFWN, shoot fresh weight in controlled environment; SFWD, shoot fresh weight in dry environment; SDWN, shoot dry weight in controlled environment; SDWD, shoot dry weight in drought environment; TFWN, total fresh weight in controlled environment; TFWD, total fresh weight in drought environment; TDWN, total dry weight in controlled environment; TDWD, total dry weight in drought environment.

(**Table 2**). For example, the three SNP loci (*Affx-109782056*, *Affx-111326878*, and *Affx-111434051*) on chromosome 2B for seedling biomass traits were in the same chromosome region as a QTL (*QSRN.cgb-2B*) controlling seminal root number (Liu et al., 2013). Two previously reported QTL for kernel number per spike (*QKNS.caas-4BS*) and spike number per unit area (*QSN.caas-4BS*) also located near the interval targeted by three SNPs (*Affx-109674658*, *Affx-88444969*, and *Affx-88596529*) on chromosome 4B for seedling biomass traits (<0.1 Mb) (Li et al., 2018). Additionally, one SNP (*Affx-88431037*) on chromosome 4D for SDWN was 1.7 Mb from *Rht-D1* controlling plant height. The close locations of the loci for root and seedling biomass traits in the current analysis and those reported previously for plant height and grain yield are in line with the strong relationship between those traits (Bai et al., 2013; El Hassouni et al., 2018). Considering that the LD block in this population is about 4 Mb, it is likely that the loci for root and seedling biomass in the current analysis are the same as those for plant height and grain yield. However, further experiments are required to confirm the genetic relationship between these loci.

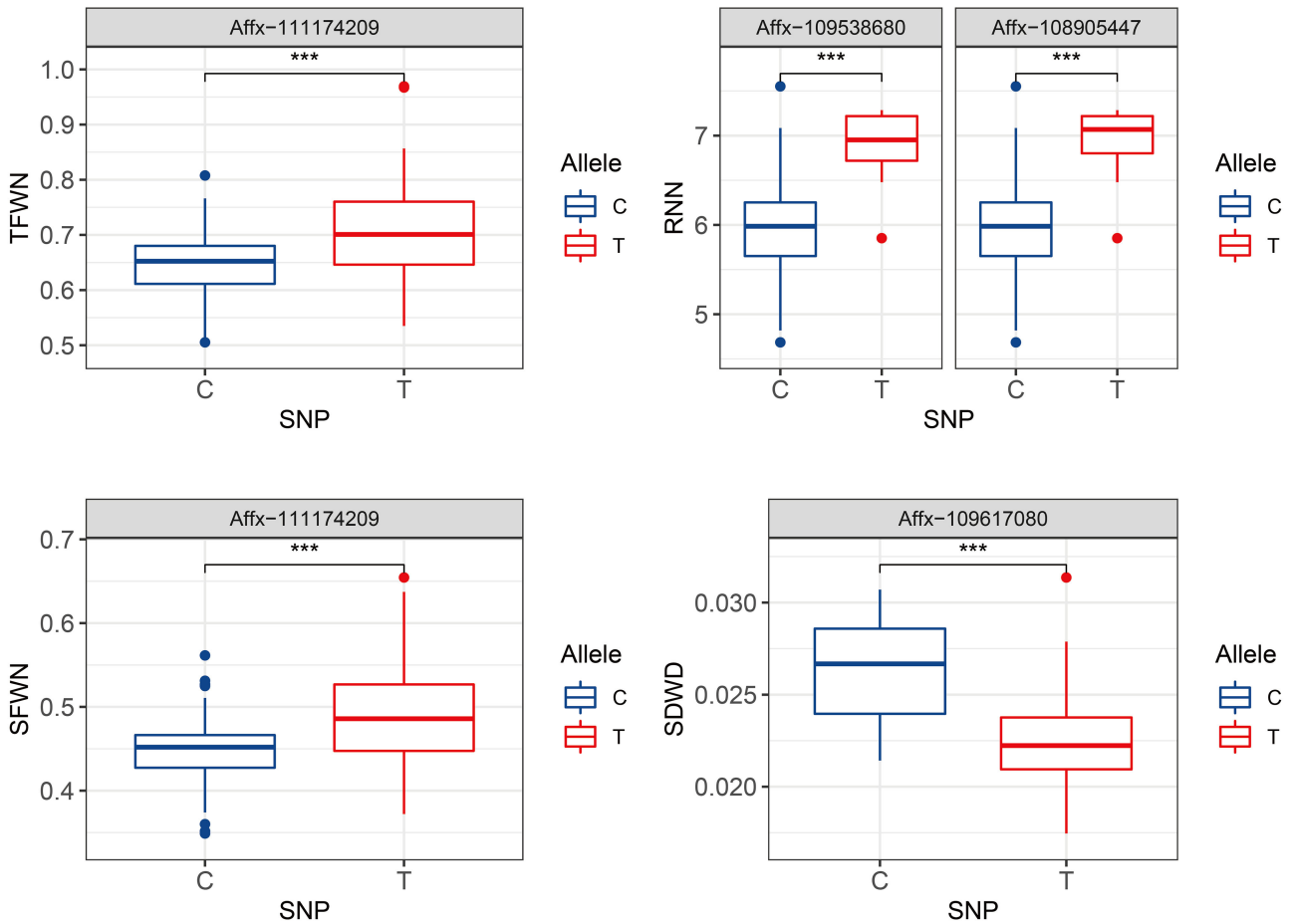

FIGURE 3 | Comparison of the allele effects of four environmentally stable SNPs on chromosomes 2D, 4D, and 6D. Trait abbreviations: TFWN, total fresh weight in controlled environment; SFWN, shoot fresh weight in controlled environment; RNN, root number in controlled environment; SDWD, shoot dry weight in drought environment. ***$p < 0.001$.

Because of the stability of the 11 SNPs under different water conditions, we investigated the potential candidate genes for these 11 SNPs. Some of the genes located within the 4-Mb LD block on each side of the 11 SNPs were known to be associated with plant growth and development based on previous literature (**Table 2**). For example, an *Arabidopsis* gene *AtOFP1* could suppress cell elongation and regulate cotyledon development in a postembryonic manner (Wang et al., 2007, 2011). In the current study, *TraesCS1B02G310200*, a wheat homologous gene of *AtOFP1*, is only about 95 kb from *Affx-109979441* on 1B controlling several seedling biomass traits and RDWN. Previous research found that a mutant of rolled and erect leaf 2 (*REL2*) gene is associated with the increased ability of rice leaves to capture light energy and exchange gas and thus increase the yield of rice (Yang et al., 2016). In this study, we found two homologous genes of the rice *REL2*, one is *TraesCS2B02G423500* near *Affx-109782056* on chromosome 2B, and the other is *TraesCS2D02G402400* near *Affx-109617080* on chromosome 2D (**Supplementary Table 3**). The expression of *OsTMK* is particularly high in regions undergoing cell division and elongation, and low in the non-growing region of the internode, implying that *OsTMK* regulates rice growth (Van Der Knaap et al., 1999; Hirose et al., 2007). *TraesCS4B02G048400*, a homologous gene of *OsTMK* was located 1 Mb away from *Affx-109674658* on chromosome 4B. Kurepa et al. (2009) showed that partial loss of function of the regulatory particle non-ATPase (RPN) subunits RPN10 and RPN12a caused a stronger defect in proteasome function and also resulted in cell enlargement and decreased cell proliferation in *Arabidopsis*. The gene *TraesCS4B02G049700* near the SNP *Affx-88444969* is homologous to the *Arabidopsis* gene *RPN12a* (**Supplementary Table 3**). Further experiments will be carried out to determine the functions of the above genes in root and seedling development.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories

and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

JM conceptualized and designed the research. IF, YG, XX, and JJ performed experiment. JM, IF, YG, and XX analyzed data and wrote the manuscript. JM, WZ, CX, and SD discussed the results and revised the manuscript. All authors have read and approved final manuscript.

FUNDING

This research was supported by National Key Research and Development Program of China (2018YFD0300501).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.663557/full#supplementary-material

REFERENCES

Alahmad, S., El Hassouni, K., Bassi, F. M., Dinglasan, E., Youssef, C., Quarry, G., et al. (2019). A major root architecture QTL responding to water limitation in durum wheat. *Front. Plant Sci.* 10, 1–18. doi: 10.3389/fpls.2019.00436

Ayalew, H., Liu, H., and Yan, G. (2017). Identification and validation of root length QTLs for water stress resistance in hexaploid wheat (*Titicum aestivum* L.). *Euphytica* 213, 1–11. doi: 10.1007/s10681-017-1914-4

Bai, C., Liang, Y., and Hawkesford, M. J. (2013). Identification of QTLs associated with seedling root traits and their correlation with plant height in wheat. *J. Exp. Bot.* 64, 1745–1753. doi: 10.1093/jxb/ert041

Beyer, S., Daba, S., Tyagi, P., Bockelman, H., Brown-Guedira, G., and Mohammadi, M. (2019). Loci and candidate genes controlling root traits in wheat seedlings—a wheat root GWAS. *Funct. Integr. Genomics* 19, 91–107. doi: 10.1007/s10142-018-0630-z

Boyer, J. S. (1996). Advances in drought tolerance in plants. *Adv. Agron.* 56, 187–218.

Cane, M. A., Maccaferri, M., Nazemi, G., Salvi, S., Francia, R., Colalongo, C., et al. (2014). Association mapping for root architectural traits in durum wheat seedlings as related to agronomic performance. *Mol. Breed.* 34, 1629–1645. doi: 10.1007/s11032-014-0177-1

Cao, P., Ren, Y., Zhang, K., Teng, W., Zhao, X., Dong, Z., et al. (2014). Further genetic analysis of a major quantitative trait locus controlling root length and related traits in common wheat. *Mol. Breed.* 33, 975–985. doi: 10.1007/s11032-013-0013-z

Chen, J., Hu, X., Shi, T., Yin, H., Sun, D., Hao, Y., et al. (2020). Metabolite-based genome-wide association study enables dissection of the flavonoid decoration pathway of wheat kernels. *Plant Biotechnol. J.* 18, 1722–1735. doi: 10.1111/pbi.13335

Christopher, J., Christopher, M., Jennings, R., Jones, S., Fletcher, S., Borrell, A., et al. (2013). QTL for root angle and number in a population developed from bread wheats (*Triticum aestivum*) with contrasting adaptation to water-limited environments. *Theor. Appl. Genet.* 126, 1563–1574. doi: 10.1007/s00122-013-2074-0

Ehlers, J., Xu, N. W., and Xu, S. (2009). Estimating the broad-sense heritability of early growth of cowpea. *Int. J. Plant Genomics* 2009. doi: 10.1155/2009/984521

El Hassouni, K., Alahmad, S., Belkadi, B., Filali-Maltouf, A., Hickey, L. T., and Bassi, F. M. (2018). Root system architecture and its association with yield under different water regimes in Durum wheat. *Crop Sci.* 58, 2331–2346. doi: 10.2135/cropsci2018.01.0076

Farooq, M., Hussain, M., and Siddique, K. H. M. (2014). Drought stress in wheat during flowering and grain-filling periods. *CRC. Crit. Rev. Plant Sci.* 33, 331–349. doi: 10.1080/07352689.2014.875291

Hirose, N., Makita, N., Kojima, M., Kamada-Nobusada, T., and Sakakibara, H. (2007). Overexpression of a type-A response regulator alters rice morphology and cytokinin metabolism. *Plant Cell Physiol.* 48, 523–539. doi: 10.1093/pcp/pcm022

Iannucci, A., Marone, D., Russo, M. A., De Vita, P., Miullo, V., Ferragonio, P., et al. (2017). Mapping QTL for root and shoot morphological traits in a durum wheat × *T. dicoccum* segregating population at seedling stage. *Int. J. Genomics* 2017. doi: 10.1155/2017/6876393

IPCC (2014). *Climate Change 2014, Synthesis Report Summary for Policymakers. Fifth assessment report (AR5)*. Washington DC.

Jin, J., Duan, S., Qi, Y., Yan, S., Li, W., Li, B., et al. (2020). Identification of a novel genomic region associated with resistance to Fusarium crown rot in wheat. *Theor. Appl. Genet.* 133, 2063–2073. doi: 10.1007/s00122-020-03577-1

Kirkegaard, J. A., Lilley, J. M., Howe, G. N., and Graham, J. M. (2007). Impact of subsoil water use on wheat yield. *Aust. J. Agric. Res.* 58, 303–315. doi: 10.1071/AR06285

Kurepa, J., Wang, S., Li, Y., Zaitlin, D., Pierce, A. J., and Smalle, J. A. (2009). Loss of 26S proteasome function leads to increased cell size and decreased cell number in Arabidopsis shoot organs. *Plant Physiol.* 150, 178–189. doi: 10.1104/pp.109.135970

Lawlor, D. W., and Cornic, G. (2002). Photosynthetic carbon assimilation and associated metabolism in relation to water deficits in higher plants. *Plant Cell Environ.* 25, 275–294. doi: 10.1046/j.0016-8025.2001.00814.x

Li, F., Wen, W., He, Z., Liu, J., Jin, H., Cao, S., et al. (2018). Genome-wide linkage mapping of yield-related traits in three Chinese bread wheat populations using high-density SNP markers. *Theor. Appl. Genet.* 131, 1903–1924. doi: 10.1007/s00122-018-3122-6

LiLin-Yin (2020). Manhattan Plot. Available at: https://github.com/YinLiLin/CMplot

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444

Liu, P., Jin, Y., Liu, J., Liu, C., Yao, H., Luo, F., et al. (2019). Genome-wide association mapping of root system architecture traits in common wheat (*Triticum aestivum* L.). *Euphytica* 215, 1–12. doi: 10.1007/s10681-019-2452-z

Liu, X., Li, R., Chang, X., and Jing, R. (2013). Mapping QTLs for seedling root traits in a doubled haploid wheat population under different water regimes. *Euphytica* 189, 51–66. doi: 10.1007/s10681-012-0690-4

Manschadi, A. M., Christopher, J., Devoil, P., and Hammer, G. L. (2006). The role of root architectural traits in adaptation of wheat to water-limited environments. *Funct. Plant Biol.* 33, 823–837. doi: 10.1071/FP06055

Reynolds, M., Dreccer, F., and Trethowan, R. (2007). Drought-adaptive traits derived from wheat wild relatives and landraces. *J. Exp. Bot.* 58, 177–186. doi: 10.1093/jxb/erl250

Sharma, S., and Carena, M. J. (2016). BRACE: A method for high throughput maize phenotyping of root traits for short-season drought tolerance. *Crop Sci.* 56, 2996–3004. doi: 10.2135/cropsci2016.02.0116

Shewry, P. R. (2009). Wheat. *J. Exp. Bot.* 60, 1537–1553. doi: 10.1093/jxb/erp058

Sun, C., Zhang, F., Yan, X., Zhang, X., Dong, Z., Cui, D., et al. (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. doi: 10.1111/pbi.12690

Szegletes, Z., Erdei, L., Tari, I., and Cseuz, L. (2000). Accumulation of osmoprotectants in wheat cultivars of different drought tolerance. *Cereal Res. Commun.* 28, 403–410. doi: 10.1007/bf03543622

Trethowan, R. M., and Pfeiffer, W. H. (2000). "Challenges and future strategies in breeding wheat for adaptation to drought stressed environments: a CIMMYT wheat program perspective," in *Molecular Approaches for the Genetic Improvement of Cereals for Stable Production in Water-Limited Environments*. A Strategic Planning Workshop (Texcoco: El Batan).

Tuberosa, R., and Salvi, S. (2006). Genomics-based approaches to improve drought tolerance of crops. *Trends Plant Sci.* 11, 405–412. doi: 10.1016/j.tplants.2006.06.003

Van Der Knaap, E., Song, W. Y., Ruan, D. L., Sauter, M., Ronald, P. C., and Kende, H. (1999). Expression of a gibberellin-induced leucine-rich repeat receptor-like protein kinase in deepwater rice and its interaction with kinase-associated protein phosphatase. *Plant Physiol.* 120, 559–569. doi: 10.1104/pp.120.2.559

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Wang, S., Chang, Y., Guo, J., and Chen, J. G. (2007). Arabidopsis ovate family protein 1 is a transcriptional repressor that suppresses cell elongation. *Plant J.* 50, 858–872. doi: 10.1111/j.1365-313X.2007.03096.x

Wang, S., Chang, Y., Guo, J., Zeng, Q., Ellis, B. E., and Chen, J. G. (2011). Arabidopsis Ovate family proteins, a novel transcriptional repressor family, control multiple aspects of plant growth and development. *PLoS ONE* 6:23896. doi: 10.1371/journal.pone.0023896

Wasaya, A., Zhang, X., Fang, Q., and Yan, Z. (2018). Root phenotyping for drought tolerance: a review. *Agronomy* 8, 1–19. doi: 10.3390/agronomy8110241

Xie, Q., Fernando, K. M. C., Mayes, S., and Sparkes, D. L. (2017). Identifying seedling root architectural traits associated with yield and yield components in wheat. *Ann. Bot.* 119, 1115–1129. doi: 10.1093/aob/mcx001

Yang, S. Q., Li, W. Q., Miao, H., Gan, P. F., Qiao, L., Chang, Y. L., et al. (2016). REL2, A gene encoding an unknown function protein which contains DUF630 and DUF632 domains controls leaf rolling in rice. *Rice* 9:6. doi: 10.1186/s12284-016-0105-6

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546

Zhu, J. K. (2002). Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* 53, 247–273. doi: 10.1146/annurev.arplant.53.091401.143329

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fgene.2021.651785

Genome-Wide Association Study of Kernel Traits in *Aegilops tauschii*

Qing Wang[1,2†], Ning Yan[1,2†], Hao Chen[1,2], Sirui Li[3], Haiyan Hu[4], Yu Lin[1,2], Haoran Shi[1,2], Kunyu Zhou[1,2], Xiaojun Jiang[1,2], Shifan Yu[1,2], Caixia Li[2], Guangdeng Chen[5], Zisong Yang[6] and Yaxi Liu[1,2]*

[1] State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Chengdu, China, [2] Triticeae Research Institute, Sichuan Agricultural University, Chengdu, China, [3] Chengdu Foreign Language School, Chengdu, China, [4] School of Life Sciences and Technology, Henan Institute of Science and Technology, Xinxiang, China, [5] College of Resources, Sichuan Agricultural University, Chengdu, China, [6] College of Resources and Environment, Aba Teachers University, Wenchuan, China

OPEN ACCESS

Edited by:
Penghao Wang, Murdoch University, Australia

Reviewed by:
Yong Jia, Murdoch University, Australia
Zehou Liu, Sichuan Academy of Agricultural Sciences, China
Junyan Feng, Sichuan Academy of Agricultural Sciences, China

***Correspondence:**
Yaxi Liu
liuyaxi@sicau.edu.cn
orcid.org/0000-0001-6814-7218

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Genomics, a section of the journal Frontiers in Genetics

Received: *11 January 2021*
Accepted: *04 May 2021*
Published: *28 May 2021*

Citation:
Wang Q, Yan N, Chen H, Li S, Hu H, Lin Y, Shi H, Zhou K, Jiang X, Yu S, Li C, Chen G, Yang Z and Liu Y (2021) Genome-Wide Association Study of Kernel Traits in Aegilops tauschii. Front. Genet. 12:651785. doi: 10.3389/fgene.2021.651785

Aegilops tauschii is the diploid progenitor of the D subgenome of hexaploid wheat (*Triticum aest*ivum L.). Here, the phenotypic data of kernel length (KL), kernel width (KW), kernel volume (KV), kernel surface area (KSA), kernel width to length ratio (KWL), and hundred-kernel weight (HKW) for 223 *A. tauschii* accessions were gathered across three continuous years. Based on population structure analysis, 223 *A. tauschii* were divided into two subpopulations, namely T-group (mainly included *A. tauschii* ssp. *tauschii* accessions) and S-group (mainly included *A. tauschii* ssp. *strangulata*). Classifications based on cluster analysis were highly consistent with the population structure results. Meanwhile, the extent of linkage disequilibrium decay distance (r^2 = 0.5) was about 110 kb and 290 kb for T-group and S-group, respectively. Furthermore, a genome-wide association analysis was performed on these kernel traits using 6,723 single nucleotide polymorphism (SNP) markers. Sixty-six significant markers, distributed on all seven chromosomes, were identified using a mixed linear model explaining 4.82–13.36% of the phenotypic variations. Among them, 15, 28, 22, 14, 21, and 13 SNPs were identified for KL, KW, KV, KSA, KWL, and HKW, respectively. Moreover, six candidate genes that may control kernel traits were identified (*AET2Gv20774800, AET4Gv20799000, AET5Gv20005900, AET5Gv20084100, AET7Gv20644900,* and *AET5Gv21111700*). The transfer of beneficial genes from *A. tauschii* to wheat using marker-assisted selection will broaden the wheat D subgenome improve the efficiency of breeding.

Keywords: *Aegilops tauschii*, candidate gene, genetic diversity, GWAS, kernel traits, SNP

INTRODUCTION

Aegilops tauschii (2n = 2× = 14, DD) is the diploid progenitor of the D subgenome of hexaploid wheat (*Triticum aestivum* L., 2n = 6× = 42, AABBDD) and a vital genetic resource for the improvement of wheat quality and yield (Dvořák et al., 1998; Ogbonnaya et al., 2013). The *A. tauschii* has a rich genetic diversity and multiple biological and abiotic resistances, including excellent genetic resources such as stress resistance (Qin et al., 2016), disease resistance (Zhang et al., 2019), and improved yield, which are uncommon in ordinary hexaploid wheat.

Hexaploid wheat arose through natural hybridization and chromosome doubling between a cultivated allotetraploid (2n = 4× = 28, AABB) and *A. tauschii* (Dvořák et al., 1998; Matsuoka, 2011). However, common wheat descends from a small number of spontaneous interspecific hybrids (Cox, 1997). Thus, there is scope for *A. tauschii* to improve wheat and increase wheat yield by artificially synthesizing hexaploids. After several collaborative long-time research efforts, the International Center for Maize and Wheat Improvement (CIMMYT) have synthesized hexaploid wheat lines by crossing elite tetraploid durum with *A. tauschii* (Ogbonnaya et al., 2013). The *A. tauschii* has several yield traits or components that may be transferred to synthetic hexaploid wheat when used as a paternal parent. Furthermore, it has previously been used to introgress yield traits into wheat, such as the large-kernel wheat Chuanmai 42 (Zhang et al., 2004) and heavy panicle Shumai 830 (Hao et al., 2019). With the rapid development of scientific research technology, identification of the linked markers in the genetic background of *A. tauschii* can enable targeted introgressions, thus making it economical.

Single nucleotide polymorphism (SNP) is a third-generation genetic marker technology. SNPs are abundant and have traits such as high frequency and good genetic stability. Currently, SNP genetic studies are widely used for kernel yield, disease resistance, and stress resistance. Genome-wide association study (GWAS) based on linkage disequilibrium (LD) has been widely adopted to identify loci significantly associated with important and complex morphological traits in several species, including *A. tauschii* (Liu et al., 2015a,b), rice (Chen et al., 2014), wheat (Lin Y. et al., 2017, 2019, 2020a; Liu Y. et al., 2017), and maize (Lu et al., 2010, 2011 Yang et al., 2014). Moreover, only a few GWAS have reported kernel size traits in *A. tauschii*. For example, using 193 *A. tauschii* accessions worldwide, 58 SSR were identified in three environments for seven grain traits (Zhao et al., 2015). Using 5,249 SNPs, a GWAS was performed for 114 *A. tauschii* germplasm, and a total of 17 SNPs associated with grain size traits distributed over all the seven chromosomes (Arora et al., 2017). However, this study aimed to investigate marker-trait associations for kernel size traits using SNPs in a core collection of 223 *A. tauschii* of diverse origin. Moreover, our objective was to scan candidate gene responses to kernel size traits. These identified genes and SNPs will provide an important research framework for cloning kernel trait genes in *A. tauschii*.

MATERIALS AND METHODS

Plant Materials and Field Experiments

A total of 223 *A. tauschii* accessions were collected by Triticeae Research Institute, Sichuan Agricultural University. These *A. tauschii* accessions were originally obtained from 17 different countries (**Supplementary Table 1**). According to morphological classification criteria (Zhao et al., 2018), 135 and 88 *A. tauschii* accessions were classified as *A. tauschii* ssp. *tauschii* and *A. tauschii* ssp. *strangulata*, respectively (**Supplementary Table 1**).

All *A. tauschii* were planted in Wenjiang, Chongzhou, and Wenjiang in 2017, 2018, and 2019, respectively. Each accession was planted in three rows. Each row's length was 1.5 m, and the space between the rows was 0.6 m, as a previous study described (Liu et al., 2015b). Spikes were harvested at physiological maturity and threshed by hand. Fifty kernels of each *A. tauschii* plant were used to evaluate six traits with three repetitions. Kernel length (KL), kernel width (KW), kernel width to length ratio (KWL), kernel surface area (KSA), and kernel volume (KV) were evaluated in all three environments, and hundred-kernel weight (HKW) was evaluated in 2018 and 2019. Kernel morphologic traits, including KL, KW, KWL, KSA, and KV, were scanned using an Epson XL scanner system (11,000 ×) (Seiko Epson Corporation, Nagano-ken, Japan) and analyzed using the Win-SEEDLE Pro 2012a image analysis system (Régent Instruments, Quebec, Canada) software. Hundred-kernel weight was calculated as two times the weight of 50 kernels.

Statistical Analysis of Phenotypic Data

Analysis of variance (ANOVA) was conducted using the "car" package in the software R 3.5.1 R Core Team (2014). As HKW was only calculated in two environments, ANOVA could not be conducted for HKW. In this study, we established selection indices involving multiple kernel traits, and a series of linear regressions were performed for all traits. We built a series of linear regressions to explain HKW and chose our predictive variables through a stepwise selection process.

The broad-sense heritability was calculated using the Smith et al. (1998) method as previous studies described (Liu Y. et al., 2017; Lin et al., 2020b; Lin Y. et al., 2021). Meanwhile, to reduce the environmental impact on kernel traits, best linear unbiased predictors (BLUP) of each trait across environments were calculated using SAS 9.2 (SAS Institute Inc., Cary, NC). Descriptive analysis, Pearson's correlation, linear regression, and clustering analyses were performed based on BLUP values for each trait using SPSS 20 (IBM, United States). Moreover, three different categories were calculated based on traits, i.e., low-, mid-, and high-performing genotypes corresponding to below, between, and above X ± SD (Standard Deviation), respectively (Zar, 2010; Abdel-Ghani et al., 2012), where X represent mean values of each trait. Meanwhile, Shannon–Weaver diversity index (*H'*) was calculated based on BLUP values for six kernel traits using the formula.

$$H' = -\sum_{i=1}^{n} P_i Ln(Pi)$$

Where P_i is the number of materials in the *i* level of a specific trait in the total percentage of copies, and Ln is the natural logarithm (Hutcheson, 1970).

Genotyping and Genetic Diversity Analysis

Genomic DNA from each *A. tauschii* samples was extracted from the young leaves using the CTAB method

(Murray and Thompson, 1980). All *A. tauschii* samples were genotyped by Illumina 10K SNP array, and the gathered SNPs were mapped onto the *A. tauschii* reference genome v4.0 (Aet v4.0[1]) to obtain the physical location (Luo et al., 2017). Then, the mapped SNPs with minor allele frequency (MAF) >5% and missing data <20% were retained for further analysis. Finally, a total of 6,723 polymorphic markers were obtained and used for population structure, kinship, and association analysis. Genetic diversity was evaluated using polymorphism information content (PIC), as PIC = $1-\Sigma(\mathrm{Pi})^2$, where Pi is the proportion of the population carrying the allele (Botstein et al., 1980).

Population Structure, Kinship and Linkage Disequilibrium Analysis

Population structure was analyzed using the Bayesian inference program STRUCTURE 2.3.4 based on the linkage ancestry model (Pritchard et al., 2000; Falush et al., 2007). A total of 10,000 burn-in iterations followed by 10,000 Markov Chain Monte Carlo iterations for K = 1–10 clusters were used to identify the optimal range of K, performing 10 runs per K. The optimal value of K was determined using STRUCTURE HARVESTER (Earl and vonHoldt, 2012) based on the Evanno method (Evanno et al., 2005). The CLUMPP (Jakobsson and Rosenberg, 2007) was used to determine the best comparison among five repeated samples. Kinship was estimated using 6,723 markers in TASSEL 3.0 (Bradbury et al., 2007). The LD squared allele frequency correlation (r^2), which contains both mutational and recombination history, as evaluated for linked/syntenic loci ($p < 0.001$). The LD analyses was conducted separately for the T-group and S-group, respectively. The LD estimates between marker pairs were obtained using TASSEL 3.0, the mean r^2 over different genetic distances was calculated for the T-group and S-group, respectively.

Genome-Wide Association Analysis and Candidate Gene Prediction

Genome-wide association analysis was performed based on 6,723 SNPs using mean value of each environment and the BLUP values of each trait in Tassel 3.0 based on a mixed linear model (MLM) (Bradbury et al., 2007). The significance threshold was set at p -value < 0.001, correspondingly $-\log_{10}(p)$ = 3.00 as previous studies (Liu J. et al., 2017; Ye et al., 2019; Fu et al., 2020). Manhattan and Quantile-Quantile plots of GWAS results were plotted in R 3.5.1 (R Core Team., 2014).

Based on Aet v4.0, putative genes in 10 Kb upstream and downstream of the significant SNPs were selected and then annotated using KEGG Orthology Based Annotation System 3.0 (KOBAS 3.0) (Xie et al., 2011; Arora et al., 2017; Wu et al., 2017). Arabidopsis and rice were used as background species. Candidate genes were identified according to the homologous function.

[1]https://www.ebi.ac.uk/ena/browser/view/GCA_002575655.1

RESULTS

Marker Distribution and Population Structure Analysis

A total of 6,723 polymorphic SNPs was mapped on the *A. tauschii* reference genome Aet v4.0 with MAF >5%, missing data <20%. The 6,723 SNPs were evenly distributed on seven chromosomes of *A. tauschi* (**Supplementary Figure 1**). The number of SNPs ranged from 784 for chromosome 4D to 1,231 for chromosome 2D (**Supplementary Table 2**). The marker density ranged from 0.53 to 0.70 Mb for each chromosome (chromosomes 2 and 6D, respectively) (**Supplementary Table 2**). The PIC ranged from 0.10 to 0.50, with an average value of 0.42 for the whole subgenome (**Supplementary Table 2**), indicating a high polymorphism of SNPs.

Based on the population structure analysis, $K = 2$ was selected. Thus, the whole panel was divided into two groups (**Supplementary Table 1**). Group 1 (S-group) contained 84 *A. tauschii*, including 83 *A. tauschii* ssp. *strangulata* and one *A. tauschii* ssp. *tauschii*. Group 2 (T-group) contained 139 *A. tauschii*, including 137 of *A. tauschii* ssp. *tauschii* and two *A. tauschii* ssp. *strangulata*. Meanwhile, the LD analyses were conducted separately for the T-group and S-group two lineages. The mean r^2 values gradually decreased with increasing pairwise distance. The extent of LD decay distance ($r^2 = 0.5$) was about 110 and 290 kb for T-group and S-group, respectively (**Supplementary Figure 2**).

Phenotypic Variation and Cluster Analysis

The ANOVA results for 223 *A. tauschii* samples are listed in **Table 1**. All kernel traits showed significant ($p < 0.001$) differences among genotypes and environments, except for HKW. The coefficients of variation of the six kernel traits among three environments ranged from 8.49 to 48.99% (**Supplementary Table 3**). The heritability ranged from 0.74 for KL to 0.87 for KW, indicating medium to high heritability (**Table 2**). Based on BLUP values, coefficient of variation of six kernel traits ranged from 5.13 to 23.49% (**Table 2**). The minimum, maximum, and average values of KL, KW, KV, KSA, KWL, and HKW in the S-group were significantly ($p < 0.01$) higher than those in the T-group (**Supplementary Table 4**), and there were significant differences between the two subspecies (**Figure 1**). Results indicated that the six kernel traits in S-group exhibited higher *H'* values than those in T-group, and S-group subspecies had a wider diversity range than those in T-group subspecies. Regarding the phenotypic distribution of six kernel traits based on BLUP values, all traits frequency distribution was continuous (**Supplementary Figure 3**), indicating that kernel traits were quantitative and controlled by multiple genes.

Based on the BLUP value, correlation analysis for each trait showed significant correlations among traits, with correlation coefficients ranging from 0.27 (between KL and KW) to 0.98 (between KSA and KV) (**Table 3**). All correlations were positive, except for that between KL and KWL. KW showed medium-to-high correlations with KV, KSA, KWL, and HKW, while KL

TABLE 1 | The analysis of variance of six kernel traits of 223 *Aegilops tauschii* among three environments.

	Type III sum of square		Mean square		*F*-values		Significance†	
	Environment	Genotype	Environment	Genotype	Environment	Genotype	Environment	Genotype
DF	2	222	2	222	2	222	2	222
KL	29.18	82.05	14.59	0.37	150.61	3.82	***	***
KW	25.33	50.49	12.67	0.23	442.70	7.95	***	***
KV	662.19	475.14	331.10	2.14	798.22	5.16	***	***
KSA	10556.28	5703.71	5278.14	25.69	933.59	4.54	***	***
KWL	1.49	2.13	0.74	0.01	542.37	7.00	***	***
HKW	–	–	–	–	–	–	–	–

DF, degrees of freedom; KL, kernel length; KW, kernel width; KV, kernel volume; KSA, kernel surface area; KWL, kernel width to length ratio; HKW, hundred-kernel weight.
† ****, significant at $p < 0.001$, respectively. KL, KW, KV, KSA, and KWL were measured for 3 years, while HKW was measured for 2 years.*

TABLE 2 | Descriptive analysis, coefficient of variation, heritability, and Shannon–Weaver diversity index (*H′*) of six kernel traits based on BLUP values among the 223 *Aegilops tauschii*.

Trait	Mean ± SD	CV%	Min	Max	Heritability	*H′*		
						223 accessions	T-group†	S-group
KL (mm)	5.02 ± 0.26	5.13	4.31	5.77	0.74	0.85	0.76	0.98
KW (mm)	2.23 ± 0.24	10.70	1.83	2.89	0.87	0.89	0.79	0.93
KV (mm^3)	2.88 ± 0.68	23.49	1.84	5.01	0.81	0.80	0.76	0.77
KSA (mm^2)	17.75 ± 2.27	12.77	13.71	24.81	0.78	0.81	0.83	0.85
KWL (/)	0.45 ± 0.05	10.71	0.37	0.59	0.86	0.81	0.79	0.90
HKW (g)	0.82 ± 0.16	20.05	0.49	1.27	0.80	0.86	0.72	0.85

CV, coefficient of variation; KL, kernel length; KW, kernel width; KV, kernel volume; KSA, kernel surface area; KWL, kernel width to length ratio; HKW, hundred-kernel weight; SD, standard deviation.
† *T-group and S-group were divided based on population structure analysis.*

showed low-to-medium correlations with KV, KSA, KWL, and HKW (**Table 3**). Additionally, KW and KL showed a medium-to-high correlation with HKW. These results indicated that KW and KL were the major determinants of kernel architecture and weight in *A. tauschii*, showing that the same loci may orchestrate the control of these traits, indicating that SNPs identified in our study may play pleiotropic effects. The final model based on HKW phenotypic variation explained 76.8% of variability with KV, KL, KSA, and KW, verifying the above results (**Supplementary Table 5**). Meanwhile, cluster analysis (Ward's method) grouped the 223 *A. tauschii* into two clusters (**Supplementary Table 1**). Results showed a high consistency of classification results by population structure.

Marker Trait Associations for Kernel Size Traits

GWAS was performed on all six traits using 6,723 SNPs among three environments. A total of 141 significant SNPs were identified for six kernel traits with phenotypic variation explained (PVE) ranging from 4.82 to 17.14% (**Supplementary Table 6**). The highest number of markers was detected for KV(78), which was followed by kernel volume KW (43), KSA (42), KWL (34), KL (26), HKW (21) (**Supplementary Table 6**). Based on BLUP values, GWAS was performed on all six traits using 6,723 SNPs by MLM. A total of 66 significant SNPs was identified for six kernel traits with phenotypic variation explained (PVE) ranging from 4.82 to 13.36% (**Table 4** and **Supplementary Table 6**), and these markers were distributed on all seven chromosomes (**Figures 2**, **3**, **Table 4**, and **Supplementary Table 6**). In order to reduce environment effects, significant SNPs detected using BLUP values were used for further analysis.

Based on BLUP values, 15 significant SNPs for KL were detected with PVE, ranging from 4.85 to 9.02%, these SNPs were distributed on chromosomes 1D, 2D, and 7D (**Table 4** and **Supplementary Table 6**). For KW, 28 significant SNPs were detected with PVE, ranging from 4.83 to 8.73%. These SNPs were distributed on chromosomes 2D, 3D, 4D, 5D, and 7D. For KV, 22 significant SNPs were detected with PVE, ranging from 4.82 to 13.02%. These SNPs were distributed on chromosomes 2D, 4D, 5D, and 7D. For KSA, 14 significant SNPs were detected with PVE, ranging from 4.87 to 13.36%. These SNPs were distributed on chromosomes 2D, 4D, 5D, and 7D. For KWL, 21 significant SNPs were detected with PVE, ranging from 4.86 to 9.66%. A total of 13 significant SNPs were detected for HKW with PVE ranging from 4.82 to 7.05%. These SNPs were distributed on chromosomes 3D, 4D, 5D, 6D, and 7D. These SNPs were distributed on all seven chromosomes (**Table 4** and **Supplementary Table 6**). The *contig17143_54*, located on chromosome 5D at 538.15 Mb, was strongly associated with KV with 13.02% PVE (**Supplementary Table 6**). The *contig 17143_54*, located on chromosome 5D at 538.15 Mb, was most

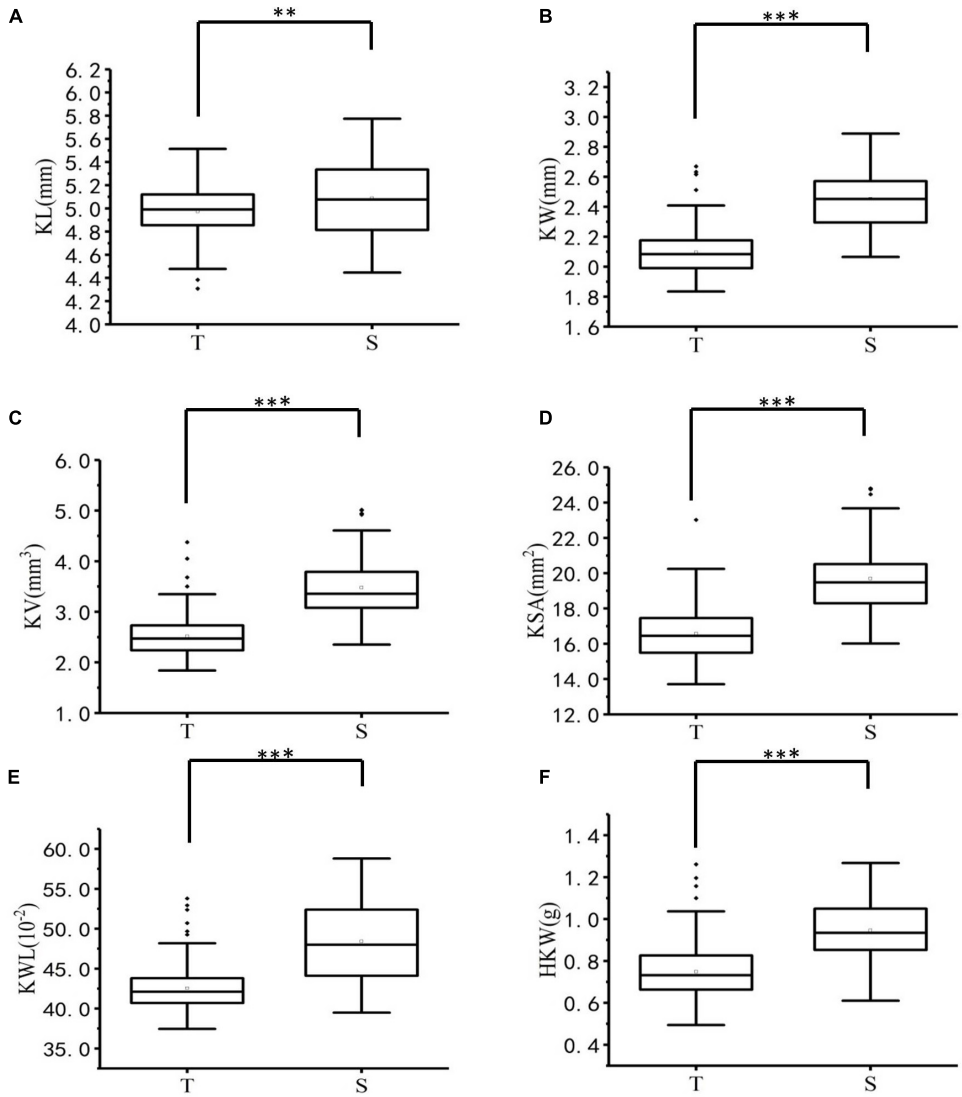

FIGURE 1 | Boxplots of six kernel characters of T-group and S-group. **, and *** denote significance at $p < 0.01$ and $p < 0.001$, respectively. **(A)** KL, kernel length (mm), **(B)** KW, kernel width (mm), **(C)** KV, kernel volume (mm^3), **(D)** KSA, kernel surface area (mm^2), **(E)** KWL, kernel width to length ratio (%), **(F)** HKW, hundred-kernel weight (g).

significant for KSA with 13.36% PVE, while *contig67633_66*, located on chromosome 6D at 406.04 Mb, was most significant with KL with 9.66% PVE. Twenty-six loci of six kernel traits showed pleiotropy, e.g., *contig17143_54,* located on chromosome 5D at 538.15 Mb, was significantly related to KW, KSA, KV, and HKW; *GDRF1KQ01CJ4KM_378*, located on chromosome 7D at 246.23 Mb, was significantly related to HKW, KSA, KL, KV, KW, and *F1BEJMU01CNNGZ_79*, located on chromosome 4D at 453.78 Mb, was significantly related to KW, KWL, KV.

Candidate Genes That May Be Linked to Kernel Traits

Based on Aet v4.0, putative genes in 10 Kb upstream and downstream of the significant SNPs were homologous comparison using KOBAS 3.0. A total of 38 predicted genes were selected. Thirty-six and 38 genes were homologous to arabidopsis and rice, respectively (**Supplementary Table 7**). Six predicted genes, included *AET2Gv20774800*, *AET4Gv20799000*, *AET5Gv20005900*, *AET5Gv20084100*, *AET7Gv20644900,* and *AET5Gv21111700*, were homologous to *MST1* (Takeda et al., 2001; Mao et al., 2011), *MAC3B* (Monaghan et al., 2009; Li S. et al., 2018), *ETR1* (Yin et al., 2017), *ZAR1* (Guo et al., 2013; Yu et al., 2016), *NAC047* (Kunieda et al., 2008; Mathew et al., 2016), *EXPA7* (Lizana et al., 2010; Jadamba et al., 2020), respectively (**Table 5**). These genes (*MST1*, *MAC3B*, *ETR1*, *ZAR1*, *NAC047*, *EXPA7*) could affect embryo development, cause seed surface atrophy, increase the number of cells to increase organ size, or affect kernel size through ethylene response. Thus, the six *A. tauschii* genes

TABLE 3 | The correlation analysis of the six kernel traits based on best linear unbiased prediction (BLUP) values.

Trait	KL	KW	KV	KSA	KWL	HKW
KL	1					
KW	0.27**	1				
KV	0.47**	0.96**	1			
KSA	0.61**	0.90**	0.98**	1		
KWL	−0.30**	0.83**	0.67**	0.53**	1	
HKW	0.56**	0.79**	0.84**	0.84**	0.46**	1

*KL, kernel length; KW, kernel width; KV, kernel volume; KSA, kernel surface area; KWL, kernel width to length ratio; HKW, hundred-kernel weight. ** represented significance at $p < 0.01$ respectively.*

TABLE 4 | Significant SNP markers identified for six kernel-related traits by genome-wide association study based on best linear unbiased prediction (BLUP) values.

Trait	Number	Chromosome	Mean $-\log_{10}(p)$	$-\log_{10}(p)$ range	Mean PVE (%)	PVE range (%)
KL	15	1D/2D/7D	3.70	3.02–5.11	6.14	4.85–9.02
KW	28	2D/3D/4D/5D/7D	3.54	3.01–5.16	5.84	4.83–8.73
KV	22	2D/4D/5D/7D	3.74	3.02–6.64	6.22	4.82–13.02
KSA	14	2D//4D/5D/7D	4.22	3.05–6.83	7.13	4.87–13.36
KWL	21	1D/2D/3D/4D/5D/6D/7D	4.14	3.03–5.65	6.92	4.86–9.66
HKW	13	3D/4D/5D/6D/7D	3.22	3.01–3.74	5.27	4.82–7.05

KL, kernel length; KW, kernel width; KV, kernel volume; KSA, kernel surface area; KWL, kernel width to length ratio; HKW, hundred-kernel weight; PVE, phenotypic variation explained.

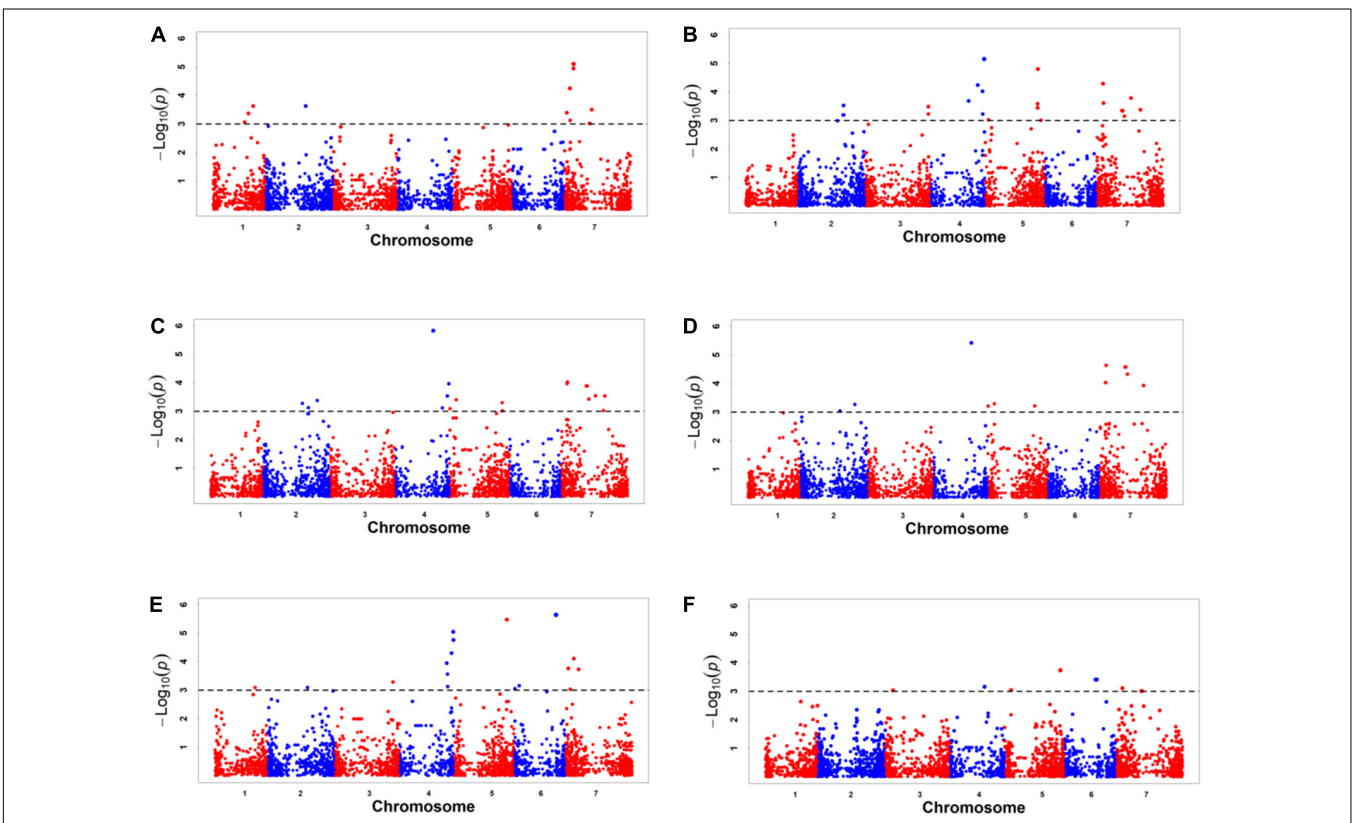

FIGURE 2 | Manhattan plots of genome-wide association study results for six kernel traits based on BLUP value. **(A)** KL, kernel length, **(B)** KW, kernel width, **(C)** KV, kernel volume, **(D)** KSA, kernel surface area, **(E)** KWL, kernel width to length ratio, **(F)** HKW, hundred-kernel weight.

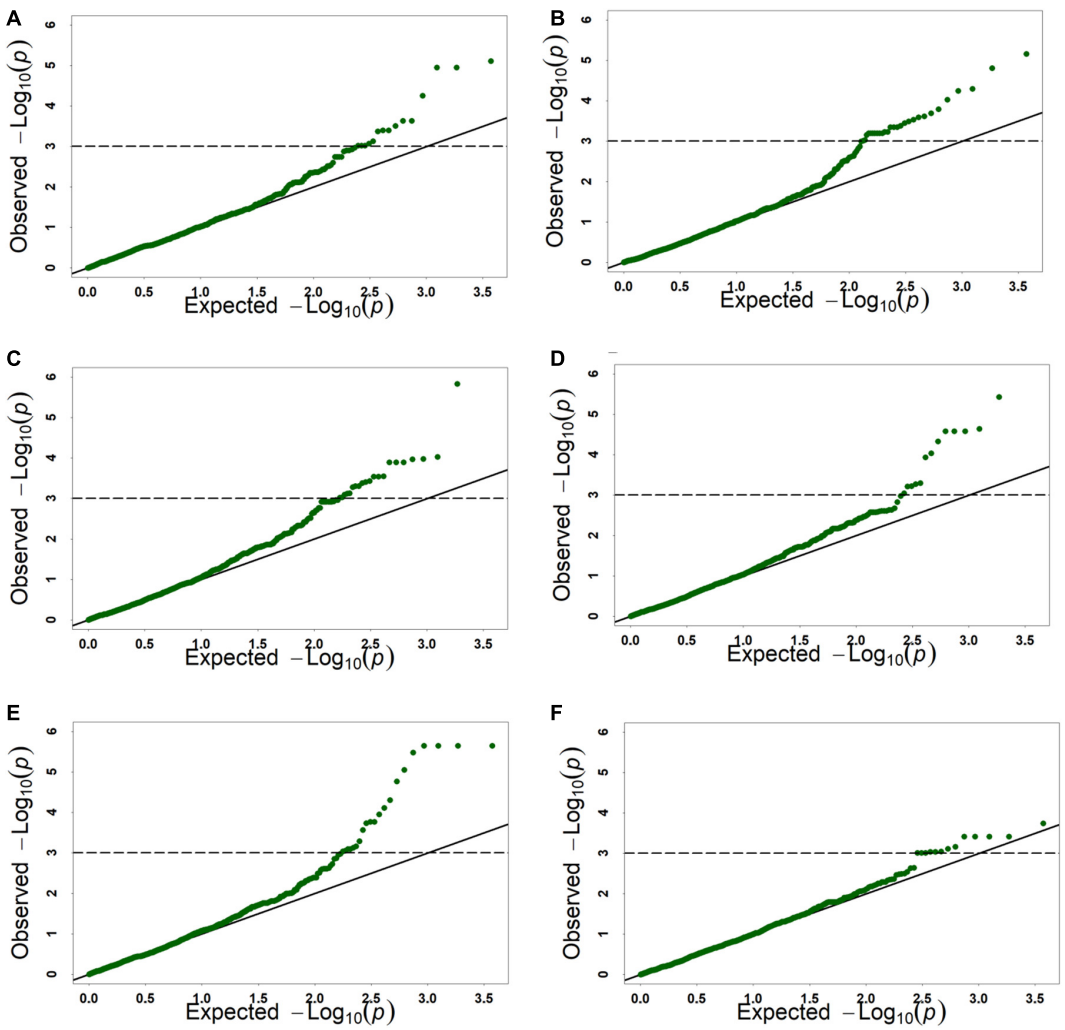

FIGURE 3 | The Q-Q plots of genome-wide association study results for six kernel traits based on BLUP value. **(A)** KL, kernel length, **(B)** KW, kernel width, **(C)** KV, kernel volume, **(D)** KSA, kernel surface area, **(E)** KWL, kernel width to length ratio, **(F)** HKW, hundred-kernel weight.

TABLE 5 | Candidate genes identified for six kernel traits.

A. tauschii gene	Marker	Chr	Position (Mb)	Trait	Rice gene	Arabidopsis gene	Putative candidate genes
AET2Gv20774800	*GBF1XID01 D2CAC_283*	2D	438.95	KW		*MST1*	Monosaccharide transporters (*MST1*)
AET4Gv20799000	*GA8KES401 CWBR7_178*	4D	501.57	KW, KWL		*MAC3B*	U-Box Proteins (*MAC3B*)
AET5Gv20005900	*be405667Contig 1ATwsnp1*	5D	2.89	KV, KSA		*ETR1*	Ethylene receptor protein (*ETR1*)
AET5Gv20084100	*F5XZDLF01A U4HH_125*	5D	32.51	KW		*ZAR1*	RLK/Pelle kinase family (*ZAR1*)
AET7Gv20644900	*GDRF1KQ01C J4KM_378*	7D	246.23	KW, KL,HKW, KV,KSA		*NAC047*	*NAC* Family Proteins
AET5Gv21111700	*contig1 7143_54*	5D	538.15	KW, KV, KSA, HKW	*EXPA7*		Expansin genes (*EXPA7*)

Chr, chromosome; KL, kernel length; KW, kernel width; KV, kernel volume; KSA, kernel surface area; KWL, kernel width to length ratio; HKW, hundred-kernel weight.

maybe directly or indirectly regulate kernel growth or regulate kernel size.

DISCUSSION

The improvement of common wheat has gone through the cross between landraces and main popularized varieties to the cross between elite varieties now. However, wheat has been affected by domestication and selection of long-term backbone parents, and genetic "evolutionary bottlenecks" have appeared, which leads to a decrease in yield. Wheat kernel traits are the most important factor affecting yield, and excellent kernel traits greatly increase yield. However, *A. tauschii* is one of the ancestral species of hexaploid wheat and the donor species of D subgenome. It has a lot of valuable genes and a rich genetic diversity, and a vital genetic resource for the improvement of wheat quality and yield (Dvořák et al., 1998; Ogbonnaya et al., 2013). The purpose of this research is to dig out the excellent genes that regulate kernel size in *A. tauschii*, and lay a foundation for the transfer of common wheat and the broadening of genetic diversity. There have been many successful cases of introducing *A. tauschii* genes into wheat through hybridization. Such as Chuanmai 42, Shumai 969, and Shumai 830 etc. thereinto, Chuanmai 42 is a large and heavy spike cultivar with a large kernel (Zhang et al., 2004; Duan et al., 2006), and Shumai 830, is also a heavy spike cultivar (Hao et al., 2019). These successful wheat varieties suggest the considerable potential of *A. tauschii* for wheat improvement, especially for breeding cultivars with large and heavy spikes. However, the aim of this study was to discover genes in *A. tauschii* that regulate kernel size, introducing the significant target markers that affect kernel traits directly into wheat would accelerate the breeding of target varieties and save time.

In this study, 223 *A. tauschii* were divided into T-group and S-group subgroups through population STRUCTURE. T-group and S-group representing *A. tauschii* ssp. *tauschii* and *A. tauschii* ssp. *strangulata,* respectively. The results of this study are completely consistent with previous studies (Arora et al., 2017). Population structure is one of the most important factors affecting LD (Flint-Garcia et al., 2003). Thus, the LD analyses was conducted separately for the T-group and S-group, respectively. Only one previous study reported LD decay distance in *A. tauschii*. The LD decay distance was reported at 9.8 and 2.7 cM for T-group and S-group, respectively (Arora et al., 2017). This study firstly reported the LD decay distance based on physical distance. The LD decay distance was highest in the T-group (approximately 110 kb) and S-group (approximately 290 kb), the average LD decay distance was approximately 200 kb. In wheat, the LD decay distance was 250 kb for D subgenome, consistent with our results (Long et al., 2019).

The present study aimed to identify significant markers for kernel size trait in *A. tauschii*, the D subgenome donor of hexaploid wheat. Significant ($p < 0.001$) differences were noted among genotypes and environments; *A. tauschii* showed high diversity, indicating high research and utilization value, and it could make a major contribution to broadening the genetic diversity of the wheat D subgenome. In the present study, heritabilities of kernel traits were medium to high. In previous studies, moderate or high heritabilities were also observed in *A. tauschii* (Arora et al., 2017), consistent with our study. In some reports in wheat, moderate to high heritability were also observed (Kuchel et al., 2007; Liu Y. et al., 2017). These results indicated that the kernel size-related traits was more controlled by genetic factors. We used correlation and linear regression analyses revealed a significant positive correlation between KL, KW, KV, and HKW, and HKW increased with increasing KL, KW, and KV. Previous studies have found that, in the tetraploid and hexaploid wheat, *TaGW2-A1* mutations could increase KW and KL, thereby increasing yield by increasing TKW (Simmonds et al., 2016). The correlations between these trait points to a causal relationship between kernel size and weight because longer and wider kernels can accumulate more starch and, therefore, have greater kernel weight (Duan and Sun, 2005). Our study showed the consistent results. Thus, in the process of breeding, breeders could pay attention to discovering varieties with long or wide kernels. The identified *A. tauschii* accessions with long or wide kernels could be used in further breeding through SHW to broaden the genetic diversity of wheat.

In the present study, cluster analysis results were highly consistent (83%) with the population structure results. Similar results have been reported in *A. tauschii* and wheat (Liu et al., 2015b; Liu Y. et al., 2017). A previous cluster analysis performed using 29 morphological traits in 322 *A. tauschii* showed 72% consistency with the population structure results (Liu et al., 2015b). This result may be caused by an intermediate type between *A. tauschii* ssp. *tauschii* and *A. tauschii* ssp. *strangulata* subspecies (Zhao et al., 2018). Indeed, Kihara et al. (1965) reported that intermediate forms and hybrids existed between the two subspecies. Due to the presence of intermediate type or hybrids, there may be uncertainty in morphological identification, resulting in differences between morphology- and genotype-based identification. Besides, morphological traits may also be more easily affected by the ecological environment. Plant growth will be affected to varying degrees in different ecological environments, affecting plant growth and, ultimately, kernel size (Wang et al., 2009; Liu et al., 2015b). Thus, the morphological traits clustering is roughly correct, but there will be some classification errors. Compared with genotype data classification results, cluster analysis could correctly classify most *A. tauschii*.

For GWAS results, the threshold is set to Bonferroni correction method a = 0.01 or 0.05. Because MLM is too strict and can lead to over-correction. The Bonferroni correction method a = 1 to reduce negative errors caused by overcorrection (Yang et al., 2014). The *p*-value was 1.49×10^{-4} for the 6,723 SNPs, with a corresponding $-\log_{10}(p) = 3.80$. However, only 22 significant SNPs were identified. This is due to overcorrection caused by MLM. In addition, $-\log_{10}(p) = 3.00$ is also commonly set as a threshold (Liu J. et al., 2017; Ye et al., 2019; Fu et al., 2020) to reduce the negative false rate. Meanwhile, it has also been successfully applied in GWAS in wheat (Liu J. et al., 2017; Ye et al., 2019; Fu et al., 2020). This indicates that this threshold is a frequently used empirical value and the results are reliable. In this study, the threshold setting as $-\log_{10}(p) = 3.00$, 66

significant loci were identified for kernel traits using 6,723 SNPs markers by GWAS. To the best of our knowledge, to date, only two studies have reported QTL for kernel characteristic traits in *A. tauschii* based on GWAS (Zhao et al., 2015; Arora et al., 2017). Owing to the lack of publicly available marker sequence information, our results were compared to those of previous studies based on chromosomes. In previous studies for KL, significant loci were identified on chromosomes 1D, 2D, 5D, and 6D (Zhao et al., 2015; Arora et al., 2017), while in this study, 11 significant markers were identified on chromosome 7D; these may represent novel loci. For KW, we detected 28 significant markers on chromosomes 2D, 3D, 4D, 5D, and 7D. Significant markers on these chromosomes were also found in previous studies (Zhao et al., 2015; Arora et al., 2017). For KSA, 14 significant markers were identified on chromosomes 2D, 4D, 5D, and 7D, and loci on chromosome 4D may be novel according to a comparison with previous research (Zhao et al., 2015). For HKW, loci were identified on all chromosomes in previous studies, except for 7D (Zhao et al., 2015; Arora et al., 2017). In this study, significant markers were identified on chromosome 7D, as well as on chromosomes 3D, 4D, 5D, and 6D. Besides, 21 significant markers for KWL were identified on all seven chromosomes, and 22 for KV were identified on chromosomes 2D, 4D, 5D, and 7D. GWAS is an important and effective approach for wheat breeding by helping in the design of hybrid crosses. The identified significant markers/variants can be designed for use as molecular markers in wheat breeding directly.

To further serve the breeding of target varieties with desirable kernel traits, six candidate genes were identified based on homologous functions in arabidopsis and rice. *AET2Gv20774800* was flanking the SNP marker *GBF1XID01D2CAC_283*. This gene was homologous to the *MST1* gene in arabidopsis; they are also called sulfurtransferases 1 (*STR1*). Of note, a mutation of *STR1* alone resulted in a shrunken seed phenotype. The shrunken seed phenotype was associated with delayed/arrested embryo development (Mao et al., 2011). In addition, the *MST1* family gene *OsMST5* plays an important role in early seed development in rice (Takeda et al., 2001). *AET5Gv20084100* was flanking SNP marker *F5XZDLF01AU4HH_125* for KW on chromosome 5D. It was homologous to the *ZAR1* gene in arabidopsis, which belongs to the *RLK/Pelle* kinase family (Yu et al., 2016). Maize *ARGOS1* (*ZAR1*) transgenic alleles increase hybrid maize yield, as *ZAR1* increased plant and organ size primarily through increasing cell numbers (Guo et al., 2013). *AET5Gv20005900* was flanking SNP marker *be405667Contig1ATwsnp1* for KV and KSA on chromosome 5D, and it was homologous to *ETR1* in arabidopsis, respectively. In Rice, a reduction of *ETR2* expression could increase the thousand-seed weight (Wuriyanghan et al., 2009). However, enhanced ethylene response also may be related to a larger or heavier kernel (Yin et al., 2017). The starch and protein of rice kernel determine factors of seed dry weight and size (Duan and Sun, 2005; Yin et al., 2017). Thus, *AET5Gv20005900* may presumably affect starch accumulation in *A. tauschii*, finally affecting kernel size. The *AET4Gv20799000* gene related to KW and KWL was predicted on the 4D chromosome. It was homologous to the *MAC3B* gene in arabidopsis, which belongs to the *U-box* family. *U-box* is a ubiquitin ligase activity-related protein domain in plants (Li and Li, 2014; Li S. et al., 2018). Ubiquitin ligases have been identified as key factors of seed size control in plants (Li N. et al., 2018). For example, grain width gene *GW2* encodes a *RING-type E3* ubiquitin ligase and controls kernel width and weight in rice (Matsuoka and Ashikari, 2007). Finally, *AET7Gv20644900* was homologous to *NAC047* in arabidopsis and belonged to the *NAC* gene family. It was reported that *NAC2* regulates embryogenesis, affecting seed shapes in arabidopsis (Kunieda et al., 2008; Mathew et al., 2016). As *NAC2* and *NAC047* belong to the same family, we speculate that these candidate genes may affect embryo development, kernel size, and yield. The genes identified by GWAS can be speeds up selective breeding using *CRISPR-Cas9* system, which is a powerful tool for rapid and effective genetic improvement and allows several QTL/genes to be edited precisely and simultaneously or even novel alleles to be created.

CONCLUSION

This in-depth study of *A. tauschii* provides new insight into its potential role in wheat improvement. Six kernel traits, including KL, KW, KWL, KSA, KV, and HKW, were evaluated among 223 *A. tauschii* over 3 years. *H*' was in the range 0.80-0.89, showing that *A. tauschii* had high diversity. Kernel traits showed medium to high heritability (0.74-0.87), and correlation and linear regressions analyses showed that HKW increased with increasing KL, KW, and KV. Kernel size traits affected kernel weight and, subsequently, yield. Our research results revealed that there are favorable varieties with longer and wider kernels in both subspecies of *A. tauschii*. Based on BLUP values, a total of 66 significant SNPs was identified using GWAS, and six candidate genes were identified as potential genetic drivers of these yield-related traits. The identified SNPs/genes will speed up the wheat breeding by MAS and genome-editing technology. It is expected that the excellent target gene from the D subgenome can be successfully introduced into wheat, so as to increase the yield of wheat and broaden the genomic resources of wheat.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

QW and NY drafted and revised the manuscript and contributed to data analysis. HC, HH, YL, HS, KZ, XJ, and SY performed the phenotypic evaluation and helped with data analysis. SL helped to perform linkage disequilibrium analysis and revise the manuscript. CL, GC, and ZY helped to draft the manuscript. YL designed and coordinated the study and revised the manuscript. All authors have read and approved the final manuscript for publication.

FUNDING

This study was supported by the National Natural Science Foundation of China (91731305 and 31771794), the National Key Research and Development Program of China (2016YFD0101004 and 2017YFD0100900), the International Science and Technology Cooperation Program of the Bureau of Science and Technology of Chengdu China (No. 2015DFA306002015-GH03-00008-HZ), and the Sichuan Science and Technology Program (2019YFN0150).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.651785/full#supplementary-material

Supplementary Figure 1 | The distribution of 6723 SNP markers on seven chromosomes of *Aegilops tauschii*.

Supplementary Figure 2 | Linkage disequilibrium (LD) decay plot for T-group and S-group.

Supplementary Figure 3 | Frequency distribution of six kernel trait in *Aegilops tauschii*.

Supplementary Table 1 | Genotype classification, morphological classification and cluster analysis of 223 *Aegilops tauschii* accessions.

Supplementary Table 2 | Distribution of SNP markers on chromosomes and polymorphism information content (PIC) of diverse chromosomes.

Supplementary Table 3 | Descriptive analysis, coefficient of variation of six kernel traits among environments.

Supplementary Table 4 | Comparison of six kernel characters between T-group and S-group.

Supplementary Table 5 | Multiple linear stepwise regression to explain hundred-kernel weight (HKW) from other kernel traits based on BLUP values.

Supplementary Table 6 | Through genome-wide association analysis of various environmental traits in 223 materials, significant SNPs were identified for kernel traits.

Supplementary Table 7 | Genes Homologous annotation using KOBAS 3.0.

Supplementary Table 8 | The six kernel traits measured in multiple trials.

REFERENCES

Abdel-Ghani, A. H., Kumar, B., Reyes-Matamoros, J., Gonzalez-Portilla, P. J., Jansen, C., Martín, J. S., et al. (2012). Genotypic variation and relationships between seedling and adult plant traits in maize (*Zea mays* L.) inbred lines grown under contrasting nitrogen levels. *Euphytica* 189, 123–133. doi: 10.1007/s10681-012-0759-0

Arora, S., Singh, N., Kaur, S., Bains, N. S., Uauy, C., and Poland, J. (2017). Genome-wide association study of grain architecture in wild wheat *Aegilops tauschii*. *Front. Plant Sci.* 8:886. doi: 10.3389/fpls.2017.00886

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 323, 314–331. doi: 10.1016/0165-1161(81)90274-0

Bradbury, P., Zhang, Z., Kroon, D., Casstevens, T. M., Ramdoss, Y., and Buckler, E. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 2319, 2633–2635. doi: 10.1093/bioinformatics/btm308

Chen, W., Gao, Y., Xie, W., Gong, L., Lu, K., Wang, W., et al. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.*46, 714–721. doi: 10.1038/ng.3007

Cox, T. S. (1997). Deepening the wheat gene pool. *J. Crop Product.* 1, 1–25. doi: 10.1300/J144v01n01_01

Duan, M., and Sun, S. (2005). Profiling the expression of genes controlling rice grain quality. *Plant Mol. Biol.* 59, 165–178. doi: 10.1007/s11103-004-7507-3

Duan, X., Zhang, Y., and Yang, W. (2006). A breakthrough high-yield, high-quality wheat variety Chuanmai 42 with high resistance to stripe rust [in China]. *Sichuan Agric. Sci. Technol.* 1:13.

Dvořák, J., Luo, M., Yang, Z., and Zhang, H. (1998). The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. *Theor. Appl. Genet.* 97, 657–670. doi: 10.1007/s001220050942

Earl, D., and vonHoldt, B. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv. Genet. Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 148, 2611–2620. doi: 10.1111/J.1365-294X.2005.02553.X

Falush, D., Stephens, M., and Pritchard, J. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7, 574–578. doi: 10.1111/j.1471-8286.2007.01758.x

Flint-Garcia, S., Thornsberry, J., and Buckler, E. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–374.

Fu, L., Wu, J., Yang, S., Jin, Y., Liu, J., Yang, M., et al. (2020). Genome-wide association analysis of stem water-soluble carbohydrate content in bread wheat. *Theor. Appl. Genet.* 133, 2897–2914. doi: 10.1007/s00122-020-03640-x

Guo, M., Rupe, M. A., Wei, J., Winkler, C., Goncalves-Butruille, M., Weers, B. P., et al. (2013). Maize ARGOS1 (ZAR1) transgenic alleles increase hybrid maize yield. *J. Exp. Bot.* 65, 249–260. doi: 10.1093/jxb/ert370

Hao, M., Zhang, L., Zhao, L., Dai, S., Li, A., Yang, W., et al. (2019). A breeding strategy targeting the secondary gene pool of bread wheat: introgression from a synthetic hexaploid wheat. *Theor. Appl. Genet.* 132, 2285–2294. doi: 10.1007/s00122-019-03354-9

Hutcheson, K. (1970). A test for comparing diversities based on the shannon formula. *J. Theor. Biol.* 29, 151–154. doi: 10.1016/0022-5193(70)90124-4

Jadamba, C., Kang, K., Paek, N., Lee, S. I., and Yoo, S. (2020). Overexpression of rice expansin7 (Osexpa7) confers enhanced tolerance to salt stress in rice. *Int. J. Mol. Sci.* 21:454. doi: 10.3390/ijms21020454

Jakobsson, M., and Rosenberg, N. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 2314, 1801–1806. doi: 10.1093/bioinformatics/btm233

Kihara, H., Yamashita, K., and Tanaka, M. (1965). *Morphological, Physiological, Genetical and Cytological Studies in Aegilops and Triticum Collected From Pakistan, Afghanistan and Iran*. Kyoto: Comity of the Kyoto University Scientific Expedition to the Karakorum and Hindukush.

Kuchel, H., Williams, K., Langridge, P., Eagles, H. A., and Jefferies, S. (2007). Genetic dissection of grain yield in bread wheat. I. QTL analysis. *Theor. Appl. Genet.* 115, 1029–1041. doi: 10.1007/s00122-007-0629-7

Kunieda, T., Mitsuda, N., Ohme-Takagi, M., Takeda, S., Aida, M., Tasaka, M., et al. (2008). NAC family proteins NARS1/NAC2 and NARS2/NAM in the outer integument regulate embryogenesis in *Arabidopsis*. *Plant Cell Online* 20, 2631–2642. doi: 10.1105/tpc.108.060160

Li, N., and Li, Y. (2014). Ubiquitin-mediated control of seed size in plants. *Front. Plant Sci.* 5:332. doi: 10.3389/fpls.2014.00332

Li, N., Xu, R., Duan, P., and Li, Y. (2018). Control of grain size in rice. *Plant Reproduct.* 31, 237–251. doi: 10.1007/s00497-018-0333-6

Li, S., Liu, K., Zhou, B., Li, M., Zhang, S., Zeng, L., et al. (2018). MAC3A and MAC3B, two core subunits of the MOS4-associated complex, positively influence miRNA biogenesis. *Plant Cell* 30, 481–494. doi: 10.1105/tpc.17.00953

Lin, Y., Chen, G., Hu, H., Yang, X., Zhang, Z., Jiang, X., et al. (2020a). Phenotypic and genetic variation in phosphorus-deficiency-tolerance traits in Chinese wheat landraces. *BMC Plant Biol.* 20:330. doi: 10.1186/s12870-020-02492-3

Lin, Y., Jiang, X., Hu, H., Zhou, K., Wang, Q., Yu, S., et al. (2021). Qtl mapping for grain number per spikelet in wheat using a high-density genetic map. *Crop J.* doi: 10.1016/j.cj.2020.12.006

Lin, Y., Jiang, X., Tao, Y., Yang, X., Wang, Z., Wu, F., et al. (2020b). Identification and validation of stable quantitative trait loci for grain filling rate in common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 133, 2377–2385. doi: 10.1007/s00122-020-03605-0

Lin, Y., Liu, S., Liu, Y., Liu, Y., Chen, G., and Xu, J. (2017). Genome-wide association study of pre-harvest sprouting resistance in Chinese wheat founder parents. *Genet. Mol. Biol.* 40, 620–629. doi: 10.1590/1678-4685-GMB-2016-0207

Lin, Y., Yi, X., Tang, S., Chen, W., Wu, F., Yang, X., et al. (2019). Dissection of phenotypic and genetic variation of drought-related traits in diverse chinese wheat landraces. *Plant Genome* 12, 1–14. doi: 10.3835/plantgenome2019.03.0025

Liu, J., He, Z., Rasheed, A., Wen, W., Yan, J., Zhang, P., et al. (2017). Genome-wide association mapping of black point reaction in common wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 17:220. doi: 10.1186/s12870-017-1167-3

Liu, Y., Lin, Y., Gao, S., Li, Z., Ma, J., Deng, M., et al. (2017). A genome wide association study of 23 agronomic traits in Chinese wheat landraces. *Plant J.* 91, 861–873. doi: 10.1111/tpj.13614

Liu, Y., Wang, L., Deng, M., Li, Z., Lu, Y., Wang, J., et al. (2015a). Genome-wide association study of phosphorus-deficiency-tolerance traits in *Aegilops tauschii*. *Theor. Appl. Genet.* 128, 2203–2212. doi: 10.1007/s00122-015-2578-x

Liu, Y., Wang, L., Mao, S., Liu, K., Lu, Y., Wang, J., et al. (2015b). Genome-wide association study of 29 morphological traits in *Aegilops tauschii*. *Sci. Rep.* 5:15562. doi: 10.1038/sre15562

Lizana, X. C., Riegel, R., Gómez, L. D., Herrera, J. C., Isla, A., McQueen-Mason, S. J., et al. (2010). Expansins expression is associated with grain size dynamics in wheat (*Triticum aestivum* L.). *J. Exp. Bot.* 61, 1147–1157. doi: 10.1093/jxb/erp380

Long, L., Mao, X., Wang, J., Chang, X., Reynolds, M., and Jing, R. (2019). Genetic dissection of drought and heat-responsive agronomic traits in wheat. *Plant Cell Environ.* 42, 2540–2553. doi: 10.1111/pce.13577

Lu, Y., Xu, J., Yuan, Z., Hao, Z., Xie, C., Li, X., et al. (2011). Comparative LD mapping using single SNPs and haplotypes identifies QTL for plant height and biomass as secondary traits of drought tolerance in maize. *Mol. Breed.* 30, 407–418. doi: 10.1007/s11032-011-9631-5

Lu, Y., Zhang, S., Shah, T., Xie, C., Hao, Z., Li, X., et al. (2010). Joint linkage–linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. *Proc. Natl. Acad. Sci U.S.A.* 107, 19585–19590. doi: 10.1073/pnas.1006105107

Luo, M., Gu, Y., Puiu, D., Wang, H., Twardziok, S., Deal, K., et al. (2017). Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* 551, 498–502. doi: 10.1038/nature24486

Mao, G., Wang, R., Guan, Y., Liu, Y., and Zhang, S. (2011). Sulfurtransferases 1 and 2 play essential roles in embryo and seed development in *Arabidopsis thaliana*. *J. Biol. Chem.* 286, 7548–7557. doi: 10.1074/jbc.M110.182865

Mathew, I. E., Das, S., Mahto, A., and Agarwal, P. (2016). Three rice NAC transcription factors heteromerize and are associated with seed size. *Front. Plant Sci.* 7:1638. doi: 10.3389/fpls.2016.01638

Matsuoka, M., and Ashikari, M. (2007). A quantitative trait locus regulating rice grain width. *Nat. Genet.* 39, 583–584. doi: 10.1038/ng0507-583

Matsuoka, Y. (2011). Evolution of polyploid triticum wheats under cultivation: the role of domestication, natural hybridization and allopolyploid speciation in their diversification. *Plant Cell Physiol.* 525, 750–764. doi: 10.1093/pcp/pcr018

Monaghan, J., Xu, F., Gao, M., Zhao, Q., Palma, K., Long, C., et al. (2009). Two Prp19-Like U-Box proteins in the MOS4-associated complex play redundant roles in plant innate immunity. *PLoS Pathog.* 5:e1000526. doi: 10.1371/journal.ppat.1000526

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 819, 4321–4325. doi: 10.1093/nar/8.19.4321

Ogbonnaya, F., Abdalla, O., Mujeeb-Kazi, A., Kazi, A. G., Xu, S., Gosman, N., et al. (2013). Synthetic hexaploids: harnessing species of the primary gene pool for wheat improvement. *Plant Breed. Rev.* 37, 35–122. doi: 10.1002/9781118497869.CH2

Pritchard, J., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 1552, 945–959. doi: 10.1093/genetics/155.2.945

Qin, P., Lin, Y., Hu, Y., Liu, K., Mao, S., Li, Z., et al. (2016). Genome-wide association study of drought-related resistance traits in *Aegilops tauschii*. *Genet. Mol. Biol.* 39, 398–407. doi: 10.1590/1678-4685-GMB-2015-0232

R Core Team (2014). *R: A Language and Environment for Statistical Computing. MSOR Connections, 1*. Vienna: R Core Team.

Simmonds, J., Scott, P. J., Brinton, J., Mestre, T. C., Bush, M., Blanco, A. D., et al. (2016). A splice acceptor site mutation in TaGW2-A1 increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains. *Theor. Appl. Genet.* 129, 1099–1112. doi: 10.1007/s00122-016-2686-2

Smith, S. E., Kuehl, R. O., Ray, I., Hui, R., and Soleri, D. (1998). Evaluation of simple methods for estimating broad-sense heritability in stands of randomly planted genotypes. *Crop Sci.* 38, 1125–1129. doi: 10.2135/cropsci1998.0011183X003800050003x

Takeda, T., Toyofuku, K., Matsukura, C., and Yamaguchi, J. (2001). Sugar transporters involved in flowering and grain development of rice. *J. Plant Physiol.* 158, 465–470. doi: 10.1078/0176-1617-00358

Wang, R. X., Zhang, X. Y., Ling, W., Wang, R., Huang, L., Yang, G. X., et al. (2009). QTL analysis of grain size and related traits in winter wheat under different ecological environments. *Entia Agric. Sin.* 42, 398–407.

Wu, F., Yang, X., Wang, Z., Deng, M., Ma, J., Chen, G., et al. (2017). Identification of major quantitative trait loci for root diameter in synthetic hexaploid wheat under phosphorus-deficient conditions. *J. Appl. Genet.* 58, 437–447. doi: 10.1007/s13353-017-0406-5

Wuriyanghan, H., Zhang, B., Cao, W., Ma, B., Lei, G., Liu, Y., et al. (2009). The ethylene receptor ETR2 delays floral transition and affects starch accumulation in rice[W]. *Plant Cell Online* 21, 1473–1494. doi: 10.1105/tpc.108.065391

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identifycation of enriched pathways and diseases. *Nucleic Acids Res.* 39(suppl_2), W316–W322. doi: 10.1093/nar/gkr483

Yang, N., Lu, Y., Yang, X., Huang, J., Zhou, Y., Ali, F., et al. (2014). Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genet.* 10:e1004573. doi: 10.1371/journal.pgen.1004573

Ye, X., Li, J., Cheng, Y., Yao, F., Long, L., Wang, Y., et al. (2019). Genome-wide association study reveals new loci for yield-related traits in Sichuan wheat germplasm under stripe rust stress. *BMC Genom.* 20:640. doi: 10.1186/s12864-019-6005-6

Yin, C., Zhao, H., Ma, B., Chen, S., and Zhang, J. (2017). Diverse roles of ethylene in regulating agronomic traits in rice. *Front. Plant Sci.* 8:1676. doi: 10.3389/fpls.2017.01676

Yu, T., Shi, D., Jia, P., Tang, J., Li, H., Liu, J., et al. (2016). The arabidopsis receptor kinase ZAR1 is required for zygote asymmetric division and its daughter cell fate. *PLoS Genet.* 12:e1005933. doi: 10.1371/journal.pgen.1005933

Zar, J. H. (2010). Biostatistical analysis. *Q. Rev. Biol.* 18, 797–799.

Zhang, C., Huang, L., Zhang, H., Hao, Q., Lyu, B., Wang, M., et al. (2019). An ancestral NB-LRR with duplicated 3′UTRs confers stripe rust resistance in wheat and barley. *Nat. Commun.* 10:4023. doi: 10.1038/s41467-019-11872-9

Zhang, Y., Yang, W. Y., Hu, X. R., Yu, Y., Zou, Y. C., and Li, Q. M. (2004). Anaylsis of agronomic characters of new wheat variety Chuanmai 42 derived from synthetics (*Triticum duru Aegilops tauschii*). *Southwest China J. Agric. Sci.* 17, 141–145.

Zhao, J., Wang, H., Zhang, X., Du, X., Li, A., and Kong, L. (2015). Association analysis of grain traits with SSR markers between *Aegilops tauschii* and hexaploid wheat (*Triticum aestivum* L.). *J. Integr. Agric.* 14, 1936–1948. doi: 10.1016/s2095-3119(15)61070-x

Zhao, L., Ning, S., Yi, Y., Zhang, L., Yuan, Z., Wang, J., et al. (2018). Fluorescence in situ hybridization karyotyping reveals the resence of two distinct genomes in the taxon *Aegilops tauschii*. *BMC Genom.* 19:3. doi: 10.1186/s12864-017-4384-0

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 July 2021
doi: 10.3389/fgene.2021.661276

Genomic Prediction for Whole Weight, Body Shape, Meat Yield, and Color Traits in the Portuguese Oyster *Crassostrea angulata*

Sang V. Vu[1,2,3], Wayne Knibb[1], Cedric Gondro[4]*, Sankar Subramanian[1,2], Ngoc T. H. Nguyen[3], Mobashwer Alam[5], Michael Dove[6], Arthur R. Gilmour[7], In Van Vu[3], Salma Bhyan[1,2], Rick Tearle[8], Le Duy Khuong[9], Tuan Son Le[10] and Wayne O'Connor[1,2,6]**

[1] GeneCology Research Centre, University of the Sunshine Coast, Sippy Downs, QLD, Australia, [2] School of Science, Technology and Engineering, University of the Sunshine Coast, Sippy Downs, QLD, Australia, [3] Northern National Broodstock Center for Mariculture, Research Institute for Aquaculture Number 1, Hai Phong, Vietnam, [4] Department of Animal Science, College of Agriculture and Natural Resources, Michigan State University, East Lansing, MI, United States, [5] Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Saint Lucia, QLD, Australia, [6] NSW Department of Primary Industries, Port Stephens Fisheries Institute, Taylors Beach, NSW, Australia, [7] Consultant, Orange, NSW, Australia, [8] School of Animal and Veterinary Science, The University of Adelaide, Adelaide, SA, Australia, [9] Faculty of Environment, Ha Long University, Uong Bi, Vietnam, [10] Research Institute for Marine Fisheries, Ngo Quyen, Hai Phong, Vietnam

OPEN ACCESS

Edited by:
Penghao Wang, Murdoch University, Australia

Reviewed by:
Ross Houston, University of Edinburgh, United Kingdom
Francesco Tiezzi, North Carolina State University, United States

***Correspondence:**
Sang V. Vu
vuvansangria1@gmail.com
Cedric Gondro
gondroce@msu.edu
Wayne O'Connor
wayne.o'connor@dpi.nsw.gov.au

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *30 January 2021*
Accepted: *07 June 2021*
Published: *08 July 2021*

Citation:
Vu SV, Knibb W, Gondro C, Subramanian S, Nguyen NTH, Alam M, Dove M, Gilmour AR, Vu IV, Bhyan S, Tearle R, Khuong LD, Le TS and O'Connor W (2021) Genomic Prediction for Whole Weight, Body Shape, Meat Yield, and Color Traits in the Portuguese Oyster Crassostrea angulata. Front. Genet. 12:661276. doi: 10.3389/fgene.2021.661276

Genetic improvement for quality traits, especially color and meat yield, has been limited in aquaculture because the assessment of these traits requires that the animals be slaughtered first. Genotyping technologies do, however, provide an opportunity to improve the selection efficiency for these traits. The main purpose of this study is to assess the potential for using genomic information to improve meat yield (soft tissue weight and condition index), body shape (cup and fan ratios), color (shell and mantle), and whole weight traits at harvest in the Portuguese oyster, *Crassostrea angulata*. The study consisted of 647 oysters: 188 oysters from 57 full-sib families from the first generation and 459 oysters from 33 full-sib families from the second generation. The number per family ranged from two to eight oysters for the first and 12–15 oysters for the second generation. After quality control, a set of 13,048 markers were analyzed to estimate the genetic parameters (heritability and genetic correlation) and predictive accuracy of the genomic selection for these traits. The multi-locus mixed model analysis indicated high estimates of heritability for meat yield traits: 0.43 for soft tissue weight and 0.77 for condition index. The estimated genomic heritabilities were 0.45 for whole weight, 0.24 for cup ratio, and 0.33 for fan ratio and ranged from 0.14 to 0.54 for color traits. The genetic correlations among whole weight, meat yield, and body shape traits were favorably positive, suggesting that the selection for whole weight would have beneficial effects on meat yield and body shape traits. Of paramount importance is the fact that the genomic prediction showed moderate to high accuracy for the traits studied (0.38–0.92). Therefore, there are good prospects to improve whole weight, meat yield, body shape, and color traits using genomic information. A multi-trait selection program using the genomic information can boost the genetic gain and minimize inbreeding in the long-term for this population.

Keywords: genetic improvement, genomic selection, high density SNP markers, prediction accuracy, genetic parameters

INTRODUCTION

The Portuguese oyster, *Crassostrea angulata*, is becoming an important aquaculture mollusk species in various parts of the world, including Asia and Europe (Grade et al., 2016; Vu et al., 2017a; Gagnaire et al., 2018). The hatchery and production technologies are well established for this species. A genetic improvement program was established in Vietnam in 2014 using stock from three hatcheries to form the base population (Vu et al., 2017b). Initially, the breeding program focused solely on increasing the whole weight at harvest using traditional family selection approaches. The genetic evaluation of the first three generations (generation interval is 1 year) of the program showed a considerable genetic gain of around 6.0% per generation in the whole weight at harvest (Vu et al., 2019). Other traits such as meat yield, body shape ("cup" and "fan" ratios), and color of the shell and mantle have more recently been evaluated and incorporated into the breeding program due to an increasing market demand for these traits (Vu et al., 2019, 2020). Seafood color, especially for oysters that are usually sold live or in the half-shell, affects consumers' preferences and their willingness to buy because certain colors are perceived to be associated with better flavor and quality (Kahn and Wansink, 2004; Xing et al., 2018). In addition, shell shape has been reported to be related to meat yield in mollusk species (Gimin et al., 2004). Misshapen animals fetch a lower price than well-formed oysters (Lee et al., 2004). Previous studies showed that the heritability for meat yield, body shape, and color traits varied from 0.13 to 0.57 and has potential for genetic improvement (Vu et al., 2019, 2020). Conventional selection, however, is less efficient for traits that cannot be measured in living animals (Meuwissen et al., 2013). This is the case for meat yield and mantle color as these traits require slaughter before measurements can be taken. The current breeding program relies on measurements of meat yield and mantle color from the siblings of the selection candidates and on correlated traits. As a result, selection accuracy has been low and genetic progress has been slow for these traits. An alternative would be to add genomic selection for these difficult-to-measure traits to the breeding program; it is considered a more efficient way to speed up the rate of genetic progress and increase the prediction accuracy (Muranty et al., 2014).

Genomic selection is now widely used in many agricultural species, and it can significantly improve traits that are determined by many loci with small effects (Robledo et al., 2018b). Genomic selection can be used to estimate the breeding values of individuals with no pedigree information and can more accurately predict breeding values than traditional selection based on pedigree records because it uses genetic markers to build a genomic relationship matrix to be used to estimate the breeding values for individuals that better approximate *realized* relationships (Goddard and Hayes, 2007). The advantages of genomic selection over traditional pedigree-based approaches in terms of the accuracy of the predictions for polygenic traits have been well established for aquaculture species (Vallejo et al., 2017; Gutierrez et al., 2018, 2019; Hollenbeck and Johnston, 2018; Joshi et al., 2019). In addition, genomic selection is extremely useful for exploring within-family variability to increase the rates of genetic gain and to reduce generation intervals (Georges et al., 2019). In aquaculture species, the accuracy of genomic prediction reported for growth, carcass, and meat quality traits as well as disease resistance traits ranged between 0.16 and 0.83 (Castillo-Juárez et al., 2015; Zenger et al., 2018; Joshi et al., 2019; Liu et al., 2019; Yoshida et al., 2019). Genomic prediction has not yet been widely adopted for mollusk breeding, but there are a few examples of its use for traits such as growth and disease resistance in Pacific oysters, *Crassostrea gigas* (Gutierrez et al., 2018, 2019), and pearl quality in pearl oysters, *Pinctada maxima* (Jones et al., 2017). To our knowledge, there are no reports on the genomic prediction for whole weight in Portuguese oysters nor for meat yield (soft tissue weight and condition index), body shape, and color (shell and mantle) in any other oyster species.

The purpose of this study was to (i) estimate the genomic heritability and genetic correlations for meat yield, color traits, harvest whole weight, and body shape and (ii) evaluate the accuracy of genomic prediction for these traits in Portuguese oysters.

MATERIALS AND METHODS

Oyster Samples

The Portuguese oysters used in this study originated from a breeding program to improve the whole weight at harvest in Vietnam (Vu et al., 2019). The oysters were produced in a hatchery and then raised in the open ocean from spat (2–4 mm) until harvest. The oyster families were separately raised in the ocean. Tissue samples were collected from the oysters at harvest after about a 9-month period of culture. Information on the animals included sire, dam, spawning date, harvest date, and rearing condition. Phenotypic measurements taken on the oysters included whole weight, body shape, meat yield, and color traits. Tissue samples were collected from a total of 647 oysters: 188 oysters representing 57 full-sib families from the first generation and 459 oysters representing 33 full-sib families from the second generation. The number per family ranged from two to eight oysters for the first and 12–15 oysters for the second generation. Most of the oysters from the second generation were progeny of the first generation of oysters (Vu et al., 2021b). All tissue samples were preserved in 80% ethanol, stored at −80°C, and sent to Diversity Array Technology Pty. Ltd., Canberra, Australia, for sequencing.

Traits Studied

Growth

The oyster shells were cleaned in water before taking the measurements. The whole weight of oysters at harvest was recorded using electronic scales with an accuracy of 0.01 g.

Body Shape

The cup and fan ratios were used as indicators of the body shape of the oysters. The cup ratio was calculated by dividing the shell width by the shell depth, and the fan ratio was calculated by dividing the shell length by the shell depth (Walton et al., 2013).

Meat Yield

Electronic scales with an accuracy of 0.01 g were used to measure the soft tissue weight. The condition index was calculated as soft tissue weight multiplied by 100/(whole weight minus shell weight) (Lawrence and Scott, 1982).

Color

The shell and mantle colors of oysters at harvest were measured using a FRU colorimeter WR10 8 mm model (Giang et al., 2019) *via* a CIELab L*a*b colorimeter which determines the color as a number under a device-independent 3D color model (Leon et al., 2006). The three dimensions in the model are L* (whiteness), a* (redness), and b* (yellowness). The values of L* range from 0 for black to 100 for white; a* values are negative for green and positive for red; and b* values are negative for blue and positive for yellow (b* = zero, neutral color) (Giang et al., 2019).

DNA Extraction, Library Construction, and Genotyping

The frozen oyster mantle samples were sequenced using DArTSeq™ (Diversity Arrays Technology Pty. Ltd., Canberra, Australia) at a high marker density of 2.5 million sequences per sample. Genomic DNA was extracted and purified by Diversity Array Technology Pty. Ltd. (**Supplementary Table S1**). The SNP development and analysis were as described by Kilian et al. (2012). Briefly, DNA samples were digested with *Pst*I–*Sph*I (Kilian et al., 2012; Vu et al., 2021a) and then ligated using two different adaptors with two different restriction enzyme overhangs. The *Pst*I-compatible adapter consists of Illumina's flow-cell attachment, a sequencing primer, and a "staggered" varying-length barcode region (Elshire et al., 2011). The reverse adapter included a flow-cell attachment region and a *Sph*I-compatible overhang sequence. The PCR reaction used to amplify the mixed fragments of *Pst*I–*Sph*I following the reaction conditions comprised of an initial denaturation for 1 min at 94°C, followed by 30 cycles of 94°C for 20 s, 58°C for 30 s, and 72°C for 45 s, and a final extension at 72°C for 7 min.

The PCR products from each sample were fed to the c-Bot (Illumina) bridge PCR, followed by sequencing 77 nucleotide single-end reads on an Illumina Hiseq 2500. Reads with inaccurate barcode sequences were filtered out. Approximately 2,500,000 sequences per barcode/sample were used in marker calling. Finally, identical sequences were collapsed into "fastqcoll files," which were "groomed" to correct for low-quality bases. The "groomed" fastqcoll files were used in a secondary pipeline incorporating DArT SNP and SilicoDArT (presence/absence of restriction fragments in representation), called analysis algorithm DArTsoft14, which clustered all tags from all libraries at a threshold of 3 for SNP calling. Technical parameters, especially the balance of read counts for the allelic pairs, were used to parse into separate SNP loci. Additional selection criteria included analysis of approximately 1,000 controlled cross-populations and testing for Mendelian distribution of alleles to assign consistency scores to classify high-quality/low-error rate markers. Calling quality was assured by a high average read depth per locus (the average across all markers was over 30 reads/locus).

The sample and marker statistics are given in **Supplementary Tables S2,S3**, respectively. There were 18,849 SNP markers identified from genotyping by sequencing. Quality control was conducted with a call rate of 50% of SNP presence in the samples, leaving 13,048 SNP markers retained for analysis. The missing genotypes were imputed using the SVS Suite software with a default setting, changing them to average values (Bozeman, 2010).

Estimation of Narrow Sense Heritability and Genetic Correlations

Univariate linear mixed models were used to estimate the narrow sense heritabilities of the traits. The fixed effects fitted in the model were generation (age is a linear covariate nested within generation), sex determined at harvest (age based on spawning and harvest dates), and their interactions. The additive genetic effect and the residuals were used as random factors.

Heritability was calculated as the ratio of σ_a^2 on σ_p^2, where σ_a^2 is the additive genetic variance and σ_p^2 is the phenotypic variance calculated as $\sigma_p^2 = \sigma_a^2 + \sigma_e^2$ and σ_e^2 is the residual variance.

Genetic correlations between traits were estimated using a bivariate version of the same model (Mrode, 2014). All analyses were conducted using the average information restricted maximum likelihood procedures in the SVS Suite software (Bozeman, 2010).

Genomic Prediction

The accuracy of genomic prediction was estimated for all traits by fivefold cross-validation analysis (training set—80% and validation set—20%) with five replicates for each trait using gBLUP through the SVS Suite software (Bozeman, 2010). The gBLUP method assumes equal distribution and variance for individual SNP effects. The genome-wide complex trait analysis method was used to calculate the normalized genomic relationship matrix in our analysis (Yang et al., 2011). The gBLUP model is written in matrix notation as follows:

$$y = m + X\text{ß} + Ma + e$$

where y is the vector of phenotypic observations (whole weight, body shape, meat yield, and color traits), m = overall mean, and X is the incidence matrix consisting of fixed effects (generation, sex, age, and interactions) in ß. The matrix M is the incidence matrix of genetic effects, and the genetic values are $g = Ma$, such that $g_{n\times 1} \sim N(0n, G\sigma_g^2)$ and the genomic relationship matrix is $G_{n\times n} = \frac{MM'}{\sum_{j=1}^{m} 2p_jq_j}$ (VanRaden, 2008), in which $p_j's$ are the minor allele frequencies of the SNP genotypes ($q_j = 1 - p_j$) and $\sigma_g^2 = \sum_{j=1}^{m} 2p_jq_j\sigma_a^2$. Using the genomic best linear unbiased prediction (GBLUP) (VanRaden, 2008), genetic values were then fitted (on discovery population) and predicted (on validation population) by solving the mixed model equation.

The allele substitution effects (ASE) and fixed effect coefficients obtained from iterations and k-folds of the cross-validation analysis that gave the largest R^2-value were used to

predict the phenotypes of individual animals with the following model:

$$\hat{y} = X\hat{ß} + M\hat{\alpha}$$

where $\hat{y}$ is the predicted phenotypes, X is the fixed effects matrix, $\hat{ß}$ is the fixed effect coefficient, M is the genotype matrix, and $\hat{\alpha}$ is the ASE values.

The fivefold cross-validation using the gBLUP method was conducted using the SVS Suite (Bozeman, 2010). The actual and predicted phenotypes were compared using linear regression analysis. The coefficient of determination (R^2) from the regression analysis was used to assess the predictive ability of the gBLUP model for the traits studied. Furthermore, the animal effects were also predicted by traditional BLUP using the pedigree-based numerator relationship matrix for comparison with the genomic-based predictions described above. The pedigree summary of the oyster population in this study is given in **Supplementary Table S4**.

Population Genetics

Nucleotide diversity was estimated using the method developed by Nei and Li (1979). Using an in-house computer program, the genotype file obtained from DArTSeq was converted to the variable call format (VCF). This file was used as an input for the online software *VCF2PopTree* (Subramanian et al., 2019) to obtain the pairwise divergence matrix of all combinations of oyster individuals. This matrix was then used to construct a neighbor-joining tree using the software MEGA (Kumar et al., 2018). Fixation index (F_{ST}) was calculated using the nucleotide diversities of the total and sub-populations (Nei, 1973).

RESULTS

Phenotypic Data

The summary statistics of the traits are shown in **Table 1**. The average whole weight at harvest calculated from the 647 oysters in this study was 51.07 g, with the data range within three standard deviations of the mean (**Table 1**). There was considerable variation in whole weight, with a wide range from 15.71 to 69.90 g. The shell and mantle color had a significant yellow component (mean b* significantly greater than zero) and was dark (L around 25% but with considerable variation). The average cup and fan ratios indicated a typical depth/width/length ratio of 1:1.6:2.9. The largest variation levels were also found in color traits of tissues and shells, especially for shell color a*, with values from −2.07 to 54.96, and mantle color L*, from 7.24 to 63.70.

Heritability and Genetic Correlation

The genomic heritability (h^2), the proportion of total variance explained by the SNP markers, ranged from 0.14 to 0.77 (**Table 2**). The heritability of the shell color traits ranged from moderate to high (0.14–0.54), and the estimates of heritability were 0.13 for a* (red), 0.26 for b* (yellow), and 0.34 for L* (lightness). For the mantle, these traits had heritabilities of 0.54 for red, 0.49 for yellow, and 0.34 for lightness. The heritability estimates for meat yield traits were relatively high (0.43–0.77). Similarly, the estimate of heritability for whole weight was quite high at 0.45. However, the heritability estimates for body shape using the indicators of cup and fan ratios were low to moderate at 0.24 and 0.33, respectively.

Genetic Correlations

The positive genetic correlation between whole weight and soft tissue weight was high at 0.63, suggesting that selection for improved whole weight can result in desirable changes in soft tissue weight. However, low or insignificant genetic correlations were observed between whole weight and the other traits: color traits (−0.14 to 0.34), condition index (0.18), and body shape (0.08–0.14), implying that the selection for growth would result in little or no changes in these traits in this population. In addition, the genetic correlations between whole weight and all color traits were not only small, but they all had large standard errors and were therefore not significant. The genomic genetic correlations (r_g) of whole weight with the other traits are presented in the **Table 2**.

Accuracy of Genomic Prediction

Table 3 shows the prediction accuracy of genomic selection for the traits using the fivefold cross-validation method. High prediction accuracies using GBLUP were found for meat yield and color traits, where measurements cannot be taken on living animals. The accuracies were 0.71 for soft tissue weight and 0.49 for condition index. For the color traits, the accuracies ranged between 0.38 and 0.92. The GBLUP prediction accuracy was comparatively high for the harvest whole weight at 0.74; similarly, the body shape traits also showed a high accuracy, with 0.40 for cup ratio and 0.63 for fan ratio. Across all traits, the GBLUP predictions had higher prediction accuracies than the traditional BLUP prediction accuracies.

The association test of predicted and actual phenotypes for four predominant traits are presented in **Figure 1A** for the whole weight, **Figure 1B** for soft tissue, **Figure 1C** for mantle color a*, and **Figure 1D** for mantle color b*. The R^2 values varied from 0.37 for the mantle color L* to 0.91 for the shell color a*.

DISCUSSION

Our results have made five important discoveries that can be used to improve the genetic progress of the current breeding program for Portuguese oysters. Specifically, they answer the following questions.

Is There Any Genetic Variation in the Traits Studied?

Our analysis indicated a large genetic variation, with estimates of genomic heritability falling in the range of 0.14–0.77. Our results were somewhat higher than those we have previously obtained with the traditional pedigree-based relationship matrix in the same population (Vu et al., 2019). The genomic heritability for meat yield and body shape, in particular, was much larger than that reported using a pedigree-based estimation (Vu et al., 2019).

TABLE 1 | Summary of the phenotypic data.

Groups	Traits	Unit	*n*	Mean	SD	Min	Max
Growth	Whole weight	g	647	51.07	16.13	15.71	69.90
Body shape	Cup ratio	Ratio	361	1.61	0.33	0.74	2.60
	Fan ratio	Ratio	361	2.93	0.72	1.46	5.07
Meat yield	Soft tissue weight	g	490	9.55	3.13	1.05	19.92
	Condition index	Index	284	67.78	21.65	14.29	99.93
Color	Shell color a*	CIELab	306	2.35	3.71	−2.07	54.96
	Shell color b*	CIELab	306	11.58	4.25	3.99	43.02
	Shell color L*	CIELab	306	23.63	12.30	6.49	69.71
	Mantle color a*	CIELab	448	4.40	3.46	0.16	24.60
	Mantle color b*	CIELab	448	10.11	4.01	1.02	23.30
	Mantle color L*	CIELab	448	24.21	8.73	7.24	63.70

TABLE 2 | SNP heritability (h^2) and genetic correlations (r_g) of whole weight with other traits.

Groups	Traits	Heritability		Genetic correlation	
		h^2	SE	r_g	SE
Growth	Whole weight	0.45	0.06		
Body shape	Cup ratio	0.24	0.10	0.14	0.24
	Fan ratio	0.33	0.08	0.08	0.15
Meat yield	Soft tissue weight	0.43	0.07	0.63	0.10
	Condition index	0.77	0.07	0.18	0.15
Color	Shell color a*	0.14	0.07	−0.13	0.22
	Shell color b*	0.26	0.09	−0.02	0.13
	Shell color L*	0.34	0.09	0.34	0.19
	Mantle color a*	0.54	0.06	−0.13	0.13
	Mantle color b*	0.49	0.07	−0.14	0.15
	Mantle color L*	0.34	0.09	0.14	0.19

TABLE 3 | Genomic prediction accuracy of the traits with GBLUP and traditional BLUP.

Factors		Single variate linear mixed model	
Groups	Traits	GBLUP	BLUP
Growth	Whole weight	0.74	0.66
Body shape	Cup ratio	0.40	0.29
	Fan ratio	0.63	0.55
Meat yield	Soft tissue weight	0.71	0.51
	Condition index	0.49	0.22
Color	Shell color a*	0.92	0.46
	Shell color b*	0.62	0.65
	Shell color L*	0.83	0.66
	Mantle color a*	0.87	0.67
	Mantle color b*	0.89	0.44
	Mantle color L*	0.38	0.14

Similarly, the estimates of heritability based on SNP markers for mantle color traits (h^2 = 0.34–0.54) were around twice those based on pedigree (h^2 = 0.15–0.33) (Vu et al., 2020). These differences between genomic and pedigree heritability estimates of meat yield, body shape, and color traits are most likely due to errors in pedigree recording between spawning and harvesting. These results suggest that genomic selection can be extremely efficient for traits that cannot be measured directly on the selection candidates especially since oyster breeding involves a few large families, and pedigree information is difficult to record accurately. Unfortunately, there have been no other studies using a genomic selection approach to estimate heritability for meat yield, color, and body shape traits in mollusk species that could be used for comparison. Finally, the estimation of heritability for whole weight in this study was significantly higher than that reported in the Pacific oyster by Gutierrez et al. (2018) (h^2 = 0.45

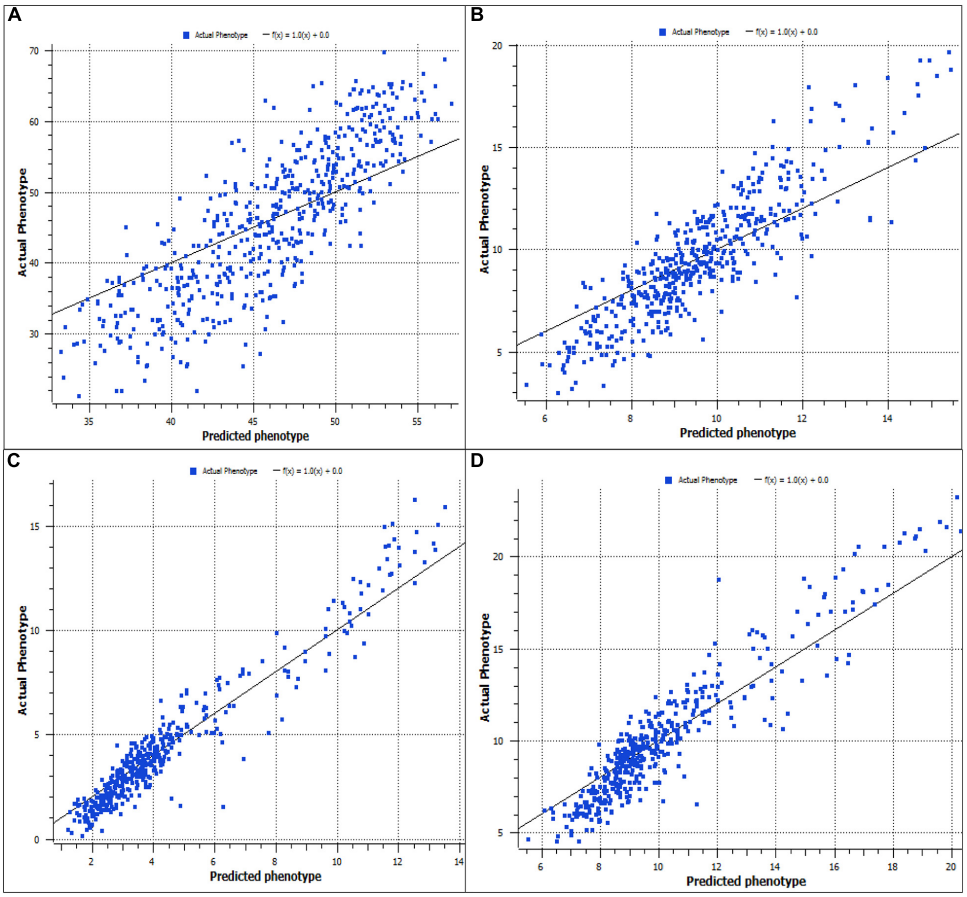

FIGURE 1 | Prediction accuracy for whole weight **(A)**, soft tissue weight **(B)**, mantle color a **(C)**, and mantle color b **(D)**.

vs. 0.35, respectively). However, estimated heritability in genomic selection can also depend on species, sample size, marker density, relationship of reference and validation population, and statistical methods for analysis (Georges et al., 2019). It should be noted that our population is highly inter-related, and the high estimates probably only pertain to selection within that population.

Does This Genomic Analysis Predict That Selection for Harvest Whole Weight Will Affect Body Shape, Meat Yield, and Color?

The genetic correlations obtained with the GBLUP analysis were generally similar to those obtained from the traditional BLUP method (Vu et al., 2019). The high and positive genetic correlation between harvest whole and soft tissue weight indicates that these two traits are controlled by some common sets of genes. This result is in line with the previous estimate by Vu et al. (2019) using the BLUP method (0.63 for GBLUP vs. 0.50 for BLUP). Therefore, soft tissue weight will increase along with the harvest whole weight under the selection for harvest whole weight. Meanwhile, the small positive genetic correlations found between harvest whole weight and body shape show that there is some/limited potential to improve these traits by selection for harvest whole weight, and any genetic progress will be very slow. Therefore, trait groups such as whole weight, body shape, and meat yield can improve in the same direction, and this suggests that the selection for one trait (harvest whole weight) will lead to an improvement in these other traits. In contrast, the negative and positive but low genetic correlations between whole weight and color traits found in this study agree with those reported in the same population using the pedigree BLUP method (Vu et al., 2020). Consequently, our results suggest that no potential exists to improve color traits by selection for harvest whole weight. Taken together, this study provides fundamental information to better understand the genetic architecture of quantitative complex traits in Portuguese oysters.

Is Genomic Selection Reliable for the Traits Studied?

Our study reports, for the first time, the predictive accuracy of genomic selection for color and meat yield traits in a mollusk species. The predictive accuracies using the GBLUP model for meat yield and color traits were significantly higher than those

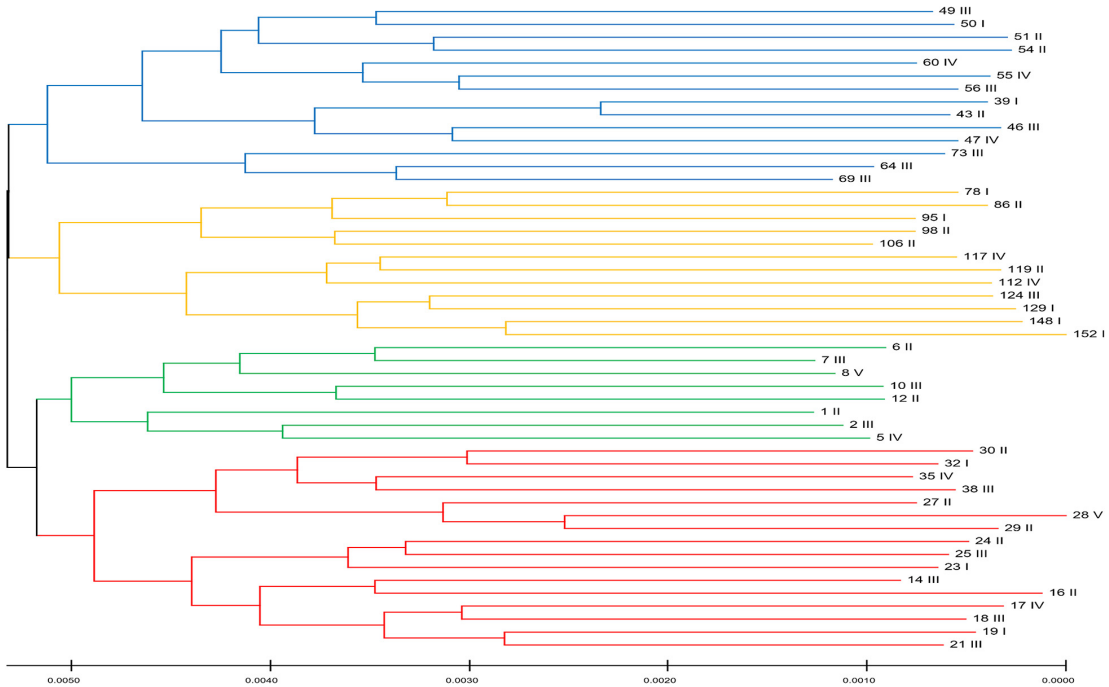

FIGURE 2 | The structure of the population for this Portuguese oyster population in the first generation. Colors represent four distinct clusters based on genetic similarities. Roman numerals represent the whole weight category that the individual belong to: I: 30–40, II: 40–50, III: 50–60, IV: 60–70, and V: >70 g. This figure demonstrates that the oysters in the population do not cluster based on their whole-body weight but only based on their genetic relationship.

obtained from the pedigree BLUP model, suggesting that the genomic selection for these traits is more efficient than the pedigree-based selection. The predictive accuracies using the GBLUP method for meat yield and color traits are in good agreement with those reported in banana shrimp (Nguyen et al., 2019). However, these accuracies were higher than those obtained for growth-related traits in the Pacific oyster (Gutierrez et al., 2018), Pacific white leg shrimp (Wang et al., 2017), and yellow drum (Liu et al., 2019). In addition, the accuracies from our study are high due to the close relationships of the sampled oysters. The differences among the studies could originate from the SNP marker density, the set of SNP markers used for the analysis (Liu et al., 2019), the training population size (Andonov et al., 2017; Zenger et al., 2018; Georges et al., 2019), the relationship between the training and validation datasets (Hayes and Goddard, 2001), the relationships between individuals in the reference population (Hayes et al., 2009), and the heritability of the traits (Daetwyler et al., 2008; Georges et al., 2019). To the best of our knowledge, no studies have reported the prediction accuracy of genomic selection for body shape in any other aquaculture species. Collectively, across aquaculture species, our estimates of accuracies for whole weight, body shape, meat yield, and color traits in Portuguese oysters fall in the range observed in other aquaculture species (Zenger et al., 2018). The high levels of prediction accuracy open new selection opportunities to improve meat yield and color traits in the Portuguese oyster.

Does Population Stratification Affect the Results of This Study?

To quantify the genetic variation in the data, we estimated the nucleotide diversity, which produced a value of 0.009, suggesting 0.9% variation in the population. This estimate was slightly less than that observed for the Pacific oyster (1.2%) (Zhang et al., 2012). One of the confounding factors in genomic selection method is the structure of the populations, owing to the effects of genetic drift (Goddard and Hayes, 2007). Some of the genetic markers can, by chance alone, correlate with the traits due to the familial or pedigree structure of the populations compared. To examine the genetic relatedness among the first-generation animals, we selected 50 individuals, excluding their siblings. We categorized these oysters based on their whole-body weight as I: 30–40, II: 40–50, III: 50–60, IV: 60–70, and V: >70 g. We then constructed a neighbor-joining tree (**Figure 2**), which revealed four distinct clusters (four colors) in the population based on their genetic similarities. The fixation index (F_{ST}) of 0.13 among these clusters suggested the existence of a population structure. However, it is evident from **Figure 2**, which shows that the whole weight of the members of each cluster varies significantly within a cluster. Therefore, these results suggest that the population structure observed in the data does not correlate with the phenotypic trait (whole weight) evaluated in this study. Hence, population stratification is unlikely to influence the results of this study.

Are There Any Prospects for Using Genomic Selection in This Portuguese Oyster Population?

Genomic selection programs can be improved by using updated estimates of genetic parameters and benefiting from the higher predictive accuracy for quantitative complex traits, especially meat yield and color traits, in the Portuguese oyster. The favorably positive genetic correlations between whole weight, body shape, and meat yield traits found in this study suggest that there is little need to use a selection index to improve these related traits. To further illustrate, the heritability and genetic correlations from our study were used in SelAction (Rutten et al., 2002) to simulate the selection process and indicated that the traits whole weight, meat yield, and body shape can be improved simultaneously. The selection for enhancing whole weight results in beneficial changes in the other traits. However, it is necessary to use a selection index to improve color traits as well as weight traits due to the low genetic correlations between whole weight and color traits.

Collectively, the results from this study showed that there are prospects for the application of genomic selection to improve a range of complex traits in the Portuguese oyster. Consequently, further studies should be carried out to collect further information on predictive power and potential genetic gain. Firstly, the training population size needs to be increased to obtain a more accurate prediction (Goddard, 2009), i.e., we need to increase our sample size from the rather small population of 647 oysters used in this study to a much larger reference population which represents more families and generations. One strategy can be a mixture of low- and high-density panels, where the broodstock is sequenced at a higher marker density and the offspring are sequenced at a lower density, and then the lower density panels are imputed up to the higher density for the genomic prediction (Yoshida et al., 2018; Tsairidou et al., 2020). This strategy allows a larger number of samples to be sequenced, thus a larger training population size. This has been done at medium density in Pacific oysters for growth-related traits or in salmon (Gutierrez et al., 2018). Secondly, the use of a genome-wide association study—informative SNPs instead of random SNPs—was able to assist in improving the predictive accuracy (Liu et al., 2019). Therefore, the use of a smaller number of informative SNPs may help to reduce the cost of genotyping in this population. Thirdly, the prediction equations were evaluated in the reference population in which both phenotypes and genotypes were available; validation should be carried out in a breeding population where the selection candidates are the training population, and validation is on subsequent generations. Fourthly, individual grouping methods such as mean weight of family based on genomic information should be considered in Portuguese oysters to improve the accuracy for further studies. These individual grouping methods have been shown to increase the predictive accuracy and resulted in a higher genetic gain (Chu et al., 2019).

From the discussion above, the optimization of genomic selection will reduce the generation interval and increase the genetic gain in this population. Lowering the generation interval may result in an increase in inbreeding rate in the long-term. Therefore, to balance between inbreeding rate and genetic gain, genetic improvement programs have applied optimal genomic selection methodology and strategies (Nielsen et al., 2011; Gorjanc and Hickey, 2018; Robledo et al., 2018a). For the Portuguese oyster sector, if genomic selection is integrated into the breeding program, it could reduce the generation interval below 1 year per generation. The lower generation interval contributes to a faster life cycle, resulting in a higher economic revenue due to saving labor and time. In addition, estimates of breeding values using genomic information are more accurate than those using physical tags in the pedigree, where errors may occur during spawning, larval rearing, or pedigree management. Therefore, genomic selection will allow a more accurate selection of oysters to become parents in the next generations. However, a major impediment to widely using genomic selection in the aquaculture industry is the cost of genotyping and phenotyping the selection candidates. The former can be dealt with by using low-density SNP panels that can balance between prediction accuracy and sequencing costs to achieve the accuracy needed for genomic selection (Lillehammer et al., 2013; Gutierrez et al., 2018; Tsairidou et al., 2020; Boudry et al., 2021; Vu et al., 2021a).

CONCLUSION

The predictive accuracy using genomic selection was relatively high for all the traits studied. In addition, the estimates of heritability in the meat yield traits such as soft tissue weight and condition index were high, while those of color traits such as shell and mantle color were from low to high. The genomic selection for improvement of whole weight leads to desirable changes in other traits, such as meat yield and body shape traits, but will not affect color traits. Future breeding programs should combine all traits into a selection index to bring about higher revenues for aquaculture farmers and entrepreneurs.

DATA AVAILABILITY STATEMENT

The DNA sequence data used in this study is available and can be accessed through the USC Research Bank under doi 10.25907/00038. The phenotypic data is available by request from the corresponding authors.

ETHICS STATEMENT

There are no requirements on ethics approval for mollusk species in Vietnam.

AUTHOR CONTRIBUTIONS

SV, WK, and WO'C conceived and designed the experiments and selected the samples for sequencing. SV and NN prepared, drafted, and edited the manuscript. SV, CG, MA, and SB analyzed the data. SS involved in the population structure analysis. TL and LK prepared the oyster tissue samples. SS, CG, MA, MD, AG, RT, IV, and WO'C participated in editing the manuscript. All authors read and approved the manuscript.

FUNDING

This work was funded by the ACIAR project "Enhancing bivalve production in northern Vietnam and Australia" (FIS/2010/100) and a John Allwright Fellowship at the GeneCology Research Centre of the University of the Sunshine Coast, Queensland, Australia.

ACKNOWLEDGMENTS

The authors thank Phan Thi Van, Director of the Research Institute for Aquaculture No. 1 (RIA1), MARD, Vietnam, and the oyster team of the Northern National Broodstock Centre for Mariculture, RIA1, MARD, who conducted the breeding and grow-out experiments. Special thanks go to Nguyen Hong Nguyen at the University of the Sunshine Coast for providing general comments on the first manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.661276/full#supplementary-material

REFERENCES

Andonov, S., Lourenco, D., Fragomeni, B., Masuda, Y., Pocrnic, I., Tsuruta, S., et al. (2017). Accuracy of breeding values in small genotyped populations using different sources of external information—a simulation study. *J. Dairy Sci.* 100, 395–401. doi: 10.3168/jds.2016-11335

Boudry, P., Allal, F., Aslam, M. L., Bargelloni, L., Bean, T. P., Brard-Fudulea, S., et al. (2021). Current status and potential of genomic selection to improve selective breeding in the main aquaculture species of international council for the exploration of the sea (ICES) member countries. *Aquac. Rep.* 20:100700. doi: 10.1016/j.aqrep.2021.100700

Bozeman, M. (2010). *Golden Helix, Inc. SNP & Variation Suite (Version 7. x)*.

Castillo-Juárez, H., Campos-Montes, G. R., Caballero-Zamora, A., and Montaldo, H. H. (2015). Genetic improvement of Pacific white shrimp *[Penaeus (Litopenaeus) vannamei]*: perspectives for genomic selection. *Front. Genet.* 6:93.

Chu, T. T., Bastiaansen, J. W., Berg, P., and Komen, H. (2019). Optimized grouping to increase accuracy of prediction of breeding values based on group records in genomic selection breeding programs. *Genet. Sel. Evol.* 51:64.

Daetwyler, H. D., Villanueva, B., and Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 3:e3395. doi: 10.1371/journal.pone.0003395

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379

Gagnaire, P.-A., Lamy, J.-B., Cornette, F., Heurtebise, S., Dégremont, L., Flahauw, E., et al. (2018). Analysis of genome-wide differentiation between native and introduced populations of the cupped oysters *Crassostrea gigas* and *Crassostrea angulata*. *Genome Biol. Evol.* 10, 2518–2534. doi: 10.1093/gbe/evy194

Georges, M., Charlier, C., and Hayes, B. (2019). Harnessing genomic information for livestock improvement. *Nat. Rev. Genet.* 20, 135–156. doi: 10.1038/s41576-018-0082-2

Giang, C. T., Knibb, W., Ninh, N. H., and Nguyen, N. H. (2019). Prospects for genetic improvement in objective measurements of body colour in Pacific Whiteleg Shrimp (*Litopenaeus vannamei*). *J. Mar. Sci. Eng.* 7:460. doi: 10.3390/jmse7120460

Gimin, R., Mohan, R., Thinh, L., and Griffiths, A. (2004). The relationship of shell dimensions and shell volume to live weight and soft tissue weight in the mangrove clam, *Polymesoda erosa* (Solander, 1786) from northern Australia. *NAGA WorldFish Center Quarterly* 27, 32–35.

Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136, 245–257. doi: 10.1007/s10709-008-9308-0

Goddard, M., and Hayes, B. (2007). Genomic selection. *J. Anim. Breed. Genet.* 124, 323–330.

Gorjanc, G., and Hickey, J. M. (2018). AlphaMate: a program for optimizing selection, maintenance of diversity and mate allocation in breeding programs. *Bioinformatics* 34, 3408–3411. doi: 10.1093/bioinformatics/bty375

Grade, A., Chairi, H., Lallias, D., Power, D. M., Ruano, F., Leitao, A., et al. (2016). New insights about the introduction of the Portuguese oyster, *Crassostrea angulata*, into the North East Atlantic from Asia based on a highly polymorphic mitochondrial region. *Aquat. Living Resour.* 29:404. doi: 10.1051/alr/2016035

Gutierrez, A. P., Matika, O., Bean, T. P., and Houston, R. D. (2018). Genomic selection for growth traits in Pacific oyster (*Crassostrea gigas*): potential of low-density marker panels for breeding value prediction. *Front. Genet.* 9:391.

Gutierrez, A. P., Symonds, J., King, N., Steiner, K., Bean, T. P., and Houston, R. D. (2019). Potential of genomic selection for improvement of resistance to Ostreid Herpes virus in Pacific oyster (*Crassostrea gigas*). *BioRxiv [Preprint]* doi: 10.1101/754473

Hayes, B. J., Visscher, P. M., and Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* 91, 47–60. doi: 10.1017/s0016672308009981

Hayes, B., and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genet. Sel. Evol.* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819

Hollenbeck, C. M., and Johnston, I. A. (2018). Genomic tools and selective breeding in molluscs. *Front. Genet.* 9:253.

Jones, D., Toole, P., Khatkar, M., Raadsma, H., Jerry, D., and Zenger, K. (2017). Developing a genomic selection breeding program for complex pearl colour traits within the silver-lipped pearl oyster. *Proc. Assoc. Advancement Anim. Breed. Genet.* 22, 537–540.

Joshi, R., Skaarud, A., de Vera, M., Alvarez, A. T., and Ødegård, J. (2019). Genomic prediction for commercial traits using univariate and multivariate approaches in Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 516:734641. doi: 10.1016/j.aquaculture.2019.734641

Kahn, B. E., and Wansink, B. (2004). The influence of assortment structure on perceived variety and consumption quantities. *J. Consum. Res.* 30, 519–533. doi: 10.1086/380286

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., et al. (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol. Biol.* 888, 67–89. doi: 10.1007/978-1-61779-870-2_5

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lawrence, D., and Scott, G. (1982). The determination and use of condition index of oysters. *Estuaries* 5, 23–27. doi: 10.2307/1352213

Lee, D.-J., Xu, X., Lane, R. M., and Zhan, P. (2004). "Shape analysis for an automatic oyster grading system," in *Two-and Three-Dimensional Vision Systems for Inspection, Control, and Metrology II*, Vol. 5606, ed. K. G. Harding (Bellingham: International Society for Optics and Photonics).

Leon, K., Mery, D., Pedreschi, F., and Leon, J. (2006). Color measurement in L* a* b* units from RGB digital images. *Food Res. Int.* 39, 1084–1091. doi: 10.1016/j.foodres.2006.03.006

Lillehammer, M., Meuwissen, T. H., and Sonesson, A. K. (2013). A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. *Genet. Sel. Evol.* 45, 1–8.

Liu, G., Dong, L., Gu, L., Han, Z., Zhang, W., Fang, M., et al. (2019). Evaluation of genomic selection for seven economic traits in yellow drum (*Nibea albiflora*). *Mar. Biotechnol.* 21, 806–812. doi: 10.1007/s10126-019-09925-7

Meuwissen, T., Hayes, B., and Goddard, M. (2013). Accelerating improvement of livestock with genomic selection. *Annu. Rev. Anim. Biosci.* 1, 221–237. doi: 10.1146/annurev-animal-031412-103705

Mrode, R. A. (2014). *Linear Models for the Prediction of Animal Breeding Values*. Wallingfort: CABI.

Muranty, H., Jorge, V., Bastien, C., Lepoittevin, C., Bouffier, L., and Sanchez, L. (2014). Potential for marker-assisted selection for forest tree breeding: lessons from 20 years of MAS in crops. *Tree Genet.* 10, 1491–1510. doi: 10.1007/s11295-014-0790-5

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. U. S. A.* 70, 3321–3323. doi: 10.1073/pnas.70.12.3321

Nei, M., and Li, W.-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* 76, 5269–5273. doi: 10.1073/pnas.76.10.5269

Nguyen, N. H., Phuthaworn, C., and Knibb, W. (2019). Genomic prediction for disease resistance to Hepatopancreatic parvovirus and growth, carcass and quality traits in *Banana shrimp Fenneropenaeus merguiensis*. *Genomics* 112, 2021–2027. doi: 10.1016/j.ygeno.2019.11.014

Nielsen, H., Sonesson, A., and Meuwissen, T. (2011). Optimum contribution selection using traditional best linear unbiased prediction and genomic breeding values in aquaculture breeding schemes. *J. Anim. Sci.* 89, 630–638. doi: 10.2527/jas.2009-2731

Robledo, D., Matika, O., Hamilton, A., and Houston, R. D. (2018a). Genome-wide association and genomic selection for resistance to amoebic gill disease in Atlantic salmon. *G3* 8, 1195–1203. doi: 10.1534/g3.118.200075

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., and Houston, R. (2018b). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev. Aquac.* 10, 670–682. doi: 10.1111/raq.12193

Rutten, M., Bijma, P., Woolliams, J., and Van Arendonk, J. (2002). SelAction: software to predict selection response and rate of inbreeding in livestock breeding programs. *J. Heredity* 93, 456–458. doi: 10.1093/jhered/93.6.456

Subramanian, S., Ramasamy, U., and Chen, D. (2019). VCF2PopTree: a client-side software to construct population phylogeny from genome-wide SNPs. *PeerJ* 7:e8213.

Tsairidou, S., Hamilton, A., Robledo, D., Bron, J. E., and Houston, R. D. (2020). Optimizing low-cost genotyping and imputation strategies for genomic selection in Atlantic salmon. *G3* 10, 581–590. doi: 10.1534/g3.119.400800

Vallejo, R. L., Leeds, T. D., Gao, G., Parsons, J. E., Martin, K. E., Evenhuis, J. P., et al. (2017). Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genet. Sel. Evol.* 49:17.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Vu, S. V., Gondro, C., Nguyen, N. T. H., Gilmour, A. R., Tearle, R., Knibb, W., et al. (2021a). Prediction accuracies of genomic selection for nine commercially important traits in the Portuguese oyster (*Crassostrea angulata*) using DArT-Seq technology. *Genes* 12:210. doi: 10.3390/genes12020210

Vu, S. V., Knibb, W., O'Connor, W., Nguyen, N. T., Van In, V., Dove, M., et al. (2020). Genetic parameters for traits affecting consumer preferences for the Portuguese oyster, *Crassostrea angulata*. *Aquaculture* 526:735391. doi: 10.1016/j.aquaculture.2020.735391

Vu, S. V., Premachandra, H. K. A., O'Connor, W., Nguyen, N. T. H., Dove, M., Vu, I. V., et al. (2021b). Development of SNP parentage assignment in the Portuguese oyster *Crassostrea angulata*. *Aquac. Rep.* 19:100615. doi: 10.1016/j.aqrep.2021.100615

Vu, V. I., O'Connor, W., Vu, V. S., Phan, T. V., and Knibb, W. (2017a). Resolution of the controversial relationship between Pacific and Portuguese oysters internationally and in Vietnam. *Aquaculture* 473, 389–399. doi: 10.1016/j.aquaculture.2017.03.004

Vu, V. I., Vu, V. S., O'Connor, W., Phan, T. V., Dove, M., Knibb, W., et al. (2017b). Are strain genetic effect and heterosis expression altered with culture system and rearing environment in the Portuguese oyster (*Crassostrea angulata*)? *Aquac. Res.* 48, 4058–4069. doi: 10.1111/are.13227

Vu, V. S., Knibb, W., Nguyen, T. H. N., Vu, V. I., O'Connor, W., Dove, M., et al. (2019). First breeding program of the Portuguese oyster *Crassostrea angulata* demonstrated significant selection response in traits of economic importance. *Aquaculture* 518:734664. doi: 10.1016/j.aquaculture.2019.734664

Walton, W., Rikard, F., Chaplin, G., Davis, J., Arias, C., and Supan, J. J. A. (2013). Effects of ploidy and gear on the performance of cultured oysters, *Crassostrea virginica*: survival, growth, shape, condition index and Vibrio abundances. *Aquaculture* 414, 260–266. doi: 10.1016/j.aquaculture.2013.07.032

Wang, Q., Yu, Y., Li, F., Zhang, X., and Xiang, J. (2017). Predictive ability of genomic selection models for breeding value estimation on growth traits of Pacific white shrimp *Litopenaeus vannamei*. *Chinese J. Oceanol. Limnol.* 35, 1221–1229. doi: 10.1007/s00343-017-6038-0

Xing, D., Li, Q., Kong, L., and Yu, H. (2018). Heritability estimate for mantle edge pigmentation and correlation with shell pigmentation in the white-shell strain of Pacific oyster, *Crassostrea gigas*. *Aquaculture* 482, 73–77. doi: 10.1016/j.aquaculture.2017.09.026

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011

Yoshida, G. M., Carvalheiro, R., Lhorente, J. P., Correa, K., Figueroa, R., Houston, R. D., et al. (2018). Accuracy of genotype imputation and genomic predictions in a two-generation farmed Atlantic salmon population using high-density and low-density SNP panels. *Aquaculture* 491, 147–154. doi: 10.1016/j.aquaculture.2018.03.004

Yoshida, G. M., Lhorente, J. P., Correa, K., Soto, J., Salas, D., and Yáñez, J. M. (2019). Genome-wide association study and cost-efficient genomic predictions for growth and fillet yield in Nile tilapia (*Oreochromis niloticus*). *G3* 9, 2597–2607. doi: 10.1534/g3.119.400116

Zenger, K. R., Khatkar, M. S., Jones, D. B., Khalilisamani, N., Jerry, D. R., and Raadsma, H. W. (2018). Genomic selection in aquaculture: application, limitations and opportunities with special reference to marine shrimp and pearl oysters. *Front. Genet.* 9:693.

Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., et al. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490, 49–54.

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.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 12 July 2021
doi: 10.3389/fvets.2021.666767

Whole-Genome Comparative Analysis Reveals Association Between *Salmonella* Genomic Variation and Egg Production Systems

Hamid Reza Sodagari[1†], *Shafi Sahibzada*[1*†], *Ian Robertson*[1], *Ihab Habib*[1,2*] and *Penghao Wang*[3*]

[1] School of Veterinary Medicine, College of Science, Health, Engineering and Education, Murdoch University, Murdoch, WA, Australia, [2] Department of Veterinary Medicine, College of Food and Agriculture, United Arab Emirates University, Al Ain, United Arab Emirates, [3] Medical, Molecular and Forensic Sciences, College of Science, Health, Engineering and Education, Murdoch University, Murdoch, WA, Australia

OPEN ACCESS

Edited by:
Filippo Biscarini,
National Research Council (CNR), Italy

Reviewed by:
Yosra A. Helmy,
The Ohio State University,
United States
Maria Pia Franciosini,
University of Perugia, Italy
Sadjia Bekal,
Institut National de Santé Publique du Québec, Canada

***Correspondence:**
Shafi Sahibzada
sshafi@murdoch.edu.au
Penghao Wang
p.wang@murdoch.edu.au
Ihab Habib
i.habib@uaeu.ac.ae

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Veterinary Science

Received: 11 February 2021
Accepted: 08 June 2021
Published: 12 July 2021

Citation:
Sodagari HR, Sahibzada S, Robertson I, Habib I and Wang P (2021) Whole-Genome Comparative Analysis Reveals Association Between Salmonella Genomic Variation and Egg Production Systems. Front. Vet. Sci. 8:666767. doi: 10.3389/fvets.2021.666767

Non-typhoidal *Salmonella*, particularly *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium), is the predominant endemic serovar in the Australian egg production industry and is one of the most frequently reported serovars in foodborne infections in Australia. This study was conducted to investigate the genomic characteristics of *Salmonella* isolated from retail table eggs in Western Australia and to identify the impact of production systems on genomic characteristics of *Salmonella* such as virulence and antimicrobial resistance. A total of 40 non-typhoidal *Salmonella* isolates [*S.* Typhimurium isolates ($n = 28$) and *Salmonella* Infantis isolates ($n = 12$)] sourced from retail eggs produced by different production systems (barn-laid, cage, and free-range) in Western Australia were sequenced by whole-genome sequencing. The isolates were *de novo* assembled, annotated, and analyzed. The results indicated an association between *Salmonella* genomic variation and the system used to raise poultry for egg production (p-value < 0.05). All but one of the *S.* Infantis isolates were recovered from eggs collected from poultry raised under barn and cage production systems. A higher proportion (83.3%) of *S.* Typhimurium isolates were recovered from the eggs produced by free-range production system as compared with those produced under barn (76.9%) and cage production systems (53.3%). Our analysis indicated that *Salmonella* isolated from the eggs produced by barn and cage production systems had more virulence genes than the isolates of the free-range produced eggs. A low carriage of antimicrobial-resistant gene was detected in the isolates of this study. We have built a *Salmonella* genomics database and characteristics-linked gene pools to facilitate future study, characterization, and tracing of *Salmonella* outbreaks.

Keywords: *Salmonella*, genomics, egg production system, antimicrobial resistance, virulence, genome environment interaction

BACKGROUND

Salmonella are gram-negative bacteria known for more than 100 years to cause foodborne illness in humans. *Salmonella* are considered host-adapted to many wild and domesticated animals, based on their international distribution and high prevalence in poultry, pigs, and sheep (1). Among 2,600 identified *Salmonella* serovars (2), non-typhoidal serovars have been recognized as the source of 550 million foodborne illnesses annually in the world with 230,000 deaths every year in the world (3); however, many of these *Salmonella*-associated illnesses are preventable with appropriate interventions. Foodborne disease surveillance can be used to gather evidence to help identify emerging strains and resistance that could help in designing appropriate control measures and to evaluate the efficacy of interventional efforts.

In Australia, an estimated 5.4 million cases of foodborne disease occur annually, costing an estimated $1.2 billion per year (4). A number of different *Salmonella* serovars have been isolated in humans and food animals in Australia; however, *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium) is the predominant endemic serovar in the Australian egg production industry and is the most frequently reported serovar in foodborne infections (5, 6). Between 2011 and 2014, 128 outbreaks of *S.* Typhimurium in humans resulted in 2,343 cases with 347 requiring hospitalizations due to consumption of raw eggs and raw egg-related products in various food preparation settings across Australian states/territories, particularly in New South Wales, Queensland, and Victoria (7). Recently, a significant preference of Australian consumers for eggs from cage-free production systems (free-range and barn-laid) has been observed, due to the perception that non-cage production systems produce safer and higher-quality eggs (8) and to public concerns for animal welfare (9).

Over the past decade, Western Australia (WA) has experienced a higher rate of occurrence of human *Salmonella* infection, for both *S.* Typhimurium and non-Typhimurium infections, when compared with the overall national average (10). Better strategies for preventing and managing *Salmonella* outbreaks are of critical importance, and these require improved understanding of the genetic characteristics linked to the mutation, serovar, virulence, and antimicrobial resistance of *Salmonella*. Genomic and metagenomic researches on bacteria play an important role in public health and food safety. The increasing use of whole-genome sequencing (WGS) of *Salmonella* isolates from outbreaks provides a rapid, highly accurate, and discriminatory source tracing and identification of strains. However, there are limited surveillance data and resources available in WA, which have resulted in a lack of detailed understanding of the *Salmonella* outbreaks that occurred in WA.

There have been studies in the USA on the safety of non-cage egg production systems, and it has been hypothesized that non-cage production systems may lead to higher likelihood of exposure to pathogens by chickens, including *Salmonella* (11, 12), due to higher exposure of layer flocks to the outdoor environment, pests, and wildlife vectors. The recent expansion of non-cage production systems has raised concerns about the potential increase in pathogens,such as *Salmonella*, on eggs arising from contaminated farm environments (13). We isolated *Salmonella* strains from eggs available through retail supermarkets in metropolitan Perth, the capital of WA, as previously described (14). Apart from confirming serovar diversity and multilocus sequence types (MLSTs), we conducted WGS to form a better understanding of the genomic mechanisms of the *Salmonella* isolates. In total, we sequenced 40 isolates consisting of three different egg production systems (barn-laid, caged, and free-range). The goals of the study include (1) understanding the genomic characteristics (serovar and MLST) of the *Salmonella* from the WA egg industry and (2) assessing the association between production systems and the genomic characteristics of *Salmonella* such as mutation, virulence, and antimicrobial resistance. This study also indicated the likelihood impact of the environment in driving variations in *Salmonella* genomes and thus forms a basis for devising more effective guidelines and recommendations for industry to better manage the risks of public health in *Salmonella* outbreaks.

MATERIALS AND METHODS

Sampling

We isolated *Salmonella* strains from egg samples (each containing one dozen eggs, totalling 2,400 eggs of 200 dozen packages) purchased from different supermarket chains across Perth residential suburbs. The sample size was calculated as 101 using previously described pooled sampling with Epitools online for uncertain test sensitivity and specificity (15). The estimated true prevalence was considered at precision level of 5% and desired confidence level of 95%, assuming 90% sensitivity and specificity of the test. Subsequently, we collected 200 pool samples to further minimize bias and to represent diversity of retail chain. Details of sample collection and *Salmonella* isolation were described in our previous investigation (14). The proportion of samples from each production system was targeted at ∼50% free-range (93 dozen packages), 30% cage-laid (68 dozen packages), and 20% barn-laid (39 dozen packages). These sampling ratios of the production systems were chosen in an effort to reflect the recent egg production demand proportion in Australia (16).

Isolation, Identification, and Serotyping of *Salmonella*

Salmonella isolation was performed according to the ISO 6579-1:2017 standard and followed the procedure described by (17). In summary, the bag containing crushed shells was weighed and mixed a corresponding volume of Buffered Peptone Water (BPW) (Oxoid, Hampshire, UK) to obtain sample-to-diluent ratio at 1:9. Then the mixture was homogenized in a stomacher for 1 min. The bag containing the pooled egg contents of the same sample unit was first blended in a stomacher for 2 min in 25 ml and was homogenized for 1 min with 225 ml of BPW. Both of the homogenized crushed shells and contents were incubated at 37°C for 48 h. After pre-enrichment of the incubated homogenate, 1 ml was inoculated into Muller-Kauffman Tetrathionate Novobiocin Broth (MKTTn) (Oxoid, Hampshire, UK), and also from the same homogenate 0.1 ml was spotted (three drops) on the surface

of Modified Semi-solid Rappaport Vassiliadis (MSRV) (Oxoid, Hampshire, UK). MKTTn was incubated for 24 h at 37°C, while MSRV was incubated at 41.5°C and checked after 24 h for a migration zone (turbid, white halo, with radius larger than 10 mm). MSRV plates with no migration zone after 24 h were checked again after 48 h. Streaks from both MKTTn broth and MSRV media were applied on Xylose Lysine Deoxycholate (XLD) agar (Oxoid, Hampshire, UK) and Brilliant Green (BGA) agar (Oxoid, Hampshire, UK), which were then incubated at 37°C for 24 h. Presumptive (up to five) colonies with suspected *Salmonella* morphology were selected from both selective media and transferred into Nutrient Agar (Oxoid, Hampshire, UK) plates. After incubating Nutrient Agar plates at 37°C for 24 h, well-isolated colonies were confirmed to species level using matrix-assisted laser desorption ionization–time-of-flight mass spectrometry (MALDI-TOF MS) using the Microflex instrument (Bruker Diagnostics, Berlin Germany). All confirmed *Salmonella* isolates (up to five isolates per positive sample) were sent for serotyping (Kauffmann-White-Le Minor scheme) by a nationally accredited reference laboratory (PathWest Laboratory, Perth, WA, Australia). Isolates from confirmed positive egg samples were stored at −80°C till further use. A total 40 non-typhoidal *Salmonella* isolates were selected for WGS based on diversity in production systems.

Whole-Genome Sequencing

DNA was extracted using the BIOLINE DNA extraction kit (ISOLATE II, Genomic DNA Kit) according to the manufacturer's instructions. Library preparation was performed using an Illumina NexTera® XT library preparation kit (Illumina, San Diego, CA, USA) as per manufacturer's instructions. The library preparations were sequenced on an Illumina NextSeq platform using a mid-output 2 × 150 kit.

Genome Assembly

All the raw sequencing reads were analyzed by using FastQC (version 0.10.1) and MultiQC ver 1.8 (18) to check for read quality. To alleviate contamination from adaptor and vector sequences, the raw read sequences were examined against a comprehensive Illumina adaptor sequence and contaminant library created in-house; and adaptor sequences were removed from the reads when detected. If a read had more than 30 base pair (bp) contaminant sequences, reads were considered problematic and discarded. The raw reads were further processed to remove PCR artifact: only one copy was retained for exactly stacked reads.

The filtered reads were *de novo* assembled using SPAdes software ver 3.11.1 (19) in paired-end mode. Default parameters for SPAdes were used to generate contigs; error correction and Kmer sizes were set to auto. The contigs files were scaffolded by using SSPACE ver 3.0 (20), minimum scaffold size was set to 300 bp, and expected insert size was set to 150 bp with minimum allowed error of 50%.

All read data generated in this study have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive, and the 40 whole-genome-sequenced *Salmonella* isolate accession numbers are in a continuous serial between SAMN12097892 and SAMN12097931 (project accession number PRJNA549805).

Serotype, Multilocus Sequence Type, and Virulence Gene Identification

The assembly genomes were uploaded to the Centre for Genomic Epidemiology (http://www.genomicepidemiology.org/) to screen for serotype using SeqSero 1.2 (21), MLST by using MLST ver 1.8 (22), and the contents of virulence genes by VirulenceFinder ver 2.0 (23). The antibiotic resistance genes were also identified by using ResFinder ver 3.1 (23) and NCBI's curated Bacterial Antimicrobial Resistance Reference Gene Database (NCBI, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA313047).

Genome Annotation and Comparative Analysis

Prokka ver 1.1.2 (24) was used to annotate the gene and microRNA contents of the genome assemblies with bacteria kingdom as annotation model. Standard annotation files, including GFF and GBK annotation files, were generated by running Prokka. *Salmonella* pathogenicity islands (SPIs) were detected with SPIFinder ver 1.0 with default parameters (25) and by BLAST against known SPIs. Prophage regions were identified with PHASTER ver 1.0 (26). Pseudogenes were determined with Pseudofinder ver 0.10 (27) with standard parameters and length 0.8. Genomic rearrangements were detected with progressive Mauve ver 2.4.0, with standard parameters (28). Roary pipeline ver 3.11.2 (29) was used to perform whole-genome comparison and generate a pan-genome of the sequenced *Salmonella* isolates at default parameter set. Scoary ver 1.6.16 (30) and R ver 3.3.2 (9) were used to evaluate the significance of association of the genes identified in the pan-genome with different factors, including serovar and production systems.

Phylogenetic Analysis

From the phylogenetic analysis, single copy gene families, and multigene families that are conserved among species are identified. Lineage-specific genes, which may contribute to species-specific phenotypes, are determined. Phylogenetic tree of *Salmonella* isolates included in the pan-genome analysis was inferred using maximum likelihood approach (PhyML ver 3.3 under smart model selection) (31) on the single copy orthologous genes identified by Roary with 100 bootstrap replicates. The phylogenetic trees were visualized by using R package "ggtree" ver 2.2.4 (32). The phylogeny was inferred as unrooted and without using outgroups.

Pathway

Pathway analysis was performed by using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (33). Annotated genes were supplied to KEGG database through the online interface to map to functional pathways. Identified virulence genes and antimicrobial-resistant genes were searched to identify the pathways they are likely to be involved with by using KEGG database. The associated pathways were visualized by KEGG online access interface.

Statistical Analysis

The basic genome assemblies' statistics were calculated in R using custom R scripts. The association analysis between *Salmonella* pan-genes and other egg production factors including production system and serotype were performed using Scoary (https://github.com/AdmiralenOla/Scoary). Scoary employs several procedures to filter and perform the association (30). Bonferroni-adjusted p-values $\leq$0.05 were used as significance cutoff for Scoary-based association results. Fold change was also used as the secondary criterion in determining the significance of the association. All the downstream statistical analysis and visualization were performed in R using custom R scripts. R packages "ggplots" (34) were used for exploratory analysis; package "ggtree" was used for plotting the phylogenetic tree; "pathfindR" (35) was used for pathway analysis.

RESULTS

Genome Assembly and Annotation

All 40 *Salmonella* isolates were assembled to a near-complete form with an average genome size of 4,825,017 bp. The genomes were assembled at high continuity with an average N50 of 257,277 bp, the largest N50 achieved for a single isolate is around 460,000 bp, and the average number of scaffolds is 76. The complete list of the basic statistics of the assemblies of all isolates is given in **Supplementary Table 1**. The detailed statistics of the assembly of one sample S30, a randomly selected retail isolate from an egg produced in a caged production system, is given in **Supplementary Figure 1** as an example to demonstrate the quality of the assembly.

A pan-genome of *Salmonella* isolates was constructed by genome comparison among all 40 retail isolates of all three production systems. The genic sections of the isolate genomes were identified to be quite reserved. In total 5,604 genes were identified across the 40 retail *Salmonella* isolates, while 38 isolates share 3,980 core genes, and 1,145 genes were identified shared between six and 38 isolates. In comparison, 479 genes were identified as "cloud genes" that were unique to individual isolates (**Figure 1**). The functional annotation indicated that the most abundant functional pathways of the bacteria include amino acid transport and metabolism; energy production and conversion; translation and transcription-related activities including ribosomal structure and biogenesis, replication, recombination, and repair; posttranslational modification, protein turnover, and chaperones; and cell wall or membrane-related biogenesis, signaling, and transport. The overview of the functional categories of the gene annotations is given in **Figure 2**.

A local sequence database of *Salmonella* was constructed, which hosts genome sequences, protein sequences, and annotations. The web-based interface was provided, which enables concurrent enquiries and searches.

Serotype Drives the Main Genomic Divergence

Serotype analysis indicated that the retail isolates consisted of two different serovars: 12 isolates were identified as *Salmonella* Infantis and 28 isolates as *S.* Typhimurium. Interestingly, all

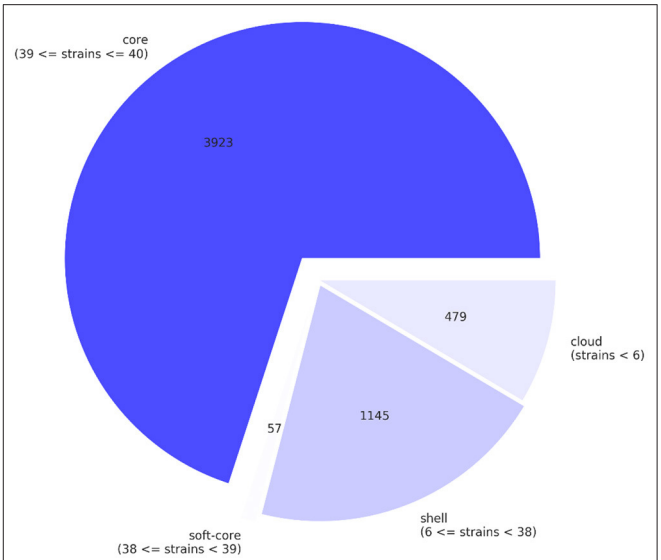

FIGURE 1 | The pan-genome of the retail *Salmonella* isolates, core genes, shell genes, and cloud genes were identified by comparing the isolates.

of the *S.* Infantis isolates except one were recovered from samples collected from barn and cage production systems, while *S.* Typhimurium isolates were recovered from samples from the free-range production system (n = 11) as well as barn and cage (n = 17) (**Supplementary Table 1**). An obvious genetic divergence between these two serotypes can be observed, and this contributes to the majority of the total genomic variations presented by the isolates. Firstly, the genome sizes of *S.* Infantis isolates were consistently smaller than the genome sizes of *S.* Typhimurium isolates (**Figure 3**). Secondly, the MLST analysis indicates that *S.* Infantis isolates have more MLSTs given their smaller genomes compared with *S.* Typhimurium isolates (**Supplementary Table 1**). The phylogenetic analysis using maximum likelihood approach on the core gene sets shared by the isolates consistently identified these two serotypes to have different origins. The inferred evolutionary tree is given in **Supplementary Figure 2**. Based on the gene presence and absence, in total, 924 genes were identified to be differentially associated with the serotypes with Bonferroni-adjusted p-values smaller than 0.05. The most consistent genes include *S.* Infantis unique genes (3-oxoacyl-[acyl-carrier-protein] reductase FabG, 50S ribosomal protein L36 2, aldo-keto reductase IolS, and arsenic resistance transcriptional regulator ArsR2, etc.) and *S.* Typhimurium unique genes (2-dehydro-3-deoxy-6-phosphogalactonate aldolase, 2-nitroimidazole transporter, carbohydrate diacid regulator, 3-hexulose-6-phosphate isomerase, 60 kDa SS-A/Ro ribonucleoprotein, abequosyltransferase RfbV, acetylornithine deacetylase, antirestriction protein KlcA, antitoxin CcdA, DinJ, VapB, aspartate aminotransferase, and so on). A heatmap displaying the genetic distribution of the identified differential genes is given in **Figure 4A**.

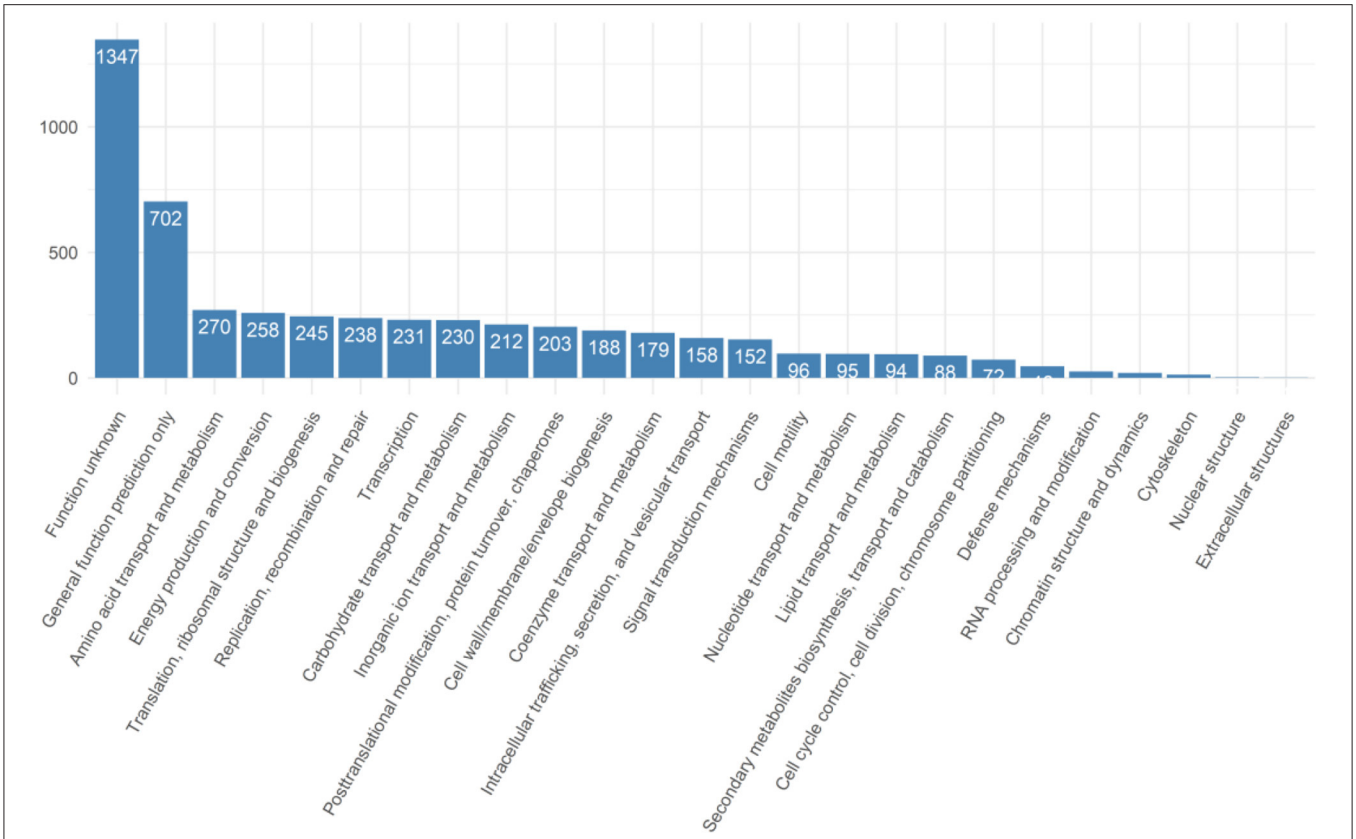

FIGURE 2 | The functional annotation of the genes identified from collected isolates.

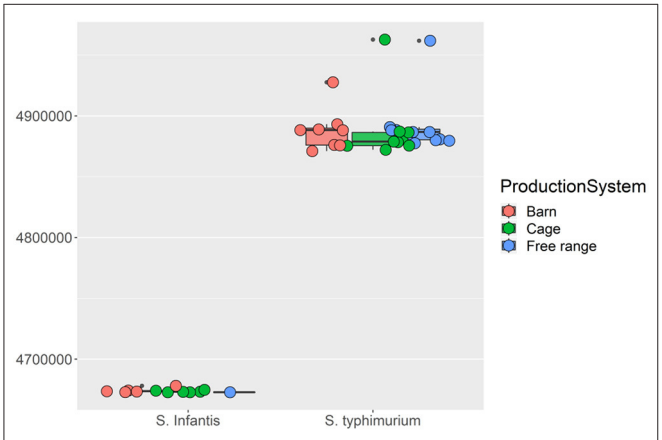

FIGURE 3 | The boxplots of the isolates genome sizes. The *Salmonella* Infantis and *Salmonella* Typhimurium serotypes were grouped separately by the three different egg production systems, namely, barn, cage, and free-range.

Production System Leads to Genetic Variations

Apart from the serotype difference, the second most important source of the genomic divergence among the isolates was identified to be associated with the egg production systems. Around 40 genes were identified to be able to distinguish the egg production systems by their presence or absence with an adjusted p-value <0.05 and 33 of which have odd ratios >2, which is given in **Supplementary Table 2**. A heatmap showing the genetic difference presented by the key genes is shown in **Figure 4B**. The most dominant divergence in gene compositions among the three production systems was observed in the free-range isolates. Our results indicated that the *Salmonella* isolated from eggs sourced from barn and cage production systems were genetically less diverse than the ones sourced from free-range systems. The most distinguishing genes of the isolates recovered from the eggs produced by free-range production system from the isolates recovered from the eggs produced by barn and cage systems include DNA translocase FtsK, L-aspartate oxidase, Macrolide export ATP-binding/permease protein MacB, exodeoxyribonuclease 8, and lysozyme RrrD. These genes were identified to be associated with translocating genes, regulating gene expression, cell wall division and formation, sugar metabolism, and phosphorylation in bacteria (36–42). This demonstrates that the *Salmonella* isolated from eggs source from barn and cage production systems have undergone different mutations that have somehow altered their genomic compositions.

FIGURE 4 | The heatmap of the genomic composition comparison by different factors. **(A)** The heatmap comparing *Salmonella* Infantis and *Salmonella* Typhimurium serotypes on the genes identified to be differentially associated with the serotypic difference. **(B)** The heatmap comparing different egg production systems on the genes identified to be differentially associated with barn, cage, and free-range production systems.

Production System Associated With Virulence

Searching the assembled genomes against virulence gene database (43), we identified four genes that are associated with virulence, i.e., *SpvB*, *SpvC*, *sseL*, and *PhoQ*, belonging to the *Salmonella* plasmid virulence gene family and are mainly involved with the pathogen's toxin and promote the survival and rapid growth of *Salmonella* in the host. Interestingly, these virulence genes were identified in the isolates from barn and cage produced eggs, and free-range isolates were largely free of these virulence genes.

PhoQ controls expression of more than 40 genes and is required for intracellular survival, cationic antimicrobial peptides resistance, and stimulation of cytokine secretion. *sseL* gene is a deubiquitinase required for macrophage killing and virulence. The associated genes and their biological pathways are given in **Supplementary Figure 3**, where the identified virulence genes are marked in red. From **Supplementary Figure 3**, the detailed regulation mechanisms of the virulence genes can be seen. The *Spv* family genes are involved in the *Salmonella* infection pathway. *SpvB* and *SpvC* are involved in the actin depolymerization by regulating F-actin.

Our analysis has observed little evidence of antimicrobial resistance genes from the *Salmonella* isolates after comprehensive searching through the NCBI's curated Bacterial Antimicrobial Resistance Reference Gene Database, while two isolates were estimated to contain a β-lactamase resistance gene family by ResFinder.

DISCUSSION

Our *Salmonella* assemblies are of consistent quality across all isolates, capturing well both gene space and repetitive segments across the genome. In total, 1,440 universal orthologs were searched against all chromosomes, with around 94.4% identified by BUSCO (benchmark universal single-copy orthologs) on average (44). This is on par with "Gold Standard" reference genome assemblies, including human (*Homo sapiens* GRCh38) at 95.5%, corn at 97.0% (*Zea mays* W22v2), and (*Arabidopsis thaliana*) at 95.7% (TAIR10). We have constructed a genome sequence database consisting of *S.* Infantis and *S.* Typhimurium isolates collected from retail eggs. Our finding is the first effort in building a comprehensive collection of isolates in WA by collecting samples from retail companies in metropolitan Perth. This study can serve as a valuable genetic resource and roadmap for studying *Salmonella*, tracing infection, and epidemiological surveillance of outbreaks in WA.

Our analysis demonstrated that the egg production system has a reasonable association with the genomic characteristics of the pathogens; different production systems may lead to genetically different pathogens and virulence. We identified around 40 genes that are unique and relevant to the individual production systems; particularly free-range strains appear to be quite genetically divergent from barn and cage strains, which share more similarity in their genomic compositions. The genetic difference of isolates from free-range and the other two systems was significant. Eggs sourced from free-range enterprises were mainly identified to have *S.* Typhimurium, while *S.* Infantis strains were most likely isolated in barn-laid and caged eggs. Our results only identified a handful of virulence genes, but more were found to be present in the *Salmonella* isolates collected from barn and cage production systems compared with the free-range isolates. Previous investigations also identified virulence genes in various *Salmonella* isolates recovered from cage farms (45, 46). However, limited information is available on virulence typing of *Salmonella* serovars isolated from free-range environments (13). There are four virulence genes that were identified almost exclusively from cage and barn eggs. Although the presence of virulence genes may indicate a higher likelihood to cause serious infections in people who contract these strains, the detailed mechanisms of these virulence genes are not fully understood as yet. In recent years, consumer demand for free-range eggs has been fast growing due to the conceived high quality (8). However, limited molecular studies have been performed to scientifically evaluate the safety of free-range eggs. Our results indicated that egg production systems can potentially lead to specific genetic variations in the *Salmonella* isolates. Further research is required to assess how different management practices related to these production systems impact upon the genetic variation (i.e., mutation and virulence) of *Salmonella* serovars.

The difference in genomic profiles and virulence between free-range and the other production systems is likely to be driven by the environmental and management factors. Free-range chicken farms have a lower stocking density than the other two systems with a high stocking density shown to be associated with higher stress, lower immune response, and higher chances of infection (47). These events could provide favorable conditions for bacteria to evolve and develop or sustain virulence factors. In contrast, chickens raised under a free-range production system are more likely to interact with farm's outside environments and intense human contact, as constant interaction is necessary during hand egg collection, cleaning, and monitoring, and herding stock. These events increase the likelihood of uptake of more diverse genes (virulence and resistance). This is very informative and presented a serious argument if certain chicken density requirement be implemented as a regulation for public health policymakers. It would be very interesting to perform more thorough study to evaluate specific environmental factors of different production systems. This will help to identify the most important hazards for fostering bacteria virulence and resistance. Although it is well-known in the plant domain, i.e., cereal crops, canola, and *Arabidopsis*, that the environment plays an equally important role in the genotypes for controlling phenotypes, how the environment affects bacterial genomes is relatively less understood. The results of our study demonstrated that different types of production system for egg laying poultry are associated with the genomic compositions of *Salmonella*.

Our finding indicated very low carriage of antimicrobial resistance genes by all the *Salmonella* strains. No antimicrobial resistance genes have been also identified from *S.* Sofia isolated from chicken meat in a previous study in Australia (48). Finding very few resistance genes in our study provided evidence of the significance of environment in driving bacteria

mutations, which might be because of WA is geographically isolated from the rest of Australia and the world or might be due to the strict conservative approach to registration of antibiotics for use in food-producing animals in Australia. As future work, it is recommended to incorporate management strategies and design new ways of quantitatively measure key environmental condition at chicken farms. In this way, the association between the bacteria genetic mutations and the environmental factors can be better understood and modeled. To correlate human *Salmonella* isolates with egg *Salmonella* isolates will be another interesting direction for future research. This kind of integrative analysis may lead to novel methods for predicting the virulence of unseen strains and the resistance to antibiotics so that the treatment and the management can be more effective.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This research was supported by Murdoch University International Postgraduate Research Scholarship awarded to HS.

ACKNOWLEDGMENTS

We would like to thank the staff of the Antimicrobial Resistance and Infectious Disease Laboratory at Murdoch University for their assistance in WGS analysis of the isolates.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2021.666767/full#supplementary-material

REFERENCES

1. Ferrari RG, Rosario DKA, Cunha-Neto A, Mano SB, Figueiredo EES. Worldwide Epidemiology of *Salmonella* serovars in animal-based foods: a meta-analysis. *Appl Environ Microbiol.* (2019) 85:e00591–19. doi: 10.1128/AEM.00591-19
2. Jajere S. A review of *Salmonella enterica* with particular focus on the pathogenicity and virulence factors, host specificity and antimicrobial resistance including multidrug resistance. *Vet World.* (2019) 12:504–21. doi: 10.14202/vetworld.2019.504-521
3. Mooijman KA. The new ISO 6579-1: a real horizontal standard for detection of *Salmonella*, at last! *Food Microbiol.* (2018) 71:2–7. doi: 10.1016/j.fm.2017.03.001
4. Angulo FJ, Kirk MD, McKay I, Hall GV, Dalton CB, Stafford R, et al. Foodborne Disease in Australia: the OzFoodNet Experience. *Clin Infect Dis.* (2008) 47:392–400. doi: 10.1086/589861
5. Moffatt CR, Musto J. *Salmonella* and egg-related outbreaks. *Microbiol Aust.* (2013) 34:94–8. doi: 10.1071/MA13033
6. Chousalkar K, Gole V, Caraguel C, Rault JL. Chasing *Salmonella* Typhimurium in free-range egg production system. *Vet Microbiol.* (2016) 192:67–72. doi: 10.1016/j.vetmic.2016.06.013
7. Moffatt CR, Musto J, Pingault N, Miller M, Stafford R, Gregory J, et al. *Salmonella* Typhimurium and outbreaks of egg-associated disease in Australia, 2001 to 2011. *Foodborne Pathog Dis.* (2016) 13:379–85. doi: 10.1089/fpd.2015.2110
8. Scott AB, Singh M, Toribio JA, Hernandez-Jover M, Barnes B, Glass K. Comparisons of management practices and farm design on Australian commercial layer and meat chicken farms: cage, barn and free range. *PLoS ONE.* (2017) 12:e0188505. doi: 10.1371/journal.pone.0188505
9. RSPCA Australia. *The Welfare of Layer Hens in Cage and Cage-Free Housing Systems.* (2016). Available online at: http://www.animalwelfarestandards.net.au/files/2015/07/FINAL_2016-08-The-welfare-of-layer-hens-in-cage-and-cage-free-housing-systems-FINAL.pdf (accessed November 4, 2020).
10. Ford L, Glass K, Veitch M, Wardell R, Polkinghorne B, Dobbinset T, et al. Increasing Incidence of *Salmonella* in Australia, 2000-2013. *PLoS ONE.* (2016) 11:e0163989. doi: 10.1371/journal.pone.0163989
11. LayJr DC, Fulton RM, Hester PY, Karcher DM, Kjaer JB, Mench JA, et al. Hen welfare in different housing systems. *Poult Sci.* (2011) 90:278–94. doi: 10.3382/ps.2010-00962
12. Daigle C, Siegford J. Welfare Quality parameters do not always reflect hen behaviour across the lay cycle in non-cage laying hens. *Anim Welf.* (2014) 23:423–34. doi: 10.7120/09627286.23.4.423
13. Gole VC, Woodhouse R, Caraguel C, Moyle T, Rault JL, Sexton M. Dynamics of *Salmonella* shedding and welfare of hens in free-range egg production systems. *Appl Environ Microbiol.* (2017) 83:e03313–6. doi: 10.1128/AEM.03313-16
14. Sodagari HR, Baraa Mohammed A, Wang P, O'Dea M, Abraham S, Robertson I, et al. Non-typhoidal *Salmonella* contamination in egg shells and contents from retail in Western Australia: serovar diversity, multilocus sequence types, and phenotypic and genomic characterizations of antimicrobial resistance. *Int J Food Microbiol.* (2019) 308:108305. doi: 10.1016/j.ijfoodmicro.2019.108305
15. Sergeant ESG. *Epitools Epidemiological Calculators [Online].* Ausvet Pty Ltd (2015). Available online at: http://epitools.ausvet.com.au/ (accessed May, 2015)
16. Government of Western Australian (2018). *Media Statement: Call for Phase-Out of Battery Cages*. Available online at: https://www.mediastatements.wa.gov.au/Pages/McGowan/2018/03/Call-for-phase-out-of-battery-cages.aspx. (accessed April 2, 2018).
17. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing (2020). Available online at: https://www.R-project.org/.
18. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* (2016) 32:3047–8. doi: 10.1093/bioinformatics/btw354
19. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* (2012) 19:455–77. doi: 10.1089/cmb.2012.0021
20. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics.* (2011) 27:578–9. doi: 10.1093/bioinformatics/btq683
21. Zhang S, Yin Y, Jones MB, Zhang Z, Deatherage Kaiser BL, Dinsmore BA, et al. Salmonella serotype determination utilizing high-throughput genome sequencing data. *J Clin Micobiol.* (2015) 53:1685–92. doi: 10.1128/JCM.00323-15

22. Bartual SG, Seifert H, Hippler C, Domínguez Luzon MA, Wisplinghoff H, Rodríguez-Valera F. Development of a multilocus sequence typing scheme for characterization of clinical isolates of *Acinetobacter baumannii*. *J Clin Microbiol.* (2005) 43:4382–90. doi: 10.1128/JCM.43.9.4382-4390.2005
23. Clausen P, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics.* (2018) 19:307. doi: 10.1186/s12859-018-2336-6
24. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* (2014) 30:2068–9. doi: 10.1093/bioinformatics/btu153
25. Roer L, Hendriksen RS, Leekitcharoenphon P, Lukjancenko O, Kaas RS, Hasman H, et al. Is the evolution of *Salmonella enterica* subsp. *enterica* linked to restriction-modification systems? *mSystems.* (2016) 1:e00009–16. doi: 10.1128/mSystems.00009-16Arndt,
26. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* (2016) 44:W16–21. doi: 10.1093/nar/gkw387
27. Syberg-Olsen M, Husnik F. *Pseudofinder, GitHub Repository*. (2018). Available online at: https://github.com/filip-husnik/pseudo-finder/.
28. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE.* (2010) 5:e11147. doi: 10.1371/journal.pone.0011147
29. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. (2015) 31:3691–3. doi: 10.1093/bioinformatics/btv421
30. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* (2016) 17:238. doi: 10.1186/s13059-016-1108-8
31. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. (2010) 593:307–21. doi: 10.1093/sysbio/syq010
32. Yu G, Smith D, Zhu H, Guan Y, Lam TL. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol.* (2017) 8:28–36. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* (2000) 28:27–30. doi: 10.1093/nar/28.1.27
33. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag (2016).
34. Ulgen E, Ozisik O, Sezerman OU. pathfindR: an r package for comprehensive identification of enriched pathways in omics data through active subnetworks. *Front. Genet.* (2019) 10:858. doi: 10.3389/fgene.2019.00858
35. Begg KJ, Dewar SJ, Donachie WD. A new *Escherichia coli* cell division gene, ftsK. *J Bacteriol.* (1995) 177:6211–22. doi: 10.1128/jb.177.21.6211-6222.1995
36. Aussel L, Barre FX, Aroyo M, Stasiak M, Stasiak AZ, Sherratt D. FtsK is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. *Cell.* (2002) 108:195–205. doi: 10.1016/S0092-8674(02)00624-4
37. Chow C, Hegde S, Blanchard JS. Mechanistic characterization of *Escherichia coli* l-aspartate oxidase from kinetic isotope effects. *Biochemistry.* (2017) 56:4044–52. doi: 10.1021/acs.biochem.7b00307
38. Marinoni I, Nonnis S, Monteferrante C, Heathcote P, Härtig E, Böttger LH, et al. Characterization of L-aspartate oxidase and quinolinate synthase from *Bacillus subtilis*. *FEBS J.* (2008) 275:5090–107. doi: 10.1111/j.1742-4658.2008.06641.x
39. Greene NP, Kaplan E, Crow A, Koronakis V. Antibiotic resistance mediated by the MacB ABC transporter family: a structural and functional perspective. *Front Microbiol.* (2018) 9:950. doi: 10.3389/fmicb.2018.00950
40. Handa N, Kobayashi I. Type III restriction is alleviated by bacteriophage (RecE) Homologous recombination function but enhanced by bacterial (RecBCD) function. *J Bacteriol.* (2005) 187:7362–73. doi: 10.1128/JB.187.21.7362-7373.2005
41. Barrios-Villa E, Martínez de la Peña CF, Lozano-Zaraín P, Cevallos MA, Torres C, Torrese AG, et al. Comparative genomics of a subset of adherent/invasive *Escherichia coli* strains isolated from individuals without inflammatory bowel disease. *Genomics*. (2020) 112:1813–20. doi: 10.1016/j.ygeno.2019.10.013
42. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids*. (2005) 33:D325–8. doi: 10.1093/nar/gki008
43. Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. In: Kollmar M, editor. *Gene Prediction*. New York, NY: Humana (2019). Available online at: https://link.springer.com/protocol/10.1007%2F978-1-4939-9173-0_14
44. Manning J, Gole V, Chousalkar K. Screening for *Salmonella* in backyard chickens. *Prev Vet Med.* (2015) 120:241–5. doi: 10.1016/j.prevetmed.2015.03.019
45. McWhorter AR, Davos D, Chousalkar KK. Pathogenicity of *Salmonella* strains isolated from egg shells and the layer farm environment in Australia. *Appl Environ Microbiol*. (2015) 81:405–14. doi: 10.1128/AEM.02931-14
46. Estevez I. Density allowances for broilers: where to set the limits? *Poult Sci.* (2007) 86:1265–72. doi: 10.1093/ps/86.6.1265
47. Abraham S, O'Dea M, Sahibzada S, Hewson K, Pavic A, Veltman T, et al. Escherichia coli and Salmonella spp. isolated from Australian meat chickens remain susceptible to critically important antimicrobial agents. *PLoS ONE.* (2019) 14:e0224281. doi: 10.1371/journal.pone.0224281
48. WHO. *WHO's First Ever Global Estimates of Foodborne Diseases Find Children Under 5 Account for Almost One Third of Deaths*. News release GENEVA: WHO (2015). Available online at: https://www.who.int/news-room/detail/03-12-2015-who-s-first-ever-global-estimates-of-foodborne-diseases-find-children-under-5-account-for-almost-one-third-of-deaths

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership