BMC Genomics

RESEARCH

Open Access

Genome-wide association study (GWAS) uncovers candidate genes linked to the germination performance of bread wheat (*Triticum aestivum* L.) under salt stress

Saeideh Javid[1,4], Mohammad Reza Bihamta[1*], Mansour Omidi[1], Ali Reza Abbasi[1], Hadi Alipour[2], Pär K. Ingvarsson[3] and Peter Poczai[4]

Abstract

Background Improving the germination performance of bread wheat is an important breeding target in many wheat-growing countries where seedlings are often established in soils with high salinity levels. This study sought to characterize the molecular mechanisms underlying germination performance in salt-stressed wheat. To achieve this goal, a genome-wide association study (GWAS) was performed on 292 Iranian bread wheat accessions, including 202 landraces and 90 cultivars.

Results A total of 10 and 15 functional marker-trait associations (MTAs) were detected under moderate (60 mM NaCl) and severe (120 mM NaCl) salinity, respectively. From genomic annotation, 17 candidate genes were identified that were functionally annotated to be involved in the germination performance of salt-stressed wheat, such as *CHX2*, *PK2*, *PUBs*, and *NTP10*. Most of these genes play key roles in DNA/RNA/ATP/protein binding, transferase activity, transportation, phosphorylation, or ubiquitination and some harbored unknown functions that collectively may respond to salinity as a complex network.

Conclusion These findings, including the candidate genes, respective pathways, marker-trait associations (MTAs), and in-depth phenotyping of wheat accessions, improve knowledge of the mechanisms responsible for better germination performance of wheat seedlings under salinity conditions.

Keywords Abiotic stress, Germination performance, Salt tolerance, Wheat

Background

Common wheat (*Triticum aestivum* L.) provides calories and carbohydrates for about 25% of the world's population [1]. This, along with the rapidly growing global human population and climate change, highlights the importance of sustainable wheat production to meet future nutritional demands [2]. Rapid climate change is expected to increase soil salinity in many areas due to, for example, saltwater intrusion following sea-level rise or excessive use of groundwater resources. Therefore, soil

*Correspondence:
Mohammad Reza Bihamta
mrghanad@ut.ac.ir
[1] Department of Agronomy and Plant Breeding, University of Tehran, Karaj, Iran
[2] Department of Plant Production and Genetics, Urmia University, Urmia, Iran
[3] Department of Plant Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden
[4] Botany and Mycology Unit, Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland

salinity is one of the greatest global challenges to wheat germination in arid and semi-arid regions [3].

Crop yield is adversely affected by salt stress in nearly 20% of the world's arable lands and in about 50% of irrigated lands. Wheat also suffers yield loss when grown in saline soils [4]. One of the impacts of high salt concentrations is disturbance in germination and seedling growth by decreased water accessibility, osmotic stress, oxidative stress, nutritional disorders, and Na^+ and Cl^- ion toxicity [5]. Thus, it is important to understand the genetic basis of salt adaptation-related mechanisms in wheat during the early stages of breeding programs.

Earlier research has shown that different crop growth stages may display varying levels of salinity tolerance or sensitivity [6]. Crops can be sensitive to salinity at one growth stage but tolerant at another, i.e., the impacts of salt tolerance at various growth stages are generally independent of each other [6]. Some studies have revealed that breeding practices for salinity tolerance at germination stages are not useful for improving salinity tolerance in adult crops [7], while others have shown the opposite results [8]. Regardless of whether germination tolerance results in adult plant tolerance, it is more important to focus on the germination performance that allows healthy seedlings to develop under salinity conditions [6].

The advent of next-generation sequencing (NGS) approaches has enabled cost-efficient genotyping-by-sequencing, which is a useful tool to facilitate genetic dissection of complex traits in non-model organisms [9]. Association mapping overcomes many of the restrictions of classic quantitative trait loci (QTL) mapping and helps identify minor genetic factors underlying complex traits [9]. QTLs identified through association mapping can be directly utilized in marker-assisted selection for improving genetic gain [10].

To date, genome-wide association studies (GWAS) have been adopted to explore marker-trait associations (MTAs) and candidate genes affecting growth and development under salinity stress in alfalfa [11], sesame [12], rapeseed [13], cotton [14], barley [9], and rice [15]. For example, Shi et al. [16] explored salinity tolerance during rice germination via GWAS and identified a 164 kb genomic region located on chromosome 2 (Chr2) that harbors two nitrate transporter genes (*NRT2.1/2*) and controls vigor index. A genomic region associated with germination time was also identified on Chr1 that contains some QTLs for total K^+ concentration, total Na^+ uptake, and Na^+: K^+ ratio. Based on GWAS analysis of root length in rice, Yu et al. [17] identified the Chr4-located gene *OsMADS31*, a MADS-box transcription factor that is down-regulated by salinity, and suggested it is involved in stress tolerance during germination. In a GWAS on dry weight, Li et al. [12] identified the gene *SiMLP31* in *Sesamum indicum* L. that encodes a major latex-like protein, homologous to the *MLP31* gene in Arabidopsis. This gene plays a key role in the salicylic acid synthesis, which improves germination and seedling vigor under salinity stress. However, little is known on QTLs linked to germination performance under salinity, corresponding genetic mechanisms, and specific genes in wheat.

In the present study, we aimed to identify trait-associated markers that can be directly or indirectly linked to the germination performance of salt-stressed wheat. This work will provide a foundation for future wheat breeding programs aimed at improving the germination performance of salt-stressed wheat by exploring salinity-responsive pathways, putative candidate genes, and the MTAs identified in our GWAS (Fig. 1).

Results

Phenotypic assessment

Analysis of Variance (ANOVA) revealed that the effect of genotype, salinity, and their interaction on the studied traits was significant at the level of 1%. Data for normal, moderate (60 mM NaCl), and severe (120 mM NaCl) salinity conditions in 292 wheat accessions were analyzed separately (Table 1). The maximum and minimum values of measured traits were obtained under normal and severe salinity conditions, respectively. Phenotypic observations revealed that most germination-related traits under salinity indicated lower performance than those in the normal conditions, suggesting that salt stress significantly limits seed germination and thereby growth, as previously reported by Yu et al. [17] for rice. The greatest variation was recorded for the germination percentage (GP) in all conditions (SD 13.76, 16.60, and 19.47 for normal, moderate, and severe salinity, respectively), indicating the potential of this trait to be employed in selection-assisted breeding.

Pearson correlation coefficients were estimated for all conditions. A significant positive correlation was observed between all traits ($P<0.01$), while the ratio of shoot to root length was negatively correlated with root length ($P<0.01$) (Table 2). These results indicate that all traits evaluated in the current study can be adopted for GWAS analysis.

Genotypic assessment

SNP genotyping

A total of 566 439 207 unique reads were obtained after sequencing, with nearly 80% being non-redundant. After de-duplication and alignment steps, 133 039 SNPs were called, of which 10 938 had a minor allele

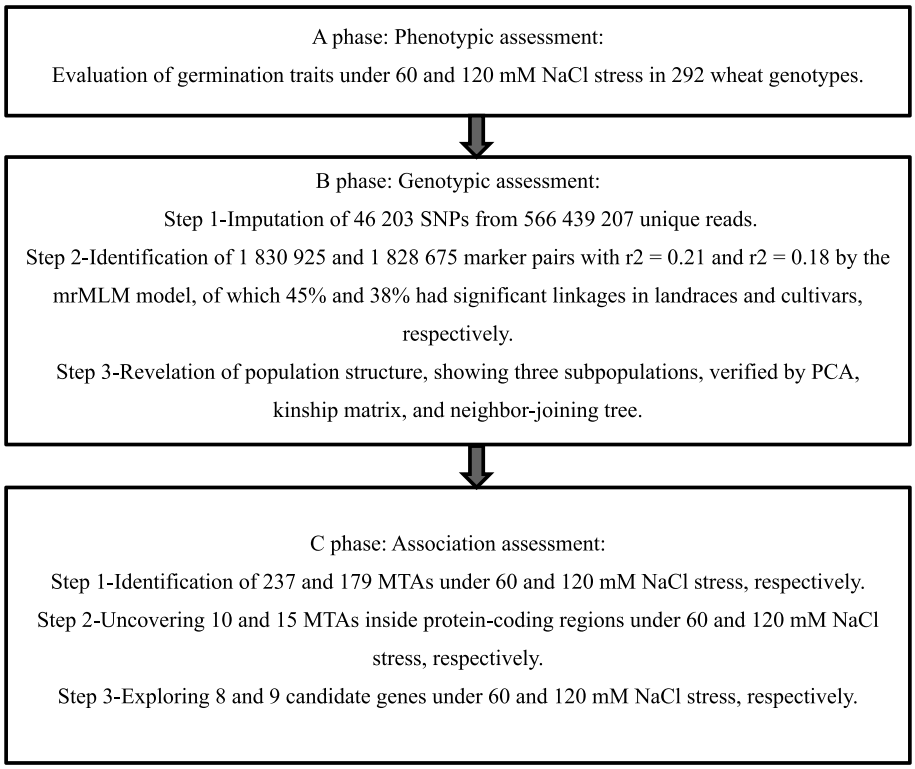

Fig. 1 Flowchart of a step-by-step process to analysis GWAS and exploring candidate genes linked to wheat germination

Table 1 Descriptive statistics for germination-related traits of Iranian bread wheat accessions under normal, moderate, and severe salinity conditions

Conditions									
Normal				Moderate stress			Severe stress		
Trait	Range	Mean	Std. Dev	Range	Mean	Std. Dev	Range	Mean	Std. Dev
RL	2.3–16.5	9.24	2.22	1.6–14.17	6.96	1.94	1.37–12.83	4.93	1.58
ShL	0.66–14.9	7.03	2.97	0.5–13.53	5.75	2.64	0.16–10	4.24	1.87
PL	3.1–29.22	16.27	4.30	3–26.47	12.72	4.10	1.97–22.66	9.17	3.12
WW	0.1–1.41	0.69	0.21	0.11–1.42	0.61	0.19	0.07–1.01	0.48	0.16
DW	0.009–0.13	0.06	0.01	0.015–0.11	0.05	0.01	0.008–0.58	0.04	0.02
ShL-RL	0.048–2.76	0.78	0.38	0.2–3.43	0.83	0.37	0.07–2.91	0.87	0.35
GI	0.28–0.97	0.74	0.13	0.12–0.93	0.65	0.15	0.09–0.87	0.57	0.16
GR	3.3–22	13.26	3.38	2–19.33	10.96	3.26	1.25–16	8.96	3.05
GE	0.36–1	0.84	0.13	0.16–1	0.76	0.16	0.12–1	0.68	0.19
GP	36–100	86.09	13.76	16–100	77.84	16.60	12–100	69.78	19.47
SV	1.36–28.05	14.32	5.01	2.96–26.47	10.23	4.51	0.39–21.75	6.68	3.34

Abbreviations: *RL* Root length, *ShL* Shoot length, *PL* Plomule length, *WW* Wet weight, *DW* Dry weight, *ShL_RL* Ratio of shoot length to root length, *GI* Germination index, *GR* Germination rate, *GE* Germination energy, *GP* Germination percentage, *SV* Seedling vigor, *Std. Dev* standard deviation

frequency (MAF) > 1%, heterozygosity < 10%, and missing data < 10%. The final data set included 46 203 imputed SNPs, which were used in all subsequent analyses.

Linkage disequilibrium

Linkage disequilibrium (LD) varies both within and between chromosomes and sub-genomes and usually decreases with increasing distance between SNPs. Based on the mrMLM model, 1 830 925 marker pairs (MP) with

Table 2 Correlation coefficients between the germination-related traits of Iranian bread wheat accessions

Traits	RL	ShL	PL	WW	DW	ShL_RL	GI	GR	GE	GP	SV
RL	1										
ShL	.606**	1									
PL	.889**	.903**	1								
WW	.699**	.743**	.805**	1							
DW	.482**	.496**	.546**	.629**	1						
ShL-RL	-.190**	.588**	.235**	.222**	.163**	1					
GI	.516**	.547**	.594**	.615**	.447**	.203**	1				
GR	.528**	.512**	.580**	.585**	.423**	.141**	.934**	1			
GE	.485**	.530**	.567**	.594**	.430**	.213**	.971**	.841**	1		
GP	.485**	.537**	.571**	.600**	.438**	.227**	.970**	.837**	.988**	1	
SV	.837**	.871**	.953**	.816**	.558**	.246**	.772**	.722**	.755**	.761**	1

Abbreviations: *RL* Root length, *ShL* Shoot length, *PL* Plomule length, *WW* Wet weight, *DW* Dry weight, *ShL_RL* Ratio of shoot length to root length, *GI* Germination index, *GR* Germination rate, *GE* Germination energy, *GP* Germination percentage, *SV* Seedling vigor, *Std. Dev* standard deviation

*$p < 0.05$

**$p < 0.01$

$r^2 = 0.21$ (as average squared allele frequency correlation) were identified in wheat cultivars, of which 38% had significant linkages at $P < 0.001$ (Table 3). All MPs indicated a distance < 10 cM. The highest and lowest numbers of marker pairs were found in the B (949 425, 51.85%) and D (206 175, 11.26%) genomes, respectively. The highest LD was found between marker pairs located on Chr4A (0.37) followed by Chr1D (0.29) (Table 3).

Implementing a similar assessment using the wheat landraces identified 1 828 675 marker pairs with an average $r^2 = 0.18$, which is lower than in cultivars. A larger fraction of MPs had significant LD (836 400, 45.74%) in landraces. The highest and lowest MPs were found in the B (928 125) and D (233 075) genomes, respectively. Linkage disequilibrium was strongest between marker pairs in Chr4A (0.32) followed by Chr2A (0.25) (Table 3).

On the other hand, the lowest estimated values for Ho, Hs, and Ht were observed in the D genome (0.026, 0.269, and 0.290, respectively) compared with the A and B genomes. However, the Fst values corresponding to each of the three genomes were different (0.074, 0.080, and 0.061 in genomes A, B, and D, respectively) (Table 4).

Population kinship and structure matrix

The population structure also revealed the maximum value of ΔK for K = 3, showing that the wheat accessions can be divided into three subpopulations (Fig. 2). Thus, the analysis of population structure identified three subpopulations with varying levels of admixture.

In the principal component analysis, PC1 and PC2 explained 16.95% and 6.39% of the genotypic variation, respectively (Fig. 3). Clear subpopulations could be identified based on the first two PCs, suggesting three subpopulations with admixed accessions falling between the main three subpopulations. As the panel of wheat cultivars and landraces have subpopulations, the PCA and kinship matrix were performed as variance–covariance.

A cluster analysis of the kinship matrix showed that the SBP-I subgroup harbors 110 accessions (105 landraces and 5 cultivars), the SBP-II harbors 38 accessions (28 landraces and 10 cultivars), and the SBP-III harbors 144 accessions (69 landraces and 75 cultivars) (Fig. 4). The Shahpassand, Shahi, Azar, Roshan, and Rayhani cultivars all displayed high admixture levels. A neighbor-joining tree of all accessions also clearly exhibited clustering into three subgroups (Supplementary 1 Fig. S1).

MTAs for germination-related traits

Collectively, across all traits, 659 MTAs were discovered at a significance threshold (cut-off) of −log10 (*P*-value) ≥ 3.0 ($P \leq 0.001$). A total of 237 and 179 highly significant MTAs were identified under moderate and severe salinity conditions, respectively (Table 5). The highest number of marker-trait associations was found in the A and B sub-genomes under severe and moderate stresses, respectively. The D sub-genome displayed the lowest number of MTAs in all three conditions.

Genomic annotation

Of 659 marker-trait associations, 10 and 15 MTAs were located inside protein-coding regions under moderate and severe salinity conditions, respectively (Tables 6 and 7). These regions mainly encode proteins responsible for several biological processes in the stressed crops, including oxidation–reduction (i.e., monooxygenase and oxidoreductase activity, Fe- and heme-binding),

Table 3 The SNP pairs and their LD (r^2) and distance (cM) per chromosomes and genomes of Iranian bread wheat cultivars and landraces

Ch	Cultivars				Landraces				Total			
	TNSP	r^2	Distance (cM)	NSSP	TNSP	r^2	Distance (cM)	NSSP	TNSP	r^2	Distance (cM)	NSSP
1A	85575	0.148218	1.7377	27125 (31.7%)	92925	0.112764	1.5964	33515 (36.07%)	110025	0.109029	1.3525	48826 (44.38%)
2A	118025	0.292156	0.9742	57858 (49.02%)	123175	0.297454	0.9444	68675 (55.75%)	135275	0.256551	0.8608	79620 (58.86%)
3A	83675	0.159365	2.5764	25903 (30.96%)	73525	0.136413	2.9397	28144 (38.28%)	95125	0.132082	2.2800	44477 (46.76%)
4A	114925	0.371766	1.5136	57774 (50.27%)	108375	0.376224	1.6121	65451 (60.39%)	128375	0.322641	1.3876	78844 (61.42%)
5A	59375	0.169369	2.3835	18718 (31.53%)	58475	0.150278	2.4165	24007 (41.06%)	70475	0.135122	2.0086	31970 (45.36%)
6A	85175	0.181387	1.4878	29645 (34.8%)	84425	0.181735	1.5010	40176 (47.59%)	97625	0.161099	1.2981	51977 (53.24%)
7A	128575	0.234215	1.3445	49426 (38.44%)	126575	0.214252	1.3660	63357 (50.05%)	148075	0.195064	1.1677	78080 (52.73%)
1B	131075	0.206251	1.0638	49717 (37.93%)	133525	0.157517	1.0413	63,803 (47.78%)	149175	0.156549	0.9350	79917 (53.57%)
2B	165475	0.198105	0.8592	66129 (39.96%)	155625	0.177663	0.9135	78,536 (50.46%)	185625	0.157919	0.7659	101594 (54.73%)
3B	176175	0.245726	0.8766	78363 (44.48%)	170925	0.221549	0.9040	89150 (52.16%)	199775	0.212639	0.7742	118862 (59.5%)
4B	51325	0.1455	2.5168	13477 (26.26%)	43025	0.1018	3.0028	12311 (28.61%)	58725	0.117756	2.2066	23396 (39.84%)
5B	134225	0.204683	1.4332	55633 (41.45%)	134675	0.14301	1.4493	56285 (41.79%)	150925	0.151374	1.2942	80074 (53.06%)
6B	158275	0.205457	0.7884	66108 (41.77%)	164475	0.139023	0.7587	71582 (43.52%)	188775	0.139448	0.6610	98910 (52.4%)
7B	132875	0.156677	1.1024	41160 (30.98%)	125875	0.129711	1.1575	50573 (40.18%)	148625	0.122897	0.9885	69532 (46.78%)
1D	37075	0.294821	4.4091	16539 (44.61%)	40975	0.232567	3.8321	19755 (48.21%)	47275	0.24563	3.4847	25602 (54.16%)
2D	48025	0.23446	2.2455	16275 (33.89%)	52825	0.169092	2.0486	20548 (38.9%)	67125	0.187305	1.6133	30724 (45.77%)
3D	25475	0.143085	6.2861	5413 (21.25%)	30125	0.174879	5.3156	11411 (37.88%)	35525	0.128602	5.2147	10004 (28.16%)
4D	10275	0.167587	10.5662	2189 (21.3%)	10375	0.14746	10.7135	3543 (34.15%)	12125	0.1343	9.1793	4233 (34.91%)
5D	22375	0.155406	9.3377	5503 (24.59%)	24825	0.142184	8.3614	8953 (36.06%)	30325	0.136465	6.9287	12067 (39.79%)
6D	28475	0.142966	5.3691	6844 (24.04%)	33475	0.14123	4.5658	12606 (37.66%)	36875	0.12788	4.1511	15587 (42.27%)
7D	34475	0.208327	5.7957	10809 (31.35%)	40475	0.153099	4.9473	14019 (34.64%)	44975	0.155443	4.4523	17504 (38.92%)
A genome	675325	0.235213	1.6204	266449 (39.45%)	667475	0.223484	1.6427	323325 (48.44%)	784975	0.197227	1.4032	413794 (52.71%)
B genome	949425	0.20158	1.0837	370587 (39.03%)	928125	0.160951	1.1104	422240 (45.49%)	1081625	0.156707	0.9550	572285 (52.91%)
D genome	206175	0.205106	5.3432	63572 (30.83%)	233075	0.170391	4.7074	90835 (38.97%)	274225	0.168573	4.1317	115721 (42.2%)
Whole genome	1830925	0.214383	1.7613	700608 (38.27%)	1828675	0.184979	1.6731	836400 (45.74%)	2140825	0.173084	1.5262	1101800 (51.47%)

TNSP Total number of SNP pairs, *NSSP* Number of significant SNP pairs ($P < 0.001$)

Table 4 Diversity indexes of 292 wheat genotypes using 43203 SNPs

Chromosomes	No. SNP	GeneDiversity	Heterozygosity	PIC	MAF	Ho	Hs	Ht	Fst
1A	2226	0.313	0.031	0.256	0.222	0.029	0.287	0.320	0.084
2A	2731	0.376	0.030	0.296	0.298	0.030	0.357	0.383	0.061
3A	1928	0.320	0.032	0.260	0.235	0.031	0.302	0.334	0.084
4A	2593	0.357	0.025	0.284	0.274	0.024	0.320	0.357	0.091
5A	1435	0.317	0.032	0.259	0.222	0.031	0.307	0.327	0.055
6A	1978	0.332	0.031	0.269	0.243	0.029	0.315	0.338	0.059
7A	2987	0.337	0.030	0.272	0.247	0.029	0.308	0.340	0.079
1B	3009	0.341	0.031	0.275	0.248	0.029	0.323	0.348	0.062
2B	3738	0.346	0.031	0.278	0.258	0.030	0.326	0.355	0.073
3B	4021	0.346	0.029	0.279	0.253	0.028	0.315	0.363	0.110
4B	1200	0.296	0.027	0.244	0.209	0.026	0.287	0.313	0.072
5B	3044	0.350	0.031	0.282	0.258	0.030	0.326	0.363	0.087
6B	3801	0.342	0.029	0.276	0.253	0.028	0.314	0.347	0.081
7B	2998	0.313	0.032	0.257	0.219	0.031	0.303	0.327	0.061
1D	971	0.302	0.029	0.248	0.212	0.026	0.280	0.300	0.059
2D	1368	0.270	0.027	0.227	0.178	0.026	0.248	0.270	0.065
3D	736	0.271	0.026	0.227	0.181	0.024	0.250	0.271	0.059
4D	268	0.314	0.028	0.256	0.227	0.027	0.282	0.311	0.078
5D	632	0.284	0.027	0.237	0.191	0.026	0.264	0.292	0.076
6D	763	0.319	0.032	0.259	0.229	0.030	0.296	0.313	0.048
7D	925	0.303	0.029	0.248	0.215	0.027	0.278	0.298	0.055
Unknown	851	0.315	0.051	0.258	0.221	0.049	0.299	0.325	0.069
A genome	15878	0.339	0.030	0.273	0.252	0.029	0.316	0.345	0.074
B genome	21811	0.338	0.030	0.273	0.247	0.029	0.316	0.349	0.080
D genome	5663	0.291	0.028	0.241	0.201	0.026	0.269	0.290	0.061
Whole genomes	43352	.332	0.030	0.269	0.243	0.029	0.310	0.340	0.075

PIC polymorphic information content, *MAF* minor allele frequency, *Ho* observed heterozygosity, *Hs* discovered heterozygosity, *Fst* Fixation index

chromosome organization (i.e., ATP and protein binding), transmembrane transport (i.e., solute:proton antiporter activity), response to ethylene (i.e., ethylene receptor activity and ethylene binding), transcription regulation (i.e., transcription factor activity), protein phosphorylation (i.e., kinase and phosphatase activity), ubiquitination (i.e., ubiquitin-protein transferase activity), and phylloquinone biosynthesis (methyltransferase activity). The strongest associations are summarized in Tables 6 and 7.

In plants cultivated under normal conditions, 11 SNPs fell within coding genes located on Chr2A, 5A, 6A, 7A, 1B, 7B, and 6D that were related to GI, GE, GP, GR, PL, RL, SL, SV, DW, and WW traits. In moderately salinity-treated plants, 10 SNPs were related to genes responsible for GI, GE, GP, PL, RL, SL, SV, DW, WW, and ShL-RL traits. These SNPs were located on Chr1A, 2A, 1B, 2B, 3B, 5B, and 6B and no significant SNP was located in the D sub-genome. In the severe salinity-treated plants, 15 markers located on Chr 5A, 6A, 7A, 1B, 2B, 4B, 6B, 2D, and 5D were identified for all traits. The Quantile–quantile and Manhattan plots of highly associated SNPs for germination-related characteristics are shown in Fig. 5 and Supplementary1 Fig S2.

Putative candidate genes

Putative candidate genes were identified for germination performance in salt-stressed wheat by annotating the function of discovered genes and their respective homologs in rice (Tables 8 and 9). Eight and nine common genes were identified for the moderate and severe salinity stress treatments, respectively. These genes encode protein products involved in DNA/RNA/protein binding, ATP binding, transferase activity, transportation, phosphorylation, and ubiquitination under moderate and severe salinity. The gene ontology of these genes indicated that they are responsible for "response to stresses", "metabolic processes", and "transcription". It

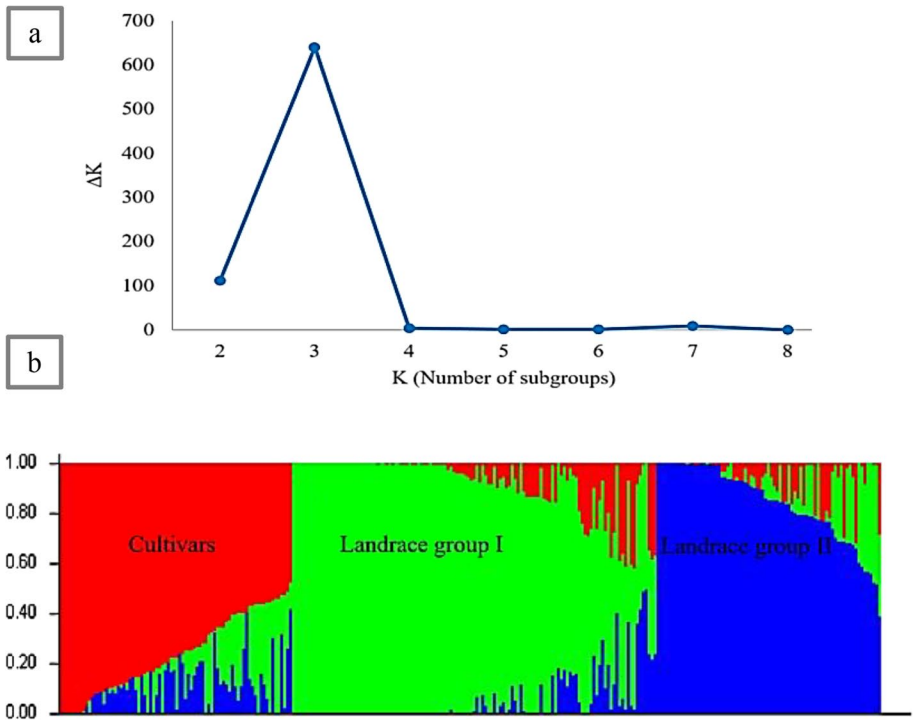

Fig. 2 The number of subpopulations in the wheat panel based on ΔK values (**a**), A structure plot of 292 wheat cultivars and landraces determined by K = 3 (**b**)

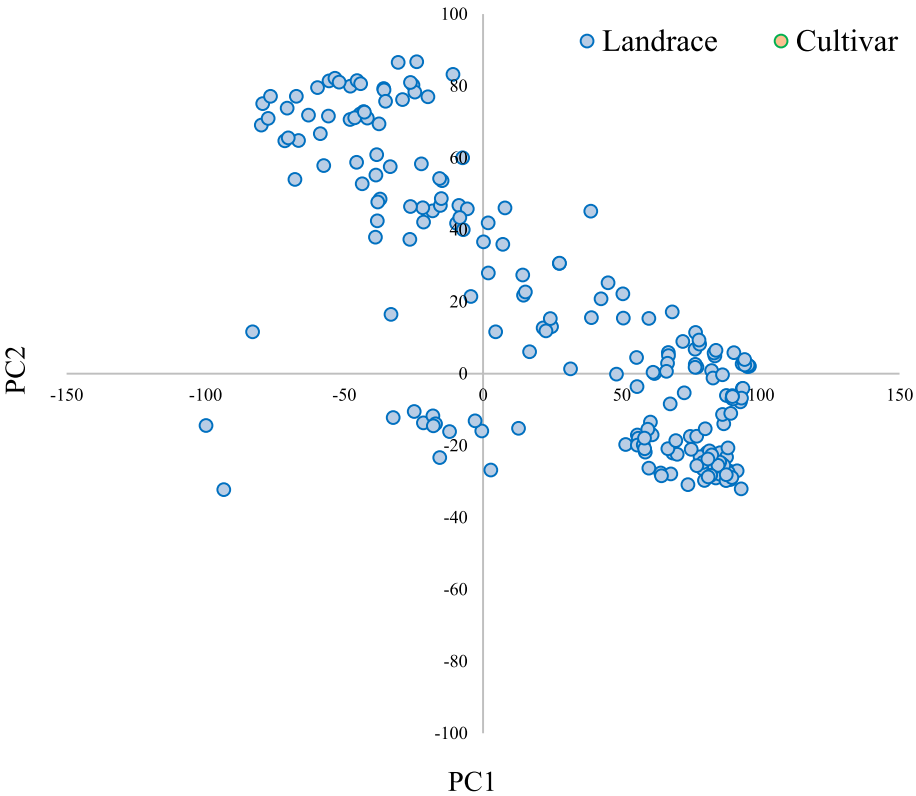

Fig. 3 Principal component analysis of Iranian bread wheat accessions using 46,203 markers

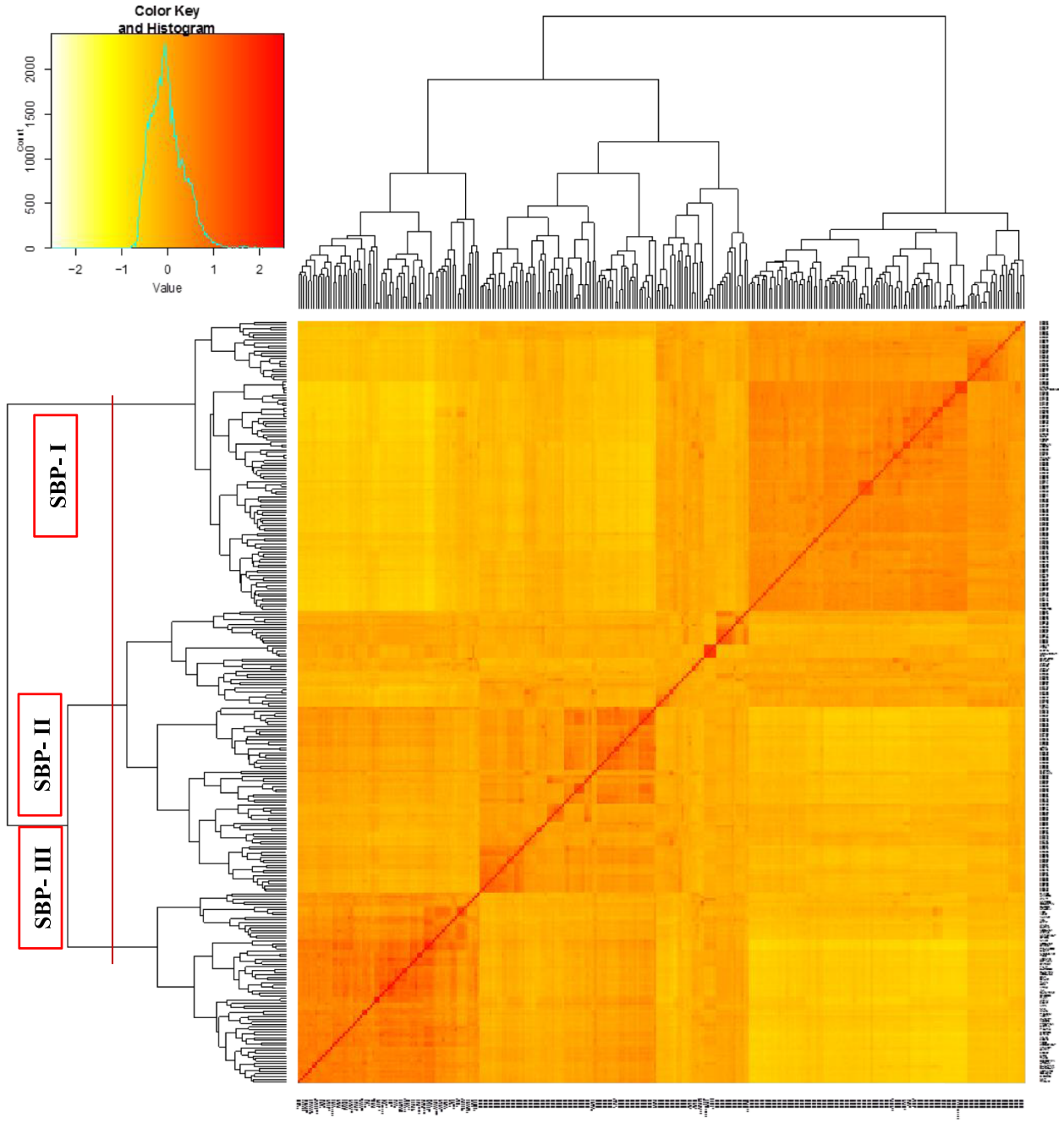

Fig. 4 Kinship matrix-based cluster analysis of Iranian wheat accessions using 46,203 markers

is worth noting that the genomic regions associated with seedling salt tolerance are problematic for comparison across various studies because of differences in the mapping population and marker platforms and in the absence of a consensus map for comparing genomic locations. Based on the rice reference genome, the oxidative phosphorylation (Fig. 6) and fatty acid elongation pathways were discovered (Supplementary1 Fig S3).

Discussion

The mechanisms underlying better germination performance of salt-stressed wheat remain almost largely unexplored. To better understand wheat germination responses to high-salt environments, we sought to identify alleles and genes linked with salt tolerance during germination using a diverse panel of accessions. Association mapping was a useful approach in this

Table 5 The number of marker-trait associations (MTAs) for germination-related traits of Iranian bread wheat accessions under moderate and severe salinity conditions

Genome	Moderate Stress	Severe Stress
Genome A	76	76
Genome B	129	67
Genome D	32	36
MTAs	237	179

study, as it exploits historical recombination events and provides high-resolution maps to identify candidate genes in response to salt stress in wheat [18].

In this study, we used a total of 11 interconnected germination traits, which were measured under normal and two salinity conditions. These traits have been successful for genomic mapping of salinity endurance during seedling germination in alfalfa [11], sesame [12], rapeseed [13], cotton [14], barley [9], and rice [15].

Seed germination is considered the first and most crucial stage in crop growth and development [13]. Seed germination begins with water imbibition, which is suppressed by salt stress, thus directly interfering with germination progression [11]. Similar to our observations on the detrimental impact of salinity on germination-related traits, researchers have previously highlighted that salinity can delay germination initiation and hence decrease seed vigor [15]. A significant, positive correlation was observed between all traits, reflecting a similar impact of salinity. Similarly, Naveed et al. [19] found a significant, positive correlation between root length and germination rate (GR) at the germination stage of rice.

From our findings, most SNPs were identified in the B and A sub-genomes while a lower number of SNP markers were found in the D sub-genome. This is consistent with previous reports [20]. The same trend was observed for linked marker pairs, where B sub-genome related MPs were nearly four times more common than those from the D sub-genome. The higher variation observed in the A and B sub-genomes may be a consequence of the following two factors: gene flow from *T. turgidum* L. (but not *Ae. tauschii* Coss.) to bread wheat or the older evolutionary background of the A and B sub-genomes [21]. In addition, it is possible that a bottleneck due to strong selection in ancestral 6×landraces during the improvement of cultivars has differentially affected the D sub-genome [22]. This bottleneck reduced the effective population size, which in turn increased the rate of rare allele loss. Moreover, the higher fraction of low-frequency alleles observed in the D sub-genome indicates a decrease in allelic variants [22]. Linkage disequilibrium and SNP distance across the B and A sub-genomes were much lower than in the D sub-genome. The fact that cultivars exhibited higher LD in contrast to landraces, especially in the D sub-genome, is presumably a consequence of numerous rounds of selection during breeding for key crop characteristics [23]. Overall, population relatedness, genetic drift, recombination, mating systems, mutation, and selection are all the main factors influencing the pattern and extent of LD [24].

A total of 10 and 15 significant, functional MTAs were detected under moderate (60 mM NaCl) and severe (120 mM NaCl) salinity, respectively. Aligning the sequences surrounding these SNPs to the reference genomes showed that most of these genes are responsible for response to ethylene (i.e., ethylene receptor activity and ethylene binding), oxidation–reduction (i.e., monooxygenase and oxidoreductase activity), chromosome organization (i.e., ATP and protein binding), transcription regulation (i.e., transcription factor activity), protein ubiquitination (i.e., ubiquitin-protein transferase activity), transmembrane transport (i.e., solute:proton antiporter activity), and protein phosphorylation (i.e., kinase and phosphatase activity). Such associations have also been demonstrated in earlier reports [9, 11, 13–15].

To date, several genes and QTLs linked to salinity tolerance at the germination stage have been identified by using linkage and association mapping in several plants and crops. In this study, we successfully identified eight and nine candidate genes for mediating tolerance to moderate and severe salinity stress, respectively. These genes encode proteins or enzymes involved in DNA/RNA/protein binding, ATP binding, transferase activity, transportation, phosphorylation, and ubiquitination. Similar functional roles have been discovered for salt-responsive genes during seed germination in cotton [14], rapeseed [13], alfalfa [11], rice [15], and barley [9].

One of the main consequences of salt stress is ionic challenge, which arises due to excess levels of Na^+. Since this ion interferes with K^+ homeostasis, cytosolic Na^+/K^+ balance is a key mechanism mediating salinity tolerance. Maintaining this homeostatic balance requires functional K^+ and Na^+ ion channels, transporters, or both [25]. Our work led to the identification of a salt-responsive gene, cation/H antiporter 2 (*CHX2*), which regulates Na^+ and K^+ homeostasis. This gene is regulated differentially in various tissues and during different growth stages in response to salinity [25]. Shi et al. [16] also identified a genomic region associated with germination time located on Chr1, near a QTL for total K^+ concentration, total Na^+ uptake, and Na^+: K^+ ratio.

Protein phosphorylation is the most widespread post-translational modification and can affect all cellular processes, such as metabolic reactions, translation,

Table 6 Description of highly significant MTAs for germination-related traits of Iranian wheat accessions exposed to the moderate salinity stress

Marker	Sequence	Trait	Ch	Position (bp)	MAF[a]	Molecular function	Biological process
rs10987	TGCAGCAAAGTCACATGTAGTACT ACGTACTCAACTTAAACCTGGGGG AGGGAGTAGCGGTGTG_12	Dry weight	3B	610,115,198–610,117,224	0.41	methyltransferase activity, 2-phytyl-1,4 naphthoquinone methyltransferase activity	phylloquinone biosynthetic process
rs51559	TGCAGGGTATATCAGCATCAGTTG ACTTTTCTTCCACTTTTGTGTTTTACT AGATGGAGATGCG_19	Germination energy	2A	705,812,562–705,821,732	0.08	protein binding,ATP binding	chromosome organization
rs28502	TGCAGCGACGGTGGCGGCCTGGTC GGCGTTGTGGAGCACGCGCCGAGA TCGGAAGAGCGGGATC_16	Germination index	1B	36,844,664–36,850,968	0.19	monooxygenase activity, iron ion binding, oxidoreductase activity, acting on paired donors with incorporation or reduction of molecular oxygen, heme binding	oxidation–reduction process
rs51559	TGCAGGGTATATCAGCATCAGTTG ACTTTTCTTCCACTTTTGTGTTTTACT AGATGGAGATGCG_19	Germination percentage	2A	705,812,562–705,821,732	0.08	protein binding, ATP binding	chromosome organization
rs39391	TGCAGCTGGAAGTGCAGGATGAGG GTGATGCTTTCGCCATCCAATGCC GACCACATGGTACCAT_42	Plomule length	6B	632,179,674–632,182,484	0.09	protein kinase activity, protein binding, ethylene receptor activity, ethylene binding	phosphorelay signal transduction system, response to ethylene
rs18853	TGCAGCATAATTTATTGTTGCAGGCTA AGTAGCATCGTGCTGTTAATCGCA ACTTTGTGCCATA_39	Root length	5B	2,462,678–2,510,604	0.06	protein kinase activity, ATP binding	protein phosphorylation
rs45886	TGCAGGCAAGGGCGCCAGGGCGGC CCACACAGGGTGGTAACAGAGTTG AGGAGGGAATAGGGGT_40	Shoot length	1A	616,821,401–616,824,347	0.42	ubiquitin-protein transferase activity, protein binding, ligase activity	protein ubiquitination
rs17383	TGCAGCAGCTCAAATTCCTCAGTG TTCAAAAATGCAAGATGGTGCAGT TGGTAGAGATAGATGC_16	Ratio of shoot length to root length	1A	459,624,435–459,627,780	0.46	protein binding	
rs18853	TGCAGCATAATTTATTGTTGCAGGCTA AGTAGCATCGTGCTGTTAATCGCA ACTTTGTGCCATA_39	Seedling vigor	5B	97,940,658–97,950,133	0.06	transferase activity, transferring acyl groups	GPI anchor biosynthetic process
rs15343	TGCAGCACGTGCTGCTCTACGAGG GCCACGACGACACGCCGCCCTTCG ACGCTATCGTTCAGTC_63	Wet weight	2B	749,888,161–749,894,706	0.42	transferase activity, transferring acyl groups other than amino-acyl groups	

[a] *MAF* minor allele frequency

Table 7 Description of highly significant MTAs for germination-related traits of Iranian wheat accessions exposed to the severe salinity stress

Marker	Sequence	Trait	Ch	Position (bp)	MAF*	Molecular function	Biological process
rs10035	TGCAGGCAGCGATTCGATGGCTGC TGCACACCATCCTGCCAACCTCCT AGTCGGCGGCAAGCTG_13	Dry weight	5D	256,463,138–256,485,526	0.05	Protein kinase, phosphoprotein phosphatase activity, ATP binding, phosphatase activity cation binding, metal ion binding	protein phosphorylation
rs3046	TGCAGAATTTCTCTTTCAGAAAAG ATGGGGAACTGAGCCTGGATTCCT CCCGTCACTGTCAGCA_61	Germination energy	7A	593,059,525–593,063,135	0.1	oxidoreductase activity	oxidation–reduction process
rs16308	TGCAGCAGCACGCCCGAGACCAGC AAGGGCTCCGTCCTCTCGCTTTCT CGCTCCGACAGCCGCG_12	Germination energy	2B	357,918,017–357,924,451	0.46	DNA-binding, transcription factor activity, sequence-specific DNA binding	oxidation–reduction process, DNA-templated regulation of secondary shoot formation
rs16308	TGCAGGAACTCCGCGAGATTTAGG GCCTCCCAGTCAGGCCCCAACGCC TCGCGAACGAAAGTCC_53	Germination index	2B	357,918,017–357,924,451	0.46	DNA-binding, transcription factor activity	
rs16825	TGCAGCAGCCGAAACAAAGTGCAA CGACAGCGCGAGCTCGGGCAGGTC GGGTGGCAAGATGAGC_28	Germination index	1B	379,150,934–379,159,267	0.09	DNA binding, transcription factor activity	transcription, DNA-templated regulation of transcription
rs3046	TGCAGAATTTCTCTTTCAGAAAAG ATGGGGAACTGAGCCTGGATTCCT CCCGTCACTGTCAGCA_61	Germination percentage	7A	593,059,525–593,063,135	0.1	oxidoreductase activity	oxidation–reduction process
rs1619	TGCAGAACTTCACTGCGTCACACA ACAACTTCAGCGGCAGGCTGCCAT CCTTCCTCGAGAATTG_55	Germination rate	4B	657,270,914–657,273,316	0.28	protein binding	
rs25188	TGCAGCCGCCAAGACCGCACTCGG CAGTCGACGCCCGAGATCGGAAGA GCGGGATCACCGACTG_12	Plomule length	5A	97,940,658–97,950,133	0.31	transferase activity, transferring acyl groups	GPI anchor biosynthetic process
rs5312	TGCAGAGACACGAGTCGGTTCAGC GCCAGCATCCCCATGGACGAGCAG ACGACCATACCTTCCT_53	Root length	6A	23,635,590–23,638,481	0.10	protein binding	
rs56460	TGCAGTACGTACGTAAGGGGGGGT TTAGATCACATGCACCTGGTACCT GAGAGCTTCTGACGCG_10	Root length	1B	16,426,498–16,428,373	0.41	transferase activity, transferring acyl other than amino-acyl groups	
rs61435	TGCAGTGGCCTGTAGGACCTGCAA TTATCTTGATAACTACAACTCCCA GGCTGAATATATCCAG_48	Shoot length	5D	529,826,163–529,827,625	0.12	protein kinase activity, ATP binding	protein phosphorylation
rs45003	TGCAGGAGTGCTCCATATGTGGGC ACTACAGACGGAAAGGCATGTGGT CATCATCCAACTACGC_13	Ratio of shoot length to root length	2D	613,324,693–613,325,883	0.11	protein binding	
rs30686	TGCAGCGCGCGCGATGCACCGTGG GGATCCTCGTGGAGCGGCGCCTCC CCGAGATCGGAAGAGC_13	Seedling vigor	6B	473,514,839–473,517,451	0.26	solute:proton antiporter activity	cation transport, transmembrane transport
rs49434	TGCAGGGAACCACGACAACACTAG GGACAGGTCCGCGACGCAGGAGAG GCATGGGTGCACAATG_12	Seedling vigor	5A	405,948,741–405,952,278	0.41	DNA binding	

Table 7 (continued)

Marker	Sequence	Trait	Ch	Position (bp)	MAF*	Molecular function	Biological process
rs56460	TGCAGTACGTACGTAAGGGGGGGT TTAGATCACATGCACCTGGTACCT GAGAGCTTCTGACGCG_10	Wet weight	1B	16,426,498–16,428,373	0.26	transferase activity, transferring acyl other than amino-acyl groups	

**MAF* Minor Allele Frequency

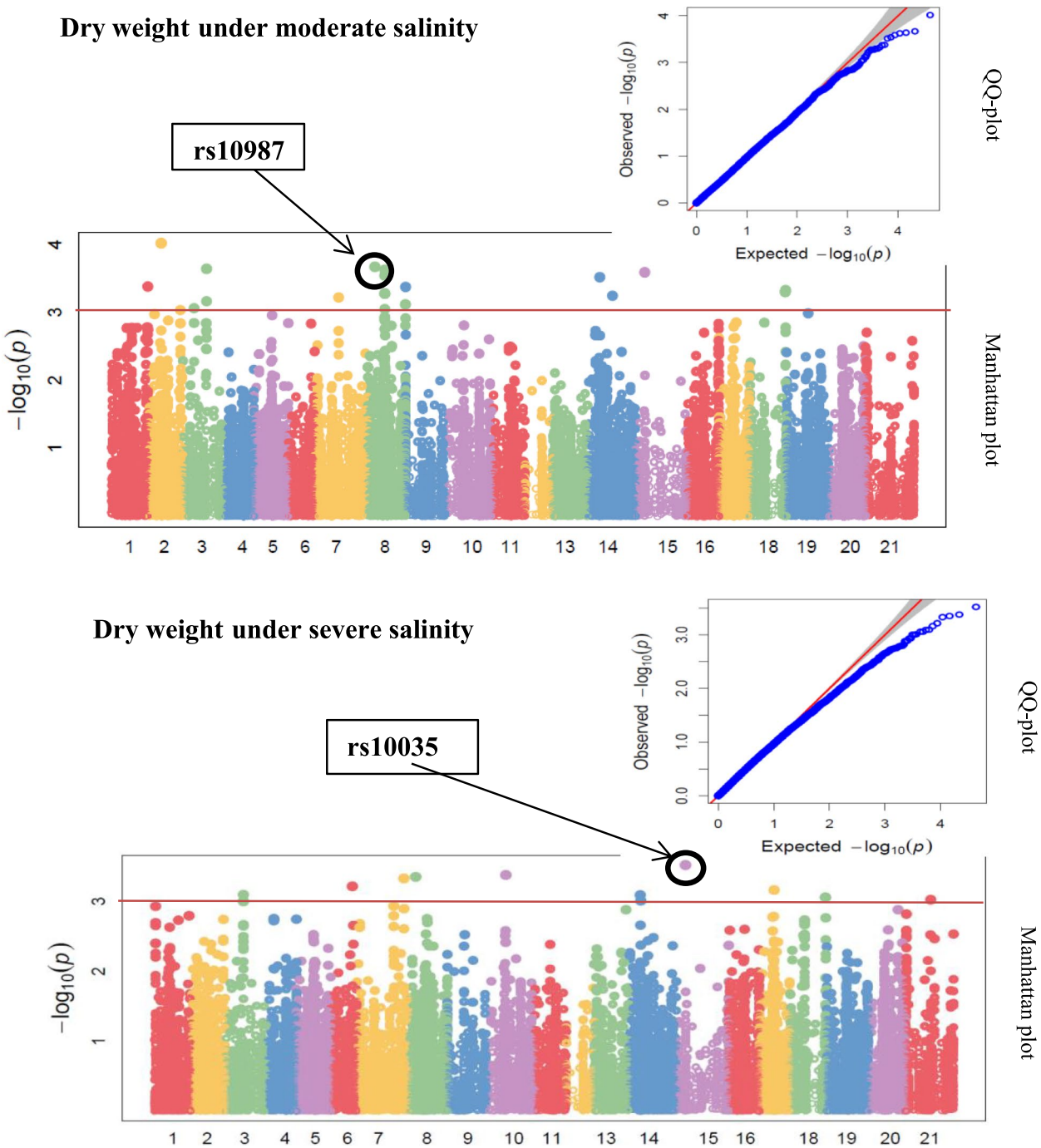

Fig. 5 The mrMLM-based Manhattan and QQ-plots of highly associated haplotypes for germination-related traits (herein, dry weight) under moderate and severe salinity conditions. X axis represents chromosome number [1)1A, 2)1B, 3)1D, 4)2A, 5)2B, 6)2D, 7)3A, 8)3B, 9)3D, 10)4A, 11)4B, 12)4D, 13)5A, 14)5B, 15)5D, 16)6A, 17)6B, 18)6D, 19)7A, 20)7B, and 21)7D] and Y axis represents $-\log_{10}$(p), respectively. Line in red is referred to the cut-off threshold

transcription, signal perception, and transduction under salt stress [26]. Generally, phosphorylation and dephosphorylation catalyzed by kinase and phosphatase enzymes can regulate protein configuration and modify their intracellular localization, function, enzyme activity, substrate specificity, and structure stability [26]. We identified a salt-responsive gene, protein kinase 2 (*GMPK2*), which regulates salt stress tolerance [26]. Luo

Table 8 Annotation of genes harbouring significant trait-associated SNPs across all chromosomes in Iranian wheat accessions exposed to the moderate salinity stress

Marker	Ch	Position (bp)	*p*-value	FDR	Gene ID in wheat	Homolog gene ID in rice	Description
rs10987	3B	22764	0.0002139714	0.9958341268	TraesCS3D02G526300	Os07g0587100	transferase activity, glycosyltransferase activity
rs51559	2A	74319	0.0003315372	0.9784545081	TraesCS2A02G457500	OsWD40-29 Os01g0924300	protein binding
rs28502	1B	17065	0.0002066565	0.9808556034	TraesCS2B02G069400	Os08g0326500	hydrolase activity, hydrolyzing O-glycosyl compounds, hydrolase activity, acting on glycosyl bonds, glucan endo-1,3-beta-D-glucosidase activity
rs51559	2A	74319	0.0003578241	0.9791934890	TraesCS2A02G457500	OsNTP10 Os10g0188300	nucleotidyltransferase activity, RNA uridylyltransferase activity
rs39391	6B	52377	0.0002147662	0.9994740214	TraesCS6B02G360200	Os01g0611900	protein binding
rs18853	5B	80923	0.0001298730	0.9988121296	TraesCS4B02G004000	Os05g0388500	RNA binding, mRNA binding, structural constituent of ribosome rRNA binding
rs17383	1A	85455	0.0001310011	0.7098953430	TraesCS1D02G386700	Os04g0623700	RNA binding, 7S RNA binding, endoplasmic reticulum signal peptide binding
rs45886	1A	616,821,401–616,824,347	0.0110322938	0.9999893491	TraesCS6A02G419200	OsPUB	protein ubiquitination
rs15343	2B	86479	0.0001863726	0.9950479430	TraesCS2B02G554900	OsGELP63 Os05g0209600	hydrolase activity, acting on ester bonds

Table 9 Annotation of genes harbouring significant trait-associated SNPs across all chromosomes in Iranian wheat accessions exposed to the severe salinity stress

Marker	Ch	Position (bp)	*p*-value	FDR	Gene ID in wheat	Homolog gene ID in rice	Description
rs10035	5D	256,463,138–256,485,526	0.0003028373	0.9882921528	TraesCS6D02G394300	Os02g0260700	protein binding
rs3046	7A	593,059,525–593,063,135	0.0001452906	0.9757731237	TraesCS2B02G109900	GMPK2 Os04g0546300	protein kinase activity, protein serine/threonine kinase activity, ATP binding
rs16308	2B	357,918,017–357,924,451	0.0003154564	0.9896520404	TraesCS5B02G005700	-	-
rs1619	4B	657,270,914–657,273,316	0.0006252700	0.9889360765	TraesCS4B02G372200	Os10g0469300	protein binding
rs25188	5A	97,940,658–97,950,133	0.0002054657	0.9999669788	TraesCS7D02G157100	Os02g0458900	hydrolase activity, acting on ester bonds
rs5312	6A	23,635,590–23,638,481	0.0001494948	0.9996324274	TraesCS6A02G046000	OsF-box220 Os04g0479800	protein binding
rs61435	5D	529,826,163–529,827,625	0.0004237373	0.9998367235	TraesCS5D02G502300	OsCutA1 Os10g0378300	copper ion binding
rs45003	2D	613,324,693–613,325,883	81616	0.4179819854	TraesCS2D02G525100	Os04g0667900	-
rs30686	6B	473,514,839–473,517,451	0.0002362291	0.9986558102	TraesCS6A02G418500	OsCHX02 Os08g0550600	solute:proton antiporter activity
rs56460	1B	16,426,498–16,428,373	0.0003256200	0.9994840151	TraesCS1B02G033700	Os12g0511900	ADP binding

et al. [5] also identified a Chr1-located SNP that controlled expression of a serine/threonine-phosphoprotein phosphatase and was associated with dry weight under salinity stress. Moreover, Yu et al. [11] identified 14 functional genes linked to 23 SNPs in alfalfa using a GWAS approach. Marker S2_46544981 on Chr2 was linked to a Ser/Thr protein kinase with a role in the ABA pathways in response to salinity. Mwando et al. [9] also performed GWAS during germination on a salt-tolerance index in 350 barley accessions using 24 138 SNPs and

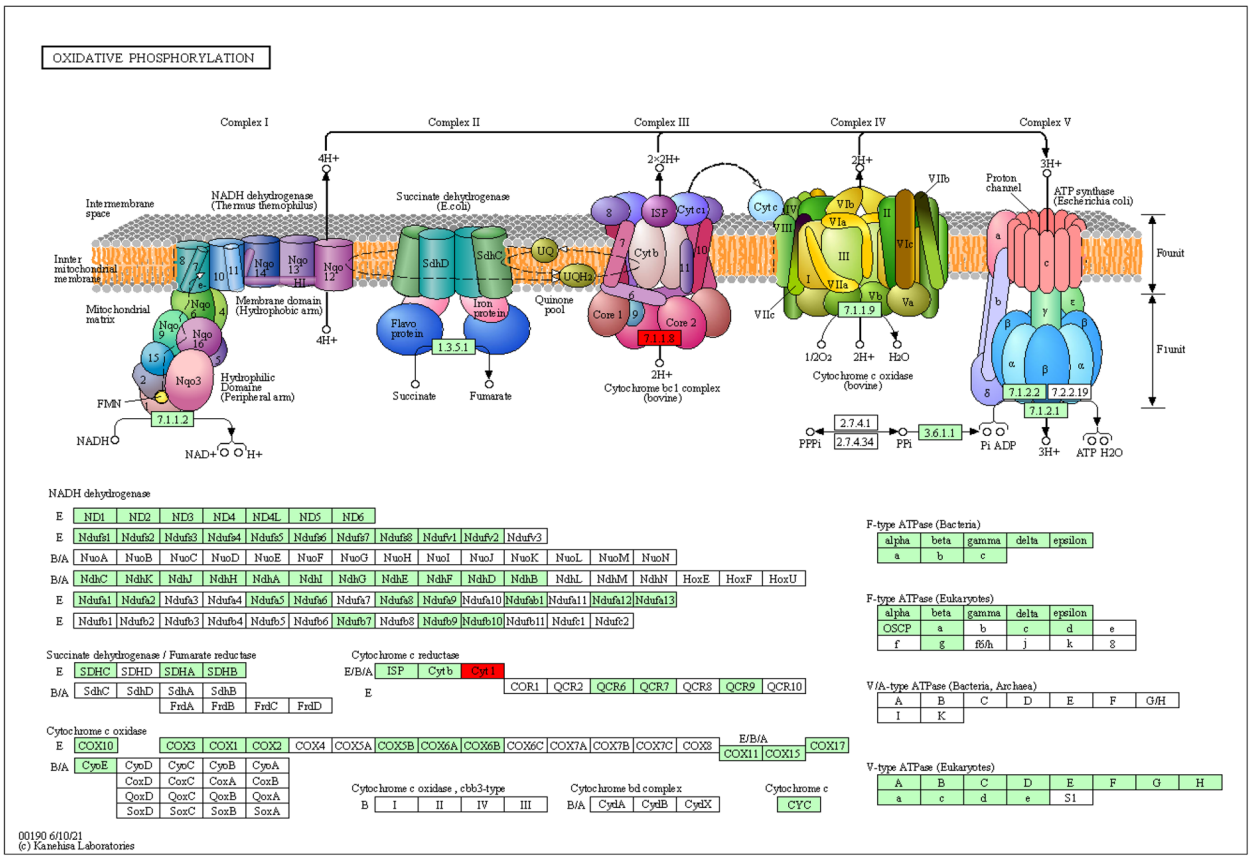

Fig. 6 The KEGG pathway of oxidative phosphorylation. Reference pathway: this is the original version; white boxes are hyperlinked to KO, ENZYME, and REACTION entries in metabolic pathways; they are hyperlinked to KO entries in non-metabolic pathways. Reference pathway (KO): blue boxes are hyperlinked to KO entries that are selected from the original version. Reference pathway (EC): blue boxes are hyperlinked to ENZYME entries that are selected from the original version. Reference pathway (Reaction): blue boxes are hyperlinked to REACTION entries that are selected from the original version

DArTseq markers. They identified 19 quantitative trait nucleotides (QTNs) and several candidate genes, such as a protein kinase, which was regulated by microRNAs and transcription factors and played a key role in salinity tolerance.

In ubiquitination, *U-box E3* ligases have key functions in degradation of proteins via post-translational modifications in stress responses [27]. We identified a salt-responsive *U-box E3 Ub* ligases *(OsPUBs)* gene, which has previously been demonstrated to regulate salinity response [28]. Tan et al. [13] also used association mapping for germination percentage (GP) and germination index (GI) in 520 *Brassica napus* L. accessions using 52 157 SNPs. They reported a significant SNP, rs9466, located 2.2 kb away from a stress-responsive gene that encodes an E3 ubiquitin-protein responsible for control of proline biosynthesis.

Transferase activity, including catalysis of the transfer of a group, e.g., a nucleotide or glycosyl from one compound to another, plays a key role in cellular mechanisms regulating plant responses to salt stress [29]. Our study detected a transferase enzyme, rice nucleotidyl transferase protein 10 *(OsNTP10)*, which uridylates small RNAs and stimulates their degradation and plays a role in salt tolerance [29].

Apart from the processes discussed above, several other processes are important for wheat response to salt stress and have been considered in various studies. For instance, Yuan et al. [14] detected 17 SNPs located near or within 98 candidate genes from 13 different genomic regions, with 35 genes involved in *Gossypium hirsutum* L. reactive to salinity. Based on results from RNA-seq, the eight salt-responsive genes were shown to be homologous to known salt-tolerance genes, such as *GAPN* (engaged in GAPDH activity), *BGAL3* (engaged in carbohydrate metabolism), *HVA22E* (response to abscisic acid), *RR23* (engaged in CK-mediated signaling), *CHIT1* (chitotriosidase 1), *ERF2* (ethylene-responsive TF), *CLPB1* and *HSP18.2* (heat shock protein), *WAK2* (wall-associated receptor kinase 2), and *CUT1* (involved in

fatty acid biosynthesis). Based on a GWAS of RL in rice, Yu et al. [17] also found the Chr4-located *OsMADS31*, a MADS-box transcription factor that is down-regulated by salinity and is involved in stress tolerance during germination.

Theoretical modeling studies on regulatory networks are helpful for understanding regulatory mechanisms underlying salt-stressed wheat. As a reliable method of theoretical modeling, machine learning is an alternative to classic statistical procedures in GWAS analysis [30]. The use of machine learning algorithms in association mapping has been reported for the first time by Szymczak, et al. [30], who elucidated random forests, Bayesian network analysis, and artificial neural networks on human diseases. Recently, Yoosefzadeh-Najafabadi et al. [31] used machine learning-based GWAS for exploring QTLs responsible for soybean yield and successfully identified the respective MTAs. It seems that using the machine learning algorithms may improve our understanding of regulatory networks involved in wheat response to salinity [32].

As mentioned above, combining mathematical modeling with experimental analysis is important for better understanding of regulatory mechanisms underlying stresses. Li et al. [33] used ordinary differential equation-based modeling to explore cellular regulatory mechanisms and reveal the stochastic dynamics of this signaling system by the potential landscape theory (PLT), which is a recent tool for uncover unknown regulatory mechanisms. They successfully found the first landscape of signaling-induced cell death. It seems that search for stress-related enzymatic activities and use the PLT can shed light on salt-induced signaling. Discovering stress-related secondary metabolites by deep learning is another stress research topic of interest. Sun et al. [34] employed graph convolutional network with graph attention network (GCNAT) alongside experimental analysis for prediction the disease-metabolite associations. They found that GCNAT can be a powerful tool to predict the correlations between metabolites and obesity, colorectal cancer, and Alzheimer's diseases. As a result, metabolite profiling of salt-stressed wheat alongside using neural networks provide an opportunity to detect salt-responsive natural products. Recently, mathematical modeling has been suggested to uncover competition of mRNA droplet pattern. Xu et al. [35] revealed the control mechanisms of protein phase-separated pattern formation. This successful attempt reflects a new approach to determine the droplet patterns of salt-induced mRNA. Although these approaches were not used in this study, there is great potential in using them in future studies focused on the relationship between salinity stress and wheat responses such as stress-related miRNA [36].

Taken together, our results provide novel insights into the genetic basis of wheat germination performance under salt stress. Moreover, the candidate genes identified in this study were located in genomic regions displaying strong associations with germination-related characteristics and are thus potential targets for wheat breeders. Breeding for germination performance under salt stress is challenging as both tolerance and germination are polygenic traits and are controlled by multiple genes. A possible solution may be to pyramid the candidate genes in future studies.

Confirmation of GWAS results by using gene expression analysis can add a screening layer to the corresponding analysis steps, so that the expression of key genes related to stress, uncovered by GWAS, can be checked again under stressful conditions. Considering the limitations of this research, it is suggested that the genes identified in this study be further evaluated by gene expression techniques such as Real-Time PCR, RNAseq, etc.

Conclusion

In this study, association mapping of germination-related properties was performed on a genetic panel of 292 Iranian wheat accessions. A total of 25 significant, functional MTAs were located within protein-coding regions in the salt-stressed seedlings. From gene ontology, 17 candidate genes were identified and shown to be involved in mechanisms regulating germination performance under salt stress. These findings provide further knowledge of the molecular control of wheat germination performance under salt-stress conditions.

Material and methods

Plant materials

The plant material used in this study consisted of 90 cultivars and 202 landraces of Iranian bread wheat. These seeds were provided by the Seed and Plant Improvement Institute and the University of Tehran, Karaj, Iran. Further details on these 292 wheat accessions can be found in Supplementary 1 Tables S1 and S2. These samples are available at USDA and CIMMYT with USDA PI numbers and CIMMYT numbers (Supplementary 1 Table S1 and S2), respectively. The authors declare that all permissions or licenses were obtained to collect wheat plants from the University of Tehran, Iran.

Experimental design, treatments, and phenotyping

Wheat accessions were evaluated for germination performance under salt stress using the following three salinity levels: 0 (control), 60 (moderate stress), and 120 (severe stress) mM NaCl. The factorial experiment was performed in a completely randomized design (CRD) with two replicates in the laboratory of the Department

of Farming and Plant Breeding, Tehran University, Iran. ANOVA was performed on genotype and treatment. The first factor included the 292 Iranian bread wheat accessions and the second factor included three salinity levels (0, 60, and 120 mM NaCl).

To perform the experiment, two Petri dishes (diameter 90 mm and thickness 15 mm) were prepared for each treatment. Petri dishes were sterilized in an autoclave at 150 °C for 2 h. Two layers of Whatman qualitative filter paper No. 1 were placed inside each Petri dish. These filter papers were previously sterilized in an autoclave at 75 °C for 2 h. Each Petri dish was regarded as a replicate and consisted of 25 seeds. Seeds were surface-sterilized in 1% NaCl for 10 min, followed by washing with 70% ethanol for 30 s, and then immediately washed three times with distilled water. The parafilm-sealed Petri dishes were placed in a germinator for 8 days at 8/16 h of light/darkness, temperature 25 °C, and humidity 85%. The number of germinated seeds was counted after 72 h of incubation. Seeds were considered germinated after emergence of about 2 mm root. The germination rate was recorded daily until no further germinated seeds were observed for two consecutive counts. On the eighth day, root and shoot length (RL and ShL), plant length (PL, the sum of root and shoot length), the ratio of root to shoot length (ShL-RL), and seedling wet and dry weight (WW and DW) were measured. Several seed germination-related indices were also calculated from the measurements (Table 10).

The datasets derived from normal, moderate, and severe salinity conditions were analyzed using analysis of variance (ANOVA) to score significant differences ($P<0.01$) between landraces and varieties by SAS 9.4. SPSS Statistics 21.0 was used to calculate the descriptive statistics of phenotypic data.

Genotyping and SNP calling

Genotyping-by-sequencing (GBS) was employed to genotype all 292 wheat accessions [37]. The GBS libraries were developed and sequenced as described by Alipour et al. [38]. Sequencing reads were qualified and trimmed by FastQC and Trimmomatic to 64 bp, respectively, and grouped into sequence tags. SNPs were explored by NCBI-BLAST in default settings, which permit mismatches up to 3 bp. To call SNPs, the UNEAK pipeline was used in TASSEL [39], which is a software package to uncover linkage disequilibrium, evolutionary patterns, and trait associations. To avoid false-positive SNPs originating from sequencing errors, SNPs with a missing rate > 10% across samples, a minor allele frequency (MAF) < 1%, and heterozygosity > 10% were excluded. Missing data were imputed using the LD KNNi method in TASSEL [39]. For the SNP calling, we used the W7984 wheat genome as the reference genome [40].

Table 10 Seed germination indexes and their calculation formula

Index	Calculation formula
Germination Energy	$GE = \frac{nt_4}{N} \times 100$
Germination Index	$GI = \frac{(7n_1+6n_2+5n_3+4n_4+3n_5+2n_6+1n_7)}{7\times N}$
Seedling Vigor Index	$SI = \frac{GP\times LSh}{100}$
Germination Rate	$GR = \sum\left(\frac{n_1}{t_1} + \frac{n_2}{t_2} + \frac{n_i}{t_i}\right)$
Germination Percentage	$GP = \frac{n}{N} \times 100$

$n_1, n_2, \ldots, n_7$ represent the number of germinated seeds after 1, 2,…, 7 days and N represents the total number of seeds

Population genetic analysis

Population structure inference was performed using an admixture model implemented in Structure software [41]. The assumptive number of subpopulations (K) was regarded from K = 1 to K = 10 and 10 000 burn-in steps were followed by 10 0000 MCMC steps. The most likely K value was determined using the ΔK method in Structure Harvester [42]. The matrix of population structure, i.e. Q, was calculated for the whole population from the Structure analyses for the best value of K [39]. The kinship matrix (K) was calculated with the EMMA algorithm using the GAPIT package in R software [43]. A principal component analysis (PCA) was also performed using the Tidyverse package in R. A neighbor-joining tree was constructed based on a pairwise distance matrix by Jaccard index and visualized by Archaeopteryx to determine the relationship between landraces and cultivars. This software visualizes, analyzed, and edits potentially highly annotated and large phylogenetic trees [https://www.phylosoft.org/archaeopteryx/].

Linkage disequilibrium

LD among SNPs was estimated based on the values of observed/expected allele frequencies in TASSEL V.5. The full matrix option was utilized for estimating the distribution of LD for each subpopulation and in the whole association panel (WAP). The pairwise LD was determined using the squared correlation coefficient of alleles (r^2) as it is less sensitive to marginal allele frequencies. LD decay was also calculated for each chromosome and sub-genome based on the theoretical expectation of r^2 (see [44] for details).

GWAS analysis

GLM and MLM single-locus models adopt a genome scan test with one SNP at a time while needing multiple corrections (e.g., Bonferroni) for managing false

positives. This process is too conservative and may lead to loss of actual associations, which are fundamental for the traits of interest [45]. Moreover, single-locus models cannot simultaneously estimate all marker impacts, and thereby cannot present a proper model for genetic mapping the quantitative properties, which are modulated by the cumulative act of numerous genes [45]. To overcome these challenges, a multi-locus mixed linear model (mrMLM) approach was adopted for detecting reliable associations.

Briefly, the mrMLM procedure was implemented in two steps. First, all potentially associated SNPs were included in a second model, where their effects were estimated using an empirical Bayes approach. Finally, a likelihood ratio test was adopted to evaluate all non-zero marker effects. The multi-locus model was evaluated using the mrMLM package in R [45]. We used a significance threshold (cut-off) of $-\log_{10}$ (P-value) ≥ 3.0 ($P \leq 0.001$) for identifying significant associations in the model, as reported by many authors. All SNPs that met this cut-off value were categorized as significant MTAs. GWAS results were summarized using Manhattan plots for visualizing associations between genotypes and phenotypes using the GAPIT package [46]. In this plot, the x-axis and y-axis represent the genomic position of SNPs and the $-\log_{10}$ (P-value) obtained from the F-test, respectively. A Q-Q plot was also performed to assess the distribution of p-values obtained from the GWAS analyses [23].

Identification of candidate genes

Genome sequences surrounding all significantly associated SNPs were collected and used for gene annotation with BLAST against the IWGSC RefSeq v1.0 and the IRGSP 1.0 genome references for wheat and rice, respectively [40, 47]. After alignment, genes exhibiting the highest blast score and identity percentage were selected. The molecular function and biological processes of putative genes were detected in Ensembl Plants [http://plants.ensembl.org/Triticum_aestivum/Info/Index]. The identification of putative candidate genes was evaluated according to the following two parameters: being located in the vicinity of the peak marker and having known functions and involvement in the studied traits in both wheat and rice. Moreover, the significant SNPs were utilized in enrichment analysis of gene ontology via KOBAS version 2.0 for testing in the KEGG [48–50].

Abbreviations

ANOVA	Analysis of variance
CHX2	Cation/H antiporter 2
GBS	Genotyping-by-sequencing
GMPK2	Protein kinase 2
GWAS	Genome-wide association studies
LD	Linkage disequilibrium
MAF	Minor allele frequency
MP	Marker pair
MTA	Marker-trait association
NGS	Next-generation sequencing
PC	Principal component
QTL	Quantitative trait loci
SNP	Single-nucleotide polymorphisms
WAP	Whole association panel

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-11188-z.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

We extend our deepest gratitude to Peter Poczai for their valuable feedback and suggestions, which greatly improved this manuscript. Their insights were instrumental in refining our approach. We also sincerely thank Majid Jahromi, whose assistance was invaluable, particularly in facilitating the deposition of our data into the appropriate repository.

Permission for land study

The authors declare that all land experiments and studies were performed according to authorized rules.

Authors' contributions

SJ performed the experiments and data analysis and wrote the article draft; MrB, MO, ArA, HA, Pkl and PP supervised the project and provided editorial input on the writing. All authors discussed the results and contributed to the final manuscript. The author(s) read and approved the final manuscript.

Funding

This research did not receive any specific funding.

Data availability

The dataset generated and analyzed during the current study has been deposited in the European Variation Archive (EVA) at EMBL-EBI under project accession number PRJEB83663. The data is publicly available and can be accessed at: https://www.ebi.ac.uk/eva/?eva-study=PRJEB83663.

Declarations

Ethics approval and consent to participate

Plants samples were provided by the Gene Bank of Agronomy and Plant Breeding Group. These samples are available at USDA with USDA PI number (Supplementary1 Table S1 and S2), respectively. The authors declare that the study complied with relevant institutional, national, and international guidelines and legislation for plant ethics in the methods section. The authors declare that all that permissions or licenses were obtained to collect the wheat plant from the University of Tehran, Iran.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 2 June 2023 Accepted: 27 December 2024
Published online: 06 January 2025

References

1. Nazir R, Mandal S, Mitra S, Ghorai M, Das N, Jha NK, Majumder M, Pandey DK, Dey A. CRISPR/Cas genome-editing toolkit to enhance salt stress tolerance in rice and wheat. Plant Physiol. 2022;174:e13642.
2. Hu P, Zheng Q, Luo Q, et al. Genome-wide association study of yield and related traits in common wheat under salt-stress conditions. BMC Plant Biol. 2021;21:1–20.
3. Liu Y, Liu Y, Zhang Q, et al. Genome-wide association analysis of quantitative trait loci for salinity-tolerance related morphological indices in bread wheat. Euphytica. 2018;214:176.
4. Wang M, Xia G. The landscape of molecular mechanisms for salt tolerance in wheat. Crop J. 2018;6(1):42–7.
5. Luo Z, Szczepanek A, Abdel-Haleem H. Genome-wide association study (GWAS) analysis of camelina seedling germination under salt stress condition. Agronomy. 2020;10(9):1444.
6. Negrão S, Schmöckel SM, Tester M. Evaluating physiological responses of plants to salinity stress. Ann Bot. 2017;119:1–11.
7. El-Hendawy SE, Hu YC, Yakout GM, Awad AM, Hafiz SE, Schmidhalter U. Evaluating salt tolerance of wheat genotypes using multiple parameters. Eur J Agron. 2005;22:243–53.
8. Shannon MC, Grieve CM. Tolerance of vegetable crops to salinity. Sci Hortic. 1999;78:5–38.
9. Mwando E, Han Y, Angessa TT, Zhou G, Hill CB, Zhang XQ, Li C. Genome-wide association study of salinity tolerance during germination in barley (*Hordeum vulgare* L.). Front Plant Sci. 2020;11:118.
10. Wang M, Wang S, Xia G. From genome to gene: a new epoch for wheat research? Trends Plant Sci. 2015;20(6):380–7.
11. Yu LX, Liu X, Boge W, Liu XP. Genome-wide association study identifies loci for salt tolerance during germination in autotetraploid alfalfa (*Medicago sativa* L.) using genotyping-by-sequencing. Front Plant Sci. 2016;7:956.
12. Li D, Dossa K, Zhang Y, Wei X, Wang L, Zhang Y, Liu A, Zhou R, Zhang X. GWAS uncovers differential genetic bases for drought and salt tolerances in sesame at the germination stage. Genes. 2018;9(2):87.
13. Tan M, Liao F, Hou L, Wang J, Wei L, Jian H, Xu X, Li J. Genome-wide association analysis of seed germination percentage and germination index in *Brassica napus* L. under salt and drought stresses. Euphytica. 2017;2:213.
14. Yuan Y, Xing H, Zeng W, et al. Genome-wide association and differential expression analysis of salt tolerance in *Gossypium hirsutum* L at the germination stage. BMC Plant Biol. 2019;19:394.
15. Cui YR, Zhang F, Zhou YL. The application of multi-locus GWAS for the detection of salt-tolerance loci in rice. Front Plant Sci. 2018;9:1464.
16. Shi Y, Gao L, Wu Z, et al. Genome-wide association study of salt tolerance at the seed germination stage in rice. BMC Plant Biol. 2017;17:92.
17. Yu J, Zhao W, Tong W, He Q, Yoon M-Y, Li F-P, Choi B, Heo E-B, Kim K-W, Park YJ. A Genome-wide association study reveals candidate genes related to salt tolerance in rice (*Oryza sativa*) at the germination stage. Int J Mol Sci. 2018;19(10):3145.
18. Quan X, Liu J, Zhang N, Xie C, Li H, Xia X, He W, Qin Y. Genome-wide association study uncover the genetic architecture of salt tolerance-related traits in common wheat (*Triticum aestivum* L.). Front Genet. 2021;12:663941.
19. Naveed SA, Zhang F, Zhang J, et al. Identification of QTN and candidate genes for salinity tolerance at the germination and seedling stages in rice by genome-wide association analyses. Sci Rep. 2018;8:6505.
20. Berkman PJ, Visendi P, Lee HC, Stiller J, Manoli S, Lorenc MT, Lai K, Batley J, Fleury D, Simkova H, et al. Dispersion and domestication shaped the genome of bread wheat. Plant Biotechnol J. 2013;11(5):564–71.
21. Dvorak J, Akhunov ED, Akhunov AR, Deal KR, Luo MC. 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. 2006;23(7):1386–96.
22. Chao S, Zhang W, Akhunov E, Sherman J, Ma Y, Luo MC, Dubcovsky J. Analysis of gene-derived SNP marker polymorphism in US wheat (Triticum aestivum L.) cultivars. Mol Breed. 2009;23(1):23–33.
23. Rahimi Y, Bihamta MR, Taleei A, Alipour H, Ingvarsson PK. Genome-wide association study of agronomic traits in bread wheat reveals novel putative alleles for future breeding programs. BMC Plant Biol. 2019;19:541.
24. Liu H, Zhou H, Wu Y, Li X, Zhao J, Zuo T, Zhang X, Zhang Y, Liu S, Shen Y, et al. The impact of genetic relationship and linkage disequilibrium on genomic selection. PLoS ONE. 2015;10(7):e0132379.
25. Hussain Z, Khan H, Imran M, Naeem MK, Shah SH, Iqbal A, Ali SS, Rizwan M, Ali S, Muneer MA, Widemann E, Shafiq S. Cation/proton antiporter genes in tomato: genomic characterization, expression profiling, and co-localization with salt stress-related QTLs. Agronomy. 2022;12(2):245.
26. Pan J, Li Z, Wang Q, Guan Y, Li X, Huangfu Y, Meng F, Li J, Dai S, Liu W. Phosphoproteomic profiling reveals early salt-responsive mechanisms in two foxtail millet cultivars. Front Plant Sci. 2021;12:712257.
27. Sabzehzari M, Zeinali M, Naghavi MR. CRISPR-based metabolic editing: next-generation metabolic engineering in plants. Gene. 2020;759:144993.
28. Kim DY, Lee YJ, Hong MJ, Kim JH, Seo YW. Genome wide analysis of U-Box E3 ubiquitin ligases in wheat (Triticum aestivum L.). Int J Mol Sci. 2021;22(5):2699.
29. Narjesi V, Mardi M, Hervan EM, Azadi A, Naghavi MR, Ebrahimi MA, Zali A. Analysis of quantitative trait loci (QTL) for grain yield and agronomic traits in wheat (Triticum aestivum L.) under normal and salt-stress conditions. Plant Mol Biol Reprod. 2015;33:2030–40.
30. Szymczak S, Biernacka JM, Cordell HJ, González-Recio O, König IR, Zhang H, Sun YV. Machine learning in genome-wide association studies. Genet Epidemiol. 2009;33(S1):S51–7.
31. Yoosefzadeh-Najafabadi M, Eskandari M, Torabi S, Torkamaneh D, Tulpan D, Rajcan I. Machine-learning-based genome-wide association studies for uncovering QTL underlying soybean yield and its components. Int J Mol Sci. 2022;23(10):5538.
32. Javid S, Bihamta MR, Omidi M, Abbasi AR, Alipour H, Ingvarsson PK. Genome-Wide Association Study (GWAS) and genome prediction of seedling salt tolerance in bread wheat (Triticum aestivum L.). BMC Plant Biol. 2022;22(1):581.
33. Li X, Zhang P, Yin Z, Xu F, Yang ZH, Jin J, Qu J, Liu Z, Qi H, Yao C, Shuai J. Caspase-1 and Gasdermin D afford the optimal targets with distinct switching strategies in NLRP1b inflammasome-induced cell death. Research. 2022;2022:9838341.
34. Sun F, Sun J, Zhao Q. A deep learning method for predicting metabolite–disease associations via graph neural network. Brief Bioinforma. 2022;23(4):266.
35. Xu F, Miao D, Li W, Jin J, Liu Z, Shen C, Zhang J, Shuai J, Li X. Specificity and competition of mRNAs dominate droplet pattern in protein phase separation. Phys Rev Res. 2023;5(2):023159.
36. Sabzehzari M, Naghavi MR. Phyto-miRNA: a molecule with beneficial abilities for plant biotechnology. Gene. 2019;683:28–34.
37. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE. 2011;6(5):e19379.
38. Alipour H, Bihamta MR, Mohammadi V, Peyghambari SA, Bai G, Zhang G. Genotyping-by-sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. Front Plant Sci. 2017;8:1293.
39. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23(19):2633–5.
40. Alipour H, Bai G, Zhang G, Bihamta MR, Mohammadi V, Peyghambari SA. Imputation accuracy of wheat genotyping-by-sequencing (GBS) data using barley and wheat genome references. PLoS ONE. 2019;14(1):e0208614.
41. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945–59.
42. Earl DA, Holdt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4:359–61.
43. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, et al. Efficient control of population structure in model organism association mapping. Genetics. 2008;178(3):1709–23.
44. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES. Structure of linkage disequilibrium and phenotypic associations in the maize genome. PNAS. 2001;98:11479–84.

45. -Zhang YW, Lwaka Tamba C, Wen YJ, Li P, Ren WL, Ni YL, Gao J, Zhang YM. mrMLM v4.0: An R platform for multi-locus genome-wide association studies. GPB. 2020; https://doi.org/10.1016/j.gpb.2020.06.006.
46. Tang Y, Liu X, Wang J, Li M, Wang Q, Tian F, Su Z, Pan Y, Liu D, Lipka AE, et al. GAPIT version 2: an enhanced integrated tool for genomic association and prediction. TPG. 2016. https://doi.org/10.3835/plantgenome2015.11.0120.
47. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice. 2013;6:4.
48. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28:27–30. https://doi.org/10.1093/nar/28.1.27.
49. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 2019;28:1947–51. https://doi.org/10.1002/pro.3715.
50. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49:D545–51. https://doi.org/10.1093/nar/gkaa970.

Publisher's Note