



OPEN QTL mapping and transcriptome analysis of seed germination under PEG-induced water stress in *Lactuca* spp.

Sadal Hwang[✉], Ivan Simko & Beiquan Mou

The impact of limited water availability on lettuce growth has been well documented. However, the mechanisms by which lettuce controls seed germination under water stress remain unknown. Germination percentage was evaluated in the cv. Salinas (*Lactuca sativa*) (*L. sativa*) × US96UC23 (*Lactuca serriola*) (*L. serriola*) recombinant inbred line (RIL) population and USDA germplasm collection using 10% polyethylene glycol (PEG). About 50% of both populations displayed less than 90% germination. The average broad-sense heritability (H^2) for germination percentage was 0.81 across both populations. Two quantitative trait loci (QTL) for germination percentage were identified on chromosomes 4 and 8 in the RIL population. The RNA-Seq and network analyses of wild lettuce, US96UC23, were performed using the control (distilled water, dH_2O) and treatment (10% PEG) datasets. The number of differentially expressed genes (DEGs) was 4,095. The top 20 gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were assessed by enrichment analysis. The consensus network analysis captured 44 modules. Gene networks were constructed for the top 20 hub genes in 10 significant modules from each dataset. This study comprehensively explains QTL, GO terms, KEGG pathways, and gene networks associated with lettuce seed germination under osmotic stress.

Lettuce (*L. sativa*) is one of the major vegetable crops worldwide. It was presumed that the wild ancestor of lettuce was *L. serriola* and originated in the Mediterranean Rim^{1,2}. The U.S. is the second-largest lettuce producer after China³. California is a hub state to supply over 70% of U.S. lettuce⁴. According to the U.S. Drought Monitor⁵, the Palmer Drought Severity Index has indicated exceptional drought status in over 50% of California for decades, affecting both major lettuce production areas, the Central Coast and Central Valley. Lettuce production in California requires a high volume of water from germination to harvest, even with drip irrigation^{6,7}. Due to the economic aspect of drip irrigation, farmers consume more water in sprinkler or furrow irrigation while water resources in the aquifers continue to decline.

Water stress diminishes crop yield by shifting metabolism, leading to accelerated senescence at the whole plant. Previous studies elucidated that *L. serriola* was resistant to water stress at early (four weeks after germination) and mature stages^{8–10}. The *L. serriola* has unique morphological characteristics that maintain high water use efficiency under water deficit conditions. The lateral roots and taproot of *L. serriola* grow longer to access deep soil moisture, whereas cultivated lettuce has shallow lateral roots and taproots¹¹. The leaves of *L. serriola* have many trichomes¹². After bolting, the cauline leaves of *L. serriola* adjust to the vertical direction and reduce leaf temperature and transpiration^{13,14}.

Germination is a complex process requiring enough water in optimized environmental conditions. Seed germination is sensitive to limited water acquisition^{15,16}. A severe water shortage during germination considerably affects the quality of seedlings and yield¹⁷. The increment of osmotic stress reduces the final germination percentage and the water content in the seeds¹⁸. Water stress constrains water absorption by seeds and defers protein synthesis by affecting the transfer of nutrient reserves^{19,20}. The optimal temperature range for *L. serriola* germination is 15 °C to 25 °C, with a broader range of 10 °C to 35 °C²¹. The life cycle and wide temperature range of *L. serriola* indicate that it has little primary dormancy²².

The seeds reduce the water potential (Ψ), turgor pressure, and cell elongation under water stress. PEG has been widely employed to simulate water-deficit conditions. Its heavy molecular weight and non-toxic nature maintain a consistent osmotic potential^{23–25}. Previous PEG studies demonstrated that the germination percentage of *L.*

United States Department of Agriculture, Agricultural Research Service, Sam Farr United States Crop Improvement and Protection Research Center, Salinas, CA, USA. ✉email: sadal.hwang@usda.gov

serriola reduced severely as Ψ decreased up to -0.5 MPa²⁶. The cultivated lettuce germinated at Ψ as low as -0.41 MPa²⁷.

Many genetic mapping studies identified QTL for water stress-related traits in the cv. Salinas (*L. sativa*)²⁸ $\times$ US96UC23 (*L. serriola*) interspecific RIL population^{29–31}. Most QTL studies on lettuce seed germination have been related to thermo-dormancy at high temperature^{32,33}. The locus, *Htg6.1*, was mapped and associated with the biosynthesis of abscisic acid (ABA). However, it remains unknown how seed germination under water stress is regulated in lettuce. Hence, this study aimed to (1) evaluate the germination percentage in the cv. Salinas $\times$ US96UC23 RIL population and USDA germplasm collection under osmotic stress with 10% PEG, (2) identify QTL for germination percentage in the RIL population, and (3) investigate the transcript and gene network for seed germination in US96UC23 under control (dH₂O) and treatment (10% PEG) conditions as a case study.

Results

Trait evaluation

The germination percentages between two genotypes, cv. Salinas and US96UC23, were compared at nine PEG concentrations (Fig. 1a). Two-way analysis of variance (ANOVA) indicated that the effects of genotype and PEG were significant (P -values $\leq 1.69 \times 10^{-7}$). The t -tests showed significant differences between genotypes from 4 to 15% PEG. The mean differences between genotypes were 36.6–60% at 10–15% PEG, implying a high likelihood of identifying QTL in the cv. Salinas $\times$ US96UC23 RIL population. The germination percentage between genotypes was not different at 0% PEG (dH₂O). The cotyledons from the seeds of both cv. Salinas and US96UC23 emerged with extended hypocotyl and radical at dH₂O (Supplementary Fig. 1). US96UC23 was more sensitive to water stress than cv. Salinas. The germination percentage of US96UC23 was less than 50% at 10% PEG. No germination was found at 20% PEG. The Ψ ranged from 0 to -0.55 MPa at nine PEG concentrations (Fig. 1b). The Ψ and plant available water (PAW) were -0.16 MPa and 70% at 10% PEG, respectively. The PAW was about 50% at 12% PEG ($\Psi = -0.22$ MPa) and 20% at 16% PEG ($\Psi = -0.36$ MPa). As mild water stress treatment, 10% PEG was used for the subsequent germination experiments.

The germination percentage of the RIL population and USDA germplasm collection (Supplementary Data 1) was evaluated at 10% PEG. The frequency distributions of germination percentage were analyzed for both populations (Supplementary Fig. 2). The histograms of both populations had left-skewed distributions with heavy tails. The median germination percentages in the RIL population and USDA germplasm collection were 80% and 90%, respectively. The germination percentage was less than 90% in 66% of the RILs and 44% of the USDA accessions, indicating that seed germination in lettuce was sensitive to water stress.

The germination percentage of the RIL population and USDA germplasm collection did not follow normal distributions (P -values $< 2.2 \times 10^{-16}$). Therefore, the Kruskal-Wallis (H) test was conducted in both populations (Table 1). The genotypes (RIL and accession) had a highly significant effect in both populations (P -values $\leq 9.92 \times 10^{-13}$). The broad-sense heritabilities for germination percentage were 0.825 and 0.801 in the RIL

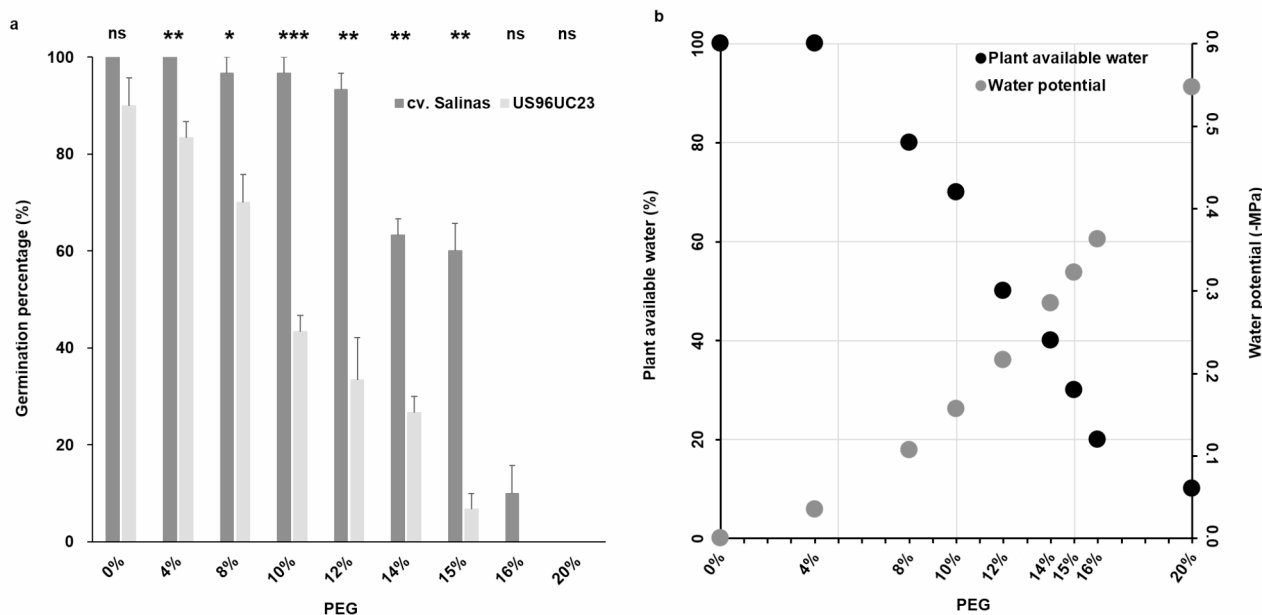


Fig. 1. (a) Comparison of germination percentages between cv. Salinas and US96UC23. The symbols on the top of the bar plot, ns, *, **, and ***, indicated $P > 0.05$ (not significant), $P < 0.05$, $P \leq 0.01$, and $P \leq 0.001$, respectively. (b) Relationship between PAW and Ψ at different PEG concentrations. The Ψ was determined using the equation²³. The PAW in sandy loam soil was estimated using the Ψ equation²³ and field capacity from previous hydrology studies^{55,56}.

Source	df ^a	Sum of squares	Mean squares	H test	P-value
cv. Salinas × US96UC23 RIL population					
RIL	212	2892687.25	13644.75	390.65	9.92e−13
Error	213	508,717	2388.34		
Total	425	3401404.25			
USDA germplasm collection					
Accession	523	40587941.88	77606.01	952.11	< 2.2e−16
Error	524	8090195.25	15439.30		
Total	1047	48678137.13			

Table 1. Kruskal-Wallis test of germination percentage at 10% PEG in the cv. Salinas × US96UC23 RIL population and USDA germplasm collection. ^aDegree of freedom.

Model ^a	QTL marker	Chromosome	Position ^b	LR ^c	LOD	R ²	Effect ^d	Favorable allele ^e	Criterion ^f
SMA	ANIM_0	4	197.6	16.22	3.53	8.6	7.14	cv. Salinas	1.03e−04
IM	ANIM_0	4	197.6	16.13	3.51	7.3	7.10	cv. Salinas	1.13e−04
CIM	ANIM_0	4	197.6	22.47	4.88	8.8	7.86	cv. Salinas	1.55e−04
	AXDH_0	8	129.4	21.30	4.63	8.3	7.66	US96UC23	1.55e−04
MIM	ANIM_0	4	197.6	16.11	3.50	7.3	7.12	cv. Salinas	BIC-M3
	ANIM_0	4	197.7	19.23	4.18	7.5	7.34	cv. Salinas	BIC-M2
	AXDH_0	8	129.5	13.53	2.94	5.7	6.10	US96UC23	BIC-M2

Table 2. QTL for germination percentage at 10% PEG in the cv. Salinas × US96UC23 RIL population. ^aSMA (single marker analysis), IM (interval mapping), and CIM (composite interval mapping) were used as single QTL models. The MIM (multiple interval mapping) was used as a multiple QTL model. ^bGenetic map position (cM) in the cv. Salinas × US96UC23 RIL population. ^cLikelihood ratio (LR) was equal to 4.6 x LOD. ^dQTL effect was defined as one-half the mean of US96UC23 alleles minus one-half the means of cv. Salinas alleles as an average effect of allele substitution or breeding value/2. ^eFavorable allele was defined as the allele giving a higher germination percentage. ^fFDR-adjusted P-values and BIC were used for single and multiple QTL models.

population and USDA germplasm collection, respectively. The narrow-sense heritability (h^2) for germination percentage was 0.63.

Genetic map analysis

The public genetic map of the cv. Salinas × US96UC23 RIL population was originally developed with 13,943 single-position polymorphism markers (SSPs) and reconstructed by discarding erroneous and redundant SSPs. The segregation ratio tests indicated that 553 SPPs on chromosomes 3 (533), 4 (1), 5 (3), 7 (3), and 8 (13) did not fit a 1:1 ratio of cv. Salinas: US96UC23 alleles (Supplementary Fig. 3). On chromosomes 3, 7, and 8, 549 SPPs showed a higher ratio of cv. Salinas allele. All segregation-distorted SPPs were discarded. Duplicate and monomorphic 8,608 SPPs were also removed. The relationship between recombination frequency (rf) and the logarithm of the odds (LOD) was evaluated in all SPP pairs (Supplementary Fig. 3). All SPP pairs were tested to find a SPP with switched alleles, and no such SPP was detected.

The final genetic map was reconstructed with 5,322 SPPs (Supplementary Fig. 4). The total length of the map was 1,969.8 cM and 1.2 times longer than the previous public one. The average distance between adjacent markers was 0.4 cM. The maximum distance between two SPPs, BULA_0 and AYXI_0, was 53.7 cM due to the deletion of segregation-distorted markers at the long arm of chromosome 3.

QTL analysis

QTL for germination percentage at 10% PEG were identified in the cv. Salinas × US96UC23 RIL population. Two QTL, ANIM_0 and AXDH_0, were identified on chromosomes 4 and 8 using different QTL models, respectively (Supplementary Fig. 4 and Table 2). ANIM_0 explained 7.3% to 8.8% of the total variance of the germination percentage. The R² values of AXDH_0 were 8.3% and 5.7% in composite interval mapping (CIM) and multiple interval mapping (MIM), respectively. The QTL effects of both QTL ranged from 6.1% to 7.86%. Cv. Salinas and US96UC23 provided favorable alleles for a higher germination percentage at ANIM_0 and AXDH_0, respectively. The significant interaction between both QTL was detected in the MIM model (LOD = 2.91) with a less stringent Bayesian information criterion (BIC). The interaction effect was 6.08%, and the R² value of the interaction was 5.3%. Considering A and B alleles originated from cv. Salinas and US96UC23, respectively, the germination percentages of four genotype groups, AA/AA, AA/BB, BB/AA, and BB/BB, were compared. (Supplementary Fig. 5). Each genotype group consisted of the first two alleles from ANIM_0 and the last two from AXDH_0. The H test showed that the germination percentages of the four groups were significantly

different ($P=2.29\text{e}-06$). The Dunn tests indicated that the germination percentage of BB/AA was about 1.5 times lower than the other three genotype groups.

Candidate genes were screened within the 95% confidence intervals (CIs) of both QTL (Supplementary Table 1). ANIM_0 was physically located between 234,370,421 bp and 234,371,251 bp on chromosome 4. No annotated lettuce gene was linked to ANIM_0. However, the homologous *Arabidopsis thaliana* (*A. thaliana*) gene, AT5G21280.1, was found in the region where ANIM_0 was positioned. AXDH_0 was located between 124,467,301 bp and 124,469,258 bp on chromosome 8. The lettuce gene, Lsat_1_v5_gn_8_85600.1 (LS21897), was linked to AXDH_0 and homologous to AT3G59530.3.

Trait evaluation of US96UC23

For RNA-Seq and network analyses, the germination percentage of US96UC23 was assessed under control (dH_2O) and treatment (10% PEG) conditions. The *t*-test indicated that the mean difference (35.3%) in germination percentage between the control and treatment groups was highly significant ($P=1.85\text{e}-06$). The average germination percentages of US96UC23 were 88% in the control group and 52.7% in the treatment group.

DEG and enrichment analyses

After filtering available fragments per kilobase of transcript per million mapped reads (FPKM), GO, and KEGG data in US96UC23, 27,245 genes were used for DEG and enrichment analyses (Supplementary Data 2). The gene IDs for 27,245 genes were created by combining the letters “LS” and a 5-digit number. These IDs were used to depict lettuce genes instead of long gene model names in the lettuce reference genome.

DEGs were identified to ascertain significantly upregulated and downregulated genes using the FPKM data from the control and treatment datasets. The FPKM data was evaluated as the normalized transcript levels (Supplementary Data 2 and Supplementary Fig. 6). Given all the replicates of US96UC23 across both datasets, 23,742 FPKM values were averagely greater than 0. The overall mean of FPKM values was 32.68, and FPKM values ranged from 0 to 30,747. The standard deviation of FPKM values ranged from 150.51 to 434.03. The *t*-test showed a significant FPKM mean difference (3.71) between the control and treatment datasets ($P=5\text{e}-04$). The average FPKM values for the control and treatment datasets were 34.59 and 30.9, respectively. The total number of DEGs was 4,095, of which 2,750 were upregulated and 1,345 were downregulated (Supplementary Data 2 and Supplementary Fig. 7).

The GO term and KEGG pathway, which were annotated in the highest number of DEGs, were identified. A total of 2,839 GO terms and 126 KEGG pathways were annotated in 27,245 genes. Three GO terms, biological process (GO:0008150), nucleus (GO:0005634), and molecular function (GO:0003674), were annotated in the largest number of DEGs in each GO class, biological process, cellular component, and molecular function, respectively (Supplementary Fig. 8). The KEGG pathway, plant hormone signal transduction (ko04075), was annotated in the largest number of DEGs (Supplementary Fig. 9).

Enrichment analysis was performed to investigate the top 20 GO terms and KEGG pathways and collect biological data, such as GO IDs and reference pathways. The top 20 GO terms were identified by GO enrichment analysis (Fig. 2a). The *P*-values of the top 20 GO terms were less than $4.79\text{e}-06$. Microtubule-based movement (GO:0007018) was the most significant ($P=3.19\text{e}-19$). Nucleus (GO:0005634) had the lowest rich factor (0.17), whereas DNA replication initiation (GO:0006270) indicated the highest rich factor (0.78). The top 20 KEGG pathways were identified by KEGG enrichment analysis (Fig. 2b). However, the *P*-values of four KEGG pathways, alpha-linolenic acid metabolism (ko00592), plant hormone signal transduction (ko04075), pyrimidine metabolism (ko00240), and selenocompound metabolism (ko00450), were slightly greater than 0.05. DNA replication (ko03030) was the most significant ($P=2.08\text{e}-08$). Plant hormone signal transduction (ko04075) had the lowest rich factor (0.17), and non-homologous end-joining (ko03450) showed the highest rich factor (0.63). A total of 1,804 DEGs were associated with the top 20 GO terms and KEGG pathways. Additionally, 228 DEGs overlapped between the top 20 GO terms and KEGG pathways (Supplementary Data 3).

Network construction

The network was constructed using the FPKM data from the control and treatment datasets in US96UC23 (Supplementary Data 4). A total of 26,071 genes, including 99.4% of DEGs, were used for network construction, excluding genes with either 0 or many missing FPKM values. The selected soft threshold power (β) was 6, considering the scale-free topology model and connectivity range (Supplementary Fig. 10). The R^2 values of topology models were 0.8 and 0.9 in the control and treatment datasets at $\beta=6$, respectively. The consensus network analysis identified 44 modules across both datasets, implying that 44 module eigengenes (MEs) were obtained from each dataset (Supplementary Data 4 and Fig. 3). The number of genes in each module ranged from 6,651 (module 1) to 47 (module 44). The 48 genes in the grey-colored module were not assigned to any of the 44 identified modules in the consensus network analysis. The module in grey was nominally designated as module 0, and the ME of module 0 was designated as ME0.

Eigengene significance, module significance, and module membership

Eigengene significance was evaluated to detect significant modules by testing the correlation coefficient between the germination percentage and ME in a module (Supplementary Table 2). Only ME0 was significantly correlated with the germination percentage in the control dataset ($P=1.4\text{e}-2$). The eigengene significance between ME0 and the germination percentage was -0.63 . The MEs of ten modules 1, 4, 6, 9, 17, 19, 24, 36, 37, and 38 were significantly correlated with the germination percentage in the treatment dataset ($P\text{-values}<0.05$), indicating that the ten modules were significant for the network study. The absolute eigengene significance in 10 MEs was greater than 0.53. Three MEs, ME9, ME36, and ME38, indicated positive eigengene significance. ME38 showed the highest correlation with the germination percentage ($P=2\text{e}-4$).

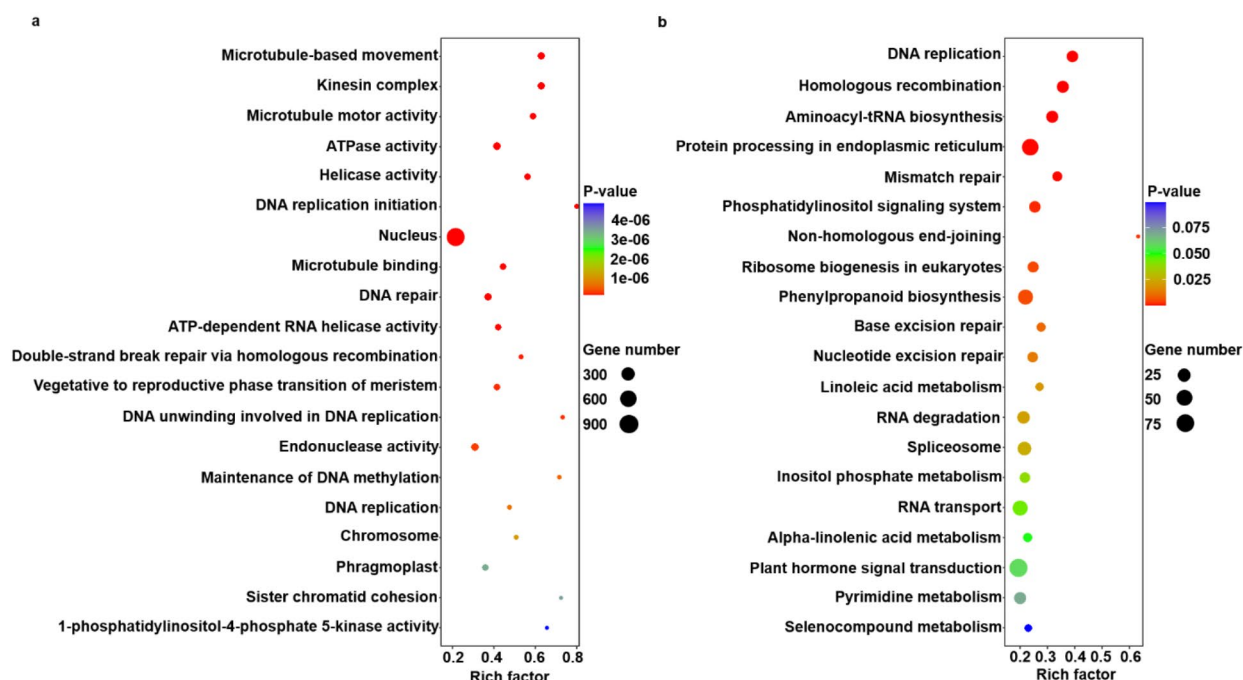


Fig. 2. (a) Top 20 GO terms in GO enrichment analysis. (b) Top 20 KEGG pathways in KEGG enrichment analysis.

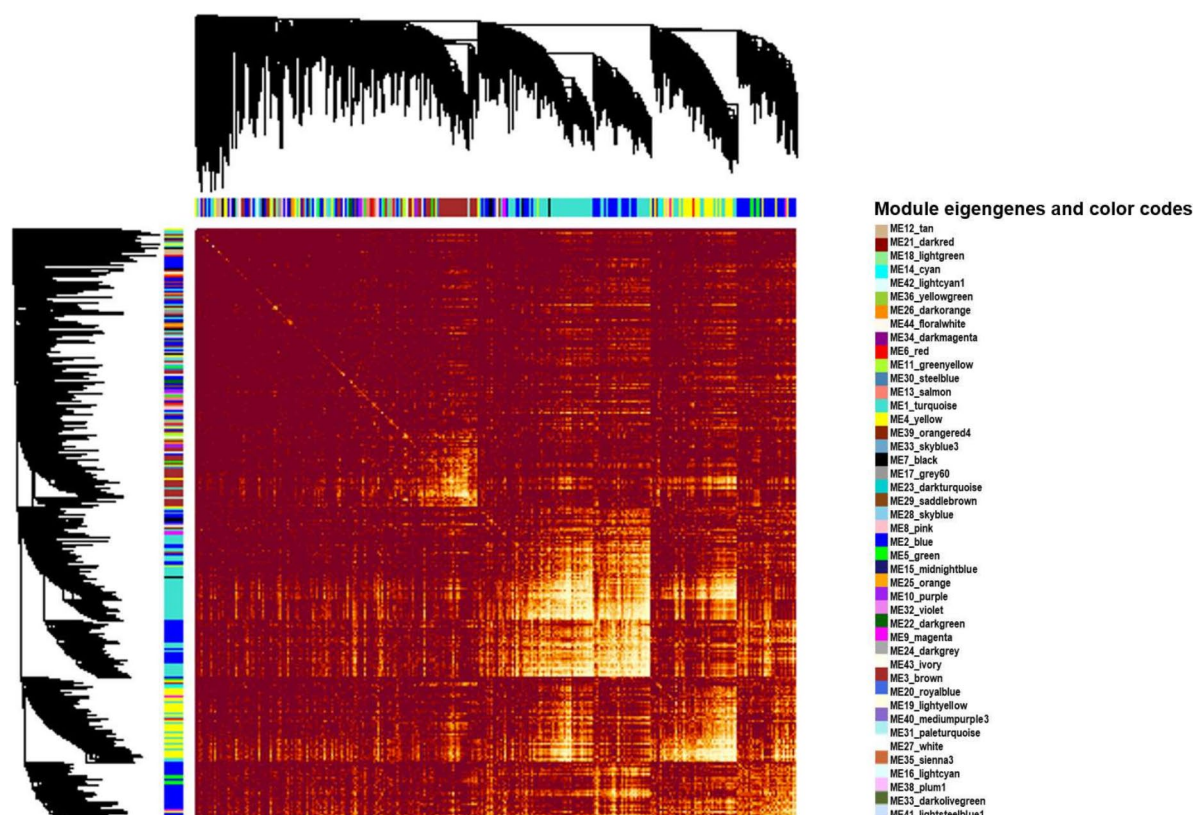


Fig. 3. Topological overlap matrix among all genes in the gene network. Darker dark red represented low overlap, while light yellow represented high overlap. The assigned modules with colors and dendrograms are shown along the left side and top. In the clustered dendrogram (tree), branch groups were densely interconnected with highly co-expressed genes. The vertical line of each branch corresponded to a gene.

Module significance was estimated as the average absolute gene significance (GS) for all genes in a module to support the relationship between eigengene significance and module significance (Supplementary Table 2). Module 0 in the control dataset showed the highest module significance (0.49). Ten modules with significant MEs exhibited higher module significance than others in the treatment dataset. The module significance of the ten modules ranged from 0.33 to 0.53. The highest module significance was 0.53 in module 38.

Module membership (MM) was evaluated to determine gene membership by testing the correlation coefficient between a gene and ME in a module (Supplementary Table 2). The MM of each gene in a module, except for module 0, was significant (P -value < 0.05). The average absolute MM across all modules was 0.72 and 0.67 in the control and treatment datasets, alluding that the genes in a module had significant MM and high connectivity.

Gene significance

The GS of a gene was evaluated to determine if the gene was significantly correlated with the germination percentage in the control and treatment datasets (Supplementary Data 4 and Supplementary Fig. 11). The total number of significant genes was 6,378, of which 1,748 DEGs (27.4%) were included. A total of 319 and 5,451 genes were significant in the control and treatment datasets, respectively. The GS of LS00896 in module 2 was -0.77 , indicating the most negative relationship ($P = 1.36e-3$) with the germination percentage in the control dataset. The GS of LS09012 in module 1 was -0.914 , with the highest correlation ($P = 1.93e-6$) in the treatment dataset. As a DEG, LS20107 in module 4 had the lowest meta P -value ($6.76e-6$) across both datasets. The Venn diagram showed that 30 significant genes commonly overlapped in each dataset and across both datasets (Supplementary Data 5 and Supplementary Fig. 12). Eight DEGs, LS04614, LS06392, LS11691, LS23172, LS01020, LS20107, LS20239, and LS21606, were observed from 30 significant genes in modules 1 (3), 2 (1), 3 (1), 4 (1), 17 (1), and 24 (1), respectively.

Gene network

Eigengene significance tests found ten significant modules in the treatment dataset (Supplementary Table 2). Although no significant module was in the control dataset, the gene networks in the ten modules of the control dataset were also considered. Therefore, gene networks were constructed for the top 20 hub genes in each of the ten modules to compare the gene networks of both the control and treatment datasets (Fig. 4 and Supplementary Data 6 and 7).

Twenty-two genes, LS03987, LS15145, LS17655, LS14143, LS24323, LS11800, LS21024, LS00870, LS01781, LS23012, LS00655, LS04439, LS18229, LS02836, LS03284, LS03518, LS05435, LS04162, LS03064, LS11217, LS07863, and LS09357, were commonly connected between the gene networks of both datasets in modules 6 (1), 17 (2), 19 (2), 24 (2), 36 (6), 37 (5), and 38 (4), respectively. Four genes, LS15145, LS21024, LS23012, and LS02836, were found to have the highest node degree and betweenness centrality in modules 17, 24, 36, and 37, respectively. These four genes had the most connections with other genes and played a critical role in exerting the interaction of other genes in the gene network.

A small number of DEGs were found to be involved in all gene networks. One DEG, LS01781, was connected between the gene networks of both datasets in module 36. In the control dataset, ten DEGs, LS21911, LS19479, LS20016, LS20239, LS07353, LS06490, LS19395, LS01781, LS16673, and LS18882, indicated the highest node degree or betweenness centrality in modules 1 (1), 6 (1), 9 (1), 17 (1), 19 (1), 24 (2), 36 (1), and 38 (2), respectively. Four DEGs, LS17190, LS07121, LS03939, and LS05949, showed the highest neighborhood connectivity in modules 4, 19, 24, and 37, respectively. These four DEGs represented the top average connectivity of all neighbors as a local centrality measure in the gene network. In the treatment dataset, 9 DEGs, LS03686, LS18593, LS03546, LS13527, LS20277, LS24081, LS05267, LS21760, and LS07225, displayed the highest node degree and betweenness centrality in modules 1 (2), 9 (2), 17 (1), 19 (1), 24 (2), and 37 (1), respectively. Thirteen DEGs, LS11197, LS25615, LS02580, LS03418, LS04782, LS07411, LS20197, LS03546, LS13527, LS12075, LS01781, LS03747, and LS07225, had the highest neighborhood connectivity in modules 1 (2), 4 (4), 6 (1), 9 (2), 17 (1), 36 (2) and 37 (1), respectively.

Based on the enrichment analysis (Fig. 2), some genes in the gene networks were annotated with the top 20 GO terms and KEGG pathways. In the control dataset, four genes, LS07173, LS05685, LS05949, and LS11854, were in modules 1, 4, 37, and 38, respectively. In the treatment dataset, seven genes, LS21135, LS13281, LS10884, LS24318, LS20277, LS07421, and LS00803, were in modules 4 (1), 6 (1), 9 (2), 17 (1), 24 (1), and 36 (1), respectively. These genes exhibited the highest node degree, betweenness centrality, or neighborhood connectivity in the gene network.

Discussion

The segregation-distorted 553 SPPs were detected by the stringent criterion in the cv. Salinas $\times$ US96UC23 RIL population (Supplementary Fig. 3). The segregation distortion might result from preferential gametic transmission in interspecific populations^{34,35}. The integrated map study demonstrated that markers in similar chromosome regions indicated segregation distortion in different F_2 interspecific populations rather than intraspecific ones³⁶. However, the direction of marker distortion in these different F_2 interspecific populations did not align with the cv. Salinas $\times$ US96UC23 RIL population. Our results have been supported by another previous study conducted on the same RIL population³⁷. Interestingly, on chromosome 3, 533 segregation-distorted SPPs were found between two flanking markers in the cv. Salinas $\times$ US96UC23 RIL population, showing a higher ratio of cv. Salinas alleles in a 122 Mb physical map interval (Supplementary Fig. 4). The segregation-distorted SPPs could affect the identification of covariates in the CIM and MIM models for QTL analysis. Therefore, all distorted markers were discarded for QTL analysis in this study.

Candidate genes in lettuce and their homologous *A. thaliana* genes were identified by narrowing down from three data sources: (1) the CIs of QTL (Supplementary Table 1), (2) duplicate DEGs between the top 20 GO terms

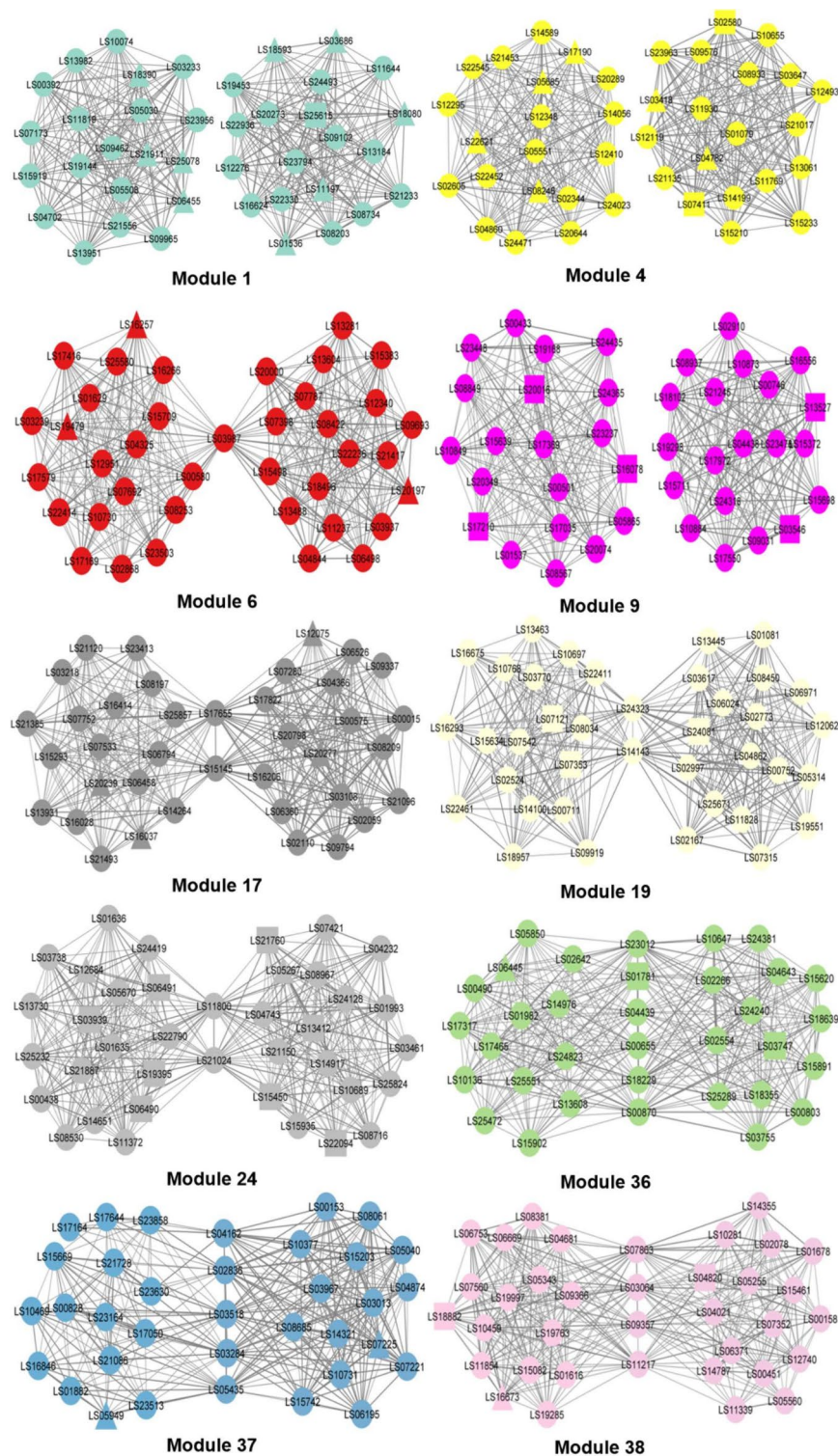


Fig. 4. Gene networks for the top 20 hub genes in 10 modules from the control and treatment datasets. The organic layout style was applied for the construction of gene networks. The gene networks of control and treatment datasets were on the left and right in each module. The node colors of gene networks corresponded to the assigned module colors. The round-shaped nodes were not DEGs. The triangle- and rectangle-shaped nodes indicated upregulated and downregulated DEGs in each gene network, respectively. The node size by node degree was not considered, highlighting node color and gene ID in each gene network. However, a degree of edge was given in each gene network. The nodes in the middle of two gene networks in each module were commonly linked between the gene networks of both datasets.

and KEGG pathways by enrichment analysis (Supplementary Data 3), and (3) network analysis (Supplementary Data 5, 6, and 7). Based on previous studies, we highlighted some notable *A. thaliana* genes associated with seed germination in each data source mentioned above.

In the CIs of QTL (Supplementary Table 1), three *A. thaliana* genes, AT3G20290.3, AT5G43020.1, and AT5G56550.1, were homologous to Lsat_1_v5_gn_4_127040.1, LS10622, and LS21902, respectively. These *A. thaliana* genes were related to stress or hormone. Endocytosis and exocytosis were complementary membrane trafficking processes to maintain plasma membrane integrity and stress tolerance³⁸. The immunolocalization analyses with pectin and actin revealed that AT3G20290.3 was endocytosed in the cell walls of the seed embryo³⁹ and root hairs⁴⁰ once the seed imbibition started. The T-DNA insertion lines revealed that AT5G43020.1 was involved in root development defects and signal transduction pathways related to hormone and abiotic stress⁴¹. AT5G56550.1 suppressed ABA signaling by regulating histone deposition at the *ABI4* promoter under environmental stress⁴². As a transcription factor (TF), *ABI4* played a pivotal role in the ABA pathway by repressing cytokinin-inducible genes and promoting gibberellin degradation at the post-germination stage⁴³.

In the list of duplicate DEGs between the top 20 GO terms and KEGG pathways by enrichment analysis (Supplementary Data 3), five *A. thaliana* genes, AT5G03730.1, AT5G58760.1, AT1G48130.1, AT2G26040.1, and AT2G38310.1, were homologous to LS01272, LS14778, LS16627, LS12480, and LS20279, respectively. These *A. thaliana* genes were related to abiotic stress or cross-talk in hormones. The C₂H₄-susceptible mutant of AT5G03730.1 reduced sensitivity to ABA, which was promoted by cytokinin in the signaling of seed germination⁴⁴. AT5G58760.1 encoded the damaged DNA binding protein 1, a component of the DETIOLATED 1 (DET1) complex. The *det1* mutant resisted osmotic and salt stress in germination⁴⁵. AT1G48130.1 played essential roles in the biosynthesis of lignin, flavonoid, and coumarin in the phenylpropanoid pathway. The mutant of AT1G48130.1 suppressed seed dormancy by reducing ABA and increasing gibberellic acid (GA) during germination⁴⁶. Both AT2G26040.1 and AT2G38310.1 transcribed the ABA-binding receptors. The overexpression of ABA receptor kinases inhibited seed germination and enhanced drought resistance⁴⁷.

Five *A. thaliana* genes, AT5G48670.1, AT2G14210.2, AT5G11460.1, AT5G66880.1, and AT1G12920.1, were related to TF or hormone signal transduction. AT5G48670.1 was homologous to LS06419 in the GS of network analysis (Supplementary Data 5). In the gene networks for the top 20 hub genes (Supplementary Data 6 and 7), four other *A. thaliana* genes were homologous to LS19285, LS06195, LS07421, and LS02836, respectively. AT5G48670.1 was a MADS-box TF in the central gene regulatory network that controlled the expression of downstream genes in cell development and function⁴⁸. AT2G14210.2 was a member of MADS-box TFs. The overexpressing lines were hypersensitive to ABA, salt, and osmotic stress in the seed germination⁴⁹. AT5G11460.1 interacted with sucrose nonfermenting-1-related protein kinase 1 (SnRK1), a central plant growth regulator. During energy starvation, the gene was induced and repressed SnRK1 by increasing the level of SnRK1 α 1 as the major subunit of SnRK1. The mutant of AT5G11460.1 accumulated more SnRK1 α 1 and inhibited germination under favorable growth conditions⁵⁰. AT5G66880.1 encoded SnRK2.3. The gene was induced by salt or osmotic stress and was involved in ABA signaling⁵¹. The mutation of substrates, which interacted with SnRK2.3, increased ABA signaling. AT1G12920.1 was plainly induced by glucose and hypersensitive to glucose in germination. The overexpressing lines enhanced sensitivity to the inhibitor of GA signal pathways⁵².

It is important to carefully select the stage and target tissue when collecting samples for RNA-Seq analysis. Like the seeds in Supplementary Fig. 1, we observed that most seeds from the RIL population and USDA germplasm collection germinated well after 4 days under control conditions. Hence, we decided to take samples 4 days after seed imbibition. Even mild water stress at 10% PEG resulted in wide variation in the germination of US96UC23 seeds, including well-germinated seeds, as shown in Supplementary Fig. 1. The relationships among MEs exhibited that the treatment dataset showed a wider range of correlation coefficients than the control dataset (Supplementary Fig. 13). Collecting all seed tissues could be a disadvantage for precise RNA-Seq analysis rather than analyzing specific tissues (e.g., root, cotyledon). This study focused on a general transcriptome analysis to understand how osmotic stress affected seed germination.

Since both QTL were identified in the controlled environmental conditions, it would be essential to conduct the field test if both QTL were stable on a multi-year or -location basis. It will be critical to confirm both QTL and identify new QTL in diverse mapping populations to apply marker-assisted or genomic selection and expand gene network information. The transcriptome analysis of US96UC23 facilitated the discovery of significant GO terms, KEGG pathways, and gene networks associated with seed germination under osmotic stress in lettuce. In future studies, identifying expression QTL (eQTL) and gene networks at the population level will help build an explicit regulatory network of seed germination in lettuce.

Methods

Germination experiment

The PEG 8000 (Sigma-Aldrich, cat. P5413) was used to make a PEG solution using autoclaved and filtered dH₂O. The filter paper with 85 mm diameter (Sigma-Aldrich, cat. WHA1001085) was drenched in 6 ml dH₂O or PEG solution. Seeds were placed on a single layer of saturated filter paper in sterilized and transparent polystyrene Petri dishes with 100 mm $\times$ 20 mm size (Sigma-Aldrich, cat. P5606). The Petri dish lid had three ventilation ribs on the underside, allowing a continuous oxygen supply. The seeds were imbibed for 4 days in the GR-41L seed germination chamber (Geneva Scientific LLC, WI, USA). The white fluorescent tube lights were turned on for 9 h during the daytime, and the irradiance intensity was medium (60 micromoles/m²/sec)⁵³. The optimal temperatures were set at 18 °C and 15 °C for the day and night²¹. Radicle emergence (RE)⁵⁴ was measured after 4 days. The seeds with RE greater than 5 mm were counted as germinated ones. The counted seed numbers were converted to the germination percentage (%).

Estimation of water potential and plant available water

To discern the degree of water stress by different PEG concentrations, Ψ and PAW were calculated. The Ψ was determined using the equation: $\Psi = 1.29C^2T - 140C^2 - 4C$ (C : PEG concentration (g/ml) and T : temperature ($^{\circ}\text{C}$))²³. The day temperature, 18 $^{\circ}\text{C}$, was applied for the variable T in the equation. The PAW in sandy loam soil was estimated using the Ψ equation²³ and field capacity from previous hydrology studies^{55,56}.

Experimental design and statistical analysis

Three germination experiments were performed with a completely randomized design (CRD) from 2022 to 2023. First, cv. Salinas and US96UC23 were compared at 0% to 20% PEG in 3 replicates. Second, the $F_{7,8}$ cv. Salinas $\times$ US96UC23 RIL population and USDA germplasm collection, consisting of 213 RILs and 524 USDA accessions (Supplementary Data 1), were evaluated at 10% PEG in 2 replicates. As parents of the RIL population, cv. Salinas and US96UC23 were also included among 524 USDA accessions. Third, for RNA-Seq and network analyses, US96UC23 was assessed in the control (dH_2O) and treatment (10% PEG) groups, with 15 replicates in each group. Thirty seeds were used per replicate in all germination experiments. The seeds of the RILs, their parents, and the USDA accessions were harvested in 2021.

All statistical analyses were performed with R (V. 4.2.1). The residuals of the CRD linear model in each germination experiment were tested for normal distribution using the `shapiro.test()`. If the estimated parameter λ in the Box-Cox procedure⁵⁷ did not result in a satisfactory transformation, the H test⁵⁸ was conducted instead of general ANOVA. The germination percentage was converted to ranked data as a categorical variable in the H test.

The `rank()` was used to rank a value in the germination percentage data. The `Kruskal.test()` was used for the H test. The Dunn test⁵⁹ was used as a *post hoc* test if the null hypothesis of the H test was rejected at $\alpha = 0.05$. The `dunnTest()` with Bonferroni-adjusted criterion was used for the multiple comparison test. The `aov()` was used for general ANOVA, and the `t.test()` was used for two sample or paired t -test. The `as.factor()` was used to convert an independent variable into a factor. The `cor.test()` was used to estimate the Pearson correlation coefficient (r) between two variables.

Heritability

The H^2 for germination percentage at 10% PEG was estimated using the mean squares of H test results in the RIL population and USDA germplasm collection. The H^2 was estimated as follows: $H^2 = \delta_g^2 / (\delta_g^2 + (\delta_e^2/r))$ (δ_g^2 : genetic variance, δ_e^2 : error variance, and r : number of replicates). Based on the Breeder's equation⁶⁰, the h^2 for germination percentage at 10% PEG was estimated using the RIL population and USDA germplasm collection. The USDA germplasm collection was considered the base population. When cv. Salinas and US96UC23 were selected from the base population, and their RIL population was generated, h^2 was estimated as follows: $h^2 = R$ (response to selection)/ S (selection differential) = $(\mu_{\text{sxu}} - \mu_g) / (\mu_a - \mu_g)$ (μ_{sxu} : the average germination percentage of 213 RILs in the cv. Salinas $\times$ US96UC23 population, μ_g : the average germination percentage of 524 USDA accessions in the USDA germplasm collection, and μ_a : the average germination percentage of cv. Salinas and US96UC23).

Genetic map construction

The public genetic map was developed in the 213 F_7 cv. Salinas $\times$ US96UC23 RIL population using 13,943 UniGene-based SPPs³⁷. The alleles of each SPP were designated with A or B in the Affymetrix array. The A and B alleles originated from cv. Salinas and US96UC23, respectively. The R/qtl package in R⁶¹ was used to identify erroneous and redundant SPPs and reconstruct the final genetic map. The `geno.table()` was used to test the segregation ratio of each SPP using the Chi-squared test. The segregation-distorted markers were removed based on a Bonferroni-adjusted criterion. The redundant markers were deleted using the `findDupMarkers()` and `drop.markers()`. The `checkAlleles()` was used to identify the marker loci with erroneously switched alleles with LOD 3.0. The map distance was tested by the maximum likelihood with the Expectation-Maximization algorithm⁶². The `est.map()` calculated the map distance between markers based on the estimated rf . The Kosambi function⁶³ was used for all linkage analyses. The `jittermap()` was used to slightly separate coincident marker positions.

QTL analysis

WinQTLCartographer (V. 2.5.011)⁶⁴ was used to identify QTL for germination percentage in 213 $F_{7,8}$ cv. Salinas $\times$ US96UC23 RIL population. The average germination percentage of each RIL was estimated using the least squares (LS) means in the CRD linear model. The LS means were used for the germination percentage data in the QTL models. As a single-QTL model, single marker analysis (SMA), interval mapping (IM), and CIM were conducted. Model 6 was chosen for CIM⁶⁵, and the number of control background markers was 5. The Forward Regression method ($\alpha = 0.05$) was applied to mitigate errors from a stepwise approach and protect against model overfitting. The single-QTL models used the false discovery rate (FDR)-adjusted P -value (Q -value)⁶⁶ to test the null hypothesis (H_0 : there is no QTL). MIM⁶⁷ was performed as a multiple-QTL model. The MIM Forward Search method identified the initial QTL. The Refine Model performed the statistical tests of identified main QTLs and their interactions. The final model was decided based on the two BIC criteria, BIC-M2 ($2\ln(n)$) and BIC-M3 ($3\ln(n)$) (n : population size). The Window size and Walk speed were 1 cM for all QTL models.

RNA-Seq analysis

All germinated seeds ($\text{RE} > 5$ mm) of US96UC23 were collected from each replicate in the control and treatment groups after 4 days as the post-germination stage. The total RNA (5 μg) of each replicate was extracted from the bulk seed sample, including all seed parts, using Trizol (Thermo Fisher, cat. 15596018). High-quality RNA samples with an RNA integrity number (> 7.0) by 2100 Bioanalyzer (Agilent, cat. G2939A) and RNA 6000 Nano

Kit (Agilent, cat. 5067-1511) were used to construct the sequencing library. One replicate was excluded from the control group due to quality control failure. The mRNA of each sample was purified from total RNA using Dynabeads™ Oligo(dT) (Thermo Fisher, cat. 61021). The purified mRNA was reduced into short fragments by the Magnesium RNA Fragmentation Module (NEB, cat. E6150). The cleaved RNA fragments were reverse-transcribed to create cDNA by SuperScript II Reverse Transcriptase (Invitrogen, cat. 1896649), which was used to synthesize U-labeled second-stranded DNAs with *E. coli* DNA polymerase I (NEB, cat. M0209), RNase H (NEB, cat. M0297), and dUTP solution (Thermo Fisher, cat. R0133). An A-base was added to the 3' ends of each strand for ligating to the indexed adapters. Each adapter contained a T-base overhang for ligating to the A-tailed fragmented DNA. Dual-index adapters were ligated to the fragments, and size selection was performed with the AMPure XP beads. After the heat-labile UDG enzyme (NEB, cat. M0280) treatment of the U-labeled second-stranded DNAs, the ligated products were amplified with PCR. The average insert size for the final cDNA library was 300 ± 50 bp. The cDNA library was sequenced with the Illumina NovaSeq™ 6000 platform, generating a total of million 2×150 bp paired-end (PE150) reads. Cutadapt (V. 1.9)⁶⁸ removed reads with adapters, poly(A) and poly(G), unknown nucleotides greater than 5%, and low-quality bases greater than 20%. The sequence quality was verified with FastQC (V. 0.11.9)⁶⁹, and clean reads were generated as FASTQ files.

The reads of all samples were mapped to the lettuce reference genome (V. 8.0) using HISAT2 (V. 2.0.4)⁷⁰. The mapped reads of each sample were assembled using StringTie (V. 1.3.4d)⁷¹. Transcripts from all samples were merged to reconstruct a comprehensive transcriptome using GffCompare (V. 0.9.8)⁷². After the final transcriptome was generated, StringTie and Ballgown⁷³ were used to estimate the expression levels of all transcripts and perform expression abundance for mRNAs by calculating FPKM. The DEG analysis between the control and treatment datasets was performed using DESeq2 with the FPKM data⁷⁴. The genes satisfying two criteria, $Q < 0.05$ ($H_0: \log_2(\text{fold change}) = 0$) and absolute $\log_2(\text{fold change}) \geq 1$, were considered DEGs.

Enrichment analysis

The genes with GO terms and KEGG pathways were annotated in the GO (<http://www.geneontology.org/>)⁷⁵ and KEGG (<https://www.kegg.jp/kegg/>)⁷⁶ database. The KEGG identifiers combined with ko and 5-digit numbers were collected for the KEGG pathway. The enrichment analysis was conducted to identify the top 20 GO terms and KEGG pathways by the hypergeometric test. The *P*-value formula of the hypergeometric test was as follows:

$$P\text{-value} = 1 - \sum_{i=0}^{m-1} \binom{m}{i} \binom{N-M}{n-i} / \binom{M}{i} \quad (N: \text{total background gene numbers annotated with all GO}$$

terms or KEGG pathways, *M*: background gene numbers annotated with a GO term or KEGG pathway, *n*: total significant DEG numbers, and *m*: significant DEG numbers in a GO term or KEGG pathway). The rich factor was calculated by dividing *m* by *M* from the *P*-value formula.

Network analysis

The R source codes and terminologies of the weighted gene co-expression network analysis (WGCNA)⁷⁷ were applied to elucidate the gene networks of seed germination in US96UC23. The FPKM data was transformed by $\log_2(\text{FPKM} + 1)$. Genes with either 0 or excessive missing FPKM values were discarded based on the zero-variance criterion. The consensus network analysis was performed using the control (dH₂O) and treatment (10% PEG) datasets. The weighted network was constructed as an unsigned network by raising the absolute correlation coefficient between a pair of genes to β . The modules were identified by hierarchical clustering, using the topological overlap matrix (TOM)-based dissimilarity and Dynamic Tree Cut. Each ME was extracted as the first principal component in a module. In the WGCNA⁷⁷, eigengene significance, module significance, MM, and GS were defined as follows: (1) eigengene significance was the correlation coefficient between the germination percentage and ME in a module, (2) module significance was the average absolute GS for all genes in a module, (3) MM was the correlation coefficient between a gene and ME in a module, and (4) GS was the correlation coefficient between germination percentage and a gene. The liberal criterion ($P < 0.05$) was applied to test eigengene significance, MM, and GS. Eigengene significance and MM were tested in each dataset. GS was tested by the *P*-value in each dataset and the meta *P*-value across both datasets. Eigengene significance was evaluated to identify significant modules for constructing gene networks. Cytoscape⁷⁸ was used to visualize the edge (interaction between a pair of genes) and node (gene) list files from WGCNA, to sort the top 20 hub genes by node degree in a module, and to estimate gene network parameters, betweenness centrality and neighborhood connectivity.

Candidate genes

The coding sequences of lettuce genes were available in the Lettuce Genome Resource (V. 8.0) (<https://lgr.genomecenter.ucdavis.edu/Home.php>). The candidate genes for QTL were screened in the 95% CIs (± 1 LOD interval) of QTL. TAIR BLASTN (V. 2.9.0+) (<https://www.arabidopsis.org/Blast/>) was used to identify *A. thaliana* genes homologous to candidate genes in lettuce. The Bit score (> 50) and E-value (< 0.05) were used as criteria to find out the best identities.

Data availability

This published article and supplementary information files include all data generated or analyzed during this study. The raw RNA sequence data (FASTQ files) for transcriptome analysis is available in the European Nucleotide Archive (ENA) with accession number PRJEB71667. The corresponding author can provide additional data upon reasonable request.

Received: 26 April 2024; Accepted: 28 October 2024

Published online: 07 November 2024

References

- Kesseli, R., Ochoa, O. & Michelmore, R. Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce (*L. sativa*). *Genome* **34**, 430–436 (1991).
- Weaver, S. E. & Downs, M. P. The biology of Canadian weeds. 122. *Lactuca serriola* L. *Can. J. Plant Sci.* **83**, 619–628 (2003).
- FAO. *FAOSTAT Statistical Database*. <http://www.fao.org/faostat/en/#home> (Food and Agriculture Organization of the United Nations, 2020).
- Geisseler, D. & Horwath, W. R. *Lettuce Production in California*. http://apps.cdca.ca.gov/frep/docs/Lettuce_Production_CA.pdf (Fertilizer Research and Education Program, 2014).
- Svoboda, M. et al. The drought monitor. *Bull. Am. Meteorol. Soc.* **83**, 1181–1190 (2002).
- Turini, T. et al. *Iceberg Lettuce Production in CALIFORNIA*. University of California, Division of Agriculture and Natural Resources, Publication No. 7215 (2011).
- Smith, R. et al. *Leaf Lettuce Production in California*. University of California, Division of Agriculture and Natural Resources, Publication No. 7216 (2011).
- Eriksen, R. L., Knepper, C., Cahn, M. D. & Mou, B. Screening of lettuce germplasm for agronomic traits under low water conditions. *HortScience* **51**, 669–679 (2016).
- Eriksen, R. L., Adhikari, N. D. & Mou, B. Comparative photosynthesis physiology of cultivated and wild lettuce under control and low-water stress. *Crop Sci.* **60**, 2511–2526 (2020).
- Knepper, C. & Mou, B. Semi-high throughput screening for potential drought-tolerance in lettuce (*Lactuca sativa*) germplasm collections. *J. Vis. Exp.* **98**, e52492 (2015).
- Gallardo, M., Jackson, L. E. & Thompson, R. B. Shoot and root physiological responses to localized zones of soil moisture in cultivated and wild lettuce (*Lactuca* spp.). *Plant Cell Environ.* **19**, 1169–1178 (1996).
- Lebeda, A. et al. Research gaps and challenges in the conservation and use of North American wild lettuce germplasm. *Crop Sci.* **59**, 2337–2356 (2019).
- Werk, K. S. & Ehleringer, J. Non-random leaf orientation in *Lactuca serriola* L. *Plant Cell Environ.* **7**, 81–87 (1984).
- Werk, K. S. & Ehleringer, J. Photosynthetic characteristics of *Lactuca serriola* L. *Plant Cell Environ.* **8**, 345–350 (1985).
- Dhanda, S. S., Sethi, G. S. & Behl, R. K. Indices of drought tolerance in wheat genotypes at early stages of plant growth. *J. Agron. Crop Sci.* **190**, 6–12 (2004).
- Reed, R. C., Bradford, K. J. & Khanday, I. Seed germination and vigor: Ensuring crop sustainability in a changing climate. *Heredity* **128**, 450–459 (2022).
- Muscolo, A., Sidari, M., Anastasi, U., Santonoceto, C. & Maggio, A. Effect of PEG-induced drought stress on seed germination of four lentil genotypes. *J. Plant Interact.* **9**, 354–363 (2014).
- Heikal, M. M., Shaddad, M. A. & Ahmed, A. M. Effect of water stress and gibberellic acid on germination of flax, sesame, and onion seeds. *Biol. Plant* **24**, 124–129 (1981).
- Dodd, G. L. & Donovan, L. A. Water potential and ionic effects on germination and seedling growth of two cold desert shrubs. *Am. J. Bot.* **86**, 1146–1153 (1999).
- Misra, A. N. Pearl millet (*Pennisetum glaucum* LR Br.) seedling establishment under variable soil moisture stress. *Acta Physiol. Plant* **16**, 101–103 (1994).
- Wu, H., Asaduzzaman, M., Shephard, A., Hopwood, M. & Ma, X. Germination and emergence characteristics of prickly lettuce (*Lactuca serriola* L.). *Crop Prot.* **136**, 105222 (2020).
- Marks, M. & Prince, S. Influence of germination date on survival and fecundity in wild lettuce *Lactuca serriola*. *Oikos* **36**, 326–330 (1981).
- Michel, B. E. Evaluation of the water potentials of polyethyleneglycol 8000 both in the presence and absence of other solutes. *Plant Physiol.* **72**, 66–70 (1983).
- Hohl, M. & Schopfer, P. Water relations of growing maize coleoptiles: Comparison between mannitol and polyethylene glycol 6000 as external osmotica for adjusting turgor pressure. *Plant Physiol.* **95**, 716–722 (1991).
- Springer, T. L. & Goldman, J. J. Seed germination of five Poa species at negative water potentials. *Am. J. Plant Sci.* **7**, 601–611 (2016).
- Mofared, E. K., Moghaddam, P. R. & Mahallati, M. N. Modeling the effects of water stress and temperature on germination of *Lactuca serriola* L. seeds. *Int. Res. J. Appl. Basic Sci.* **9**, 1957–1965 (2012).
- Kaufmann, M. R. Effects of water potential on germination of lettuce, sunflower, and citrus seeds. *Can. J. Bot.* **47**, 1761–1764 (1969).
- Ryder, E. J. ‘Salinas’ lettuce1. *HortScience* **14**, 283–284 (1979).
- Johnson, W. C. et al. Lettuce, a shallow-rooted crop, and *Lactuca serriola*, its wild progenitor, differ at QTL determining root architecture and deep soil water exploitation. *Theor. Appl. Genet.* **101**, 1066–1073 (2000).
- Hartman, Y. et al. Abiotic stress QTL in lettuce crop–wild hybrids: Comparing greenhouse and field experiments. *Ecol. Evol.* **4**, 2395–2409 (2014).
- Damerum, A. et al. The genetic basis of water-use efficiency and yield in lettuce. *BMC Plant Biol.* **21**, 1–14 (2021).
- Hayashi, E., Aoyama, N. & Still, D. W. Quantitative trait loci associated with lettuce seed germination under different temperature and light environments. *Genome* **51**, 928–947 (2008).
- Argyris, J. et al. A gene encoding an abscisic acid biosynthetic enzyme (LsNCED4) collocalizes with the high temperature germination locus *Htg6.1* in lettuce (*Lactuca* sp.). *Theor. Appl. Genet.* **122**, 95–108 (2011).
- Guo, M., Lightfoot, D. A., Mok, M. C. & Mok, D. W. S. Analyses of *Phaseolus vulgaris* L. and *P. coccineus* Lam. hybrids by RFLP: Preferential transmission of *P. vulgaris* alleles. *Theor. Appl. Genet.* **81**, 703–709 (1991).
- Li, H. et al. Construction of a high-density composite map and comparative mapping of segregation distortion regions in barley. *Mol. Genet. Genom.* **284**, 319–331 (2010).
- Truco, M. J. et al. A high-density, integrated genetic linkage map of lettuce (*Lactuca* spp.). *Theor. Appl. Genet.* **115**, 735–746 (2007).
- Truco, M. J. et al. An ultra-high-density, transcript-based, genetic map of lettuce. *G3* **3**, 617–631 (2013).
- Fan, L., Li, R., Pan, J., Ding, Z. & Lin, J. Endocytosis and its regulation in plants. *Trends Plant Sci.* **20**, 388–397 (2015).
- Pagnussat, L., Burbach, C., Baluška, F. & de la Canal, L. Rapid endocytosis is triggered upon imbibition in Arabidopsis seeds. *Plant Signal. Behav.* **7**, 416–421 (2012).
- Ovečka, M. et al. Endocytosis and vesicle trafficking during tip growth of root hairs. *Protoplasma* **226**, 39–54 (2005).
- Ten Hove, C. A. et al. Probing the roles of LRR RLK genes in *Arabidopsis thaliana* roots using a custom T-DNA insertion set. *Plant Mol. Biol.* **76**, 69–83 (2011).
- Xiao, S., Jiang, L., Wang, C. & Ow, D. W. Arabidopsis OXS3 family proteins repress ABA signaling through interactions with AFP1 in the regulation of *ABI4* expression. *J. Exp. Bot.* **72**, 5721–5734 (2021).

43. Shu, K. et al. ABI4 regulates primary seed dormancy by regulating the biogenesis of abscisic acid and gibberellins in *Arabidopsis*. *PLoS Genet.* **9**, e1003577 (2013).
44. Subbiah, V. & Reddy, K. J. Interactions between ethylene, abscisic acid and cytokinin during germination and seedling establishment in *Arabidopsis*. *J. Biosci.* **35**, 451–458 (2010).
45. Fernando, V. D. & Schroeder, D. F. *Arabidopsis* DDB1-CUL4 E3 ligase complexes in *det1* salt/osmotic stress resistant germination. *Plant Signal. Behav.* **11**, e1223004 (2016).
46. Chen, H. et al. AtPER1 enhances primary seed dormancy and reduces seed germination by suppressing the ABA catabolism and GA biosynthesis in *Arabidopsis* seeds. *Plant J.* **101**, 310–323 (2020).
47. Wang, J. et al. CARK6 is involved in abscisic acid to regulate stress responses in *Arabidopsis thaliana*. *Biochem. Biophys. Res. Commun.* **513**, 460–464 (2019).
48. Taylor-Teeple, M. et al. An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. *Nature* **517**, 571–575 (2015).
49. Lin, J. H., Yu, L. H. & Xiang, C. B. ARABIDOPSIS NITRATE REGULATED 1 acts as a negative modulator of seed germination by activating ABI3 expression. *New Phytol.* **225**, 835–847 (2020).
50. Jamsheer, K. M. et al. FCS-like zinc finger 6 and 10 repress SnRK1 signalling in *Arabidopsis*. *Plant J.* **94**, 232–245 (2018).
51. Cai, G. et al. Type A2 BTB members decrease the ABA response during seed germination by affecting the stability of SnRK2.3 in *Arabidopsis*. *Int. J. Mol. Sci.* **21**, 3153 (2020).
52. Zhou, X., Cooke, P. & Li, L. Eukaryotic release factor 1–2 affects *Arabidopsis* responses to glucose and phytohormones during germination and early seedling development. *J. Exp. Bot.* **61**, 357–367 (2010).
53. Woolley, J. T. & Stoller, E. W. Light penetration and light-induced seed germination in soil. *Plant Physiol.* **61**, 597–600 (1978).
54. Mis, S., Ermis, S., Powell, A. A. & Demir, I. Radicle emergence (RE) test identifies differences in normal germination percentages (NG) of watermelon, lettuce, and carrot seed lots. *Seed Sci. Technol.* **50**, 257–267 (2022).
55. Irmak, S. & Djaman, K. *Basic Soil and Water Resources and Irrigation Engineering/Agricultural Water Management and Related Terminology* (University of Nebraska-Lincoln, Division of the Institute of Agriculture and Natural Resources, Circular EC2009, 2015).
56. Haqiqi, I., Grogan, D. S., Hertel, T. W. & Schlenker, W. Quantifying the impacts of compound extremes on agriculture and irrigation water demand. *Hydrol. Earth Syst. Sci.* **25**, 551–564 (2021).
57. Box, G. E. & Cox, D. R. An analysis of transformations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **26**, 211–243 (1964).
58. Kruskal, W. H. & Wallis, W. A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **47**, 583–621 (1952).
59. Dunn, O. J. Multiple comparisons among means. *J. Am. Stat. Assoc.* **56**, 52–64 (1961).
60. Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* 185–187 (Prentice Hall, 1996).
61. Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
62. Lander, E. S. et al. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174–181 (1987).
63. Kosambi, D. D. The estimation of map distances from recombination values. *Ann. Eugen.* **12**, 172–175 (1944).
64. Wang, S., Basten, C. J. & Zeng, Z. B. *Windows QTL Cartographer 2.5*. <https://brcwebportal.cos.ncsu.edu/qtlcart/WQTLCart.htm> (Department of Statistics, North Carolina State University, 2011).
65. Zeng, Z. B. Precision mapping of quantitative trait loci. *Genetics* **136**, 1457–1468 (1994).
66. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Method.* **57**, 289–300 (1995).
67. Kao, C. H., Zeng, Z. B. & Teasdale, R. D. Multiple interval mapping for quantitative trait loci. *Genetics* **152**, 1203–1216 (1999).
68. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
69. Thompson, O. et al. Low rates of mutation in clinical grade human pluripotent stem cells under different culture conditions. *Nat. Commun.* **11**, 1528 (2020).
70. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
71. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
72. Pertea, G. & Pertea, M. GFF utilities: GffRead and GffCompare. *F1000Research* **9**, 304 (2020).
73. Pertea, M., Kim, D., Pertea, G., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
74. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
75. Ashburner, M. et al. Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
76. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
77. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 1–13 (2008).
78. Shannon, P. et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

Acknowledgements

Authors gratefully acknowledge the seeds and public genotype data of the cv. Salinas × US96UC23 RIL population for this research from the Dr. Richard Michelmore laboratory at the University of California-Davis Genome Center. We thank the technical support and seed increase of David Milligan, Phi Diep, Carlos Saavedra, Sun-chung Park, and Pawan Kumar. We sincerely express our gratitude for the funding provided by the California Department of Food and Agriculture (Grant Number: 20-0001-035-SF) and the USDA National Institute of Food and Agriculture (Award Number: 2021-51181-35903).

Author contributions

S.H. conducted experiments, analyzed data, performed statistical analyses, and wrote the manuscript. I.S. and B.M. obtained the funding for the research and revised the manuscript. All authors read and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-77972-9>.

Correspondence and requests for materials should be addressed to S.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024