

Transcriptomic basis for salt tolerance and disease resistance of silverleaf sunflower revealed by Iso-seq and RNA-seq

Jing Bing (✉ jingbing@nwafu.edu.cn)

Northwest Agriculture and Forestry University

Yunhe Ling

Northwest Agriculture and Forestry University

Peipei An

Northwest Agriculture and Forestry University

Enshi Xiao

Northwest Agriculture and Forestry University

Chunlian Li

Northwest Agriculture and Forestry University

Bo Song

Chinese Academy of Agricultural Sciences

Zhonghua Wang

Northwest Agriculture and Forestry University

Research article

Keywords: Helianthus argophyllus, Sunflower, Transcriptome, Salt tolerance, Disease resistance

Posted Date: November 27th, 2019

DOI: <https://doi.org/10.21203/rs.2.17906/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background Silverleaf sunflower, *Helianthus argophyllus*, is one of the most important wild species that have been usually used for the improvement of cultivated sunflower. Although a reference genome is now available for the cultivated species, *H. annuus*, its effect in helping understanding the mechanisms underlying the traits of *H. argophyllus* is limited by the substantial genomic variance between these two species.

Results In this study, we generated a high-quality reference transcriptome of *H. argophyllus* using Iso-seq strategy. This assembly contains 50,153 unique genes covering more than 91% of the whole genes. Among them, we find 205 genes that are absent in the cultivated species and 475 fusion genes containing components of coding or non-coding sequences from the genome of *H. annuus*. It is interesting that in line with the strong disease resistance observed for *H. argophyllus*, these *H. argophyllus*-specific genes are predominantly related to functions of resistance. We have also profiled the gene expressions in leaf and root under normal or salt stressed conditions and, as a result, find distinct transcriptomic responses to salt stress in leaf and root. Particularly, genes involved in several critical processes including the synthesis and metabolism of glutamate and carbohydrate transport are reversely regulated in leaf and root.

Conclusions Overall, this study provided insights into the genomic mechanisms underlying the disease resistance and salt tolerance of silverleaf sunflower and the transcriptome assembly and the genes identified in this study can serve as a complement data resources for future research and breeding programs of sunflowers.

Background

Soil salinity is one of the most important environmental stresses limiting the productivity of crops worldwide, particularly in arid and semi-arid areas. A large amount of dissolved salt in soil would decrease the ability of a plant to take up sufficient water and result in the accumulation of Na^+ and Cl^- [1], which are toxic to plant growth by decreasing photosynthesis efficiency and impairing several crucial metabolic and signaling pathways[2]. Although the mechanisms of salt tolerance have been extensively studied in many model plants, such as *Arabidopsis* and rice (see review of [3], they are poorly understood in sunflower. The tolerance to salt in sunflower is achieved by preferentially uptaking Ca^{2+} coupled with the exclusion of Na^+ , B^{3+} , Mn^{2+} and Mg^{2+} [4].

The identification of genes underlying salt tolerance would be very important for the research and breeding programs of sunflower. However, only very limited genes involved into these processes have been identified. They include genes encoding membrane G-protein-coupled receptor, cyclic nucleotide and calmodulin-regulated ion channel protein and Ca-pump protein[5]. The understanding to such an important trait would be substantially accelerated after the release of the reference genome assembly of a cultivated sunflower, *H. annuus*[6]. However, like many other crop plants, the genetic variation has been

strongly reduced in the cultivated sunflower, compared to its wild relatives[7], due to artificial selection and genetic bottlenecks during the history of domestication[8]. The reference genome can represent only a part of the gene resources. A pan-genome derived from 287 cultivated lines, 17 landraces and 189 wild accessions suggests that more than a quarter of the genes vary across genotypes and around 10% of the genes in cultivated sunflower were derived from wild species[9]. Chromosome structure also varies remarkably across species, only 64 and 83% of the genomes of *H. niveus* and *H. argophyllus* are syntenic with those of *H. annuus*[7], respectively. Many wild species have been widely used in the breeding program of the cultivated sunflower as donors of beneficial alleles. Therefore, a reference genome or transcriptome assembly of these wild germplasms is also in need for rapid identification of genes and trait improvement.

Among them, silverleaf sunflower (*H. argophyllus*), a sister species of *H. annuus*, is one of the most important donors. For example, it has been used as a donor of alleles for disease resistance[10–17] and drought tolerance[18, 19]. Overall, it is the largest donor of genes to the cultivated sunflower, with approximately 5% of the genes in the cultivated sunflower were introgressed from *H. argophyllus*. When compared to *H. annuus*, 10 inverted and 8 translocated segments were found in the genome of *H. argophyllus*. Furthermore, *H. argophyllus* also exhibited tolerance in saline soils[20] and can be used as a donor of salt tolerance alleles[21]. Although several genomic regions associated with salt tolerance had been identified in *H. argophyllus*, the genes underlying this desirable trait are poorly known.

In this study, we firstly evaluated the salt tolerance of *H. argophyllus* and compared it with that of *H. annuus*. To better understand the genomic basis for the salt tolerance, we have built a reference transcriptome of *H. argophyllus* consisting of 50,153 unique genes, 205 of which are absent in *H. annuus*, using a long-read technology, Iso-Seq. We also find 475 fusion genes composed of coding or non-coding sequences. It did not surprise us that in concert of the strong disease resistance observed for silverleaf sunflower, the *H. argophyllus*-specific and ‘fusion’ genes were majorly related to disease resistance. We further investigated its transcriptomic responses to salt stress and found 3,930 and 1,885 genes that were significantly regulated in root and leaf, respectively. The responses to salt stress in root and leaf are distinct and the former is more sensitive. Our work revealed that the disease resistance of silverleaf sunflower is facilitated by genetic innovation and the salt tolerance is more dependent on the sensitive regulation of genes. Overall, we presented a high-quality reference transcriptome for *H. argophyllus* providing insights into the disease resistance and salt tolerance. This reference transcriptome can serve as a complement data resources for future research and breeding programs of sunflowers before the reference genome is available for *H. argophyllus*.

Results

Salt tolerance assay

Although *H. argophyllus* has shown tolerance in saline soils[20], its tolerance and mechanism have not been systematically evaluated. In this study, we examined physiological and morphological phenotypes

of *H. argophyllus* and a cultivated line, sk02R, with or without salt stress in terms of height, content of chlorophyll, Na, soluble proline, peroxidase (POD), superoxide dismutase (SOD), and malonaldehyde (MDA) as well as stomatal conductance (SC) and the rate of photosynthesis and water loss. To assay the salt tolerance of *H. argophyllus*, the foldchanges of these phenotypes after the treatment of salt (25 mM NaCl, 7 days) were calculated and used for comparison between *H. argophyllus* and *H. annuus*.

Although the heights of both the lines decreased in a similar degree (foldchange = 0.75) (Figure 1a), significant differences were shown in many other physiological phenotypes (Figure 1). We measured the water loss rate (WLR) of plants every 30 min after the treatment of salt and a lower WLR at every timepoints (Figure 1b). Although generally, the WLR increases along the time of salt treatment in both the genotypes, the WLR of *H. argophyllus* increased in a much lower rate ($y = 0.0875 + 0.0043x$, $P = 0.001205$) compared to that of *H. annuus* ($y = 0.1102 + 0.0054x$, $P = 0.001389$). The lower WLR observed in *H. argophyllus* could be attributed to the decrease of SC because we found that *H. argophyllus* had shown a significantly greater decrease of SC (foldchange = 10.22, $P = 0.05$) after salt treatment compared to *H. annuus* (foldchange = 9.43, $P = 0.002$) (Figure 1c). In line with this observation, the photosynthesis rate was more greatly reduced in *H. argophyllus* (Figure 1d).

Like in many other plants[22], due to the cell damage caused by the salt stress, the level of MDA was significantly in both the genotypes (Figure 1e). We noticed that the degree of increment in *H. argophyllus* is relatively modest compared to that of *H. annuus* (Figure 1e), suggesting a lower sensitivity of *H. argophyllus*. Accumulation of reactive oxygen species (ROS), which irreversibly impair the cells[23], is another consequence that would be caused by salt stress. To cope with the excessive ROS, the activities of POD and SOD are required. The increased accumulation of SOD has been reported in many crops including cotton[24], wheat[25] and cucumber[26]. We therefore also examined the changes of POD and SOD after the treatment of salt (7 days). Our data revealed that the level of POD was significantly increased in *H. annuus* (foldchange = 4.51, $P = 0.06$) but did not change in *H. argophyllus* (foldchange = 1.04, $P = 0.59$) (Figure 1f); SOD was significantly decreased in *H. annuus* (foldchange = 0.87, $P = 0.02$) but increased in *H. argophyllus* (foldchange = 1.47, $P = 7.715e-06$) (Figure 1g); In response to salt stress, the concentration of proline was also significantly increased in both the genotypes, and it is not surprise that the increment in *H. argophyllus* (foldchange = 3.80, $P = 0.001$) was significantly lower than that in *H. annuus* (foldchange = 12.01, $P = 0.02$) (Figure 1h). These results together suggest a lower sensitivity and stronger tolerance to salt stress of *H. argophyllus*.

We further measured the content of Na⁺ in root, stem and leaf of both the genotypes before and after the treatment of salt. Despite a general increment of Na⁺ in these tissues of both the genotypes under salt stressed condition, the content in *H. argophyllus* root increased only 3.92 folds while that in *H. annuus* root increased 4.85 folds (Figure 1i). Similar patterns were also observed in stem and leaf (Figure 1i). Overall, Na⁺ is fewer accumulated in *H. argophyllus*, probably due to its ability of excluding toxic ions.

Reference transcriptome assembly

To better understand the genomic mechanisms underlying the salt tolerance of *H. argophyllus*, we first construct a high-coverage reference transcriptome assembly. We collected RNA from several tissues including leaf and root and pooled them together for Iso-seq, which finally generated 290,182 reads with an average length of 1,874 bp (Supplementary Table 1). The poly-A sequence was detected in 280,695 (96.73%) of the reads (Poly-A reads), among which 270,267 (96.28) were full-length with both the 5' and 3' primers (FL reads), 264,113 (94.09%) were full-length with nonchimerism of these ROIs (FLNC reads). We then evaluated the similarity between the *H. argophyllus* transcriptome and the reference genome of cultivated species[6] and found a very limited similarity between them. Even after a very relax filtering (Identity \geq 30%, coverage \geq 30%) and cleaning, there were only 72.16% of the reads mapped to the reference genome. This result is consistent with the reported variance between cultivated and wild sunflowers[7], and raised a caveat that the reference genome of *H. annuus* may not be the best choice for analyses in *H. argophyllus* despite they are closely related. Therefore, we assembled the reads into an assembly consisting of 50,153 unique genes, with a mean length of 1999.73 bp. We assessed the completeness of this transcriptome by querying them against the database of Benchmarking Universal Single-Copy Orthologs (BUSCO, version 3.0)[27]. Our results indicate that 91.4 % of the core embryophyta genes (CEGs) were found in this assembly, with 52.5% being single copies and 38.9% being duplicates (Supplementary Table 2), comparable to the percentage reported in the reference genome[6], suggesting a high coverage of this assembly.

Comparison between *H. argophyllus* transcriptome and *H. annuus* genome

The poor similarity between *H. argophyllus* and *H. annuus* was also confirmed by the large abundance of single nucleotide polymorphisms (SNPs) detected in *H. argophyllus* transcriptome. When compared the reads to the coding sequences of reference genome of *H. annuus*, a total of 623,076 SNPs and 99,787 indels were identified from the transcriptome of *H. argophyllus*, which is greater than that previously reported within *H. annuus* population (85,063)[28]. Averagely, the SNPs and indels occur every 140.90 and 657.76 bp (623,076 SNPs in 87,791,689 bp and 99,787 indels in 65,636,655 bp), respectively, which is comparable to that reported in *H. annuus* transcriptome (163.6 bp)[28]. As expected, we found a biased distribution of SNPs at the positions of codons, with a frequency of 9.60% (59,825 of 623,076), 7.27% (45,314 of 623,076) and 33.97% (211,717 of 623,076) for the first, second and third nucleotide, respectively (Supplementary Figure 1). On the other hand, only a third of the indels were in lengths multiple-of-three, suggesting most of the indels had resulted in frame shift of ORFs. These data together revealed that most of the SNPs and indels have resulted into changes of amino acids.

Due to the existence of present and absent variations, the dataset of SNPs cannot present the whole difference between these two species. Therefore, we further investigated their differences in genes. By clustering their genes into families, we found that 205 genes (in 7 clusters) are unique to *H. argophyllus* and 117 genes (in 8 clusters) are unique to *H. annuus*. We annotated the functions of these genes by searching in non-redundant proteins (NR) database on NCBI. The results showed that 75.12% (154 of 205

genes) of the *H. argophyllus*-specific genes are related to disease resistance (Supplementary Table 3). This explains the strong and wide disease resistance of *H. argophyllus* and the successful practice of using *H. argophyllus* to enhance the disease resistance of cultivated sunflower[10–17]. It surprised us that there are also many genes (10.25%, 12 of 117 genes) specific to *H. annuus* involved into disease resistance (Supplementary Table 4). This suggests high diversity of resistance genes, which could be due to the activities of retrotransposons, such as long terminal repeat retrotransposons (LTR), which often results in the expansion of resistance genes in plants[29, 30], after the divergence between *H. argophyllus* and *H. annuus*. Furthermore, in concert of the strong tolerance to salt as mentioned above, we also found many transporters in *H. argophyllus*-specific genes including Aquaporin-like protein (Supplementary Table 3). To explore in what biological processes the *H. argophyllus*-specific genes were involved, we performed gene ontology (GO) enrichment analysis and found that these genes were involved in 11 biological processes and were enriched into three processes, ‘mRNA processing’, ‘defense response’ and ‘transport’ (Supplementary Table 5). The GO term ‘transporter activity’ was also overrepresented in this gene set, implying an outstanding tolerance to environmental stresses. These results together suggest that *H. argophyllus* has been armed with an extra gene reservoir enabling wide resistance or tolerance to various biotic (disease) and abiotic (drought, salt) stresses.

By comparing the sequences of *H. argophyllus* unique genes and the genes in *H. annuus* genome, we found 475 genes which are composed of two parts that are distant in the genome of *H. annuus* (Table 1). This kind of genes were termed as ‘fusion’ genes in many other studies[31, 32]. Nevertheless, ‘fusion’ genes can also be ‘split’ genes depending on their origins and evolutionary trajectories. These genes ranged from 206 to 7,440 bp with a mean of 1,917.56 bp in length, comparable to that of the whole set of unique genes, suggesting they are not exactly fused genes. On the other hand, the length of their components found in the *H. annuus* genome varied from 156 to 9,444 bp, averaged at 1,318.91 bp, which was also comparable to that of other genes, suggesting they are not split genes either (Figure 2). Hereafter in this paper, for a purpose of easy description, we refer these genes to ‘fusion’ genes following the terminology in other reported studies[33]. Among the 475 ‘fusion’ genes, 171 were composed of sequences from two protein-coding genes, 55 were composed of protein-coding and non-coding genes, and 246 were from coding/non-coding genes and intergenic regions (Supplementary Table 6). Those components of non-coding genes and intergenic sequences were probably resulted by the degradation of open reading frames or pseudogenization after the split of their parent genes.

We detailly inspected the functions of these ‘fusion’ genes and found that they are involved into much more diverse biological processes (Supplementary Table 6). The ‘fusion’ genes were categorized into 606 biological processes and enriched in 277 of them (Supplementary Table 7). The roles of ‘fusion’ genes are diversified. For example, we enriched GO terms include ‘response to abiotic stimulus’, ‘pectin catabolic process’, ‘response to temperature stimulus’, ‘response to oxygen-containing compound’, ‘response to phosphate starvation’ and ‘cell wall modification’, to name only a few (Supplementary Table 7). We particularly noticed that several processes, such as ‘pectinesterase activity’, ‘response to abscisic acid’ and ‘cell wall modification’ are strongly associated with the responses to abiotic stresses[34–37]. Compared to the functional enrichment of disease resistance in *H. argophyllus*-specific genes, the ‘fusion’

genes identified from *H. argophyllus* are engaged in wide pathways and most of them are related to responses to abiotic stresses including phosphate starvation, heat, cold, and probably drought and salt stresses.

Overall, the comparison between *H. argophyllus* and *H. annuus* suggested *H. argophyllus* as an important donor of gene resources that are valuable for improvement of cultivated sunflower.

Transcriptomic responses to salt stress

Because *H. argophyllus* can potentially be used to enhance the salt tolerance of cultivated sunflower, we explored transcriptomic changes of *H. argophyllus* genes in response to salt stress. We collected leaves and roots of plants grown under normal or salt stressed conditions for RNA-seq sequencing and analyses. Reads were generated and mapped to the reference transcriptome assembly to quantify the gene expression levels under different conditions. A clustering analysis based on the normalized read counts of genes in these samples suggests high reproducibility between biological replicates (Supplementary Figure 2), which was also confirmed by a principle component analysis (PCA) (Supplementary Figure 3).

We first compared the gene expression levels before and after the treatment of salt using DESeq2 package[38], and genes with foldchange ≥ 1.5 and Q -value ≤ 0.01 were defined as significantly differentially expressed genes (DEGs). As a result, we found 1,885 DEGs in leaf (Figure 3; Supplementary Table 8), with 1,089 and 796 being significantly up- and down-regulated under salt stressed condition (Supplementary Figure 4A). Among the DEGs, 9 were *H. argophyllus*-specific genes, 2 of them were up-regulated while 7 were down-regulated in response to salt stress (Supplementary Table 9). The DEGs identified in leaf also included 36 'fusion' genes, 31 and 5 were up- and down-regulated, respectively (Supplementary Table 10). A hypergeographic test suggests that the *H. argophyllus*-specific genes tend to be down- ($P = 0.02$) but not up-regulated ($P = 0.82$) in a chance significantly larger than random occurrence; a opposite pattern was found for 'fusion' genes with a P value of $8.283374e-12$ and 0.76 for the up- and down-regulated genes. These results suggest critical but different roles played by *H. argophyllus*-specific and 'fusion' genes in response to salt stress.

Like the reports in many other plants, genes in the pathways involved into salt tolerance were significantly regulated. For example, non-specific lipid-transfer protein (nsLTP) in lipid metabolic pathway was significantly up-regulated (Supplementary Table 8). The results of GO enrichment analysis also confirmed the conclusion that processes and activities related to resistance to abiotic and biotic resistance were over-represented in the up-regulated genes (Supplementary Table 11), such as response to osmotic stress, response to water and defense response to fungus and those related to 'cell wall organization' and 'cell wall biogenesis' were over-represented in the down-regulated genes (Supplementary Table 12). Genes involved in the removal of superoxide radicals, such as *CPN20*, had been significantly up-regulated under

salt stressed condition (Supplementary Table 11), consistent with the increased activities of POD and SOD (Figure 1f and 1g). In line with the increased level of cytological proline observed in plants under salt stressed condition (Figure 1h), the genes engaged in proline biosynthesis pathway were also significantly up-regulated. Other than proline, the levels of many other amino acids may also have increased in response to salt stress because the up-regulated genes in leaf were also found to be enriched in GO terms of 'glutamate biosynthetic process (GO:0006537)', 'asparagine biosynthetic process (GO:0006529)', 'alpha-amino acid biosynthetic process (GO:1901607)' and 'arginine catabolic process (GO:0006527)'. The increased amount of amino acids may have further promoted the enhancement of translation activities (Figure 4; Supplementary Table 11).

Because root is more sensitive to salt stress, more (3,930) DEGs were identified from root (Supplementary Table 8), of which 1,925 and 2,005 were up- and down-regulated under salt stressed condition (Supplementary Figure 4B), respectively. *H. argophyllus*-specific genes were also significantly enriched in the down-regulated DEGs of root. Specifically, 2 (0.9%, $P = 0.98$) and 20 (9.8%, $P = 8.290412e-05$) *H. argophyllus*-specific genes were found in up- and down-regulated DEGs, respectively (Supplementary Table 13). On the other hand, this pattern was not shown for 'fusion' genes. For 'fusion' genes, 10 and 5 genes were up- and down-regulated (Supplementary Table 14). It is interesting that the regulation of genes in leaf and root is distinct, with only a few genes that were commonly up- or down-regulated in both the tissues (Figure 3). We therefore categorized the DEGs into eight groups, genes (1) up-regulated in leaf, (2) up-regulated in root, (3) commonly up-regulated, (4) down-regulated in leaf, (5) down-regulated in root, (6) commonly down-regulated, (7) up-regulated in leaf but down-regulated in root and (8) *vice versa* (Supplementary Table 8; Supplementary Figure 5). Our data showed that the commonly up- or down-regulated genes were enriched to biological processes that are fundamental to the machinery of salt tolerance, such as 'response to water (GO:0009415)', 'lipid transport (GO:0006869)', 'proline biosynthetic process (GO:0006561)' and 'cell wall biogenesis (GO:0042546)' (Supplementary Table 15 and Supplementary Table 16), and the genes uniquely up-regulated in root are more closely related to primary responses to abiotic stresses, resulting in an enrichment of GO terms of 'response to oxidative stress (GO:0006979)', 'response to osmotic stress (GO:0006970)' and 'response to stress (GO:0006950)' (Supplementary Table 17; Supplementary Figure 5). The genes oppositely regulated in leaf and root showed several interesting functions. For example, the genes up-regulated in leaf but down-regulated in root were enriched in GO terms related to the biosynthesis and metabolic of glutamate, which serves as a signal triggering long-distance defense[39] and mediating responses to abiotic stresses[40]; the genes up-regulated in root but down-regulated in leaf showed functional enrichment to carbohydrate transport and DNA methylation, which may have played roles in the regulation of genes (Supplementary Figure 5).

We further compared the foldchanges of the commonly regulated genes and classified them into three groups (see methods), (1) higher in leaf, (2) higher in root and (3) similar. As a result, we found 1,521, 3,717, and 190 genes belonging to these groups, respectively, suggesting a varied degree of response to salt stress of these two tissues with a stronger regulation observed in root. These results together indicate a distinct pattern of transcriptomic responses to salt stress in leaf and root.

Discussion

The reference genomes of most, if not all, of the crops were generated from a cultivated species, which can represent only a part of the gene pools due to the genetic bottlenecks during the history of domestication[8]. A pan-genome analysis of sunflower also revealed remarkable variation across different genotypes[9]. As for *H. argophyllus*, there is 83% of the chromosomes showing collinearity with the reference genome, and the percentage is even lower for *H. niveus* (64%)[7]. In concert to the genetic variation, the phenotypes, such as seed dormancy, flowering time, oil composition and content, as well as the resistance to abiotic and biotic resistance, also vary substantially across genotypes[19, 41, 42]. Therefore, to better understand the mechanisms and genomic resources underlying those desirable traits observed in the wild species, a reference genome, or at least a reference transcriptome, is needed for each valuable genotype. In this study, we generated a high-quality reference transcriptome for *H. argophyllus*, one of the most important donors of beneficial alleles to the cultivated sunflower[9], by using a long-read technique, Iso-seq, which is powerful in capturing full-length transcripts[43–46]. It is no doubt that a reference genome of *H. argophyllus* is in urgent need and some groups may have already been working on that. Before the genome sequence become available, the reference transcriptome provided in our work can serve as an alternate and be used for gene mining, and the SNPs identified from transcripts can potentially be used as markers in breeding program.

The divergence between *H. annuus* and *H. argophyllus*

The variance between wild species and the cultivated sunflower had been repeatedly described in many earlier works[7, 9, 21]. In this study, to compare the divergence between the silverleaf sunflower and the cultivated species, we sequenced the transcriptomes of *H. argophyllus* leaves and roots under normal or salt stressed conditions and identified substantial SNPs and indels. It should be noted that although RNA-seq had been proved to be efficient in identifying molecular markers associated with environment tolerance by comparing the transcriptomes between the sensitive and tolerant cultivars[47], the dataset of SNPs and indels obtained in this study can only represent a part of the whole variances between these two species due to the lack of a reference genome of *H. argophyllus*.

Structure variations between *H. argophyllus* and *H. annuus* had also been previously reported[7, 21]. Studies based on molecular markers revealed at least 10 inversion and 8 translocations in *H. argophyllus*, in comparison to *H. annuus* chromosomes[7]. However, the identification of chromosomal rearrangements heavily relies on the density of molecular marker, which could probably have resulted in an under estimation, particularly in the regions sparse of molecular markers. Large-scale rearrangements are relatively easier for identification while small rearrangement could be undetectable. One of the consequences of chromosome rearrangement is the fusion of genes at the boundaries of the rearranged segments. Indeed, we also found 475 fusion genes in *H. argophyllus*. Their functions are strongly related to several of the desirable traits of *H. argophyllus*, such as ‘disease resistance’ (Supplementary Table 6). The genomic divergence between these two species was also reflected by the presence and absence

variation of genes. We found 205 and 117 *H. argophyllus*- and *H. annuus*-specific genes, respectively. The former was highly enriched in processes related to disease resistance, consistent with the fact that *H. argophyllus* had been used as a valuable gene resources in the improvement of disease resistance of the cultivated sunflower.

Distinct salt stress responses in leaf and root

Both leaf and root had shown strong response to salt stress. It did not surprise us that root is more sensitive to the stress with 3,930 genes were significantly regulated while only 1,885 genes were regulated in leaf (Supplementary Table 8). We found that genes that have been regulated were distinct between these two tissues. We sorted the DEGs in leaf or root according to their fold changes. As a result, we found that the most strongly up-regulated genes in leaf are those related to lipid synthesis and transport and amino acid synthesis. For example, non-specific lipid-transfer protein (nsLTP) and methylsterol monooxygenase 1 (MSMO1) were frequently seen in the top ranked up-regulated DEGs (Supplementary Table 8). Lipid has been known to play critical roles in sensing salt stress[48] and mediating salt tolerance in various plants and algae[49]; nsLTP has been reported to be involved into several biological processes including the tolerance to salt in maize[50] and the null mutant of LTP in *Arabidopsis* had resulted in hypersensitivity to salt stress[51]. The most up-regulated genes in root are dehydrin-like proteins, followed by small heat shock proteins (HSP) and their cognates (Supplementary Table 8). The up-regulation of genes encoding dehydrin-like protein is line of the conclusions drawn from the studies in other crops. For example, dehydrin-like protein was accumulated in cereals in response to cold or drought stresses[52]. HSPs are widely engaged in resistance or tolerance to various environmental stresses, such as cold, heat, drought and salt[53–55], and the over-expression of HSP can significantly enhance the tolerance to drought and salt stresses[55]. The mechanism and pathway of HSP mediating salt stress tolerance is conserved because the similar responses were also observed in bacteria, such as *Clostridium botulinum*[56]. It is reasonable to believe that sunflower applied the same pathway to cope with salt stress.

On the other hand, the most down-regulated genes in root are non-symbiotic hemoglobin genes (*nsHb*) (Supplementary Table 8), which is involved in responses to several abiotic stresses including salt stress in maize[57, 58]. A study in *Arabidopsis* had shown that the expression level of *nsHb* is negatively correlated to the tolerance to salt. The over-expression of spinach *nsHb* in a transgenic *Arabidopsis* had resulted in lower tolerance to NaCl[59]. We speculate that the decreased expression of *nsHb* in *H. argophyllus* after the treatment of salt is beneficial for the plant to cope with the stress. Most of the genes top ranked in the down-regulated DEGs in leaf are hypothetical proteins without any annotated functions. It is interesting that besides these hypothetical proteins, the expression of many resistance genes dropped dramatically under salt stressed condition, which may result in a reduced resistance to diseases.

GO analysis provided a more general view of the processes that have been affected by salt stress. As expected, the processes of responses to abiotic stimulus including water, heat, inorganic substance as

well oxidative stress in root were generally activated by the treatment of NaCl. Several other related processes, such as transport, lipid and proline metabolic processes, were also up-regulated under salt stressed condition. The processes up-regulated in leaf were more concentrated on some fundamental biological processes, such as translation, mitosis, cell wall organization and carbon fixation (Supplementary Table 11). In line with the increased activities of POD and SOD (Figure 1), several processes involved in positive regulation of oxidoreductase activity were also enhanced in leaf (Supplementary Table 11). These results are consistent with earlier observations in other crops. The elongation factor had a higher abundance in cotton under salt stressed conditions[24]. We noticed that several GO terms were commonly over-represented by both the up-regulated DEGs in leaf and root. They include 'proline biosynthetic process' and 'lipid biosynthetic process' (Supplementary Table 15). Proline is one of the most important osmolytes conferring salt tolerance and can also function as a signaling molecular modulating the expression of genes involved in the recovery of cells[60]. The impairment of proline biosynthesis pathway could lead to hypersensitivity to salt stress in Arabidopsis[61]. Furthermore, we also found decreased expression of proline dehydrogenase genes under salt stressed condition (Figure 1). These results together explained the increased accumulation of proline in leaves after salt treatment.

The processes down-regulated in leaf are majorly involved in the transport of carbohydrate, such as hexose transport, glucose transport and monosaccharide transport. It does not look like a response to the salt stress, but more like a consequence caused by the impairment by salt stress. The processes of cell wall biogenesis, such as 'cell wall organization' and 'cell wall polysaccharide metabolic process', were also down-regulated (Supplementary Table 12). As a result, the metabolic processes of polysaccharide, such as sucrose, glucan, xyloglucan, in root were then subsequently affected (Supplementary Table 18; Supplementary Table 19).

Conclusions

Overall, we presented a high-quality reference transcriptome of *H. argophyllus*, which enabled us to identify a set of genes that are absent in the cultivated sunflower. These genes are related to disease resistance and salt or drought tolerance, representing beneficial gene resources that can be utilized in breeding programs. The salt tolerance of *H. argophyllus* was also measured and its transcriptomic responses to salt stress were profiled. Distinct responses were found between leaf and root of sunflower. We believe that the data provided in this study can serve as a complement data resources for future research and breeding programs of sunflowers.

Methods

Plant materials and treatments

The silverleaf sunflower *H. argophyllus* (ARG1807) obtained from the USDA North Central Regional Plant Introduction Station (NCPIS) was used in this study. The seeds were sterilized and germinated before the

seedlings were transferred to pots filled with garden soil when they were grown to two-leaf stage. Salt treatment was performed to plants at four-leaf stage by supplementing with NaCl solution (25 mM). After a salt treatment of 7 days, the leaves in the middle of the plants under salt stressed or normal conditions were taken and freeze immediately. In physiological assay, a cultivated cultivar (sk02R) maintained in our lab was also planted and treated. Three biological replicates were performed for each treatment.

Physiological assay

To compare the salt tolerance between silverleaf (ARG1807) and the cultivated sunflower (sk02R), we examined the height of plants, the content of chlorophyll, Na⁺, soluble proline, MDA, the activities of POD, SOD as well as stomatal conductance (SC) and the rate of photosynthesis and water loss. At the same time of sampling (7 days after salt treatment), the height of the plants of both the cultivars were measured and the others were measured using the frozen leaves. The activities of POD and SOD were measured as previously described[24, 62]. To measure the content of Na⁺, the leaves were incinerated and the ash was dissolved in 0.5 M HCl solution before the concentration of Na⁺ was determined as described in[63]. The content of MDA and proline was also measured according to previously described protocols[22, 64].

RNA extraction and sequencing

Total RNA was extracted using TRIzol reagent following the manufacturer's instruction (Thermo Fisher Scientific, MA, USA) and quantified using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, MA, USA). Two strategies, Iso-seq and RNA-seq, were applied in this study, the former was used to generate a reference transcriptome and the latter was used to quantify the gene expression levels under different conditions. For Iso-seq, total RNA of leaf and root with or without the treatment of salt was pooled together and reversed to cDNA using Clontech SMARTer PCR cDNA synthesis kit before they were circularized using SMRTbell Template Prep kit and sequenced on a Sequel platform (PacBio, CA, USA). Second-generation sequencing was also performed to RNA extracted from leaf or root under the normal and salt stressed conditions. The libraries were constructed using a standard protocol of Illumina as described in many other studies[29, 65] and sequenced on a HiSeq platform (Illumina, CA, USA).

Data processing

The full length reads of insert (ROIs) were firstly clustered using an algorithm of iterative clustering for error correction (ICE) to generate consensus isoforms, which were then mapped by the non-full-length and chimeric ROIs for further correction. Finally, consensus isoforms with accuracy higher than 99% were retained for further analyses. To quantify the gene expression levels of leaves or roots under normal or salt stressed conditions, the reads of RNA-seq were mapped to the reference transcriptome generated from Iso-seq ROIs as described above using bowtie2[66] with default parameters. The number of reads

hit to each transcript was counted and subjected to the identification of DEGs using DESeq2[38]. Genes with foldchange higher than 1.5 or 0.66 and adjusted P-values lower than 0.01 were defined as significantly up- and down-regulated genes. To compare the degree of regulation of DEGs, the foldchanges of genes were transferred to their logarithms of 2 and the subtracts between leaf and root were calculated. The genes with P-values higher than 0.01 were manually changed to 1 before they were transferred, which would result a value of 0 indicating no change. The DEGs with a subtract (leaf - root) higher than 2, or lower than -2, were classified into a group of 'higher in leaf', and 'higher in root', respectively, and the remaining genes were in the group of 'similar'. To detect the variance between *H. argophyllus* and the cultivated sunflower, *H. annuus*, the reads of RNA-seq were aligned to the reference genome of *H. annuus*[6] and SNPs were called using Varscan[67] with criteria of Q > 20, depth > 8, P-value < 0.01.

Annotation of transcriptome

The functions of unique genes in this reference transcriptome were annotated by homologous searching in public databases including SwissProt[68], TrEMBL[68], KEGG[69], InterPro[70] and Non-redundant protein NCBI databases[71]. The sequences of unique genes were queried against these databases using BLAST[71] and the best hit item for each unique gene was retained. Coding sequence of *H. annuus* and unique genes of *H. argophyllus* were performed the all-against-all comparisons using Blastn (version 2.2.31)[71], with an e-value cutoff of $\leq 1e-5$. OrthoMCL[72] with default parameters were used to identify gene families. The unique genes of *H. argophyllus* in *H. argophyllus*-specific gene families were selected and defined as *H. argophyllus*-specific genes.

GO analysis

The GO entries of unique genes were extracted from the output of InterPro searching results and the terms were obtained from the GO database according to the entries. GO enrichment analysis was performed for different gene sets using a hypergeometric test implemented in an R package, 'phyper'.

Abbreviations

POD: peroxidase

SOD: superoxide dismutase

MDA: malonaldehyde

SC: stomatal conductance

WLR: water loss rate

ROS: reactive oxygen species

ROIs: reads of insert

ICE: iterative clustering for error correction

BUSCO: Benchmarking Universal Single-Copy Orthologs

CEGs: core embryophyta genes

SNPs: single nucleotide polymorphisms

NR: non-redundant proteins

LTR: long terminal repeat retrotransposons

GO: Gene ontology

PCA: principle component analysis

DEGs: Differentially Expressed Genes

nsLTP: non-specific lipid-transfer protein

MSMO1: methylsterol monooxygenase 1

HSP: heat shock proteins

nsHb: non-symbiotic hemoglobin genes

KEGG: Kyoto Encyclopedia of Genes and Genomes

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no conflict of interest.

Funding

This research was funded by the National Natural Science Foundation of China (31601335); China agriculture research system (CARS-14-2-21); the Fundamental Research Funds for the Central Universities (2452016012).

Authors' Contribution

E. X., B. J. and Z. W. conceived the work; B. J. and Z. W. conducted the experiments. Y. L. and P. A. collected and prepared the samples for RNA sequencing; B. J., Y. L., C. L. and B. S. participated in the data analyses and visualization; B. J., Y. L. and E. X. wrote the manuscript. All authors have read and approved the manuscript.

References

1. Mäser P, Eckelman B, Vaidyanathan R, Horie T, Fairbairn DJ, Kubo M, Yamagami M, Yamaguchi K, Nishimura M, Uozumi NJFI: **Altered shoot/root Na⁺ distribution and bifurcating salt sensitivity in Arabidopsis by genetic disruption of the Na⁺ transporter AtHKT1.** *FEBS letters* 2002, **531**(2):157-161.
2. Munns R, Tester MJARPB: **Mechanisms of salinity tolerance.** *Annu Rev Plant Biol* 2008, **59**:651-681.
3. Deinlein U, Stephan AB, Horie T, Luo W, Xu G, Schroeder JIJTips: **Plant salt-tolerance mechanisms.** *Trends Plant Sci* 2014, **19**(6):371-379.
4. Lexer C, Welch ME, Raymond O, Rieseberg LHJE: **The origin of ecological divergence in Helianthus paradoxus (Asteraceae): selection on transgressive characters in a novel hybrid habitat.** *Evolution* 2003, **57**(9):1989-2000.
5. Lai Z, Livingstone K, Zou Y, Church S, Knapp S, Andrews J, Rieseberg LJT, Genetics A: **Identification and mapping of SNPs from ESTs in sunflower.** *Theor Appl Genet* 2005, **111**(8):1532-1544.
6. Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Brière C, Owens GL, Carrère S, Mayjonade BJN: **The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution.** *Nature* 2017, **546**(7656):148.
7. Barb JG, Bowers JE, Renaut S, Rey JI, Knapp SJ, Rieseberg LH, Burke JMJC: **Chromosomal evolution and patterns of introgression in Helianthus.** *Genetics* 2014, **197**(3):969-979.
8. Doebley JF, Gaut BS, Smith BDJC: **The molecular genetics of crop domestication.** *Cell* 2006, **127**(7):1309-1321.
9. Hübner S, Bercovich N, Todesco M, Mandel JR, Odenheimer J, Ziegler E, Lee JS, Baute GJ, Owens GL, Grassa CJNp: **Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance.** *Nature plants* 2019, **5**(1):54.
10. Heiser Jr CBJTAN: **Hybridization in the annual sunflowers: Helianthus annuus x H. argophyllus.** *The American Naturalist* 1951, **85**(820):65-72.

11. Rogers CE, Thompson TE, Seiler GJ: **Sunflower species of the United States**. 1982.
12. Miller J, Gulya TJs: **Inheritance of resistance to race 4 of downy mildew derived from interspecific crosses in sunflower**. *Crop science* 1991, **31**(1):40-43.
13. Slabaugh MB, Yu JK, Tang S, Heesacker A, Hu X, Lu G, Bidney D, Han F, Knapp SJJpbj: **Haplotyping and mapping a large cluster of downy mildew resistance gene candidates in sunflower using multilocus intron fragment length polymorphisms**. *Plant biotechnology journal* 2003, **1**(3):167-185.
14. Dussle C, Hahn V, Knapp S, Bauer EJT, genetics a: **PI Arg from Helianthus argophyllus is unlinked to other known downy mildew resistance genes in sunflower**. *Theoretical applied genetics* 2004, **109**(5):1083-1086.
15. Radwan O, Bouzidi MF, Nicolas P, Mouzeyar SJT, Genetics A: **Development of PCR markers for the PI5/PI8 locus for resistance to Plasmopara halstedii in sunflower, Helianthus annuus L. from complete CC-NBS-LRR sequences**. *Theor Appl Genet* 2004, **109**(1):176-185.
16. Seiler G, Gulya T, Marek L: **Re-collection of Helianthus argophyllus, source of the P1Arg gene for downy mildew resistance, surviving for 25 years on Daytona Beach, Florida**. In: *29th Sunflower Research Workshop, Fargo, ND: 2007*.
17. Wieckhorst S, Bachlava E, Dußle C, Tang S, Gao W, Saski C, Knapp S, Schön C-C, Hahn V, Bauer EJT et al: **Fine mapping of the sunflower resistance locus PI ARG introduced from the wild species Helianthus argophyllus**. *Theor Appl Genet* 2010, **121**(8):1633-1644.
18. Hussain MM, Rauf S, Riaz MA, Al-Khayri JM, Monneveux PJBs: **Determination of drought tolerance related traits in Helianthus argophyllus, Helianthus annuus, and their hybrids**. *Breeding science* 2017:16095.
19. Hussain MM, Rauf S, Warburton MLJPB: **Development of drought-tolerant breeding lines derived from Helianthus annuus× H. argophyllus interspecific crosses**. *Plant Breeding* 2019.
20. Richards RJP, Soil: **Increasing salinity tolerance of grain crops: is it worthwhile?** *Plant Soil* 1992, **146**(1-2):89-98.
21. Rey JI: **Silverleaf Sunflower (Helianthus Argophyllus, Torrey and Grey) in Sunflower Breeding: From Comparative Mapping to the Genetics of Salt Tolerance in the Species**. University of Georgia; 2011.
22. Wang Z, Yang C, Chen H, Wang P, Wang P, Song C, Zhang X, Wang DJPs: **Multi-gene co-expression can improve comprehensive resistance to multiple abiotic stresses in Brassica napus L.** *Plant science* 2018, **274**:410-419.
23. Sekmen AH, Ozgur R, Uzilday B, Turkan IJE, Botany E: **Reactive oxygen species scavenging capacities of cotton (Gossypium hirsutum) cultivars under combined drought and heat induced oxidative stress**. *Environmental Experimental Botany* 2014, **99**:141-149.
24. Li W, Zhao Fa, Fang W, Xie D, Hou J, Yang X, Zhao Y, Tang Z, Nie L, Lv SJFips: **Identification of early salt stress responsive proteins in seedling roots of upland cotton (Gossypium hirsutum L.) employing iTRAQ-based proteomic technique**. *Front Plant Sci* 2015, **6**:732.
25. Guo G, Ge P, Ma C, Li X, Lv D, Wang S, Ma W, Yan YJJoP: **Comparative proteomic analysis of salt response proteins in seedling roots of two wheat varieties**. *J Proteomics* 2012, **75**(6):1867-1885.

26. Du C-X, Fan H-F, Guo S-R, Tezuka T, Li JJP: **Proteomic analysis of cucumber seedling roots subjected to salt stress.** *Phytochemistry* 2010, **71**(13):1450-1459.
27. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics* 2015, **31**(19):3210-3212.
28. Bachlava E, Taylor CA, Tang S, Bowers JE, Mandel JR, Burke JM, Knapp SJ: **SNP discovery and development of a high-density genotyping array for sunflower.** *PLoS one* 2012, **7**(1):e29814.
29. Song B, Song Y, Fu Y, Kizito EB, Kamenya SN, Kabod PN, Liu H, Muthemba S, Kariba R, Njuguna JJG: **Draft genome sequence of *Solanum aethiopicum* provides insights into disease resistance, drought tolerance, and the evolution of the genome.** *GigaScience* 2019, **8**(10):giz115.
30. Kim S, Park J, Yeom SI, Kim YM, Seo E, Kim KT, Kim MS, Lee JM, Cheong K, Shin HS *et al.*: **New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication.** *Genome Biol* 2017, **18**(1):210.
31. Parker BC, Zhang WJCjoc: **Fusion genes in solid tumors: an emerging target for cancer diagnosis and treatment.** *Chin J Cancer* 2013, **32**(11):594.
32. Heyer EE, Deveson IW, Wooi D, Selinger CI, Lyons RJ, Hayes VM, O'Toole SA, Ballinger ML, Gill D, Thomas DMJNc: **Diagnosis of fusion genes using targeted RNA sequencing.** *Nature communications* 2019, **10**(1):1388.
33. Mertens F, Johansson B, Fioretos T, Mitelman FJNRC: **The emerging complexity of gene fusions in cancer.** *Nat Rev Cancer* 2015, **15**(6):371.
34. Agati G, Azzarello E, Pollastri S, Tattini MJPs: **Flavonoids as antioxidants in plants: location and functional significance.** *Plant science* 2012, **196**:67-76.
35. Petrucci E, Braidot E, Zancani M, Peresson C, Bertolini A, Patui S, Vianello AJIjoms: **Plant flavonoids –biosynthesis, transport and involvement in stress responses.** *International journal of molecular sciences* 2013, **14**(7):14950-14973.
36. Zhu J-KJC: **Abiotic stress signaling and responses in plants.** *Cell* 2016, **167**(2):313-324.
37. Yan J, He H, Fang L, Zhang AJB, communications br: **Pectin methylesterase31 positively regulates salt stress tolerance in *Arabidopsis*.** *Biochem Biophys Res Commun* 2018, **496**(2):497-501.
38. Love MI, Huber W, Anders SJGb: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome Biol* 2014, **15**(12):550.
39. Toyota M, Spencer D, Sawai-Toyota S, Jiaqi W, Zhang T, Koo AJ, Howe GA, Gilroy SJS: **Glutamate triggers long-distance, calcium-based plant defense signaling.** *Science* 2018, **361**(6407):1112-1115.
40. Forde BG, Lea PJJJoeb: **Glutamate in plants: metabolism, regulation, and signalling.** *J Exp Bot* 2007, **58**(9):2339-2358.
41. Seiler GJJFCR: **Utilization of wild sunflower species for the improvement of cultivated sunflower.** *Field Crops Research* 1992, **30**(3-4):195-230.

42. Castillo-Lorenzo E, Pritchard HW, Finch-Savage WE, Seal CEJPB: **Comparison of seed and seedling functional traits in native Helianthus species and the crop H. annuus (sunflower).** *Plant Biology* 2019, **21**(3):533-543.
43. Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy ASJNc: **A survey of the sorghum transcriptome using single-molecule long reads.** *Nature communications* 2016, **7**:11706.
44. Hoang NV, Furtado A, Mason PJ, Marquardt A, Kasirajan L, Thirugnanasambandam PP, Botha FC, Henry RJJBg: **A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing.** *BMC genomics* 2017, **18**(1):395.
45. Gupta I, Collier PG, Haase B, Mahfouz A, Joglekar A, Floyd T, Koopmans F, Barres B, Smit AB, Sloan SAJNb: **Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells.** *Nat Biotechnol* 2018, **36**(12):1197.
46. Minio A, Massonnet M, Figueroa-Balderas R, Vondras AM, Blanco-Ulate B, Cantu DJGG, Genomes, Genetics: **Iso-seq allows genome-independent transcriptome profiling of grape berry development.** *G3 (Bethesda)* 2019, **9**(3):755-767.
47. Wang D, Yang C, Dong L, Zhu J, Wang J, Zhang SJJopb: **Comparative transcriptome analyses of drought-resistant and-susceptible Brassica napus L. and development of EST-SSR markers by RNA-Seq.** *Journal of plant biology* 2015, **58**(4):259-269.
48. Jiang Z, Zhou X, Tao M, Yuan F, Liu L, Wu F, Wu X, Xiang Y, Niu Y, Liu FJN: **Plant cell-surface GIPC sphingolipids sense salt to trigger Ca²⁺ influx.** *Nature* 2019, **572**(7769):341-346.
49. You Z, Zhang Q, Peng Z, Miao XJPP: **Lipid droplets mediate salt stress tolerance in Parachlorella kessleri.** *Plant physiology* 2019, **181**(2):510-526.
50. Wei K, Zhong XJBpb: **Non-specific lipid transfer proteins in maize.** *BMC Plant Biol* 2014, **14**(1):281.
51. Pitzschke A, Datta S, Persak HJMp: **Salt stress in Arabidopsis: lipid transfer protein AZI1 and its control by mitogen-activated protein kinase MPK3.** 2014, **7**(4):722-738.
52. Borovskii GB, Stupnikova IV, Antipina AI, Vladimirova SV, Voinikov VKJBPB: **Accumulation of dehydrin-like proteins in the mitochondria of cereals in response to cold, freezing, drought and ABA treatment.** *BMC Plant Biol* 2002, **2**(1):5.
53. Zhang L, Hu W, Gao Y, Pan H, Zhang QJG, biology m: **A cytosolic class II small heat shock protein, PfHSP17. 2, confers resistance to heat, cold, and salt stresses in transgenic Arabidopsis.** *Genet Mol Biol* 2018, **41**(3):649-660.
54. Ruibal C, Castro A, Carballo V, Szabados L, Vidal SJBpb: **Recovery from heat, salt and osmotic stress in Physcomitrella patens requires a functional small heat shock protein PpHsp16. 4.** *BMC Plant Biol* 2013, **13**(1):174.
55. Feng X-H, Zhang H-X, Ali M, Gai W-X, Cheng G-X, Yu Q-H, Yang S-B, Li X-X, Gong Z-HJPP, Biochemistry: **A small heat shock protein CaHsp25. 9 positively regulates heat, salt, and drought stress tolerance in pepper (Capsicum annuum L.).** *Plant Physiol Biochem* 2019, **142**:151-162.

56. Selby K, Lindström M, Somervuo P, Heap JT, Minton NP, Korkeala HJAEM: **Important role of class I heat shock genes *hrcA* and *dnaK* in the heat shock response and the response to pH and NaCl stress of group I *Clostridium botulinum* strain ATCC 3502.** *Appl Environ Microbiol* 2011, **77**(9):2823-2830.
57. Zhao L, Gu R, Gao P, Wang GJPc, tissue, culture o: **A nonsymbiotic hemoglobin gene from maize, *ZmHb*, is involved in response to submergence, high-salt and osmotic stresses.** *Plant cell, tissue and organ culture* 2008, **95**(2):227-237.
58. Dordas CJPS: **Nonsymbiotic hemoglobins and stress tolerance in plants.** *Plant Sci* 2009, **176**(4):433-440.
59. Bai X, Long J, He X, Yan J, Chen X, Tan Y, Li K, Chen L, Xu HJSr: **Overexpression of spinach non-symbiotic hemoglobin in *Arabidopsis* resulted in decreased NO content and lowered nitrate and other abiotic stresses tolerance.** *Scientific reports* 2016, **6**:26400.
60. Szabados L, Savouré AJTips: **Proline: a multifunctional amino acid.** *Trends Plant Sci* 2010, **15**(2):89-97.
61. Székely G, Ábrahám E, Cséplő Á, Rigó G, Zsigmond L, Csiszár J, Ayaydin F, Strizhov N, Jásik J, Schmelzer EJTPJ: **Duplicated P5CS genes of *Arabidopsis* play distinct roles in stress regulation and developmental control of proline biosynthesis.** *The Plant Journal* 2008, **53**(1):11-28.
62. Chen L-S, Li P, Cheng LJP: **Effects of high temperature coupled with high light on the balance between photooxidation and photoprotection in the sun-exposed peel of apple.** *Planta* 2008, **228**(5):745.
63. Han S, Wang C-w, Jiang JJJopp: **Mitogen-activated protein kinase 6 controls root growth in *Arabidopsis* by modulating Ca²⁺-based Na⁺ flux in root cell under salt stress.** *J Plant Physiol* 2014, **171**(5):26-34.
64. de Freitas PAF, de Carvalho HH, Costa JH, de Souza Miranda R, da Cruz Saraiva KD, de Oliveira FDB, Coelho DG, Prisco JT, Gomes-Filho EJPcr: **Salt acclimation in sorghum plants by exogenous proline: physiological and biochemical changes and regulation of proline metabolism.** *Plant Cell Rep* 2019, **38**(3):403-416.
65. Wang P, Yang C, Chen H, Song C, Zhang X, Wang DJSr: **Transcriptomic basis for drought-resistance in *Brassica napus* L.** *Sci Rep* 2017, **7**:40532.
66. Langmead B, Salzberg SLJNm: **Fast gapped-read alignment with Bowtie 2.** *Nat Methods* 2012, **9**(4):357.
67. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding LJB: **VarScan: variant detection in massively parallel sequencing of individual and pooled samples.** *Bioinformatics* 2009, **25**(17):2283-2285.
68. Bairoch A, Apweiler R: **The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000.** *Nucleic Acids Res* 2000, **28**(1):45-48.
69. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic acids research* 2000, **28**(1):27-30.

70. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L: **InterPro: the integrative protein signature database.** *Nucleic acids research* 2008, **37**(suppl_1):D211-D215.
71. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403-410.
72. Li L, Stoeckert CJ, Jr., Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic genomes.** *Genome Res* 2003, **13**(9):2178-2189.

Table 1

Table 1. Statistics of identified fusion genes in *H. argophyllus* transcriptome.

| Type | Value |
|---|------------|
| Fusion gene number | 475 |
| Located in the same chromosome | 58 |
| Minimum distance between components of fusion gene (bp) | 14,029 |
| Maximum distance between components of fusion gene (bp) | 98,497,921 |

Figures

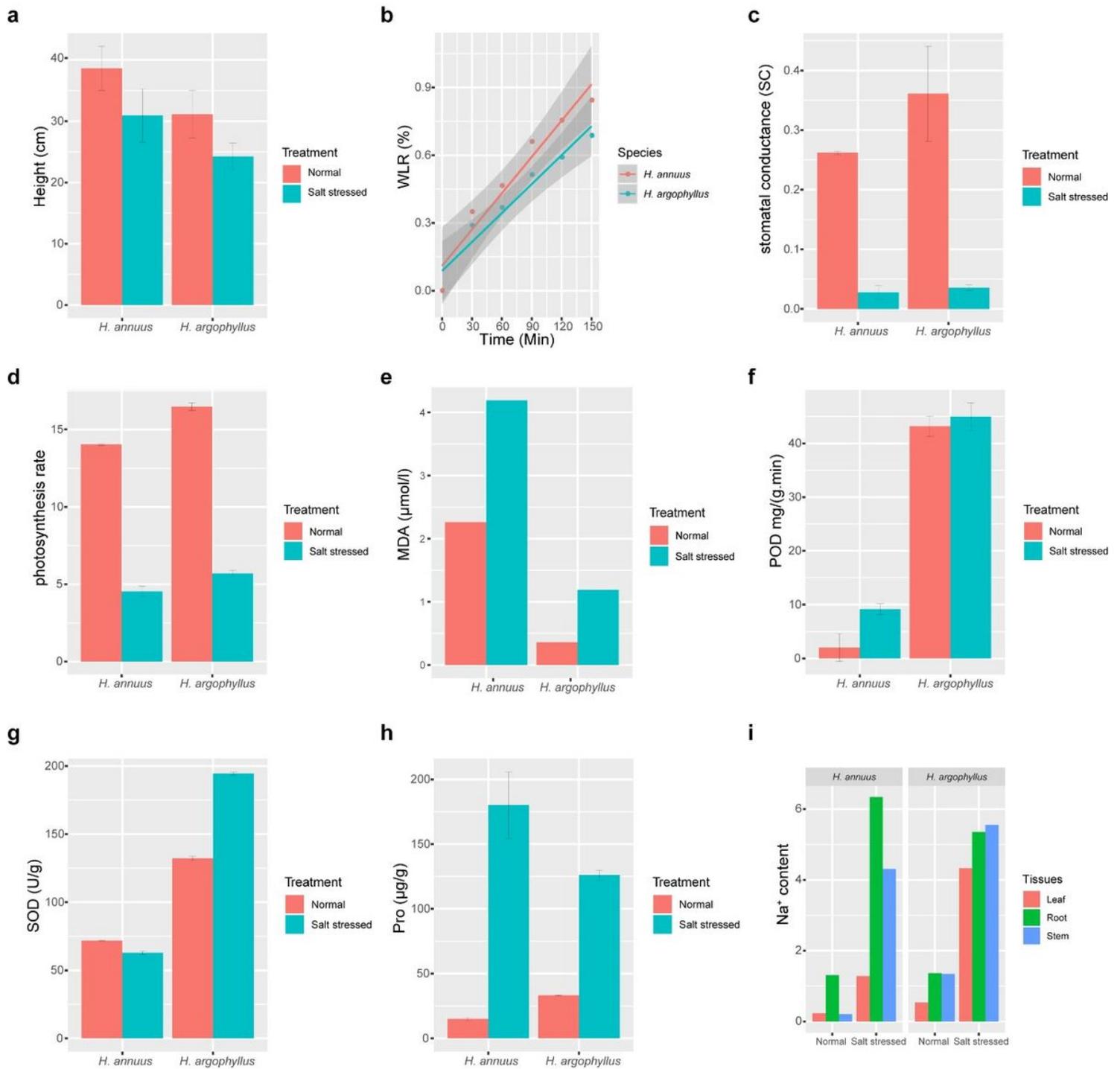


Figure 1

Physiological and morphological phenotypes of *H. argophyllum* and *H. annuus* under normal or salt stressed conditions. The differences of (a) plant height, (b) water loss rate (WLR), (c) stomatal conductance, (d) photosynthesis rate, the content of (e) MDA, (f) POD, (g) SOD, (h) proline and (i) Na^+ . The content of MDA was measured every 30 min after the treatment of salt.

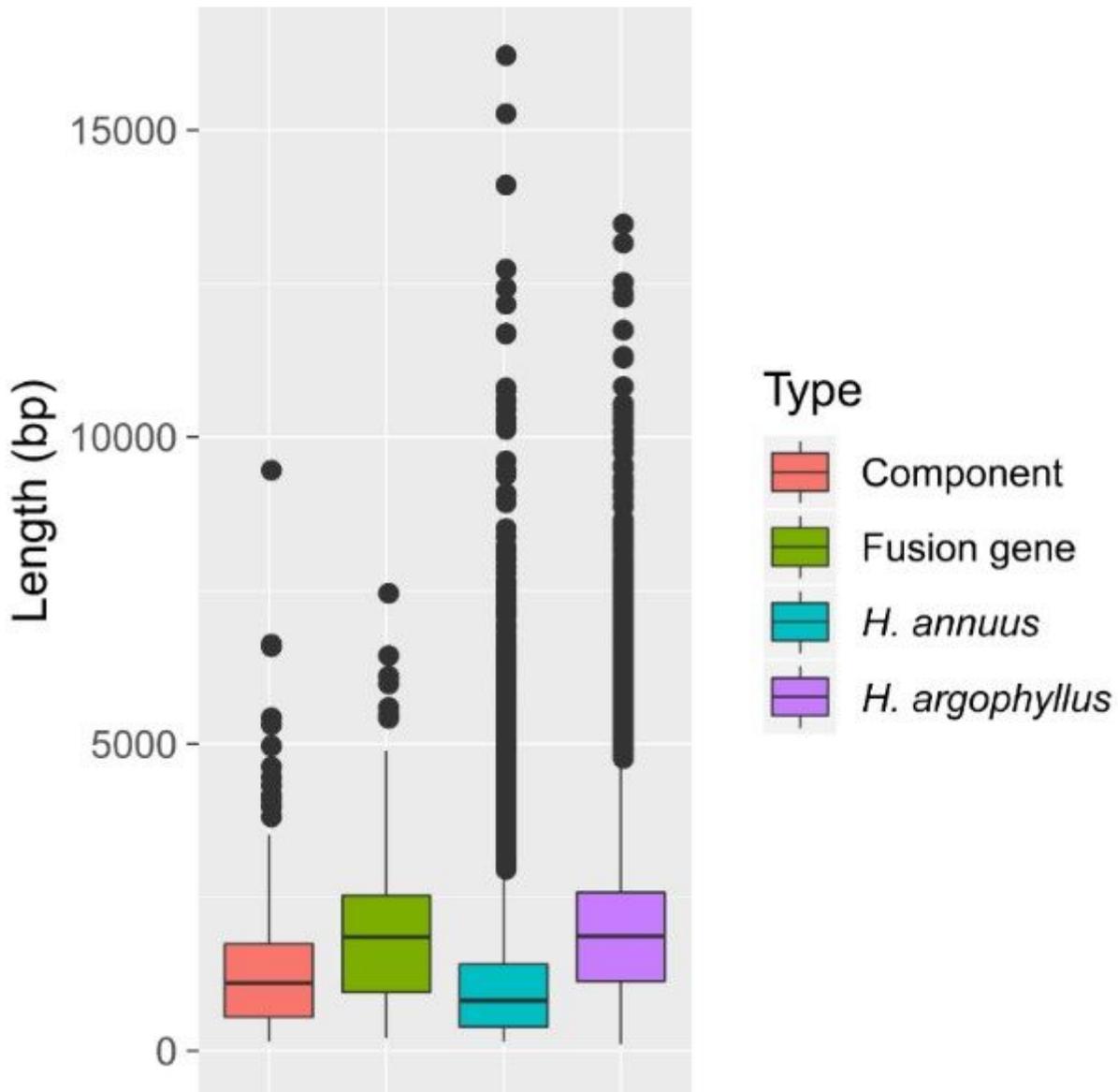


Figure 2

Length comparison between *H. argophyllus* transcriptome, the coding sequences of *H. annuus*, 'Fusion' genes of *H. argophyllus* and their components.

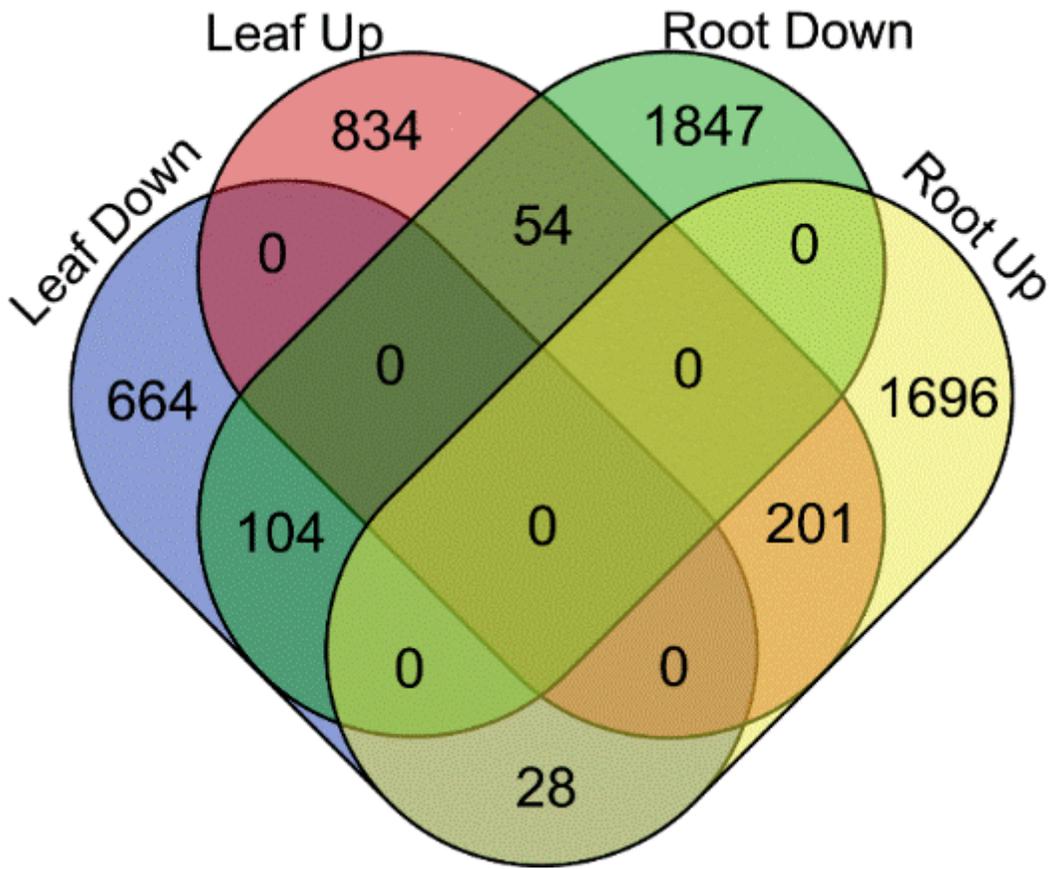


Figure 3

Venn diagram of differentially expressed genes in root and leaf under normal and salt stressed conditions. The differentially expressed genes were categorized into eight groups, genes (1) uniquely up-regulated in leaf, (2) uniquely up-regulated in root, (3) commonly up-regulated, (4) uniquely down-regulated in leaf, (5) uniquely down-regulated in root, (6) commonly down-regulated, (7) up-regulated in leaf but down-regulated in root and (8) vice versa.

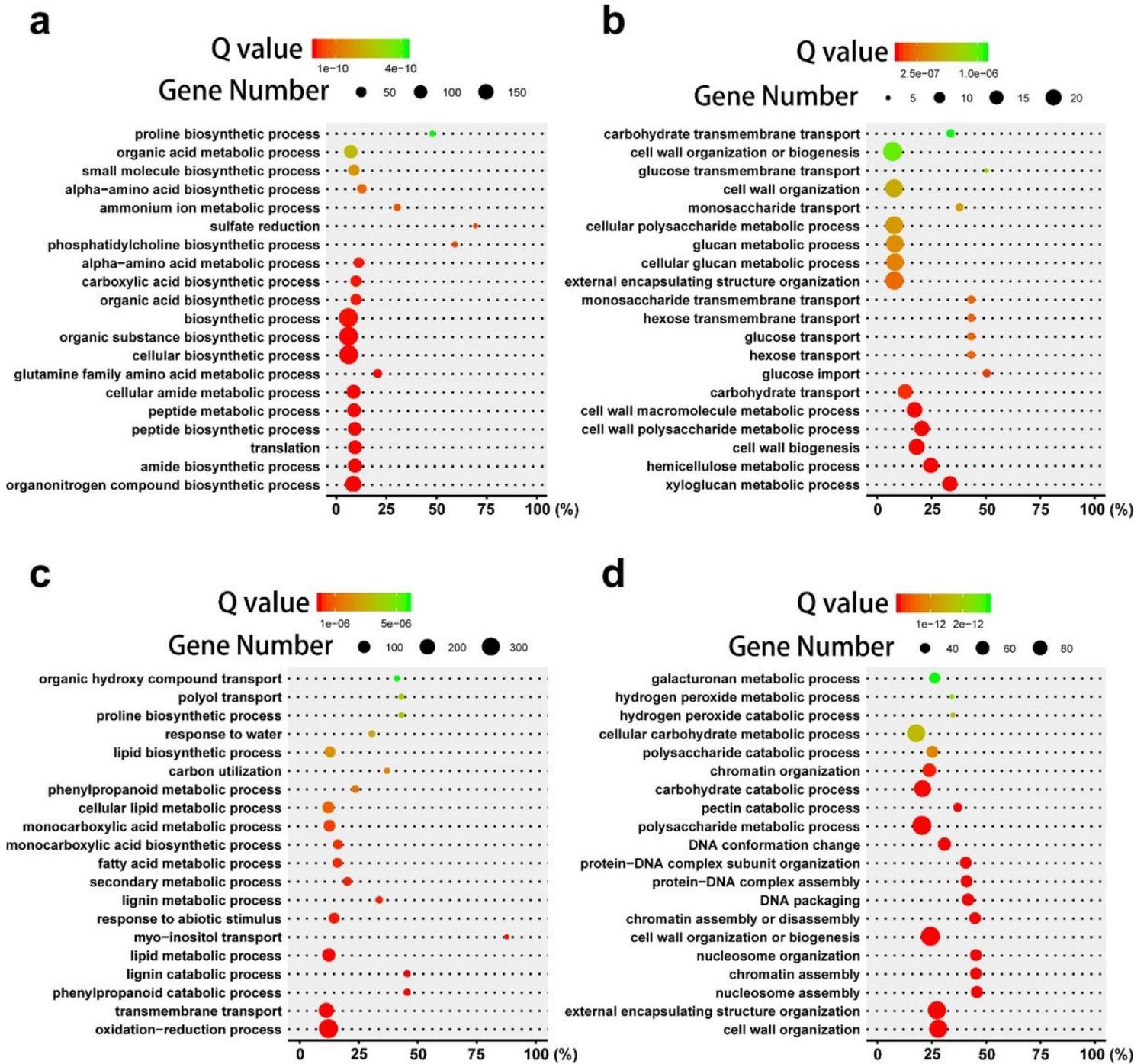


Figure 4

Go enrichment of up- and down-regulated DEGs in leaf and root. (a) Up-regulated DEGs in leaf; (b) Down-regulated DEGs in leaf; (c) Up-regulated DEGs in leaf; (d) Down-regulated DEGs in leaf.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigures.docx](#)

- [SupplementaryTables.xlsx](#)