

RESEARCH

Open Access



De novo transcriptome analysis and functional annotation of *Silybum Marianum* L. under drought stress with a focus on Silymarin synthesis and MAPK signaling pathways

Rahele Ghanbari Moheb Seraj¹, Asadollah Ahmadihah², Keyvan Esmailzadeh-Salestani^{3,4}, Vahid Shariati⁵, Mahdi Behnamian^{1*}, Neda Tariverdizadeh¹, Ali Emadi⁶ and Sara Dezhsetan⁷

Abstract

To evaluate the milk thistle transcriptome under drought stress in field conditions, irrigation was applied using a weighted method at three levels: 100% F.C every 2 days, 70% F.C every 4 days, and 40% F.C every 8 days. Sampling was performed after 8 days at the flowering stage. Plant leaves were collected for RNA-seq analysis, seeds for oily and methanolic extracts, and downstream analyses were performed. Since there was no annotated reference genome for this plant, the De novo Assembly method was implemented to assemble the transcriptome. Contigs were blasted against five databases: NT, NR, Uniprot, and protein databases of *Arabidopsis thaliana* and *Helianthus annuus*. A total of 9,517 genes (~73% of Uniprot genes) were common across all databases and selected for further analysis due to their comprehensive annotation. Then, DEGs were identified and functionally annotated using Gene Ontology (GO) analysis with the ShinyGO platform, biological pathway analysis through KEGG, and transcription factor identification via PlantTFDB. Next, silybinin content was measured using HPLC. Generally, the most repeated pathways in all treatments include the Biosynthesis of secondary metabolites and the MAPK signaling pathway. Also, most biological processes are related to the oxidation-reduction process, and response to stress, and most molecular functions are protein and mRNA binding. Our results indicate the active role of transcription factors ERF, C3H, and bHLH in drought stress tolerance. Silybin a and b showed that severe drought stress enhanced the accumulation of silybinin compared with seeds from the control. Eight differentially expressed genes (CYP86A1, CYP710A1, FATA2, LACS3, LOX2, PAL, PLA2-ALPHA, and PXG3) were used to validate the RNA-Seq data. qRT-PCR results confirmed strong consistency with the RNA-Seq findings. Finally, the genes involved in the silymarin pathway were identified, and their expression was determined through RNA-Seq data and compared with the silymarin contents.

Keywords RNA-Seq, Silybinin, Water shortage, Gene ontology, KEGG pathway, MAPK.

*Correspondence:
Mahdi Behnamian
mbehnamian@uma.ac.ir

Full list of author information is available at the end of the article



© The Author(s) 2025. **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/>.

Introduction

Milk thistle (*Silybum marianum* (L.) Gaertn) is an herbaceous plant, an annual or biennial, from the Asteraceae family, which has spread to Mediterranean climates [1]. This plant is commercially cultivated for medicinal purposes in dry rocky soils in Europe, China, Australia, Canada, and North and South America [2]. For more than 2000 years, the fruit of Milk thistle has served as a treatment for liver disorders [3]. The plant's fruit contains silymarin, a biologically active compound consisting of flavonolignans such as silybin, isosilybin, taxifolin, silydianin, and silychristin [4]. Silybin, comprising 60–70% of silymarin, is a key component with significant benefits [5]. This compound protects against a wide range of liver diseases, including toxic and viral hepatitis, cirrhosis, fatty liver, and damage to liver cells caused by various toxic substances [6]. Milk thistle has recently been characterized as having antioxidant, antidepressant, anticancer, cardioprotective, demulcent, digestive tonic, hepatoregenerative, hepatoprotective, immunostimulatory, and neuroprotective properties [7]. Therefore, as a result of the advantageous properties of the fruit of this medicinal plant, its cultivation is increasing on a global scale [1].

Plants are subjected to a range of environmental stressors, including temperature fluctuations, drought, salinity, and exposure to heavy metals, all of which have the potential to affect the growth and development of plants [8–10]. In arid and semi-arid regions, water scarcity is the primary limiting factor for plant yield [11]. Studying drought stress has gained significant importance due to the increasing global drought trend and water resource depletion. Additionally, the relative drought tolerance of milk thistle provides a valuable opportunity to investigate the molecular, phytochemical, and physiological mechanisms underlying plant adaptation to this stress. Drought is a physicochemical phenomenon that disrupts the structural integrity of biological molecules, encompassing proteins, nucleic acids, fatty acids, carbohydrates, hormones, nutrients, and ions [12]. Plant adaptation to drought involves intricate and systematic molecular mechanisms. These mechanisms encompass water channel proteins, stress-responsive proteins, transcription factors, and signaling pathways, all of which are pivotal constituents of the plant's stress response [13].

Transcriptome sequencing via RNA-seq technology offers a rapid, cost-effective, and efficient approach for unraveling the functionality of novel genes. Furthermore, it provides valuable insights into gene expression and regulation, particularly in plant species lacking complete genome sequences [14, 15]. In the absence of a good reference genome, the de novo assembly of an organism's transcriptome serves as a valuable technique to identify responsive genes or transcripts to different treatments. Additionally, it facilitates a comprehensive understanding

of the expression profile revealed by the candidate genes [16]. In a publication by Fulvio et al. [17], it was noted that more than 300 studies focusing on the medicinal properties of *S. marianum* have been documented in the PubMed database. This underscores the current scientific interest in milk thistle within international herbal plant research communities.

Although numerous studies have been conducted on *S. marianum*, few have focused on molecular investigations such as genetic diversity and gene expression analysis. Therefore, molecular studies on this plant appear to be essential [18, 19]. So far, only two transcriptome-related projects for *Silybum marianum* exist in the NCBI database. Only one (Accession: PRJNA659420), associated with this study, has examined the plant's transcriptomic response to drought stress. The milk thistle's genome documentation in the NCBI database indicates that a genome annotation file in GFF format has not yet been generated. As the plant's genome was not annotated to find and annotate differentially expressed genes (DEGs) and conduct enrichment analyses such as gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG), we sequenced the RNAs, performed de novo assembly of the reads, and created a reference transcriptome.

The main objectives of this study are as follows: [1] To perform a comprehensive transcriptome analysis of *S. marianum* to identify genes involved in the drought stress response; [2] To functionally annotate differentially expressed genes (DEGs) to elucidate their roles in drought-related biological pathways; [3] To quantify key metabolites in *S. marianum* and analyze their biosynthetic pathways; [4] To identify genes involved in the biosynthetic pathways of the selected metabolites and analyze their expression patterns using transcriptomic data [5]. To compare metabolite variations with gene expression profiles, to understand the interplay between molecular regulation and phytochemical production under drought conditions.

Results

Evaluation and cleaning of reads

Utilizing FastQC for initial data quality assessment, adapter sequences and low-quality nucleotides were excised via the Trimmomatic tool, applying a SLIDINGWINDOW: 5:20 average quality criterion and a minimum sequence length of 50 bp (MINLEN: 50). As summarized in Table 1, the vast majority of read pairs (99.98%) were retained after trimming, with only a minimal fraction of reads removed or present as singletons, indicating a high-quality dataset suitable for downstream analyses. Quality re-assessment with FastQC confirmed the efficacy of the trimming process.

Table 1 Post-trimming statistics of sequencing reads following quality trimming

Read types	Read Counts	Read percent (%)
Survived read pairs	32,515,278	99.98
Single forward reads	7974	0.02
Single reverse reads	5	< 0.01
Removed read pairs	3	< 0.01
Total input pairs	32,523,260	100

De Novo assembly and BLAST

High-quality, trimmed reads were subjected to assembly using the Trinity software, generating 159,295 contigs. These contigs underwent subsequent analysis through BLAST alignment against several databases, including the NT, NR, and Uniprot databases, as well as *Arabidopsis thaliana* and *Helianthus annuus* protein databases, to elucidate protein information. The analysis identified a total of 282,012 contigs in the NT database, 128,917 contigs in the NR database, 13,077 contigs in the Uniprot database, 73,744 contigs associated with *Arabidopsis*

thaliana, and 117,508 contigs related to *Helianthus annuus* databases.

The comparative analysis revealed that 9,517 genes were commonly found across all five examined databases, representing approximately 73% of the total genes identified in the Uniprot database. Furthermore, about 5% of the genes listed in Uniprot were also present in the four other databases, while the remaining genes were commonly present between two or three databases. For subsequent downstream analyses, we focused on contigs identified through the Uniprot database. This decision was based on several considerations: (i) Uniprot IDs have broad compatibility and overlap with contigs annotated in other databases, (ii) many commonly used tools and databases recognize and support Uniprot identifiers, and (iii) Uniprot provides more comprehensive and well-curated annotations, facilitating more reliable functional interpretation and biological insights (Fig. 1).

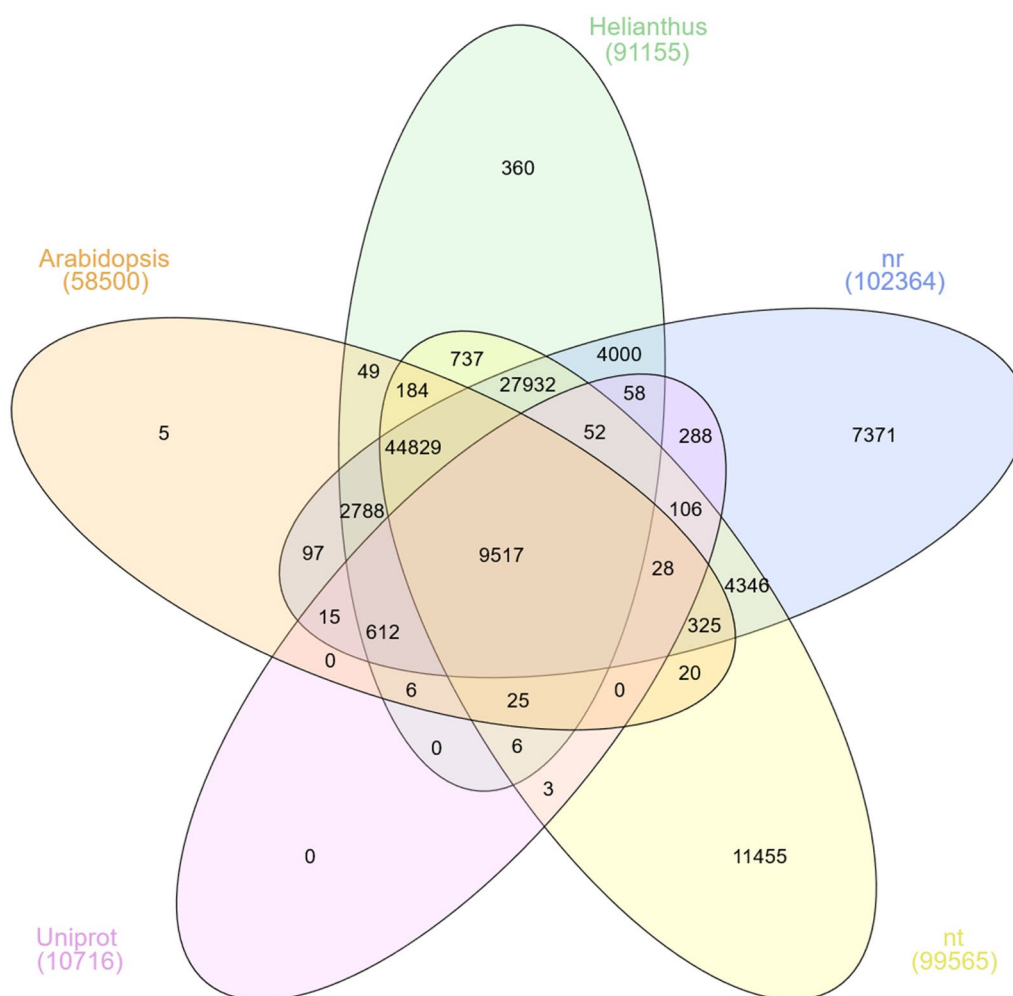


Fig. 1 Venn diagram illustrating the comparison between genes in five databases

Gene ontology classification

A GO analysis was carried out using genes from Uniprot, concentrating on biological processes, molecular function, cellular components, and protein class. The study involved selecting and listing 10 significant GO terms for each category that had a high number of genes (Sup1). The predominant biological processes identified were metabolic and cellular processes, response to stimuli, and stress. The prevalent molecular functions included catalytic activities, binding, transferase, and hydrolase activities. Enriched cell components encompassed intracellular organelles, cytoplasm, chloroplasts, and membranes. Genes obtained from Uniprot were frequently classified into protein classes such as enzymes, transporters, and signal receptors.

Identification of functional KEGG pathways

The pathways of Uniprot genes were identified and listed in sup2, including the number of genes involved in each pathway alongside KO identifications. An ensemble of 4,243 genes participates in 384 distinct pathways. There were two pathways of paramount importance in the milk thistle plant. The first pathway is the flavonoid biosynthesis pathway, which facilitates silymarin production. This pathway engages eight Uniprot genes. The second pathway is the biosynthesis pathway of unsaturated fatty acids, comprising linoleic acid, palmitic acid, oleic acid, etc. This pathway similarly includes eight Uniprot genes.

The protein families associated with these genes are systematically categorized into three groups (metabolism, genetic information processing, and signaling in juxtaposition with cellular processes) using the KEGG web server (Sup3). The metabolism category encompasses a diverse array of subgroups, including enzymes, protein phosphatases and associated proteins, protein kinases, glycosyltransferases, lipid and lipopolysaccharide biosynthesis proteins, peptidases and inhibitors, prenyltransferases, peptidoglycan biosynthesis and degradation proteins, cytochrome p450, amino acid related enzymes, and photosynthesis proteins.

The genetic information processing group was constitutive of subcategories such as transcription factors, messenger RNA biogenesis, transcription machinery, spliceosome, translation factors, ribosome biogenesis, transfer RNA biogenesis, membrane trafficking, ubiquitin system, chaperones and folding catalysts, DNA replication proteins, proteasome, chromosome and associated proteins, mitochondrial biogenesis, and DNA repair and recombination proteins. The signaling and cellular process group included subgroups of transporters, secretion system, cilium and related proteins, exosome, cytoskeleton proteins, G protein-coupled receptors, prokaryotic defense system, pattern recognition receptors, ion channels, GTP-binding proteins, CD molecules,

glycosylphosphatidylinositol (GPI)-anchored proteins, glycosaminoglycan binding proteins, and domain-containing proteins. Each subgroup underwent further subdivision, with specific gene names in the detailed components listed below. Notably, these genes corresponded with those indexed by Uniprot, underpinning their significance in the identified pathways and protein families.

Identification of DEGs

In three different treatments (70% F.C vs. F.C, 40% F.C vs. F.C, and 40% F.C vs. 70% F.C), DEGs (both up- and down-regulated) were identified. The criteria for identifying these genes were $-2 \leq \text{Log Fold Change (Log FC)} \leq +2$ and $P\text{-Value} \leq 0.01$. In the 70% F.C vs. F.C comparison, we found 4,593 up-regulated genes (Log FC ranging from +2 to +18) and 6,741 down-regulated genes (Log FC ranging from -2 to -20). In the 40% F.C vs. F.C comparison, there were 3530 up-regulated genes (Log FC range: +2 to +15) and 4793 down-regulated genes (Log FC range: -2 to -18). In the 40% F.C vs. 70% F.C comparison, we identified 5936 up-regulated genes (Log FC range: +2 to +20) and 5732 down-regulated genes (Log FC range: -2 to -18).

Comparison of gene expression in three irrigation treatments of 100%, 70% and 40% of field capacity

An analytical comparison was conducted to assess the expression levels of 159,295 genes in three different treatments. The gene expression levels ranged from Log FC zero to 690,587. To facilitate a better comparison of gene expression in the three treatments, genes with expression levels of $|\text{Log}_2 \text{FC}| \geq 2$ were excluded. The analysis proceeded with a revised dataset comprising 21,151 genes. The comparative evaluation of gene expressions under three different degrees of drought stress was visualized employing a heatmap, wherein the gene expression data were subjected to $\text{Log}_{10} \text{FC}$ to enhance clarity and interpretability. The analysis demonstrated that genes subjected to irrigation treatment at 70% of field capacity had a higher expression level than those under 100% and 40% field capacity treatments (Fig. 2), suggesting differential gene expression responses under varied irrigation regimes.

Functional KEGG pathways identification of differential expression genes

Following the identification of up- and down-regulated genes in each pairwise comparison (70% F.C vs. F.C, 40% F.C vs. F.C, and 40% F.C vs. 70% F.C), the corresponding activity pathways for these gene groups were determined. In the 70% vs. F.C pairwise comparison, 123 different activity pathways were identified for up-regulated genes, with 29 pathways found to be significant at

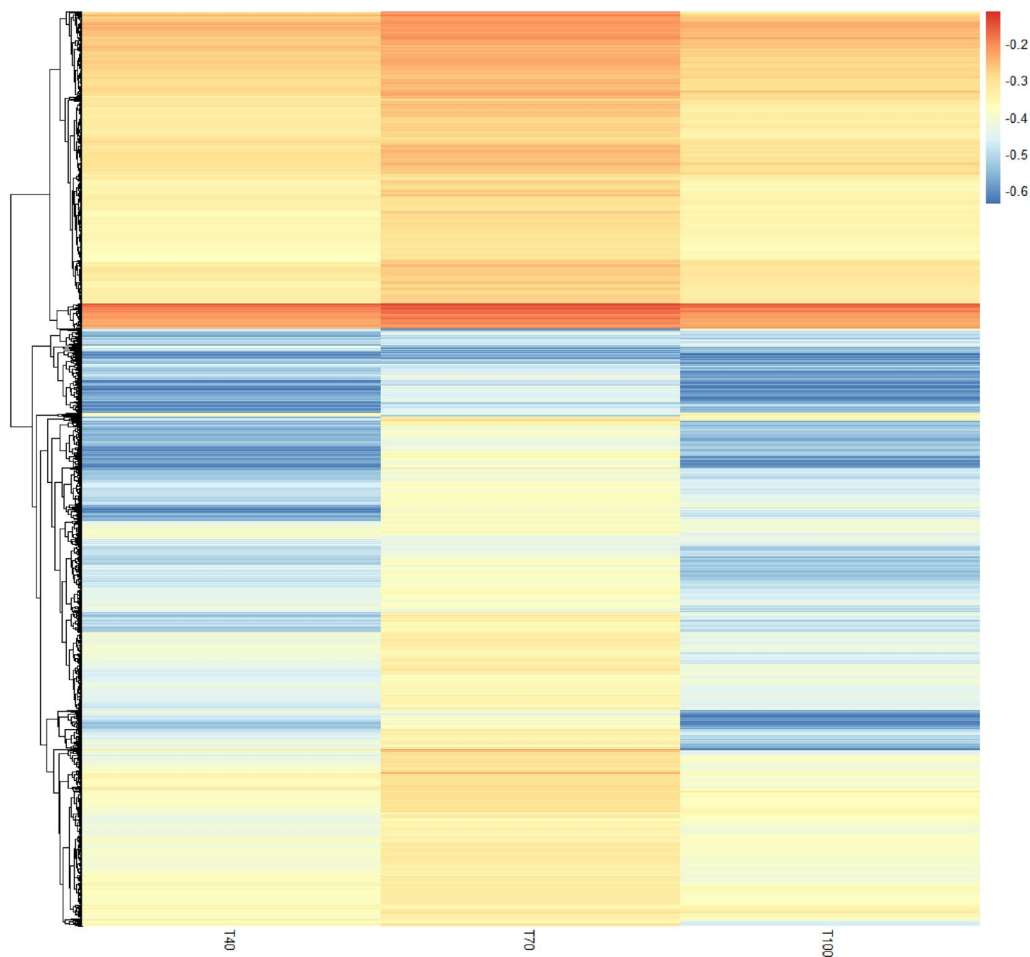


Fig. 2 Comparison of gene expression in three irrigation treatments at 100, 70 and 40% of field capacity

the ≤ 0.01 level. For down-regulated genes in the same comparison, 122 activity pathways were identified, with 30 pathways showing p -values ≤ 0.01 . In the 40% vs. F.C pairwise comparison, 125 pathways were identified for up-regulated genes, with 28 pathways significant at the ≤ 0.01 level, while for down-regulated genes, 125 activity pathways were identified, with 26 pathways significant at the p -value of ≤ 0.01 . In the 40% vs. 70% pairwise comparison, 124 different pathways were identified for up-regulated genes, with 26 pathways having a p -value of ≤ 0.01 . For down-regulated genes in the same comparison, 124 activity pathways were identified, with 46 having a p -value ≤ 0.01 . Subsequently, for each gene group, the 10 activity pathways with a p -value ≤ 0.01 and the highest number of genes were selected and listed in Table 2.

Gene ontology classification of differential expression genes

This annotation systematically classified the DEGs into three principal categories based on their functional characteristics: biological process, molecular function, and cellular component. Six different gene groups were

identified based on their expression profiles (up-regulated and down-regulated) across three pairwise comparisons of 70% F.C vs. F.C, 40% F.C vs. F.C, and 40% F.C vs. 70% F.C. The functions that were statistically significant ($P \leq 0.01$) and encompassed the highest number of genes within these groups were presented in Table 3.

Metabolic pathway analysis

The analysis linking significant DEGs with KEGG pathways identified the MAPK signaling pathway as significantly implicated (Fig. 3). This pathway began when cell membranes detected drought-related signals, subsequently regulated by PRY/PRL as soluble receptors for abscisic acid (ABA) in an ABA-dependent mechanism. The 2 C-type protein phosphatases (PP2C), responsible for the deactivation of SNF1-related protein kinases 2 (SnRK2) through dephosphorylation, were suppressed in the presence of ABA. This investigation revealed a significant upregulation in the PRY/PRL and PP2C gene expressions. The activation of SnRK2 plays a pivotal role against biotic stressors such as salinity and drought by stimulating plant defensive responses and modulating

Table 2 KEGG pathway analysis of milk Thistle Up- and Down-regulated genes under three levels of water stress

Up-regulated genes in 70 vs 100% F.C treatments			Down-regulated genes in 70 vs 100% F.C treatments		
KEGG Pathway	GenesNo.	FDR adj.	KEGG Pathway	GenesNo.	FDRadj.
Metabolic pathways	544	1.70E-45	Metabolic pathways	681	9.06E-41
Biosynthesis of secondary metabolites	264	3.41E-20	Biosynthesis of secondary metabolites	315	7.39E-15
Carbon metabolism	77	1.97E-08	Carbon metabolism	79	3.91E-04
Biosynthesis of amino acids	70	1.39E-07	Biosynthesis of amino acids	67	5.95E-03
Ribosome	69	4.79E-03	Spliceosome	64	7.84E-05
Protein processing in endoplasmic reticulum	60	8.15E-07	Protein processing in endoplasmic reticulum	60	3.37E-03
MAPK signaling pathway	53	7.44E-10	Oxidative phosphorylation	54	3.04E-04
Starch and sucrose metabolism	49	4.36E-06	Starch and sucrose metabolism	53	8.21E-04
Spliceosome	47	4.59E-04	Glyoxylate and dicarboxylate metabolism	42	3.64E-07
Plant-pathogen interaction	42	8.88E-04	mRNA surveillance pathway	38	4.18E-03
Up-regulated genes in 40 vs 100% F.C treatments			Down-regulated genes in 40 vs 100% F.C treatments		
Metabolic pathways	358	1.05E-17	Metabolic pathways	493	6.23E-28
Biosynthesis of secondary metabolites	184	4.75E-10	Biosynthesis of secondary metabolites	251	4.44E-15
Carbon metabolism	54	1.76E-04	Carbon metabolism	101	5.47E-16
Biosynthesis of amino acids	43	6.81E-03	Biosynthesis of amino acids	73	8.16E-08
Starch and sucrose metabolism	32	6.09E-03	Spliceosome	58	9.59E-07
Spliceosome	32	2.99E-03	Protein processing in endoplasmic reticulum	56	4.51E-05
Glyoxylate and dicarboxylate metabolism	31	1.06E-04	Oxidative phosphorylation	42	7.15E-04
Glycolysis/Gluconeogenesis	21	6.52E-03	Glycolysis/Gluconeogenesis	39	1.77E-05
RNA degradation	20	9.60E-03	Glyoxylate and dicarboxylate metabolism	37	4.41E-08
Glycine, serine and threonine metabolism	19	3.49E-03	mRNA surveillance pathway	35	2.99E-04
Up-regulated genes in 40 vs 70% F.C treatments			Down-regulated genes in 40 vs 70% F.C treatments		
Metabolic pathways	624	1.22E-41	Metabolic pathways	685	1.42E-63
Biosynthesis of secondary metabolites	297	6.14E-17	Biosynthesis of secondary metabolites	347	1.65E-32
Carbon metabolism	68	1.98E-03	Ribosome	141	3.81E-19
Oxidative phosphorylation	55	1.10E-05	Carbon metabolism	131	4.40E-24
Starch and sucrose metabolism	54	4.19E-05	Biosynthesis of amino acids	114	1.79E-19
Glycolysis/Gluconeogenesis	40	2.36E-04	Protein processing in endoplasmic reticulum	103	2.57E-19
Ubiquitin mediated proteolysis	40	4.32E-04	RNA transport	66	4.63E-09
MAPK signaling pathway	38	4.75E-04	Glycolysis/Gluconeogenesis	62	5.12E-13
Glyoxylate and dicarboxylate metabolism	37	2.39E-06	Oxidative phosphorylation	59	1.27E-07
Fatty acid metabolism	27	4.78E-05	MAPK signaling pathway	54	5.22E-08

transcriptional factors to increase the biosynthesis of functional proteins. The SnRK2 gene showed both up- and downregulation in our study.

The MAPK signaling cascade followed a classical pattern where MAPKKK was activated by plasma membrane receptors, which subsequently transmitted signals downstream. This led to the activation of MAPKK through the phosphorylation of the S/T-XXXXX-S/T motif, followed by MAPKK catalyzing the phosphorylation of the TXY motif in MAPK. Ultimately, MAPK activated a series of downstream enzymes, kinases, transcription factors, and additional response elements, translating extracellular environmental signals into cellular responses. However, our data indicated a reduction in MKK3 expression, whereas MPK7/14 expression varied, showing both increases and decreases.

Additionally, our findings underscored the critical role of ethylene, a hormone significantly influenced under drought stress conditions (Fig. 4). Ethylene interaction

with its receptors was mediated by a copper co-factor, presumably supplied by the copper transporters RAN1 and RTE1, which were found to be upregulated in our study—this interaction inactivated receptor functionality. The histidine kinase ETR1/ERS1 (an ethylene receptor) activated CTR1 in ethylene’s absence. CTR1 acted as a negative regulator of the MAPKK SIMKK and MPK3/6 in Milk thistle, with its inactivation upon ethylene’s presence, thereby relieving SIMKK from inhibition and allowing for the activation of MAPKs. This cascade stimulated the expression of ethylene-responsive genes through direct activation of EIN2 and EIN3 or other intermediaries.

Transcription factors identification

In the study of Milk thistle subjected to drought conditions, a comprehensive analysis of its transcriptome revealed the identification of 33 different transcription factors. These transcription factors and their expressions,

Table 3 Gene ontology analysis of milk Thistle Up- and Down-regulated genes under three levels of water stress

		Up-regulated genes in 70 vs. 100% F.C treatments		Down-regulated genes in 70 vs. 100% F.C treatments	
	Gene Ontology	Genes No.	FDR adj.	Gene Ontology	Genes No. FDR adj.
Biological process	Protein phosphorylation	181	4.14E-13	Protein phosphorylation	250 8.77E-17
	Response to cadmium ion	144	4.23E-34	Oxidation-reduction process	149 2.48E-08
	Response to abscisic acid	112	2.02E-13	Embryo development ending in seed dormancy	145 2.07E-10
	Response to cold	111	6.53E-18	Response to abscisic acid	139 1.36E-12
	Response to water deprivation	109	1.60E-19	Signal transduction	133 2.51E-10
Molecular function	Protein binding	766	7.35E-65	Protein binding	1013 2.35E-71
	mRNA binding	308	1.85E-51	mRNA binding	288 2.04E-21
	ATP binding	158	1.92E-12	ATP binding	242 3.93E-22
	Zinc ion binding	124	2.11E-08	Protein kinase activity	178 1.84E-17
	Protein kinase activity	123	7.50E-12	Zinc ion binding	153 6.25E-07
Cellular component	Nucleus	1429	3.30E-23	Nucleus	2032 1.28E-37
	Cytoplasm	895	7.35E-65	Chloroplast	1176 2.00E-52
	Chloroplast	894	6.03E-51	Cytoplasm	1060 1.01E-42
	Cytosol	876	5.70E-113	Cytosol	952 1.41E-65
	Plasma membrane	712	9.08E-60	Plasma membrane	881 8.82E-51
		Up-regulated genes in 40 vs. 100% F.C treatments		Down-regulated genes in 40 vs. 100% F.C treatments	
Biological process	Protein phosphorylation	163	3.84E-16	Protein phosphorylation	183 3.09E-12
	Oxidation-reduction process	87	1.39E-05	Embryo development ending in seed dormancy	123 2.74E-12
	Protein ubiquitination	84	2.39E-05	Oxidation-reduction process	115 1.89E-07
	Embryo development ending in seed dormancy	76	8.89E-05	Defense response to bacterium	103 7.90E-13
	Protein autophosphorylation	62	1.47E-14	Response to salt stress	93 9.89E-09
Molecular function	Protein binding	539	2.12E-33	Protein binding	759 2.40E-56
	mRNA binding	177	6.22E-17	mRNA binding	258 8.72E-30
	ATP binding	142	5.00E-15	ATP binding	203 1.60E-24
	Protein kinase activity	117	3.97E-16	RNA binding	132 5.63E-16
	Zinc ion binding	109	1.37E-09	Zinc ion binding	129 1.23E-08
Cellular component	Nucleus	1197	1.33E-29	Nucleus	1528 1.22E-30
	Chloroplast	639	2.35E-25	Cytoplasm	866 3.63E-50
	Cytoplasm	610	2.03E-27	Chloroplast	865 7.38E-38
	Cytosol	527	1.87E-35	Cytosol	798 4.99E-77
	Plasma membrane	519	2.49E-35	Plasma membrane	654 2.09E-38
		Up-regulated genes in 40 vs. 70% F.C treatments		Down-regulated genes in 40 vs. 70% F.C treatments	
Biological process	Protein phosphorylation	228	1.19E-16	Response to cadmium ion	201 3.60E-54
	Oxidation-reduction process	133	1.39E-07	Protein phosphorylation	182 5.94E-08
	Embryo development ending in seed dormancy	118	7.14E-07	Oxidation-reduction process	143 4.12E-11
	Response to abscisic acid	105	1.26E-06	Embryo development ending in seed dormancy	137 9.72E-13
	Protein ubiquitination	106	2.03E-03	Response to water deprivation	121 3.81E-19
Molecular function	Protein binding	839	4.81E-49	mRNA binding	404 6.93E-75
	mRNA binding	270	2.99E-23	ATP binding	188 2.24E-14
	ATP binding	199	1.04E-15	RNA binding	153 1.53E-18
	Protein kinase activity	161	7.79E-19	Protein kinase activity	114 7.07E-06
	Protein serine/threonine kinase activity	144	6.09E-21	Structural constituent of ribosome	105 5.62E-15
Cellular component	Nucleus	1808	2.29E-34	Nucleus	1668 3.33E-22
	Chloroplast	1040	2.00E-46	Cytosol	1168 1.31E-180
	Cytoplasm	963	4.30E-43	Cytoplasm	1150 1.15E-99
	Cytosol	823	6.02E-53	Chloroplast	1022 6.51E-49
	Plasma membrane	800	6.42E-50	Mitochondrion	663 3.34E-04

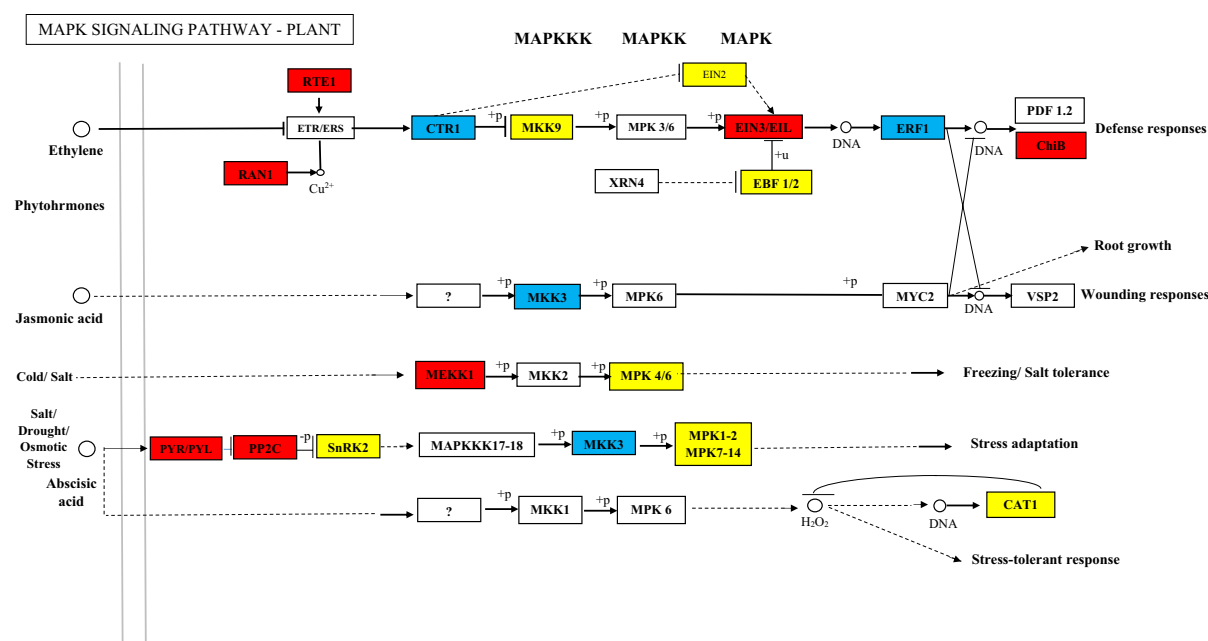


Fig. 3 MAPK signaling pathway map of Milk thistle up- and down-regulated genes under drought stress condition. The up-regulated genes are in red, the down-regulated genes are in blue, and the genes that were in both up- and down-regulated gene categories are in yellow

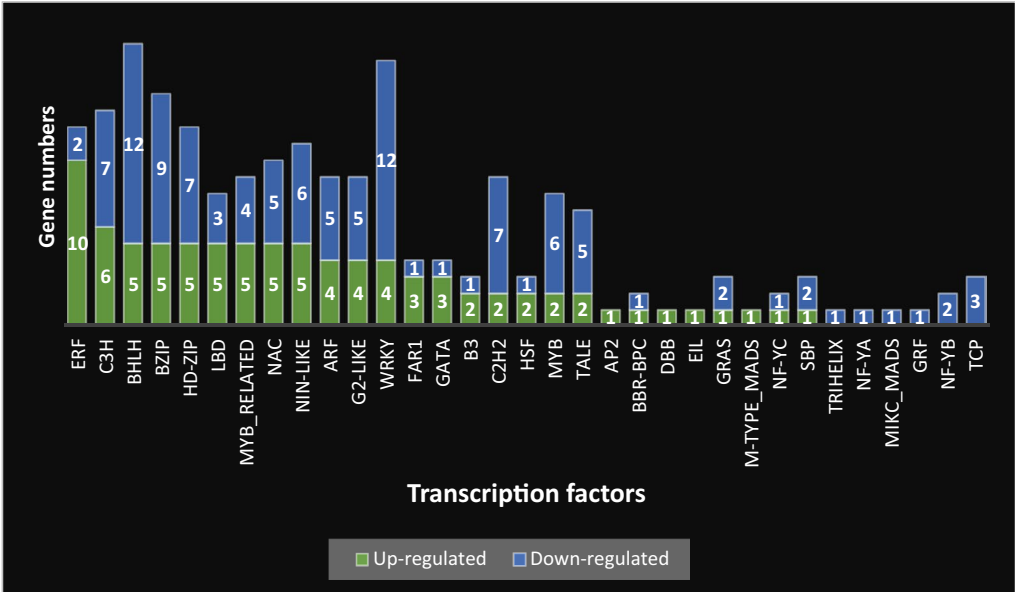


Fig. 4 Transcription factors identified in milk thistle transcriptome under water stress treatment. The green boxes correspond to up-regulated genes, and the blue boxes correspond to down-regulated genes. The numbers inside the boxes represent the number of genes

quantified as either up-regulated or down-regulated, were cataloged in Fig. 4. The ERF transcription factor category showed the most substantial increase in expression, with 10 factors being up-regulated, closely followed by the C3H with six up-regulated factors. Transcription factors BHLH, BZIP, HD-ZIP, LBD, MYB_RELATED, NAC, and NIN-LIKE demonstrated comparable levels of expression change. Among down-regulated transcription factors, the BHLH and WRKY categories showed reduced expression, each with 12 factors, followed by the BZIP category with nine down-regulated factors. The number of down-regulated transcription factors in the Milk thistle transcriptome was significantly higher than up-regulated factors. Among the transcription factors analyzed, several transcription factors, such as TRIHELIX, NF-YA, MIKC-MADS, GRE, NF-YB, and TCP, were exclusively found to be down-regulated, with

no corresponding increase in any transcription factors within these categories.

Trihelix transcription factors, characterized by a unique trihelix (helix-loop-helix-loop-helix) structure, are plant-specific and bind to GT elements in gene promoters. They play key roles in light response, growth regulation, organ development, and stress responses. This family is divided into five subgroups, with structural variations contributing to their functional diversity. Trihelix genes are involved in processes such as embryonic development, floral morphogenesis, and stomatal regulation, highlighting their importance in plant physiological regulation [20, 21]. The downregulation of the Trihelix transcription factor in our experiment may be related to its role as a negative regulator in stress responses. Some members of this family, such as ASIL1, repress growth- or stress-related genes, and under drought conditions, the suppression of such factors could be part of the plant's strategy to enhance stress tolerance [22].

On the other hand, MIKC-MADS transcription factors are a subgroup of MADS proteins in plants that regulate the growth and development of various organs such as roots, stems, flowers, and fruits. These factors bind to two specific sites on DNA by forming heterotetramers, causing DNA looping, which increases the accuracy of target gene recognition and plays a vital role in biological functions and agriculture [23]. The decreased expression of MIKC-MADS in your experiment is likely a response to drought stress. The plant reduces the expression of developmental transcription factors like MIKC-MADS to conserve energy, slow down growth, and redirect resources toward stress adaptation [24]. Together, the reduced expression of Trihelix and MIKC-MADS transcription factors underscores a coordinated regulatory adjustment by plants to optimize survival under adverse conditions.

Expression of DEGs using qRT-PCR

An analysis was conducted to evaluate the differential expression of targeted genes, focusing on both up-regulated and down-regulated genes under different irrigation conditions. This study specifically examined the expression profiles of four up-regulated genes, including PXG3, LOX2, CYP710A1, and Palmitoyl-protein thioesterase (AT5G47330), and four down-regulated genes, including FATA2, CYP86A1, LACS3, and PLA2-ALPHA under three different irrigation levels (three replicates).

Our findings revealed that PXG3 expression significantly increased by 65.119 and 59.302-fold under 40% F.C and 70% F.C irrigation levels, respectively, compared to the F.C condition. Palmitoyl-protein thioesterase (PAL) significantly increased expression at 70% F.C compared to the control (154.879 times). In contrast, there was a non-significant increase at 40% F.C (1.693 times)

compared to the control. Expression of LOX2 was significantly increased in both water-stressed conditions (40% F.C and 70% F.C), with a pronounced expression at 70% F.C (1917.49-fold increase) and 40% F.C (60.129-fold increase). Expression of CYP710A1 was diminished at both 40% and 70% F.C levels when compared to the control, with 0.419- and 0.015-times changes at expression levels, respectively.

The expression of PLA2-ALPHA increased at 40% F.C (4.332 times) and 70% F.C (3.972 times) compared to the control, but there was no statistical significance. The LACS3 demonstrated a decreased expression at both 40% and 70% F.C (0.044 and 0.068, respectively) compared to the control condition. The CYP86A1 showed a significant up-regulation at 40% F.C and 70% F.C compared to the control, higher at 70% F.C (519.147-fold increase) than 40% F.C (9.952). The expression of the FATA2 gene showed a non-significant increase at 40% F.C (2.585 times) and a reduction at 70% F.C (0.343 times) compared to F.C (Fig. 5).

Comparison of Silybin content in different levels of drought stress

To quantify silybin a and b in plant extracts, their concentrations were calculated using regression equations for different drought stress treatments, as shown in Table 4. Our results demonstrated a statistically significant difference ($p \leq 0.05$) in the silybin a and b levels between drought stress treatments and the control. Furthermore, silybin a and b levels increased significantly with escalating drought stress. Specifically, the severe stress treatment (40% F.C) caused a significant increase in silybin a and b levels (67.30 and 98.92 mg/g Grain DW, respectively) compared to the control and the 70% F.C treatment.

Interaction of gene expression with Silymarin levels

The study investigated the expression of 19 genes directly engaged in silymarin biosynthesis. Under conditions of mild drought stress (represented as 70% F.C), an upregulation in the expression of myriad genes was observed, with the exceptions being caffeine-CoA 3-O-methyltransferase (CCOMT) and Peroxidase, both of which showed a decrease in expression levels. As drought stress intensity increased to 40% F.C, alterations in gene expression were recorded. Only two genes demonstrated an increase in expression, namely Flavonoid 3'-hydroxylase (F3'H) (164) and Peroxidase (15.14), under severe drought stress conditions, whereas a significant reduction in expression was observed for genes including Tyrosine ammonia-lyase (TAL) and Phenylalanine ammonia-lyase (PAL) (7504.8), Cinnamate 4-hydroxylase (C4H) (438.9), p-Coumarate 3-hydroxylase (C3H) and p-Coumaroyl ester 3'-hydroxylase (C3'H) (717.39), Caffeic acid 3-O-methyltransferase

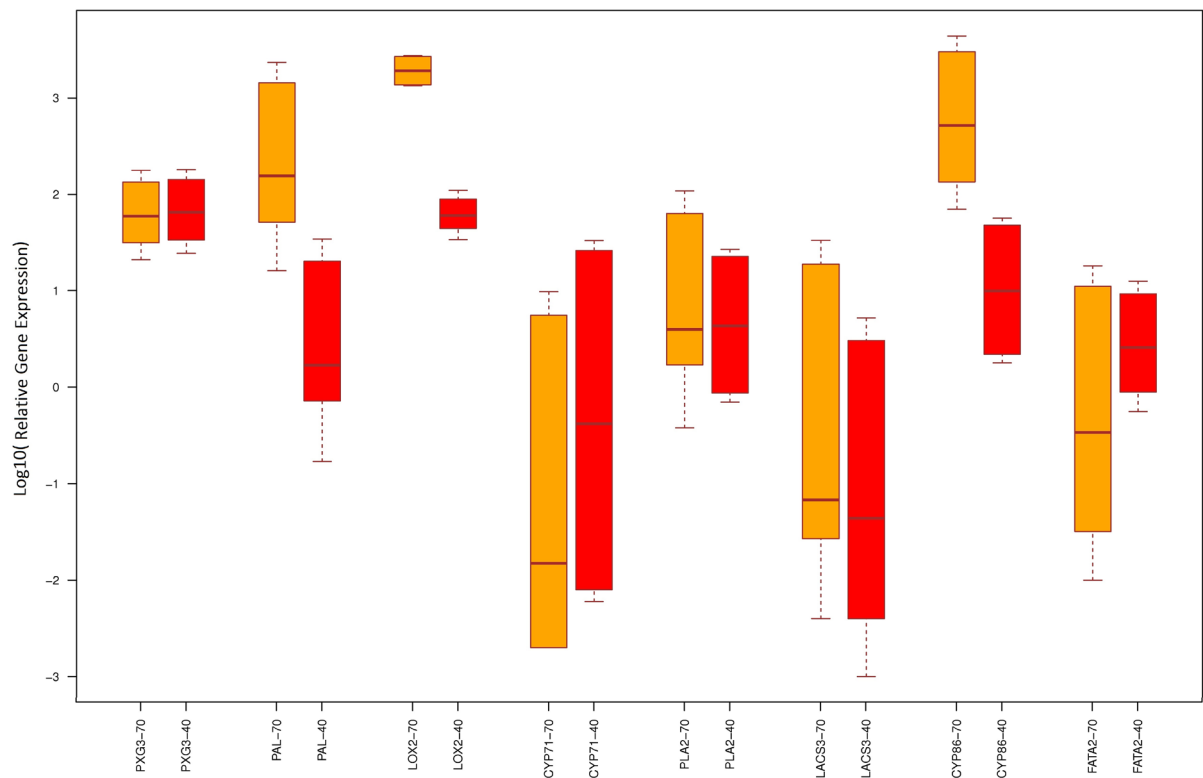


Fig. 5 Relative expression analysis of 8 hub genes (4 up-regulated and 4 down-regulated genes) at 70%FC (orange color) and 40% FC (red color) versus FC treatment (field capacity). The relative expression of genes is shown based on Log10. Because the differences in expression between the different genes were so large, logarithms were used to make them easier to display and compare better

Table 4 Silybin a and b values calculated for different levels of water stress

Water stress treatment levels	Silybin a (mg/g Grain DW)	Silybin b (mg/g Grain DW)
Field capacity	41.23 ^b	58.18 ^c
70% Field capacity	41.98 ^b	64.28 ^b
40% Field capacity	67.30 ^a	98.92 ^a

(COMT) (182), Caffeoyl-CoA 3-O-methyltransferase (CCOMT) [5], 4-Coumaric acid: Coenzyme A Ligase (4CL) (1058), Hydroxycinnamoyl-CoA Shikimate/Quinate Hydroxycinnamoyl Transferase (HCT) (1099), Caffeoyl Shikimate Esterase (CSE) (615.5), Cinnamoyl-CoA Reductase (CCR) (214), Cinnamyl Alcohol Dehydrogenase (CAD) [63], Chalcone Synthase (CHS) (5485), Homoeriodictyol/Eriodictyol Synthase (HEDS/HvCHS2) (27128), Chalcone Isomerase (CHI) (3890), Flavanone 3-Hydroxylase (F3H) (747), and Laccase (386.9) (Fig. 6). Additionally, it was observed an increase of silybin a and b content under drought stress conditions.

Discussion

This article discussed the impact of drought stress on milk thistle plants from four different perspectives.

What biological pathways and reactions are DEGs involved in?

Most genes were associated with the metabolic pathways and the biosynthesis of secondary metabolites, which became statistically significant at a high probability level (Table 4). The evolution of specific metabolic pathways deriving from primary metabolism is crucial for plants to interact with their surrounding environment [25, 26]. Pathways of secondary metabolites emerged from primary pathways, which indicates that the advent of novel enzymatic activities targeting primary metabolites has led to the production of new compounds. These compounds have improved plant adaptability to specific environmental conditions and have progressively evolved into specialized metabolites, as evidenced in the work of Weng [27].

After the pathways above, pathways associated with carbon metabolism and amino acid biosynthesis were identified with the highest number of genes. Plants' ability to navigate and adapt to ever-changing environmental conditions is partly attributable to the dynamic nature of carbon metabolism. This flexibility facilitates an efficient response to fluctuations in the availability of CO2 or the products of electron transport in response to the demand for carbon assimilation [28]. Moreover, beyond its critical role in protein synthesis, amino acid metabolism

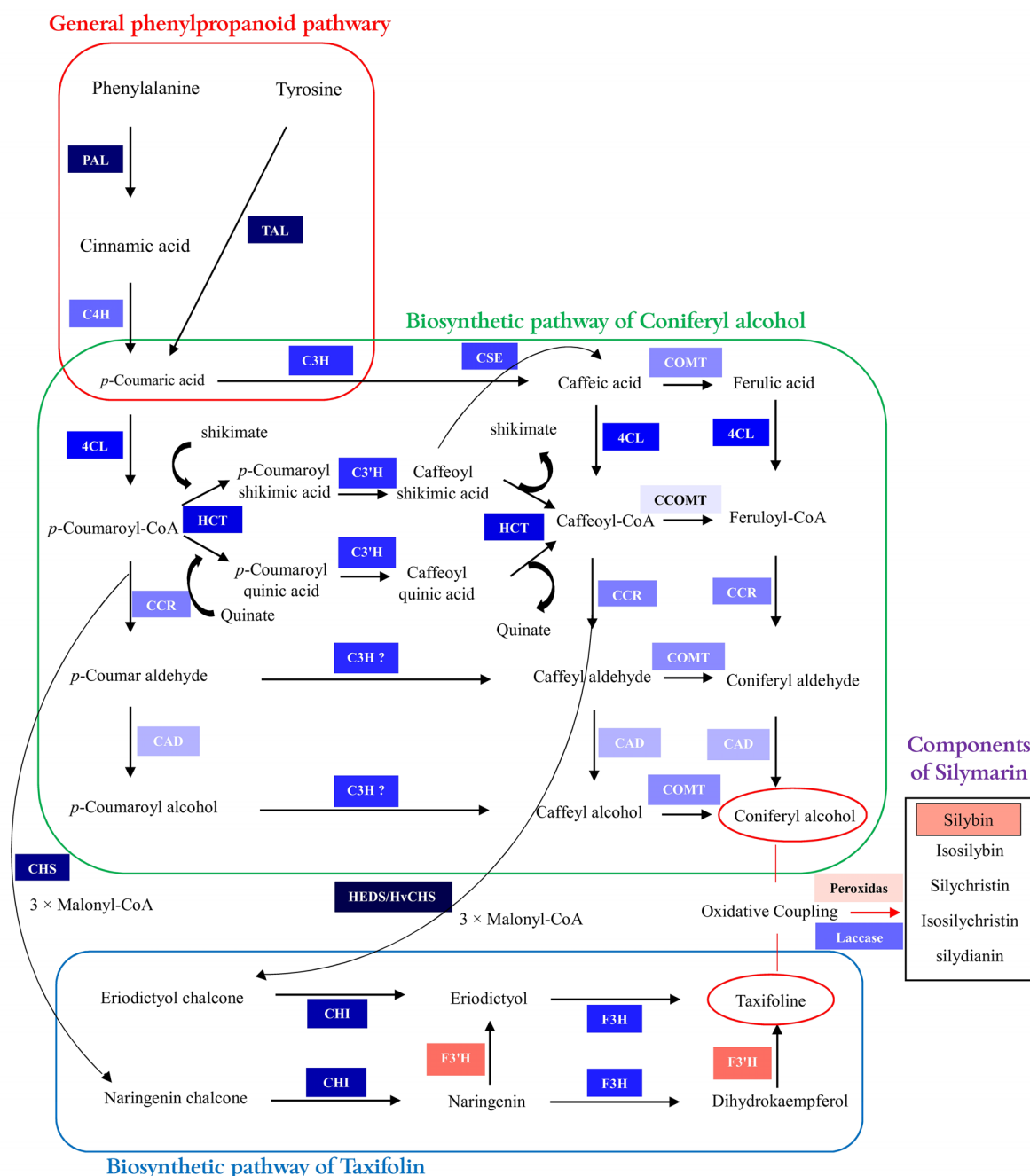


Fig. 6 Overview of silymarin biosynthesis pathway along with gene expression. Blue colored boxes indicate a decrease in gene expression and red colored boxes indicate an increased gene expression. The intensity of the color boxes indicates the amount of increase or decrease in expression

is deeply interconnected with several other metabolic processes, including energy and carbohydrate metabolism, balancing the carbon-nitrogen budget, influencing hormone and secondary metabolism, and modulating responses to stress, among others.

Gene expression involved in the protein processing pathway of the endoplasmic reticulum showed a diminished expression under severe stress conditions compared to F.C and 70% F.C; conversely, genes engaged in

the starch and sucrose metabolism pathways showed increased expression under severe stress, compared to F.C and 70% F.C. The requirement for protein folding within the endoplasmic reticulum can increase owing to various developmental scenarios or environmental conditions. The accumulation of proteins that are either unfolded or poorly folded within the endoplasmic reticulum triggers the unfolded protein response (UPR), a mechanism devised to mitigate the detrimental impact

of such proteins in plant systems [29]. Hence, a reduced expression of genes associated with protein processing could lead to increased accumulation of misfolded or unfolded proteins within the endoplasmic reticulum, leading to stress. Starch and sucrose play crucial, albeit complementary, roles in carbon storage and transport at an organismal level, with their metabolic activities responsible for a significant portion of carbon flux across most plant species [30, 31].

Genes associated with the spliceosome pathway have played a significant role under both mild and severe stress conditions, including a high number of genes. The spliceosome, a substantial RNA-protein complex, catalyzes the excision of introns from pre-mRNAs within the nucleus. Drought conditions induce considerable alterations in the spliceosome, indicative of adaptive transcriptional modifications and the stress response [32].

Expressions of genes involved in the oxidative phosphorylation and mRNA surveillance pathways demonstrated reductions under both severe (40% F.C) and mild (70% F.C) stress in comparison to control conditions (F.C). In Plants, ATP production is predominantly facilitated by mitochondrial oxidative phosphorylation (OXPHOS) and chloroplastic photophosphorylation (PHOTOPHOS) (Braun, 2020). Moreover, transcripts emanating from the eukaryotic nucleus undergo mRNA surveillance, a process ensuring that only flawless mRNAs proceed to translation [33].

Under severe stress, genes within the RNA degradation pathway and those involved in glycine, serine, and threonine metabolism showed increased expression compared to F.C. treatment. RNA degradation proteins are instrumental in removing superfluous RNA from the cells [34]. Threonine, a vital amino acid, features a close biochemical interconnection with methionine and isoleucine, and is convertible with glycine and serine [35, 36].

In the 40% F.C vs. 70% F.C pairwise comparison, the MAPK signaling pathway had a prominent activity. Moreover, the expression of genes engaged in ubiquitin-mediated proteolysis and fatty acid metabolism pathways was increased, whereas those associated with the RNA transport pathway declined. MAPKs are central to one

of plants' most comprehensively investigated signaling mechanisms, bridging stimulus perception with a range of cellular and adaptive responses [37]. The expression of genes of the MAPK pathway under drought stress in our experiment was consistent with the expression of genes of this pathway in the experiment of Hasanpour et al. [38]. The ubiquitin proteolytic system is critical in different cellular functions, comprising cell cycle regulation, control over signal transduction pathways, modulation of inflammatory and immune responses, and influence on development and differentiation processes [39]. Fatty acids biosynthesized in chloroplasts are immediately merged with glycerol to form a galactolipid, a primary constituent of the chloroplast membrane. They are also transported to the cytoplasm to join with glycerol within the endoplasmic reticulum, forming a cell membrane phospholipid. RNA transport denotes the active translocation of ribonucleic acid (RNA) molecules within the cell to different locations [40].

What biological processes, molecular functions, and cellular components are DEGs involved in?

Most genes that showed upregulation or downregulation across six different groups were engaged in the biological processes of protein phosphorylation, oxidation-reduction mechanisms, and the embryo development process ending in seed dormancy (Table 5). Protein phosphorylation, as one of the most prevalent types of post-translational modifications, is integral to a myriad of cellular functions, including but not limited to the cell cycle, signal transduction pathways, and environmental stress responses [41–43]. The cyclical nature of oxidation and reduction is fundamental for the mitigation of dormancy and quiescence, facilitating the activation of the cell cycle and inducing the genetic and epigenetic controls essential for growth and differentiation in various environmental conditions [44]. During embryo development, which ends in seed dormancy, a distinct progression of the embryo is observed from the stage of zygote formation until the end of seed dormancy [45]. Since gene expression analysis was performed at the flowering stage of the plants, this suggests that the increased expression of genes associated with embryogenesis may be attributable to this developmental stage.

The molecular functions encompassing protein binding, mRNA binding, and ATP binding were distinguished by having the highest number of up- and downregulated genes. Plant RNA-binding proteins (RBPs), characterized by their heterogeneity and being ubiquitous, control gene regulation both co- and post-transcriptionally. This coordination modulates RNA metabolism in response to internal and environmental signals, thereby regulating over 60% of the plant's transcriptome engaged in growth, development, and stress adaptation [46–51].

Table 5 Atmospheric information of Tehran (Shemiranat) from 2017.03.21 to 2017.07.22

Atmospheric information	From 2017.03.21 To 2017.04.20	From 2017.04.21 To 2017.05.21	From 2017.05.22 To 2017.06.21	From 2017.06.22 To 2017.07.22
Rainfall (mm)	60.5	76.7	0.0	26.3
Mean temperature (°C)	12.7	19.6	25.8	28.1
Moisture (%)	53	41	19	28
Wind speed (mps)	10.0	17.0	12.0	10.0

The ATP-binding cassette (ABC) transporters, a ubiquitous gene family across all life domains, engage in various biological processes. They have essential roles in metabolite transmembrane transport within plant cells through ATP hydrolysis and are instrumental in seed germination, lateral root development, stomatal operation, and environmental stress reactions [52–56]. The substrates for ABC proteins span across hormones, toxic substances, pigments, defense-important secondary metabolites, lipid molecules, and reactive oxygen species (ROS)-associated compounds [57–62]. The genes that were up- or downregulated showed predominant activity in the nucleus, cytoplasm, and chloroplasts, particularly under mild and severe drought stress conditions, highlighting these cellular locations as critical centers of gene activity.

What transcription factors are DEGs related to?

Our results showed the significant involvement of transcription factors, including ERF, C3H, bHLH, BZIP, HD-ZIP, NAC, LBD, MYB-RELATED, NIN-LIKE, and WRKY, in mediating drought stress tolerance mechanisms. Within the AP2/ERF superfamily, the Ethylene Responsive Factor (ERF) subgroup has the highest number of proteins and has garnered considerable attention for its biological functionalities. These members are crucial in regulating plant resilience against a spectrum of abiotic stressors, including drought, salinity, and cold stress. Activation of ERFs typically occurs through phosphorylation by mitogen-activated protein kinases or escape from ubiquitin-ligase enzymes, subsequently facilitating their association with nucleic acid proteins before their binding to cis-regulatory elements located in the promoter regions of genes responsive to stress [63, 64].

Numerous studies have confirmed the critical roles of AP2/ERF transcription factors, particularly the ERF subfamily, in plant responses to abiotic stresses, growth regulation, and secondary metabolism. Ma et al. [65] provided a comprehensive review highlighting the involvement of AP2/ERF members in regulating stress-responsive gene expression under drought, salinity, and cold conditions. Similarly, Licausi et al. [66] and Müller and Munné-Bosch [67] emphasized the function of ERFs as key regulatory hubs at the intersection of hormonal signaling pathways—such as ethylene and abscisic acid—and environmental adaptation. In addition to their stress-related roles, Phukan et al. [68] and Baharudin and Osman [69] reported that ERF transcription factors contribute to the regulation of secondary metabolite biosynthesis, including pathways for terpenoids and flavonoids. Furthermore, Wu et al. [63] demonstrated that ERFs activate stress-responsive pathways through binding to cis-regulatory elements in the promoters of target genes. In a

functional study on *Populus*, Kong et al. [70] showed that overexpression of the PtoERF15 gene enhances drought tolerance via activation of the jasmonic acid (JA) signaling pathway, leading to increased expression of defense-related genes. These studies underscore the multifaceted regulatory roles of ERF transcription factors in orchestrating plant physiological and molecular responses to adverse environmental conditions.

Regarding zinc finger proteins (ZFPs), the Cysteine-3Histidine (C3H) is distinguished by a Zn²⁺ motif comprising three cysteines and a single histidine residue bound to a zinc cation. This configuration is a characteristic of mRNA-binding proteins conserved across numerous eukaryotes [71, 72]. The C3H gene family has been explored in various dicots and monocots, revealing its signature motif and pivotal role in regulating plant growth, developmental processes, and response mechanisms to environments [73].

A study by Zheng et al. [74] systematically identified members of this family and examined their expression profiles under three different stress conditions: drought, heavy metal (lead) stress, and *Fusarium oxysporum* infection. Their results revealed that several CmC3H genes were significantly regulated in response to these stresses, suggesting a potential role in the plant's defense mechanisms and stress tolerance. These findings provide a theoretical foundation for future research aimed at improving biotic and abiotic stress resistance in melon [74].

Among various transcription factors involved in drought tolerance, OsZFP37, OsC3H, OsNAC94, and OsbHLH148 have been shown to exhibit significantly higher transcript levels in the drought-tolerant rice cultivar IR36 compared to MTU1010 [75]. Moreover, overexpression of these transcription factors enhances drought tolerance, while their silencing increases drought sensitivity. This evidence highlights the critical roles of C3H, NAC, and bHLH family members in orchestrating drought stress responses, suggesting similar mechanisms may be conserved across different plant species [75].

Other researchers have also confirmed the positive role of transcription factors from the C3H family and related families in drought stress responses and the regulation of secondary metabolism. For instance, Arisha et al. [76] identified genes associated with drought stress responses in hexaploid sweet potato using RNA-Seq analysis. Additionally, Jiang et al. [77] investigated the C3H gene family in *Aegilops tauschii* and demonstrated their involvement in drought stress response. Furthermore, Cheng et al. [78] analyzed the C3H gene family in wheat and explored their association with seed dormancy and germination.

Furthermore, many reports showed the involvement of other transcription factors, including bHLH [79], BZIP [80], HD-ZIP [81], NAC [82], LBD [83], MYB-RELATED [84], NIN-LIKE [85], and WRKY [86] in conferring

resilience to a multitude of abiotic stresses. These findings underscore the intricate regulatory networks underpinning plant adaptive responses to adverse environmental conditions.

In recent years, numerous studies have elucidated the role of basic helix-loop-helix (bHLH) transcription factors in regulating plant responses to environmental stresses and the biosynthesis of secondary metabolites. Rabeh et al. [87] presented a comprehensive review of transcription factor-mediated regulation of plant secondary metabolism under stress, highlighting the cross-talk between bHLH, MYB, and WRKY factors in stress signaling pathways. In a study focused on woody plants, Yan et al. [88] explored the functions and regulatory mechanisms of bHLH transcription factors and demonstrated their involvement in responses to salinity and low temperature. Gaude and Jalmi [89] further discussed the transcriptional regulation of stress-induced secondary metabolite biosynthesis, specifically the induction of such compounds under drought and UV exposure. Muhammad et al. [90] examined the diverse roles of bHLH genes in fleshy fruit-bearing species, showing their contributions not only to stress responses but also to developmental processes such as fruit ripening and pigmentation. In two separate reviews, Qian et al. [91] and Guo et al. [92] summarized the functions of bHLH transcription factors in tolerance to abiotic stresses, including drought, cold, and salt. Additionally, Wang et al. [93] identified StbHLH47 in potato (*Solanum tuberosum* L.) as a negative regulator of drought tolerance, underscoring the potential for both positive and negative regulatory roles of bHLH family members in stress adaptation.

Examination of two main products of milk thistle:

Silymarin and lipids

Silymarin

In plants, flavonolignans biosynthesis and accumulation are modulated by various environmental factors. An oxidative linkage between the phenylpropanoid derivative, coniferyl alcohol, and the flavonoid compound, taxifolin, mediates silymarin biosynthesis [94]. These two precursor molecules originate from phenylpropane and flavonoid structural units, with coniferyl alcohol and taxifolin being produced via the monolignol and flavonoid biosynthetic pathways, respectively [95].

Silymarin comprises a complex of seven flavonolignans—including silybin a, silybin b, silychristin a, silychristin b, isosilybin a, isosilybin b, and silydianin—and one flavonoid, taxifolin [94]. The compound silibinin, a combination of silybin a and b, has been shown to underlie the hepatoprotective capabilities attributed to silymarin in milk thistle [94]. Analytical quantification through HPLC demonstrated that silybinin levels in seeds of milk thistle were significantly increased under severe drought

stress compared to control conditions, suggesting stress-induced phytochemical accumulation [96]. They showed that among the various components of silymarin, silybin accumulation was significantly higher in seeds collected from plants subjected to drought stress. Furthermore, this phenotypic response was reported in some studies [97–99], indicating an adaptive silymarin biosynthetic increase as a component of the antioxidative defense mechanism against prolonged water deficit stress.

Our transcriptome analysis, utilizing RNA-Seq technology, revealed an upregulated expression of silymarin biosynthesis-related genes under mild stress conditions (70% F.C) and a downregulation under severe stress (40% F.C). This was consistent with findings from ElSayed et al. [96], where the expression of genes encoding chalcone synthase (CHS1, CHS2, CHS3), integral to the silybin biosynthetic pathway, increased under drought stress conditions. These data suggested a genetic predisposition in milk thistle towards drought stress tolerance, with seeds derived from severely stressed plants showing an increased silymarin content [99]. At the enzymatic level, the CHS and chalcone isomerase (CHI) play important roles in catalyzing the biosynthesis of naringenin through the conversion of malonyl-CoA and 4-hydroxycinnamoyl-CoA molecules, a critical initial step in the phenylpropanoid pathway leading to taxifolin biosynthesis, the flavonoid precursor of silymarin [100, 101].

It was shown that expression of CHS is induced in response to various biotic and abiotic stressors [102], a phenomenon also observed in our study, which confirms the increased expression of CHS genes in milk thistle leaves under drought conditions. Another study conducted by Drouet et al. [19] has elucidated the involvement of genes such as PAL (l-phenylalanine ammonia-lyase), CAD (cinnamyl alcohol dehydrogenase), CHS, F3H (flavanone 3-dioxygenase), F3'H (flavone 3'-hydroxylase), LAC (laccases), and POX (peroxidases) in the silymarin biosynthesis pathway. Their results showed that key genes in the phenylpropanoid and flavonoid pathways, such as F3H, F3'H, CHS, and CHI, are activated during milk thistle fruit maturation and coincide with the increase in flavonolignan accumulation. Additionally, PAL and CAD genes are involved in the production of lignan precursors, while POX and LAC genes play a role in the final coupling of these compounds. These findings indicate that these genes and enzymes closely and coordinately regulate silymarin biosynthesis in the fruit pericarp. Our experimental findings were consistent with these observations, reinforcing the hypothesis of a coordinated genetic and enzymatic regulation underpinning silymarin biosynthesis.

Lipids

Milk thistle is recognized as an oil-producing and medicinal plant, distinguished by its rich content of unsaturated fatty acids, making it an advantageous choice for edible oil within the food industry. The role of lipids, encompassing fatty acids, hydrocarbons, esters, steroids, etc., is crucial in plant cells, especially under stress conditions such as drought. This is due to their involvement in initiating defense mechanisms and mitigation processes in response to stress factors. To assess the critical role played by milk thistle lipids in improving drought tolerance, a field study was conducted to investigate the plant's response under different levels of water deficit: F.C, 70% F.C, and 40% F.C. The findings showed the significant function of lipids as key components of plant cellular structure, particularly in adverse conditions.

Previously, the lipid constituents and the genes responsible for their biosynthesis in milk thistle were identified. Subsequently, eight genes (CYP86A1, CYP710A1, FATA2, LACS3, LOX2, Palmitoyl-protein thioesterase (PAL), PLA2-ALPHA, and PXG3) were shown to have differential expression, were selected for further examination. The metabolic networks related to these genes were identified, revealing a correlation between the two networks. This enabled the grouping of specific lipids with their biosynthesizing genes for each pathway.

In plants, lipoxygenases (LOXs) are key enzymes in the fatty acid metabolism pathway and play crucial roles in regulating physiological and biochemical processes. These enzymes are present in various organs including roots, stems, leaves, flowers, and fruits, and significantly influence organ growth and development, fruit ripening, oil oxidation, and signal transduction related to biotic and abiotic stresses. Additionally, LOXs regulate secondary metabolism, impacting the production of important compounds such as flower aromas and fruit flavors [103]. Similarly, LACS genes, responsible for activating long-chain fatty acids, facilitate their entry into essential metabolic pathways like lipid synthesis and degradation. They contribute to forming protective layers in plants, reducing water loss and enhancing tolerance to stresses such as drought and salinity. Defects in these genes lead to increased water loss and decreased drought tolerance, highlighting their essential role in plant growth, development, and stress resilience [104]. Furthermore, PXG3, a peroxygenase with potential calcium-binding capability, is involved in the degradation of storage lipids within oil bodies and participates in abiotic stress-related signaling pathways. It may contribute to drought tolerance through stomatal regulation under water-deficient conditions [105].

The PLA2 enzyme hydrolyzes membrane phospholipids to release free fatty acids and lysophospholipids, playing a vital role in plant responses to environmental

stimuli. It also regulates the trafficking of PIN proteins involved in hormonal balance and suppresses the transcriptional activity of MYB30, thereby modulating plant defense responses [106]. PAL plays a key role in terminating de novo fatty acid synthesis, particularly in producing the saturated fatty acid palmitic acid (16:0), which is critical for plant growth and seed development. It may also participate in the synthesis of long-chain fatty acids [107]. The CYP710A1 gene is crucial for lignification and enhancing drought tolerance in plants. It promotes lignin accumulation in the cell wall and regulates the expression of key structural genes involved in lignin biosynthesis, thereby strengthening cell structure and improving plant resistance to water deficit conditions [108]. Likewise, CYP86A1 is a key enzyme responsible for ω -hydroxylation of fatty acids, playing an important role in suberin biosynthesis—a protective barrier in plants. It is predominantly expressed in roots and is essential for the production of suberin monomers, contributing to enhanced plant protection against environmental stresses [109]. Finally, despite extensive investigation, no reliable information was found regarding the role of the FATA2 gene in these processes. These selected genes were further utilized to validate the expression results obtained from RNA-Seq experiments through qRT-PCR, showing consistency in gene expression outcomes.

Conclusion

This study investigated the impact of drought stress on transcriptome (DEGs) and related biological pathways in milk thistle. Thousands of genes were identified, notably involved in flavonoid and lipid biosynthesis and stress response pathways (including MAPK). Additionally, transcription factors regulating drought-responsive genes were characterized. Metabolomic analysis revealed a significant increase in key metabolites such as silybin a and b, which are strongly correlated with gene expression changes, highlighting complex gene–metabolite interactions under drought conditions. Morphological and physiological data confirmed milk thistle's notable drought tolerance. Optimal irrigation at 70% field capacity with four-day intervals provided the best growth and performance. This study represents one of the first de novo transcriptome analyses of milk thistle under drought stress, simultaneously examining gene expression related to lipid and flavonolignan pathways. The results offer practical agricultural strategies and provide deeper insights into the mechanisms of drought tolerance.

Materials and methods

The general process of doing this article is presented in Fig. 7. All bioinformatics analyses were conducted using tools and databases that were available and accessed between 2019 and 2020.

Plant materials and experimental site

The milk thistle seeds were bought from Pakanbazar, Isfahan, Iran. The research site was at Shahid Beheshti University in Tehran, Iran (latitude 51.23°N, longitude 35.48°E). The area had a mountainous and moderate climate, with an average annual temperature

and precipitation of 22 °C and 145.2 mm, respectively. Monthly data on precipitation, air temperature, moisture, and wind speed were gathered from the meteorological site (Table 5). The soil consisted of one-third clay, one-third sand, and one-third leaf compost. The total area of the field encompassed 150.0 m². Furrows were carefully established for the study, and the spacing of plants within the rows was set at 0.5 m, with a 1-meter gap between the rows [110, 111].

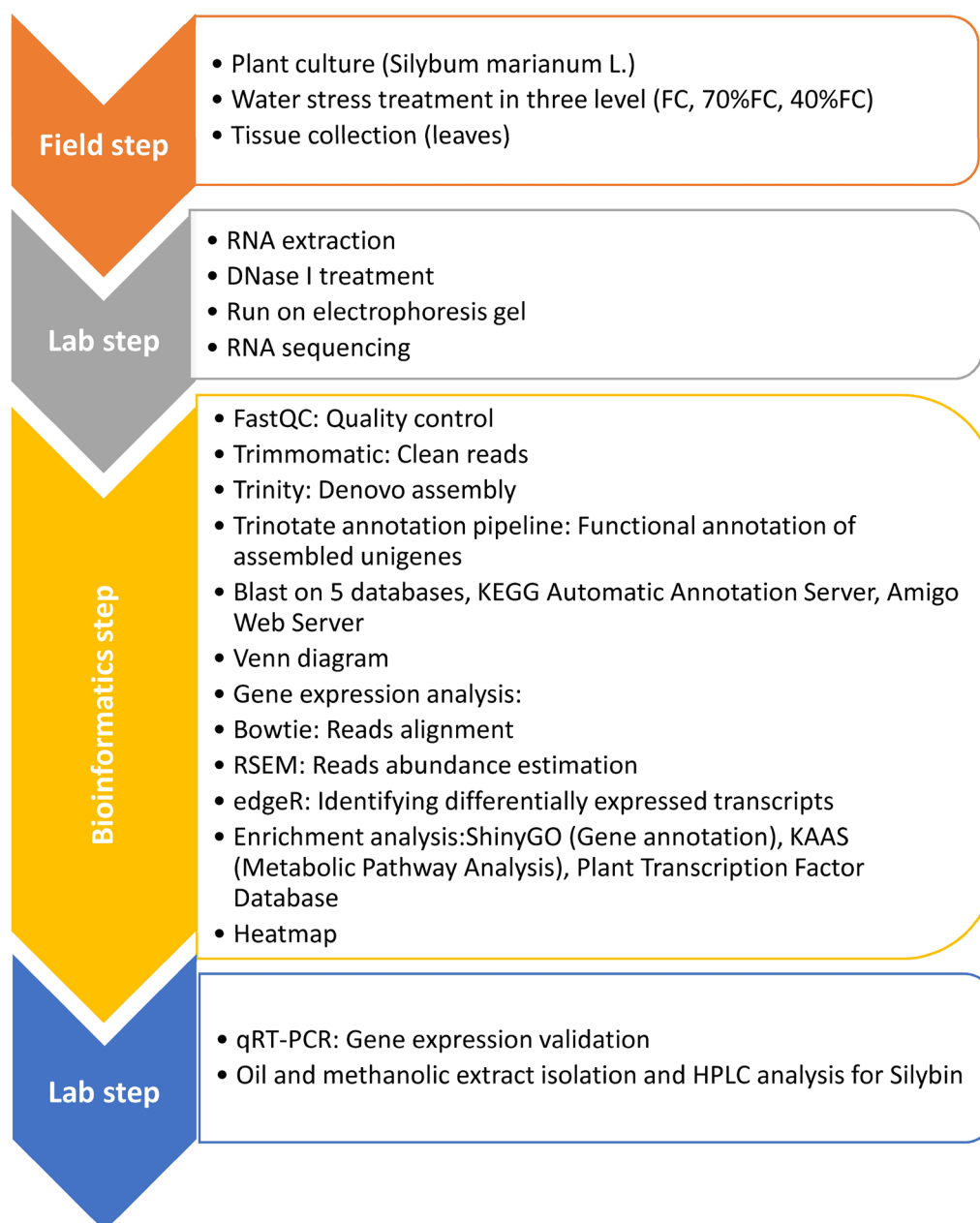


Fig. 7 Field and Lab experiment and RNA-seq analysis workflow

Field experiments and drought stress treatments

The experiment began with sowing milk thistle seeds, followed by regular irrigation every two days. When the plants reached the flowering stage, they were exposed to drought stress using a weighted method. During this period, the plants received irrigation at 100% field capacity (F.C) bi-daily, 70% F.C once every 4 days, and 40% F.C every 8 days [112]. After 8 days, leaf sampling was conducted. Four plants (four biological replicates) were randomly selected, and medium-sized leaves from the middle part of the stem were collected. These leaves were quickly wrapped in aluminum foil, immersed in liquid nitrogen, and kept in a laboratory freezer at -80°C .

RNA-Seq library construction and sequencing

The RNA was isolated from milk thistle leaves (three biological replicates) using the RNeasy Plant Mini Kit (QIAGEN, USA), following the protocols provided by the manufacturer. The isolated RNAs were subjected to DNase I treatment (Thermo Fisher Scientific, USA) to eliminate any potential DNA contamination. The 1% agarose gel and a NanoDrop 1000 spectrophotometer (Thermo Scientific, USA) were employed to evaluate the quality and concentration of the RNA samples, respectively. Further assessment of RNA quality was achieved using a QC Bioanalyzer (Agilent Technologies, BGI-Hong Kong NGS Lab), ensuring that the RNA integrity number (RIN) for each sample exceeded 7. The procedure for selecting Poly A, preparing cDNA, ligating adapters, forming clusters, and sequencing was executed at the Beijing Genomes Institute (China), following the manufacturer's protocols, utilizing standard Illumina sequencing kits. The sequencing process employed the Poly A mRNA Capture technique on the BGISEQ-500 platform, using a paired-end strategy with a read length of 100 bp and generating 6Gbp of raw data.

Processing of RNA-Seq data and de Novo assembly

The FastQC tool [113] available at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> assessed the quality of the initially sequenced reads. Subsequent trimming of low-quality segments and adaptor sequences from these reads was accomplished using the Trimmomatic software (Version 0.39) [114]. The quality control process post-trimming was once again verified through the application of FastQC, ensuring the retention of only high-quality reads. These processed reads were then pooled for de novo assembly, which was executed using the Trinity software (Release 2021-03-04) [115], according to the default parameters while setting the k-mer length at 31. The CAP3 package was employed to cluster the transcripts based on a 95% identity threshold and a minimum contiguous sequence length of 200 nucleotides to refine the assembly output further. The identification

of potential coding sequence (CDS) regions within the sequenced transcripts was facilitated using the TransDecoder tool (<http://transdecoder.github.io>), which was pivotal for predicting open reading frames (ORFs). The concluding stage involved leveraging the protein sequences derived from these predicted ORFs to delineate unigenes, enriching the understanding of the transcriptome's coding potential.

Functional annotation of milk Thistle assembled unigenes

The Trinotate annotation pipeline (<http://trinotate.github.io/>) was utilized to annotate the milk thistle transcriptome. This process included subjecting the assembled milk thistle unigenes to BLAST analysis against several databases: non-redundant nucleotide (NT) (<http://www.ncbi.nlm.nih.gov/>), non-redundant proteins (NR) (<http://www.ncbi.nlm.nih.gov/>), UniProt (Swiss-Prot and TrEMBL) [116], Arabidopsis (version TAIR11) [117], Sunflower protein (http://ftp.ensemblgenomes.org/pub/plants/release-50/fasta/helianthus_annuus/pep/) database with an E-value threshold of $10e-5$ for significance. To identify metabolic pathways associated with each unigene, Kyoto Encyclopedia of Genes and Genome (KEGG, <http://www.genome.jp/kegg/>) analysis was performed [118] using KEGG Automatic Annotation Server (KAAS, <http://www.genome.jp/kegg/kaas/>) [119]. For the functional classification of assembled unigenes, the Amigo Web Server (<http://amigo.geneontology.org>) was employed, following the established protocols [120]. Venn diagrams were generated through the InteractiVenn website based on gene trinity ID [121], facilitating a visual representation of our findings.

Differential gene expression analysis

The quantification of transcripts assembled by Trinity was performed by employing the RSEM (version 1.3.3) [122]. Initially, high-quality sequence reads were realigned to the constructed transcriptome via Bowtie2 (version 2.4.1) [115]. Subsequently, RSEM estimated the read alignment of individual transcripts. The output files from RSEM encompassed normalized expression metrics for each unique transcript and gene. The identification of differentially expressed (DE) contigs was achieved through analysis of the counts matrix generated by RSEM, utilizing the edgeR package from Bioconductor [123] in Rstudio [124]. Within this analysis, TMM normalization was applied by the edgeR package to account for variances in sample composition. The criteria for obtaining DEGs included a threshold false discovery rate (FDR) of ≤ 0.001 , alongside a requisite absolute Log ratio ≥ 2 and a minimum four-fold expression change.

Table 6 RT-qPCR primer sequence of selected candidate and reference genes and their amplification characteristics

Primer name	Primer sequence	PCR product length (bp)	TM	PCR amplification efficiency
PXG3	F: CCAGCAAACCTTGAGAAC R: GCAACGCCTTACTGATTC	207	54 54	1.95
PAL	F: AGGGTAATCTAATAGGCC R: ACTGAACTCTCCATCTGG	155	52 54	1.88
LOX2	F: CCACAGTGGAACATGTC R: ATCTTCAACCGCCATACC	255	54 54	1.94
CYP710A1	F: TTCTACCTACACTGAGC R: AGGAAGTCAAACAGGTGG	198	50 54	1.90
PLA2-ALPHA	F: ATGGGAAGTACTGTGGG R: GTGTTGCCTTTGAATGTC	194	52 52	1.99
LACS3	F: GAGATGAATTATGACGCC R: GCCACATATTCTCCTTG	256	52 50	1.91
CYP86A1	F: ACGTGACACCTCCTCCG R: CATGTTCTGTTGCCAGCG	242	57 58	1.97
FATA2	F: GTACTAGACGTGATTGG R: CTTCTGGAAATGCTAATC	177	50 49	2.00
18S rRNA	F: ATGATAACTCGACGGATCGC R: CTTGGATGTGGTAGCCGTTT	200	56 57	1.97

Enrichment analysis

The analysis of DEGs involved singular enrichment analysis (SEA) through the ShinyGO platform (<http://bioinformatics.sdstate.edu/go/>) for the identification of enriched gene ontologies (GO). These DEGs were systematically categorized into three main groups: biological processes (BP), molecular functions (MF), and cellular components (CC), with a significant threshold determined by the p-value. Further, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were determined by employing in-house scripts developed by the NIGEB Genome Center, which leveraged both the David and KEGG databases for comprehensive analysis. To identify transcription factors, the analysis turned to the resources of the Plant Transcription Factor Database (<http://planttfdb.gao-lab.org/>).

Validation of RNA-seq data using qRT-PCR

RNA extraction and cDNA synthesis

Total RNA was isolated from 200 mg of mature milk thistle seeds collected at the flowering phase, utilizing a total RNA extraction kit (RB1001, RNA, Iran). Then, DNase I (RB125A, RNA, Iran) was used to eliminate any potential genomic DNA contamination from isolated RNA. The quality and concentration of the isolated RNA were evaluated using 1% agarose gel electrophoresis and spectroscopic analysis with a NanoDrop 1000 spectrophotometer (Thermo Scientific, USA), respectively. Subsequently, five µg of the purified RNA was employed to synthesize cDNA using a cDNA Synthesis kit (RB125A, RNA, Iran), according to the guidelines provided by the manufacturer. Then, the synthesized cDNA was stored at -20 °C for future experimentation.

Primer design and qRT-PCR analysis

Primers for targeted genes were designed by employing the nucleotide sequences of closely related species, *Carthamus tinctorius* and *Helianthus annuus*, from the NCBI database. Subsequent comparative analysis was performed against our prior milk thistle assembly data employing the BlastStation tool (<https://www.blaststation.com/>), enabling the selection of sequences with high identity. Regions proximal to the terminal polyadenine sequences, targeting a sequence length between 150 and 250 base pairs were considered for designing primers. Parameters such as homodimer and heterodimer formation, stem-loop structures, GC content, and melting temperature (TM) were meticulously evaluated utilizing both Oligo 7 software (<https://www.oligo.net/>) [125] and Vector NTI® Express Designer Software (<https://vector-nti.software.informer.com/11.0/>) [126]. The 18 S rRNA gene was used as the reference gene for normalization, and its suitability had been previously tested and validated in our prior experiments. Ultimately, the synthesis of designed primers was carried out by Bioneer Company (South Korea). Detailed sequences and additional characteristics of the primers are documented in Table 6.

The amplification was conducted utilizing the Rotor-Gene 2000 apparatus (Corbett Life Science, Australia) employing SYBR® Green Real-Time PCR Master Mix (RB120, RNA, Iran). The reaction volume (20 µl) consisted of 2× SYBR Green mix (10 µL), synthesized cDNA (1 µL), forward and reverse primers (each one µL) at a concentration of 20 nmol, and up to 7 µL of RNase-free water. The thermal cycling conditions initiated with a pre-denaturation step at 95 °C for 5 min, followed by 35 cycles of amplification, which included denaturation at

95 °C for 1 min, annealing between 50 and 60 °C based on the melting temperatures of the primers for 1 min, and extension at 72 °C for 15 s. A melting curve analysis post-amplification was executed to verify the specificity of the PCR amplifications, spanning a temperature interval between 60 and 95 °C for every reaction. Cycle threshold (Ct) values and PCR efficiency were calculated using the LinRegPCR software [127]. The relative expression levels of target genes were determined using the Relative Expression Software Tool (REST), in adherence to the protocols outlined by Pfaffl et al. [128], with the efficiency of the primers (E) incorporated into the calculation. The qRT-PCR experiments were executed in triplicate, encompassing both biological and technical replicates to ensure the reliability and reproducibility of the findings.

$$\text{Gene expression ratio} = (E_{\text{Gene}})^{\Delta \text{ct Gene}} / (E_{\text{Ref}})^{\Delta \text{ct Ref}}$$

Isolation of oil and methanolic extract and Silybin measurement by HPLC

The seed samples (three biological replicates), upon drying, were finely ground into powder form. Subsequently, 10 g of this powder was used for oil extraction in a Soxhlet extractor utilizing n-hexane as the extraction solvent. This procedure was conducted at a temperature of 70 °C over six hours, and the extracted oil was subsequently stored in an amber glass container to protect against light degradation. After removing the oil, the residual powder was dried at 37 °C for one week in preparation for the methanolic extraction phase. For the extraction with methanol, 2 g of the dried, oil-free powder was combined with 200 ml of 80% methanol and subjected to agitation on a shaker for two days. Then, the resultant mixture was filtered using a filter paper and kept at a temperature of 4 °C. This extraction process was done again for the remaining residue, and the subsequent extracts were added to the first collected portion and allowed to undergo evaporation at room temperature for two weeks to yield a concentrated form.

The concentrated extracts and standard silymarin at specific concentrations were solubilized in methanol. These solutions were then introduced into the HPLC system (comprising an Agilent Infinity1260, equipped with UV/VIS and Diode Array detectors, a Merck Hitachi L-7100 pump, and an EZ Chrome software) using syringe filters of 0.2 µm pore size. A volume of 20 µl of each Silybin solution and plant sample extract was analyzed within the HPLC system. The outcome of this analytical procedure involved the examination of HPLC data and the identification of related peaks.

Identification of Silymarin biosynthesis pathway genes

To identify the genes engaged in the silymarin biosynthesis, this study began with a comprehensive review of pathways previously described in existing literature [19, 95, 129–131]. After identifying candidate genes, their sequences were extracted from plants within the same family as Milk Thistle using NCBI and Uniprot databases. Subsequently, these sequences were blasted in the Milk thistle database (created and developed in this project) using the BlastStation tool. To validate the accuracy of the sequences, a further comparative analysis was conducted on the NCBI platform. The next step involved determining the expression levels of these identified genes through analysis of our RNA-SEQ data. Finally, the open reading frame (ORF) sequences of the genes of interest were determined by employing Vector NTI software in preparation for submission to the NCBI database.

Statistical analysis

The analysis of the data was conducted utilizing R version 3.5.3 (<https://www.r-project.org/>) [132] and RStudio version 1.1.463 (<https://www.rstudio.com/>) [133]. We adopted a Completely Randomized Block Design (CRBD) approach, wherein we categorized the treatments as fixed effects and positioned the replications as a random effect within the ANOVA analysis. The Duncan test, accessed through the agricolae package, was employed at a 5% significance level to assess mean differences. A Completely Randomized Design (CRD) was used for metabolite and molecular data. Analysis of variance (ANOVA) was performed, and the mean comparisons were conducted using the Least Significant Difference (LSD) test at $p \leq 0.05$ using the agricolae package in R.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-025-07272-5>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Author contributions

Rahele Ghanbari Moheb Seraj 35%- Material preparation, data collection and analysis, writing paper. Asadollah Ahmadikeh 10%- contributed to the study conception and design, read and approved the final manuscript. Vahid Shariati 10%- Design and analysis of transcriptome analysis section. Mahdi Behnamian 10%- contributed to the study conception, read and approved the final manuscript. Keyvan Esmaeilzadeh-Salestani 10%- Advice on writing the paper, edit and scientific corrections of the paper. Neda Tariverdizadeh 10%- Design and analysis of silymarin synthesis pathway. Ali Emadi 10%- Collaboration in silymarin biosynthesis pathway genes and submit to NCBI. Sara Dezhsetan 5%- Topic selection advisor.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability

On August 26, 2020, the University of Mohaghegh Ardabili submitted the raw transcriptome sequences of Milk thistle to the SRA section of the NCBI (National Center for Biotechnology Information) international database. The Project ID was PRJNA659420, the sample ID was SAMN15905480, and the experiment accession was SRR12539246, SRR12539245, and SRR12539244 for F.C, 70% F.C, and 40% F.C, respectively. Additionally, the NCBI website received the sequence of 19 genes of the silymarin pathway identified in *Silybum marianum* L., along with their accessions. The names of the genes along with their accession numbers are as follows: C4H (PP965187), C3H (PP965188), COMT (PP965189), CCOMT (PP965190), 4CL (PP965191), HCT (PP965192), CSE (PP965193), CCR (PP965194), CAD (PP965195), CHS1 (PP965196), CHS2 (PP965197), CHS3 (PP965198), CHI (PP965199), F3H (PP965200), Iac (PP965201), HEDS/HvCHS2 (PP965202), F3'H (PP965203) and POX (PP965204). Additionally, the sequences of lipid biosynthesis pathway genes include CYP86A1, CYP710A1, FATA2, LACS3, LOX2, Palmitoyl-protein thioesterase (PAL), PLA2-ALPHA, and PXG3 were submitted to the NCBI international database (nucleotide section) under the accession numbers of MW151571, MW151572, MW151573, MW151574, MW151575, MW151576, MW151577, and MW151578, respectively.

Code availability

In the manuscript, we detailed a compilation of utilized software for de novo transcriptome assembly, as well as for pre- and post-assembly processes and transcriptome annotation. Each software program, along with its corresponding version, is meticulously cataloged within the Methods section. It is noteworthy that our research exclusively incorporated software that is publicly accessible, avoiding the use of any custom-developed code.

Declarations

Ethics approval

All the ethical principles in this study are respected.

Consent to participate

All authors' contributions are approved.

Competing interests

The authors declare no competing interests.

Consent for publication

All authors agree to publish the work.

Author details

¹Department of Horticultural Sciences, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran

²Department of Cell & Molecular Biology, Faculty of Life Sciences & Biotechnology, Shahid Beheshti University, Tehran, Iran

³Chair of Crop Science and Plant Biology, Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, Kreutzwaldi 1, Tartu 51006, Estonia

⁴Institute of Technology, University of Tartu, Nooruse 1, Tartu 50411, Estonia

⁵Department of Plant Molecular Biotechnology, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran

⁶Department of Biotechnology, Iranian Research Organization for Science and Technology (IROST), Tehran, Iran

⁷Department of Agriculture and Plant Breeding group, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran

Received: 3 September 2024 / Accepted: 19 August 2025

Published online: 28 August 2025

7. References

- Karkanis A, Bilalis D, Efthimiadou A. Cultivation of milk thistle (*Silybum marianum* L. Gaertn.), a medicinal weed. *Ind Crop Prod*. 2011;34(1):825–30.
- Andrzejewska J, Sadowska K, Mielcarek S. Effect of sowing date and rate on the yield and flavonolignan content of the fruits of milk thistle (*Silybum*

marianum L. Gaertn.) grown on light soil in a moderate climate. *Ind Crops Prod*. 2011;33(2):462–8.

- Bhattacharya S. Milk thistle (*Silybum marianum* L. Gaert.) seeds in health. Nuts and seeds in health and disease prevention. Elsevier; 2011. p. 759–66.
- Saller R, Meier R, Brignoli R. The use of silymarin in the treatment of liver diseases. *Drugs*. 2001;61:2035–63.
- Negi AS, Kumar J, Luqman S, Shanker K, Gupta M, Khanuja S. Recent advances in plant hepatoprotectives: a chemical and biological profile of some important leads. *Med Res Rev*. 2008;28(5):746–72.
- Lucini L, Kane D, Pellizzoni M, Ferrari A, Trevisi E, Ruzickova G, et al. Phenolic profile and in vitro antioxidant power of different milk thistle [*Silybum marianum* (L.) Gaertn.] cultivars. *Ind Crops Prod*. 2016;83:11–6.
- Post-White J, Ladas EJ, Kelly KM. Advances in the use of milk thistle (*Silybum marianum*). *Integr Cancer Ther*. 2007;6(2):104–9.
- Shahbazi M, Tohidfar M, Azimzadeh Irani M. Identification of the key functional genes in salt-stress tolerance of *Cyanobacterium phormidium tenue* using in silico analysis. *3 Biotech*. 2021;11(12):503.
- Ghanbari Moheb Seraj R, Tohidfar M, Azimzadeh Irani M, Esmaeilzadeh-Salestani K, Moradian T, AhmadiKhah A, et al. Metabolomics analysis of milk thistle lipids to identify drought-tolerant genes. *Sci Rep*. 2022;12(1):12827.
- Seraj RGM, Mohammadi M, Shahbazi M, Tohidfar M. Co-expression network analysis of soybean transcriptome identify hub genes under saline-alkaline and water deficit stress. *Iran J Biotechnol*. 2022;20(4): e3124.
- Munns R, Tester M. Mechanisms of salinity tolerance. *Annu Rev Plant Biol*. 2008;59(1):651–81.
- Dhanda S, Sethi G, Behl R. Indices of drought tolerance in wheat genotypes at early stages of plant growth. *J Agron Crop Sci*. 2004;190(1):6–12.
- Aslam M, Maqbool MA, Cengiz R, Aslam M, Maqbool MA, Cengiz R. Mechanisms of drought resistance. Drought stress in maize (*Zea Mays* L) effects, resistance mechanisms, global achievements and biological strategies for improvement. *Springer briefs in agriculture*. 2015:19–36.
- Vaidya K, Ghosh A, Kumar V, Chaudhary S, Srivastava N, Katudia K, et al. De novo transcriptome sequencing in *trigonella foenum-graecum* L. to identify genes involved in the biosynthesis of Diosgenin. *Plant Genome*. 2013;6(2): plantgenome2012080021.
- Misra BB. An updated snapshot of recent advances in transcriptomics and genomics of phytomedicinals. *J Postdoc Res*. 2014;2(2):1–15.
- Li F-D, Tong W, Xia E-H, Wei C-L. Optimized sequencing depth and de novo assembler for deeply reconstructing the transcriptome of the tea plant, an economically important plant species. *BMC Bioinformatics*. 2019;20:1–11.
- Fulvio F, Martinelli T, Paris R. Selection and validation of reference genes for RT-qPCR normalization in different tissues of milk thistle (*Silybum marianum*, Gaert). *Gene*. 2021;768:145272.
- Rafizadeh A, Koohi-Dehkordi M, Sorkheh K. Molecular insights of genetic variation in milk thistle (*Silybum marianum* [L.] Gaertn.) populations collected from Southwest Iran. *Mol Biol Rep*. 2018;45:601–9.
- Drouet S, Tungmunthum D, Lainé É, Hano C. Gene expression analysis and metabolite profiling of Silymarin biosynthesis during milk Thistle (*Silybum marianum* (L.) Gaertn.) fruit ripening. *Int J Mol Sci*. 2020;21(13): 4730.
- Yang J, Tang Z, Yang W, Huang Q, Wang Y, Huang M, et al. Genome-wide characterization and identification of trihelix transcription factors and expression profiling in response to abiotic stresses in Chinese willow (*Salix matsudana* Koidz). *Front Plant Sci*. 2023;14:1125519.
- Lan Y, Gong F, Li C, Xia F, Li Y, Liu X, et al. New insights into the evolution analysis of trihelix gene family in eggplant (*Solanum melongena* L.) and expression analysis under abiotic stress. *BMC Genomics*. 2024;25(1):1040.
- Gao M-J, Li X, Lui H, Gropp GM, Lydiate DD, Wei S, et al. ASIL1 is required for proper timing of seed filling in Arabidopsis. *Plant Signal Behav*. 2011;6(12):1886–8.
- Käppel S, Rümpler F, Theißen G. Cracking the floral quartet code: how do multimers of MIKCC-type MADS-domain transcription factors recognize their target genes? *Int J Mol Sci*. 2023;24(9): 8253.
- Shah L, Sohail A, Ahmad R, Cheng S, Cao L, Wu W. The roles of MADS-box genes from root growth to maturity in Arabidopsis and rice. *Agron*. 2022;12(3):582.
- Pott DM, Osorio S, Vallarino JG. From central to specialized metabolism: an overview of some secondary compounds derived from the primary metabolism for their role in conferring nutritional and organoleptic characteristics to fruit. *Front Plant Sci*. 2019;10:835.
- Shahbazi M, Tohidfar M, Irani MA, Seraj RGM. Functional annotation and evaluation of hypothetical proteins in *Cyanobacterium synechocystis* sp. PCC 6803. *Biocatal Agric Biotechnol*. 2022;39: 102246.

27. Weng JK. The evolutionary paths towards complexity: a metabolic perspective. *New Phytol.* 2014;201(4):1141–9.
28. Leegood R. Carbon metabolism. Photosynthesis and Production in a Changing Environment: A field and laboratory manual: Springer; 1993. pp. 247–67.
29. Howell SH. Endoplasmic reticulum stress responses in plants. *Annu Rev Plant Biol.* 2013;64(1):477–99.
30. Geigenberger P. Regulation of sucrose to starch conversion in growing potato tubers. *J Exp Bot.* 2003;54(382):457–65.
31. Smith AM, Stitt M. Coordination of carbon supply and plant growth. *Plant Cell Environ.* 2007;30(9):1126–49.
32. Marondedze C, Thomas L, Lilley KS, Gehring C. Drought stress causes specific changes to the spliceosome and stress granule components. *Front Mol Biosci.* 2020;6: 163.
33. Kawa D. Security notice: this plant immunity is under mRNA surveillance. *Plant Cell.* 2020. <https://doi.org/10.1105/tpc.20.00169>.
34. Sohrabi-Jahromi S, Hofmann KB, Boltendahl A, Roth C, Gressel S, Baejen C, et al. Transcriptome maps of general eukaryotic RNA degradation factors. *Elife.* 2019;8:e47040.
35. Hildebrandt TM, Nesi AN, Araújo WL, Braun H-P. Amino acid catabolism in plants. *Mol Plant.* 2015;8(11):1563–79.
36. Joshi V, Joung J-G, Fei Z, Jander G. Interdependence of threonine, methionine and isoleucine metabolism in plants: accumulation and transcriptional regulation under abiotic stress. *Amino Acids.* 2010;39:933–47.
37. Hamel L-P, Nicole M-C, Duplessis S, Ellis BE. Mitogen-activated protein kinase signaling in plant-interacting fungi: distinct messages from conserved messengers. *Plant Cell.* 2012;24(4):1327–51.
38. Hasanpour K, Aalami A, Ghanbari Moheb Seraj R, Hosseini R, Naeimi S, Esmailzadeh-Salestani K. Identification of drought-tolerant hub genes in Iranian KC-2226 genotype of *Aegilops tauschii* using transcriptomic analysis. *Sci Rep.* 2023;13(1):9499.
39. Ciechanover A, Orian A, Schwartz AL. Ubiquitin-mediated proteolysis: biological regulation via destruction. *Bioessays.* 2000;22(5):442–51.
40. Chapman KD, Ohlrogge JB. Compartmentation of triacylglycerol accumulation in plants. *J Biol Chem.* 2012;287(4):2288–94.
41. Pawson T, Scott JD. Protein phosphorylation in signaling—50 years and counting. *Trends Biochem Sci.* 2005;30(6):286–90.
42. Hunter T. Tyrosine phosphorylation: thirty years and counting. *Curr Opin Cell Biol.* 2009;21(2):140–6.
43. Cheng H, Deng W, Wang Y, Ren J, Liu Z, Xue Y. DbPPT: a comprehensive database of protein phosphorylation in plants. *Database.* 2014;2014:bau121.
44. Considine MJ, Foyer CH. Redox regulation of plant development. *Antioxid Redox Signal.* 2014;21(9):1305–26.
45. Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE. Mouse genome database (MGD) 2019. *Nucleic Acids Res.* 2019;47(D1):D801–6.
46. Dedow LK, Bailey-Serres J. Searching for a match: structure, function and application of sequence-specific RNA-binding proteins. *Plant Cell Physiol.* 2019;60(9):1927–38.
47. Martín G, Márquez Y, Mantica F, Duque P, Irimia M. Alternative splicing landscapes in Arabidopsis thaliana across tissues and stress conditions highlight major functional differences with animals. *Genome Biol.* 2021;22:1–26.
48. Li Y, Guo Q, Liu P, Huang J, Zhang S, Yang G, et al. Dual roles of the serine/arginine-rich splicing factor SR45a in promoting and interacting with nuclear cap-binding complex to modulate the salt-stress response in Arabidopsis. *New Phytol.* 2021;230(2):641–55.
49. Melo JP, Kalyna M, Duque P. Current challenges in studying alternative splicing in plants: the case of *physcomitrella patens* SR proteins. *Front Plant Sci.* 2020;11:286.
50. Esmailzadeh-Salestani K, Tohidfar M, Ghanbari Moheb Seraj R, Khaleghdoust B, Keres I, Marawne H, et al. Transcriptome profiling of barley in response to mineral and organic fertilizers. *BMC Plant Biol.* 2023;23(1):261.
51. Azad M, Tohidfar M, Ghanbari Moheb Seraj R, Mehralian M, Esmailzadeh-Salestani K. Identification of responsive genes to multiple abiotic stresses in rice (*Oryza sativa*): a meta-analysis of transcriptomics data. *Sci Rep.* 2024;14(1):5463.
52. Higgins CF. ABC transporters: physiology, structure and mechanism—an overview. *Res Microbiol.* 2001;152(3–4):205–10.
53. Suh SJ, Wang Y-F, Frelet A, Leonhardt N, Klein M, Forestier C, et al. The ATP binding cassette transporter AtMRP5 modulates anion and calcium channel activities in Arabidopsis guard cells. *J Biol Chem.* 2007;282(3):1916–24.
54. Kang J, Hwang J-U, Lee M, Kim Y-Y, Assmann SM, Martinoia E, et al. PDR-type ABC transporter mediates cellular uptake of the phytohormone abscisic acid. *Proc Natl Acad Sci U S A.* 2010;107(5):2355–60.
55. Matsuda S, Funabiki A, Furukawa K, Komori N, Koike M, Tokuiji Y, et al. Genome-wide analysis and expression profiling of half-size ABC protein sub-group G in rice in response to abiotic stress and phytohormone treatments. *Mol Genet Genomics.* 2012;287:819–35.
56. Moon S, Jung K-H. Genome-wide expression analysis of rice ABC transporter family across spatio-temporal samples and in response to abiotic stresses. *J Plant Physiol.* 2014;171(14):1276–88.
57. Jasiński M, Stukkens Y, Degand H, Purnelle B, Marchand-Brynaert J, Boutry M. A plant plasma membrane ATP binding cassette-type transporter is involved in antifungal terpenoid secretion. *Plant Cell.* 2001;13(5):1095–107.
58. Footitt S, Slocumbe SP, Larner V, Kurup S, Wu Y, Larson T et al. Control of germination and lipid mobilization by COMATOSE, the Arabidopsis homologue of human ALDP. *EMBO J.* 2002;21(12):2912–22.
59. Goodman CD, Casati P, Walbot V. A multidrug resistance-associated protein involved in anthocyanin transport in *Zea mays*. *Plant Cell.* 2004;16(7):1812–26.
60. Geisler M, Blakeslee JJ, Bouchard R, Lee OR, Vincenzetti V, Bandyopadhyay A, et al. Cellular efflux of auxin catalyzed by the Arabidopsis MDR/PGP transporter AtPGP1. *Plant J.* 2005;44(2):179–94.
61. Song W-Y, Park J, Mendoza-Cózatl DG, Suter-Grottemeyer M, Shim D, Hörtensteiner S, et al. Arsenic tolerance in Arabidopsis is mediated by two ABCG-type phytochelatin transporters. *Proc Natl Acad Sci U S A.* 2010;107(49):21187–92.
62. Do THT, Choi H, Palmgren M, Martinoia E, Hwang J-U, Lee Y. Arabidopsis ABCG28 is required for the apical accumulation of reactive oxygen species in growing pollen tubes. *Proceedings of the National Academy of Sciences.* 2019;116(25):12540–9.
63. Wu Y, Li X, Zhang J, Zhao H, Tan S, Xu W, et al. ERF subfamily transcription factors and their function in plant responses to abiotic stresses. *Front Plant Sci.* 2022;13:1042084.
64. Ouaked F, Rozhon W, Lecourieux D, Hirt H. A MAPK pathway mediates ethylene signaling in plants. *EMBO J.* 2003. <https://doi.org/10.1093/emboj/cdg131>.
65. Ma Z, Hu L, Jiang W. Understanding AP2/ERF transcription factor responses and tolerance to various abiotic stresses in plants: a comprehensive review. *Int J Mol Sci.* 2024;25(2): 893.
66. Licausi F, Ohme-Takagi M, Perata P. APETALA 2/ethylene responsive factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytol.* 2013;199(3):639–49.
67. Müller M, Munné-Bosch S. Ethylene response factors: a key regulatory hub in hormone and stress signaling. *Plant Physiol.* 2015;169(1):32–41.
68. Phukan UJ, Jeena GS, Tripathi V, Shukla RK. Regulation of Apetala2/Ethylene response factors in plants. *Front Plant Sci.* 2017;8:150.
69. Baharudin NF, Osman NI. Plant development, stress responses, and secondary metabolism under ethylene regulation. *Plant Stress.* 2023;7:100146.
70. Kong L, Song Q, Wei H, Wang Y, Lin M, Sun K, et al. The AP2/ERF transcription factor PtoERF15 confers drought tolerance via JA-mediated signaling in *Populus*. *New Phytol.* 2023;240(5):1848–67.
71. Blackshear P. Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. *Biochem Soc Trans.* 2002;30(6):945–52.
72. Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, et al. Induced NCRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature.* 2008;454(7200):126–30.
73. Liu C, Xu X, Kan J, ming Cheng Z, Chang Y, Lin J, et al. Genome-wide analysis of the C3H zinc finger family reveals its functions in salt stress responses of *pyrus Betulaefolia*. *PeerJ.* 2020;8:e9328.
74. Zheng L, Dai H, Mu Y, Li J, Cheng Y, Han J. Genome-wide identification and expression analysis of C3H gene family in melon. *Front Plant Sci.* 2025;16:1500429.
75. Sahid S, Roy C, Shee R, Datta R, Paul S. ZFP37, C3H, NAC94, and bHLH148 transcription factors regulate cultivar-specific drought response by modulating r40C1 gene expression in rice. *Environ Exp Bot.* 2023;214: 105480.
76. Arisha MH, Ahmad MQ, Tang W, Liu Y, Yan H, Kou M, et al. RNA-sequencing analysis revealed genes associated drought stress responses of different durations in hexaploid sweet potato. *Sci Rep.* 2020;10(1):12573.
77. Jiang A-L, Xu Z-S, Zhao G-Y, Cui X-Y, Chen M, Li L-C, et al. Genome-wide analysis of the C3H zinc finger transcription factor family and drought responses of members in *Aegilops Tauschii*. *Plant Mol Biology Report.* 2014;32(6):1241–56.
78. Cheng X, Cao J, Gao C, Gao W, Yan S, Yao H, et al. Identification of the wheat C3H gene family and expression analysis of candidates associated with seed dormancy and germination. *Plant Physiol Biochem.* 2020;156:524–37.

79. Liang Y, Ma F, Li B, Guo C, Hu T, Zhang M, et al. A *bHLH* transcription factor, *SibHLH96*, promotes drought tolerance in tomato. *Hortic Res.* 2022;9: uhac198.
80. Liu H, Song S, Liu M, Mu Y, Li Y, Xuan Y, et al. Transcription factor ZmNAC20 improves drought resistance by promoting stomatal closure and activating expression of stress-responsive genes in maize. *Int J Mol Sci.* 2023;24(5): 4712.
81. Li Y, Yang Z, Zhang Y, Guo J, Liu L, Wang C, et al. The roles of HD-ZIP proteins in plant abiotic stress tolerance. *Front Plant Sci.* 2022;13:1027071.
82. Liu H, Tang X, Zhang N, Li S, Si H. Role of bZIP transcription factors in plant salt stress. *Int J Mol Sci.* 2023;24(9): 7893.
83. Jiao P, Wei X, Jiang Z, Liu S, Guan S, Ma Y. ZmLBD2 a maize (*Zea mays* L.) lateral organ boundaries domain (LBD) transcription factor enhances drought tolerance in transgenic *Arabidopsis thaliana*. *Front Plant Sci.* 2022;13:1000149.
84. Lv K, Wei H, Liu G. A R2R3-MYB transcription factor gene, BpMYB123, regulates BpLEA14 to improve drought tolerance in *betula platyphylla*. *Front Plant Sci.* 2021;12:791390.
85. Hartman S. NIN-like Protein7 is controlled by oxygen and nitric oxide and contributes to stress tolerance through PROTEOLYSIS6. Oxford University Press; 2021.
86. Zhang J, Huang D, Zhao X, Zhang M, Wang Q, Hou X, et al. Drought-responsive *WRKY* transcription factor genes *IgWRKY50* and *IgWRKY32* from *Iris Germanica* enhance drought resistance in transgenic *Arabidopsis*. *Front Plant Sci.* 2022;13:983600.
87. Rabeh K, Hnini M, Oubohssaine M. A comprehensive review of transcription factor-mediated regulation of secondary metabolites in plants under environmental stress. *Stress Biology.* 2025;5(1):15.
88. Yan T, Shu X, Ning C, Li Y, Wang Z, Wang T, et al. Functions and regulatory mechanisms of bHLH transcription factors during the responses to biotic and abiotic stresses in woody plants. *Plants.* 2024;13(16):2315.
89. Gaude AA, Jalmi SK. Environmental stress induced biosynthesis of plant secondary metabolites-transcriptional regulation as a key. *Crop Design.* 2025. <https://doi.org/10.1016/j.crope.2025.100100>.
90. Muhammad N, Uddin N, Khan MKU, Ali N, Ali K, Jones DA. Diverse role of basic helix-loop-helix (bHLH) transcription factor superfamily genes in the fleshy fruit-bearing plant species. *Czech J Genet Plant Breed.* 2022. <https://doi.org/10.17221/2/2022-CJGPB>.
91. Qian Y, Zhang T, Yu Y, Gou L, Yang J, Xu J, et al. Regulatory mechanisms of bHLH transcription factors in plant adaptive responses to various abiotic stresses. *Front Plant Sci.* 2021;12:677611.
92. Guo J, Sun B, He H, Zhang Y, Tian H, Wang B. Current understanding of bHLH transcription factors in plant abiotic stress tolerance. *Int J Mol Sci.* 2021;22(9): 4921.
93. Wang P, Wu X, Li N, Nie H, Ma Y, Wu J, et al. The StbHLH47 transcription factor negatively regulates drought tolerance in potato (*Solanum tuberosum* L.). *BMC Plant Biol.* 2025;25(1):14.
94. AbouZid S, Ahmed OM. Silymarin flavonolignans: structure–activity relationship and biosynthesis. *Stud Nat Prod Chem.* 2013;40:469–84.
95. Lee JW, Zoclanclounon YAB, Jung HJ, Lee TH, Kim JG, Park GH, et al. Probable biosynthetic pathways of silymarin precursors. *Korean J Med Crop Sci.* 2022;30(5):347–56.
96. ElSayed AI, El-Hamamy MA, Rafudeen MS, Mohamed AH, Omar AA. The impact of drought stress on antioxidant responses and accumulation of flavonolignans in milk thistle (*Silybum marianum* (L.) Gaertn.). *Plants.* 2019;8(12): 611.
97. Keshavarz Afshar R, Chaichi MR, Ansari Jovini M, Jahanzad E, Hashemi M. Accumulation of silymarin in milk thistle seeds under drought stress. *Planta.* 2015;242:539–43.
98. Majidi MM, Shafiei-Koij F, Pirnajmedin F, Jami M, Radan Z. Fatty acid profile, silymarin content and production properties of milk thistle (*Silybum marianum*) germplasm under different water environments. *Crop Pasture Sci.* 2021;72(4):302–10.
99. El-Garhy HA, Khattab S, Moustafa MM, Abou Ali R, Azeiz AZA, Elhalwagi A, et al. Silybin content and overexpression of chalcone synthase genes in *Silybum marianum* L. plants under abiotic elicitation. *Plant Physiol Biochem.* 2016;108:191–202.
100. Torres M, Corchete P. Gene expression and flavonolignan production in fruits and cell cultures of *Silybum marianum*. *J Plant Physiol.* 2016;192:111–7.
101. Petrucci E, Braidot E, Zancani M, Peresson C, Bertolini A, Patui S, et al. Plant flavonoids—biosynthesis, transport and involvement in stress responses. *Int J Mol Sci.* 2013;14(7):14950–73.
102. Dao T, Linthorst H, Verpoorte R. Chalcone synthase and its functions in plant resistance. *Phytochem Rev.* 2011;10:397–412.
103. Cong L, Deng L, Yao H, Zhang Y, Li H, Wang H, et al. Responses of the lipoxygenase gene family to drought stress in broomcorn millet (*Panicum miliaceum* L.). *Genes.* 2025;16(4):368.
104. Xu Y, Fu S, Huang Y, Zhou D, Wu Y, Peng J, et al. Genome-wide expression analysis of LACS gene family implies *GhLACS25* functional responding to salt stress in cotton. *BMC Plant Biol.* 2024;24(1):392.
105. Partridge M, Murphy DJ. Roles of a membrane-bound caleosin and putative peroxigenase in biotic and abiotic stress responses in *Arabidopsis*. *Plant Physiol Biochem.* 2009;47(9):796–806.
106. Lee OR, Kim SJ, Kim HJ, Hong JK, Ryu SB, Lee SH, et al. Phospholipase A2 is required for PIN-FORMED protein trafficking to the plasma membrane in the *Arabidopsis* root. *Plant Cell.* 2010;22(6):1812–25.
107. Raffaele S, Vailleau F, Leger A, Joubès J, Miersch O, Huard C, et al. A MYB transcription factor regulates very-long-chain fatty acid biosynthesis for activation of the hypersensitive cell death response in *Arabidopsis*. *Plant Cell.* 2008;20(3):752–67.
108. Zhang Q, Ahmad N, Li Z, He J, Wang N, Naeem M, et al. CtCYP71A1 promotes drought stress tolerance and lignin accumulation in safflower and *Arabidopsis*. *Environ Exp Bot.* 2023;213: 105430.
109. Wei X, Mao L, Lu W, Wei X, Han X, Guan W, et al. Three transcription activators of ABA signaling positively regulate suberin monomer synthesis by activating cytochrome P450 CYP86A1 in kiwifruit. *Front Plant Sci.* 2020;10:1650.
110. Ghanbari Moheb Seraj R, Behnamian M, Ahmadikhah A, Shariati V, Dezhsetan S. Chitosan and salicylic acid regulate morpho-physiological and phytochemical parameters and improve water-deficit tolerance in milk thistle (*Silybum marianum* L.). *Acta Physiol Plant.* 2021;43:1–17.
111. Ghanbari Moheb Seraj R, Behnamian M, Ahmadikhah A, Esmaeilzadeh-Salestani K. Foliar application of chitosan and salicylic acid improved milk thistle's tolerance to drought stress. *Russ J Plant Physiol.* 2024;71(3):62.
112. Boutraa T. Improvement of water use efficiency in irrigated agriculture: a review. *J Agron.* 2010;9(1):1–8.
113. Andrews S. FastQC: a quality control tool for high throughput sequence data. Cambridge, United Kingdom; 2010.
114. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20.
115. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29(7):644–52.
116. UniProt. The universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021;49(D1):D480–9.
117. Lamesch P, Berardini TZ, Li D, Swarbrick D, Wilks C, Sasidharan R, et al. The *Arabidopsis* information resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 2012;40(D1):D1202–10.
118. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30.
119. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 2007;35(suppl2):W182–5.
120. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, et al. AmiGO: online access to ontology and annotation data. *Bioinformatics.* 2009;25(2):288–9.
121. Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics.* 2015;16:1–7.
122. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:1–16.
123. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 2012;40(10):4288–97.
124. Team R. RStudio: integrated development for R. RStudio, Inc, Boston, MA. 2015;700:879.
125. Rychlik W. OLIGO 7 primer analysis software. *PCR Primer Des.* 2007;402:35–60.
126. Lu G, Moriyama EN, Vector NTI. A balanced all-in-one sequence analysis suite. *Brief Bioinform.* 2004;5(4):378–88.
127. Ruijter J, Ramackers C, Hoogaars W, Karlen Y, Bakker O, Van den Hoff M, et al. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res.* 2009;37(6):e45–e.
128. Pfaffl MW, Horgan GW, Dempfle L. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* 2002;30(9):e36–e.
129. Gad D, El-Shora H, Fraternali D, Maricchiolo E, Pompa A, Dietz K-J. Bioconversion of callus-produced precursors to silymarin derivatives in *Silybum*

- marianum* leaves for the production of bioactive compounds. *Int J Mol Sci.* 2021;22(4): 2149.
130. Lv Y, Gao S, Xu S, Du G, Zhou J, Chen J. Spatial organization of silybin biosynthesis in milk thistle [*Silybum marianum* (L.) Gaertn]. *Plant J.* 2017;92(6):995–1004.
131. Falcone Ferreyra ML, Rius SP, Casati P. Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front Plant Sci.* 2012;3:222.
132. Team RC. R foundation for statistical computing. Vienna Austria. 2013;3(0):37.

133. Allaire J, Boston. MA. 2012;770(394):165–71.

Publisher's note

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