



Unravelling the molecular mechanisms of vegetative-to-reproductive transition in *Cynara cardunculus* by RNA-Seq analysis

A. Paulino^{1,2,3} · I. Fernandes² · R. C. Pires¹ · A. Usié^{1,3} · A. Faustino^{1,3,4} · J. Santos^{1,3} · T. Brás^{1,3} · D. Rosa^{1,3,5} · O. S. Paulo² · M. F. Duarte^{1,3} · L. Marum^{1,3}

Received: 22 July 2025 / Accepted: 29 December 2025 / Published online: 31 January 2026
© The Author(s) 2026

Abstract

Cynara cardunculus inflorescence plays a significant role in cheese manufacturing and human consumption due to the milk-clotting capacity and the production of edible immature capitula. Consequently, it is crucial to understand the complex cellular and molecular processes involved in the development of cardoon flowers. The current study used western wild cardoon in two developmental stages, vegetative and reproductive, to understand the molecular mechanisms mediating cardoon's vegetative-to-reproductive transition. Transcriptome profiling identified 552 differentially expressed genes (DEGs) between vegetative and reproductive stages, with 321 upregulated in stage 4 (vegetative state) and 231 in stages 5/6 (reproductive state). Gene Ontology (GO) analysis revealed DEGs involved in biological processes such as oxidation-reduction, lipid metabolism, and defense responses. Metabolic pathways, including phenylpropanoid biosynthesis and anther/pollen development, were significantly enriched. Genes associated with lignin formation, flavonoid synthesis, and acetyltransferase activity were upregulated in the vegetative stage. At the same time, those related to flower development and sesquiterpene lactone biosynthesis were prominent in the reproductive stage. Transcription factors, including MADS-box and ethylene-responsive families, were critical in regulating developmental transitions. This study provides comprehensive insights into the morphological, chemical, and molecular dynamics of *C. cardunculus* across its phenological stages, underscoring the plant's adaptability and potential for pharmaceutical and industrial applications. The findings highlight the importance of seasonal and developmental timing in optimizing the production of valuable secondary metabolites, such as cynaropicrin.

Key message

Understanding molecular changes in *Cynara cardunculus* during flowering reveals key genes and pathways, aiding flower development and metabolite production, with applications in cheese production, pharmaceuticals, and industry.

Keywords Cardoon · Inflorescence development · Defense response · Cynaropicrin · Transcriptome

✉ L. Marum
lmarum@uevora.pt

¹ Alentejo Biotechnology Center for Agriculture and Agro-Food (CEBAL), Polytechnic University of Beja (IPBeja), Beja, Portugal

² cE3c - Centre for Ecology, Evolution and Environmental Changes and Change - Global Change and Sustainability Institute, Biology Department, Faculty of Sciences, University of Lisbon, P-1749-016 Lisboa, Portugal

³ MED – Mediterranean Institute for Agriculture, Environment and Development and CHANGE – Global Change and Sustainability Institute, CEBAL, Beja, Portugal

⁴ MED – Mediterranean Institute for Agriculture, Environment and Development, Institute for Research and Advanced Training (IIFA), University of Évora, Ap. 94, Évora, Portugal

⁵ Allelopathy Group, Department of Organic Chemistry, INBIO Institute of Biomolecules, Campus de Excelencia Internacional Agroalimentario (ceiA3), University of Cádiz, Puerto Real, Cádiz, Spain

Introduction

Cynara cardunculus L., commonly known as cardoon, is a perennial herbaceous plant native to the Mediterranean Basin and comprises a complex of wild and cultivated forms that are ecologically and economically significant. Belongs to the Compositae (Asteraceae), the largest and most successful dicot botanical family (Portis et al. 2019) and is valued for its applications in the pharmaceutical, food, and bioenergy industries (Raccuia and Melilli 2007). Taxonomically, this species includes the wild type (*C. cardunculus* var. *sylvestris*) with distinctive morphological types (eastern and western wild cardoons), and two cultivated varieties, *C. cardunculus* var. *altilis* (cultivated cardoon) and *C. cardunculus* var. *scolymus* (globe artichoke) (Wiklund 1992; Rottenberg and Zohary 1996; Sonnante et al. 2007; Gatto et al. 2013).

The eastern wild cardoon (*C. cardunculus* subsp. *cardunculus*) is predominantly distributed in the eastern Mediterranean region, including Greece, Cyprus, southern Italy, Sicily, and Sardinia. This subspecies exhibits reduced plants and floral heads, as well as elongated spiny bracts, and is widely regarded as the ancestral progenitor of cultivated artichoke (*C. cardunculus* var. *scolymus*) (Rottenberg and Zohary 1996). In contrast, western wild cardoon (*C. cardunculus* subsp. *Flavescens* Wiklund), distributed throughout the western Mediterranean, including Portugal, Spain, France, and the Canary Islands consists of more robust individuals with yellow-margined bracts and fewer, shorter spines (Gatto et al. 2013; Castro et al. 2021). It is hypothesized to be genetically closer to cultivated cardoon (*C. cardunculus* var. *altilis*), possibly representing an intermediate form between wild and domesticated types (Wiklund 1992; Sonnante et al. 2008; Gatto et al. 2013).

Cultivated varieties of *C. cardunculus* are of high agricultural value. *C. cardunculus* var. *altilis* is primarily grown for its edible stalks in the Iberian Peninsula, southern France, and Italy, and is also used together with wild types in traditional cheese production as a source of plant-based milk coagulants (Sonnante et al. 2007). *C. cardunculus* var. *scolymus* (globe artichoke), characterized by large, tender floral heads and reduced spininess, is widely cultivated throughout Mediterranean and temperate regions and is a significant horticultural crop (Portis et al. 2005; Lattanzio et al. 2009; Rau et al. 2016).

C. cardunculus exhibits a complex annual growth cycle, developing rosette-form leaves in fall and winter, and transitioning to stalk emergence and flowering in late spring. Harvesting occurs in early summer, with seeds maturing by late summer (Gominho et al. 2011). Archontoulis et al. (2010) were the first to develop a simple coding system based on the Biologische Bundesanstalt, Bundessortenamt,

Chemische Industrie (BBCH) scale for *C. cardunculus* phenological growth stages, identifying several major stages, including the differentiation of the vegetative phase from stages 0 to 4, and the reproductive phase from stages 5 to 9 (Archontoulis et al. 2010). Like other flowering plants, its inflorescence development is complex and highly regulated, involving intricate molecular and cellular events and significant genetic and environmental interactions (Harder and Prusinkiewicz 2013). Furthermore, the transition from vegetative to reproductive stages involves complex genetic networks that coordinate flower development, defense responses, and secondary metabolite synthesis (Krizek and Fletcher 2005). Secondary metabolites play a crucial role in protecting floral development by mediating responses to biotic and abiotic stressors and influencing flower morphology and differentiation (Divekar et al. 2022). Our research team has described the production of secondary metabolites, such as cynaropicrin, for the first time (Ramos et al. 2013, 2019). Cynaropicrin has an enormous biological potential (Velez et al. 2012; Brás et al. 2015; Ramos et al. 2017) with bioactive properties (Eljounaidi et al. 2015), which vary significantly across developmental stages and genotypes, highlighting the importance of seasonal and phenological monitoring (Petropoulos et al. 2019). Understanding the molecular events underlying the reproductive transition in cardoon is essential for its effective breeding and conservation. As far as we know, to date, only one transcriptomic study has been carried out in cultivated cardoon, revealing the first insights of gene expression dynamics associated with the biosynthesis of valuable secondary metabolites such as phenylpropanoids, flavonoids, and sesquiterpene lactones during inflorescence development (Puglia et al. 2020).

Therefore, the present study aimed to assess for the first time the differences in gene expression patterns between the vegetative and reproductive phases using an RNA-Seq approach, providing new insights into the regulatory networks of flowering and secondary metabolites biosynthesis of wild cardoon, namely in western type.

Materials and methods

Plant material and morphological evaluation

To evaluate the developmental stage (transition from vegetative to reproductive growth), seven genotypes of western wild cardoons (*C. cardunculus* L. subsp. *flavescens* Wilk.) located in experimental fields in Beja, Portugal (37.70622, −8.08808), were monitored and morphologically evaluated using the BBCH scale (Archontoulis et al. 2010), from the beginning of spring (March) until the end

of May 2020. Based on this morphological evaluation, leaf samples of the two developmental stages (stage 4, corresponding to the development of harvestable vegetative plant parts, and stages 5/6, corresponding to the inflorescence emergence and development (stage 5) and flowering and capitulum formation (stage 6) were collected. Plant height, foliar area, and number of inflorescences metrics were used to evaluate plant growth for each seasonal period.

The plant material was collected in 2020 during two seasonal periods (March, stage 4; and May, stage 5/6) for chemical and molecular analyses. The plant material was air-dried at room temperature until dry, and then ground with a home grinder (Moulinex) for chemical procedures. At the same time, samples for RNA extraction were flash-frozen in liquid nitrogen on-site and kept at -80°C until use.

Cynaropicrin quantification

To evaluate the biochemical changes associated with flowering and capitulum formation in wild cardoon, the sesquiterpene lactone cynaropicrin was examined at stages 4 and 5/6 of development. Cynaropicrin extraction was performed according to the previously described (Brás et al. 2020). The cynaropicrin concentration of *C. cardunculus* leaf extracts was determined using High-Performance Liquid Chromatography (HPLC, Dionex Ultimate 3000, Thermo Scientific, Waltham, Massachusetts, U.S.) with a standard curve constructed from known concentrations of the compound (0.1–0.5 mg/mL). The curve showed excellent linearity ($R^2=0.997$) (Fig. S1). Ethanol absolute anhydrous for analysis (Carlos Herba Reagents, France) was used for extraction experiments, and Cynaropicrin concentration was analyzed with Milli Q water and acetonitrile ($\geq 99.9\%$), HPLC gradient grade (Fisher Chemical). Cynaropicrin ($\geq 97.5\%$ purity) was obtained from Extrasynthese, Genay, France. The results were given in mg/g of dry weight (DW).

RNA extraction and sequencing

Total RNA was extracted from *C. cardunculus* fresh leaves sampled using the RNAqueous-4PCR RNA Isolation kit (Invitrogen™), according to the manufacturer's protocol with minor modifications. The concentration and integrity of RNA were evaluated by UV–Vis spectrophotometer (NanoVueplus, GE Healthcare Life Sciences, USA), agarose gel electrophoresis (1%), and Agilent BioAnalyzer system (BioRad, USA). High-quality RNA was processed for stranded paired-end (PE) cDNA library construction (DNB-SEQ Transcriptome), which was then subjected to sequencing through the DNBseq platform to produce PE reads of 150 bps in length at BGI (Shenzhen, China).

RNA sequencing data analysis

Raw reads were subjected to a quality evaluation using FastQC v.0.11.9 (Andrews 2010) and MultiQC v.1.12 (Ewels et al. 2016) to better define the quality thresholds to apply. Pre-processing was performed with Trimmomatic v.0.39 (Bolger et al. 2014) to remove adaptor/primer contaminations and low-quality reads with the following parameters: *ILLUMINACLIP:TruSeq3-PE-3.fa:2:30:10:2:keepBothReads MINLEN:36*. Before running FastQ Screen v.0.14.0 (Wingett and Andrews 2018) to check for contaminants against the genome of the most common model organisms, the reference genome of *C. cardunculus* v.2 was downloaded from the Global Artichoke Genome Database (<http://www.artichokegenome.unito.it>, accessed on 9th March 2022) (Acquadro et al. 2020) and included in the set of genomes used for comparison in FastQ Screen. The reads surviving the quality procedures were mapped against the cardoon genome using STAR v.2.6.1 with default parameters (Dobin et al. 2013). Only the unique mapped reads were considered for gene quantification with Htseq-count v.1.99.2 (Anders et al. 2015). Statistical analysis of differential gene expression was performed with DESeq2 v.1.28.1 (Love et al. 2014), under R v.4.0.2 (R Core Team 2020) to identify significant differences in the relative abundance of expressed genes between different development stages. Benjamini and Hochberg's approach was used for controlling the false discovery rate, FDR (Benjamini and Hochberg 2000). Differentially expressed genes (DEGs) were defined as those with an absolute $\log_2$ fold-change ($\log_2\text{FC}$) ≥ 2 and an $\text{FDR} < 0.05$. The associations between DEGs and GO terms were visualized with Cytoscape v.3.9.1 (Shannon et al. 2003). Additionally, DEGs were annotated with the Combined Pathway function of OmicsBox v.3.029 (BioBam) against the Reactome and KEGG databases (2019).

Phylogenetic analysis

Protein sequences of 21 biological relevant DEGs associated with the vegetative-to-reproductive transition were retrieved from the corresponding genome annotation ($|\log_2\text{FC}| \geq 2$). The 21 genes analyzed in this study were selected as representative candidates based on extreme differential expression across functional categories, encompassing processes such as secondary metabolism, transcriptional regulation, signaling, defense, and development. This strategy allowed investigation of both their evolutionary relationships and regulatory control. The sequences were aligned using MAFFT v7 (Katoh & Standley 2013), and the alignment was manually inspected to ensure the correct positioning of conserved regions. Phylogenetic analysis was performed

using the Maximum Likelihood (ML) method implemented in IQ-TREE v2 (Nguyen et al. 2015). The resulting tree was visualized and edited using FigTree v1.4.4 (Rambaut 2018).

Protein motif identification

Conserved motifs in the 21 protein sequences were identified using the MEME Suite v5.5.0 (Bailey et al. 2009). Analyses were performed in protein mode with the zero or one occurrence per sequence (ZOOPS) model. The maximum number of motifs was set to 10, with motif widths ranging from 6 to 50 amino acids. All other parameters were kept at default values. The resulting motifs were visualized as sequence logos.

Promoter sequence and cis-regulatory element analysis

De novo motif discovery in the promoter sequences was also performed using MEME v5.5.0, in DNA mode. The ZOOPS model was applied with a maximum of 10 motifs, and motif widths ranging from 6 to 15 bp. These motifs were used to identify conserved cis-regulatory patterns among the 21 promoters. The 1500 bp promoter sequences were further analyzed using the PlantCARE database (Lescot et al. 2002) to identify putative cis-acting regulatory elements. All detected elements were classified into functional categories, including core promoter elements, hormone-responsive elements, stress-responsive elements and light-responsive elements. The total number and distribution of regulatory elements were recorded for each gene.

Correlation analysis

To assess the relationship between metabolite accumulation and transcriptional changes, we performed a metabolite–transcriptome correlation analysis using matched cynaropicrin quantification and RNA-seq expression data from the same plants. Spearman's rank correlations (ρ) were computed between cynaropicrin levels and the expression of all detected genes. P -values were adjusted using the Benjamini–Hochberg false discovery rate (FDR), and genes with $|\rho| \geq 0.70$ and $FDR < 0.05$ were considered significantly associated. Significant genes were visualised using heatmaps (pheatmap package in R).

Statistical analysis

For morphological and chemical analysis of each plant, three samples ($n=3$) were analyzed, and all the assays were carried out in triplicate. Data were analyzed with one-way

analysis of variance (ANOVA), while mean comparison was performed with a t -test ($p \leq 0.05$).

Results

Morphological evaluation

Cynara cardunculus is a common perennial plant mainly distributed in the Mediterranean. The morphological evaluation was performed during spring where the vegetative to reproductive transition stages occur, according to Archontoulis et al (2010). Based on the results obtained, two sampling points were defined: (1) in March, when the adult leaves were expanded entirely and developed, and no inflorescence was emerging, being in accordance with stage 4 of BBCH scale (Fig. 1a); and (2) in May, when the emergence of inflorescences was observed, stage 5/6 of BBCH scale (Fig. 1b, c).

At stage 4, a maximum and a minimum height of 102 cm and 22 cm, respectively, were obtained, while the foliar area ranged from a maximum of $5062 \pm 614 \text{ cm}^2$ to a minimum of $1684 \pm 371 \text{ cm}^2$ (Table S1). At stage 5/6, it was possible to observe the emergence of inflorescences in some of the genotypes sampled. The number of inflorescences (flower buds with main and secondary capitulum) varied from 0 to 12, including immature green buds and capitulum in full flowering. The range of inflorescence types/stages that emerged may indicate some lack of synchronization in flowering time. A higher plant height mean ($62 \pm 40 \text{ cm}$) was obtained in stages 5/6 when compared with stage 4 ($47 \pm 26 \text{ cm}$), while a slight decrease in the foliar area was observed from ($2540 \pm 1166 \text{ cm}^2$) to ($2414 \pm 1794 \text{ cm}^2$), in stage 4 to stage 5/6, respectively, despite no significant difference being detected ($p > 0.05$) (Fig. 2a, b).

Cynaropicrin quantification

C. cardunculus presents several important chemical compounds, including cynaropicrin, a bitter alkaloid found in various species of the *Cynara* genus, and the most representative compound in leaves (Ramos et al. 2013) with applications in both the pharmaceutical and food industries.

The results obtained from the different plants in the two collection months of 2020 show significant variability in cynaropicrin content, ranging from 15.8 ± 1.8 to 63.1 ± 1.0 , depending on the genotype and collection period (Table 1). A representative chromatogram of a cardoon sample is shown in Figure S2, where the cynaropicrin peak is clearly identified (retention time $\sim 11 \text{ min}$).

Fig. 1 Phenological growth stages of *Cynara cardunculus* based on the BBCH scale. **A** represents stage 4, while **B** and **C** represents stages 5/6

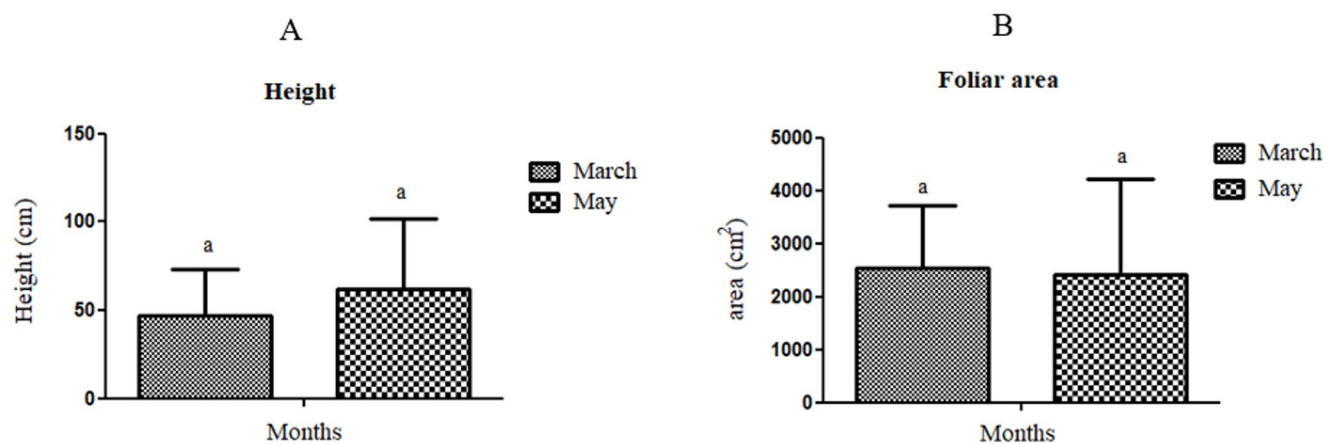


Fig. 2 Morphological evaluation. **A** represents height (cm) while **B** represents foliar area (cm²). Each bar represents the mean $\pm$ SD (t-test, $p > 0.05$) (n=7)

Table 1 Summary of *Cynara cardunculus* samples used in transcriptome study by two developmental stages (May (stage 5/6) versus March (stage 4))

Collection	Development stage	Sample ID	Cyn (mg/g DW)
March	Stage 4	1	33,2 $\pm$ 4,1
March	Stage 4	3	15,8 $\pm$ 1,8
March	Stage 4	5	20,3 $\pm$ 1,3
March	Stage 4	7	18,8 $\pm$ 0,7
March	Stage 4	9	20,2 $\pm$ 2,0
March	Stage 4	11	23,2 $\pm$ 1,2
March	Stage 4	13	33,5 $\pm$ 0,8
May	Stage 5/6	2	37,7 $\pm$ 2,6
May	Stage 5/6	4	56,1 $\pm$ 4,2
May	Stage 5/6	6	72,5 $\pm$ 3,6
May	Stage 5/6	8	63,1 $\pm$ 1,0
May	Stage 5/6	10	60,5 $\pm$ 4,2
May	Stage 5/6	12	24,9 $\pm$ 0,5
May	Stage 5/6	14	61,8 $\pm$ 6,8

The cynaropicrin concentration mean in *C. cardunculus*

plants was significantly higher in stage 5/6 compared to stage 4 ($p \leq 0,05$) (Fig. 3).

Transcriptome profiling

The sequencing run yielded approximately 47 million raw reads per sample. After applying the pre-processing procedures, 99.25% of reads were kept across the dataset, and no contamination from other organisms was detected (see Table 2 and Fig. S3). Mapping results for high-quality reads revealed that 95% of them mapped against the reference genome, with 91% being unique mapped reads (Table 2).

The number of genes expressed per sample was determined using quantification data from Htseq-count, considering only genes with a minimum abundance of 5 for further analyses. Hence, the total number of genes expressed in stage 4 was 22,446, whereas in stage 5/6 it was 22,687 (Table S2).

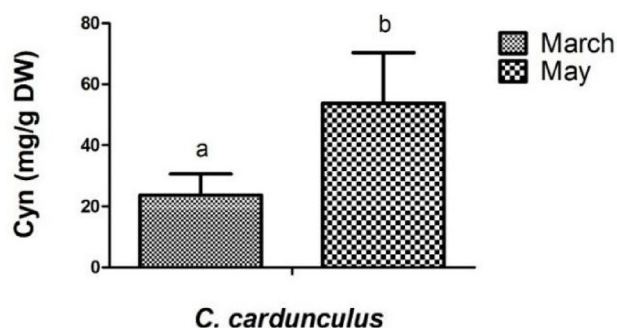


Fig. 3 Results from cynaropicrin (Cyn) content (mg/g DW) of different *Cynara cardunculus* populations compared March (reproductive stage—stage 4) to May (vegetative stage—stage 5/6) 2020. Each bar represents the mean $\pm$ SD, and different letters indicate significant differences (t-test, $p \leq 0.05$)

The PCA separates the samples from stages 4 and stage 5/6. (Fig. 4). However, the plants in stage 5/6 show a greater dispersion when compared with plants in stage 4, which could be related with in stage 4, the plants are in the vegetative development phase, whereas in stages 5/6, they transition into reproductive stages, with inflorescence emergence and development occurring in stage 5, followed by flowering and capitulum formation in stage 6.

Identification and classification of DEGs

The gene expression profiles of stage 4 and stage 5/6 were compared to determine their expression profile differences. As a result of the analyses 552 differentially expressed genes (DEGs) were identified, from which 321 were more expressed in stage 4 and 231 in stage 5/6. The 88.0% of DEGs had a functional annotation, whereas the remaining were proteins with unknown functions. Regarding Gene Ontology (GO) terms, 383 (64.4%) DEGs had at least one GO term associated. The total number of GO represented in the DEGs set was 220 (Biological Processes: 80; Molecular Function: 119; Cellular Component: 21). The GOs associated with a higher number of differentially expressed genes are represented in Table 3.

The DEGs identified were mainly involved in biological processes such as obsolete oxidation–reduction process, regulation of DNA-templated transcription, lipid metabolic process, carbohydrate metabolic process, defense response, protein phosphorylation, transmembrane transport, and proteolysis. The genes involved in the molecular function were shown to be differentially regulated by oxidoreductase activity, and binding activities including the DNA-binding transcription factor (TF) activity, hydrolase activity, protein dimerization activity, and protein kinase activity.

The annotation against Reactome and KEGG pathways resulted in 378 (68.5%) DEGs associated with at least one of those metabolic pathways. The large number of DEGs

Table 2 Summary statistics of sequencing data and mapping of reads from Ce samples, identified according to each sample ID

Sample ID	Accession number (NCBI)	Raw reads (BGI)	Clean reads (%)	% Uniquely mapped reads	% multi-mapped reads	% Unmapped reads
1	SAMN35325406	46 993 406	46 626 736 (99.22)	91.62	4.66	3.72
2	SAMN35325407	48 002 372	47 616 340 (99.20)	92.06	3.34	4.60
3	SAMN35325401	46 993 378	46 636 194 (99.24)	90.70	3.03	6.27
4	SAMN35325402	41 594 992	41 221 212 (99.10)	91.86	4.53	3.61
5	SAMN35325393	47 103 680	46 737 516 (99.22)	91.61	4.79	3.60
6	SAMN35325394	47 500 532	47 154 052 (99.27)	88.39	4.25	7.36
7	SAMN35325385	46 865 244	46 500 376 (99.22)	90.99	4.53	4.47
8	SAMN35325386	47 757 296	47 354 516 (99.16)	90.98	4.49	4.53
9	SAMN35325379	47 036 012	46 789 390 (99.48)	92.12	3.88	3.99
10	SAMN35325397	46 988 390	46 617 012 (99.21)	92.28	3.72	3.99
11	SAMN35325391	47 758 670	47 399 190 (99.25)	90.58	5.03	4.38
12	SAMN35325383	47 779 774	47 400 222 (99.21)	88.42	2.72	8.86
13	SAMN35325378	47 308 424	47 077 464 (99.51)	92.48	4.57	2.95
14	SAMN35325381	47 607 874	47 249 124 (99.25)	90.06	4.11	5.83

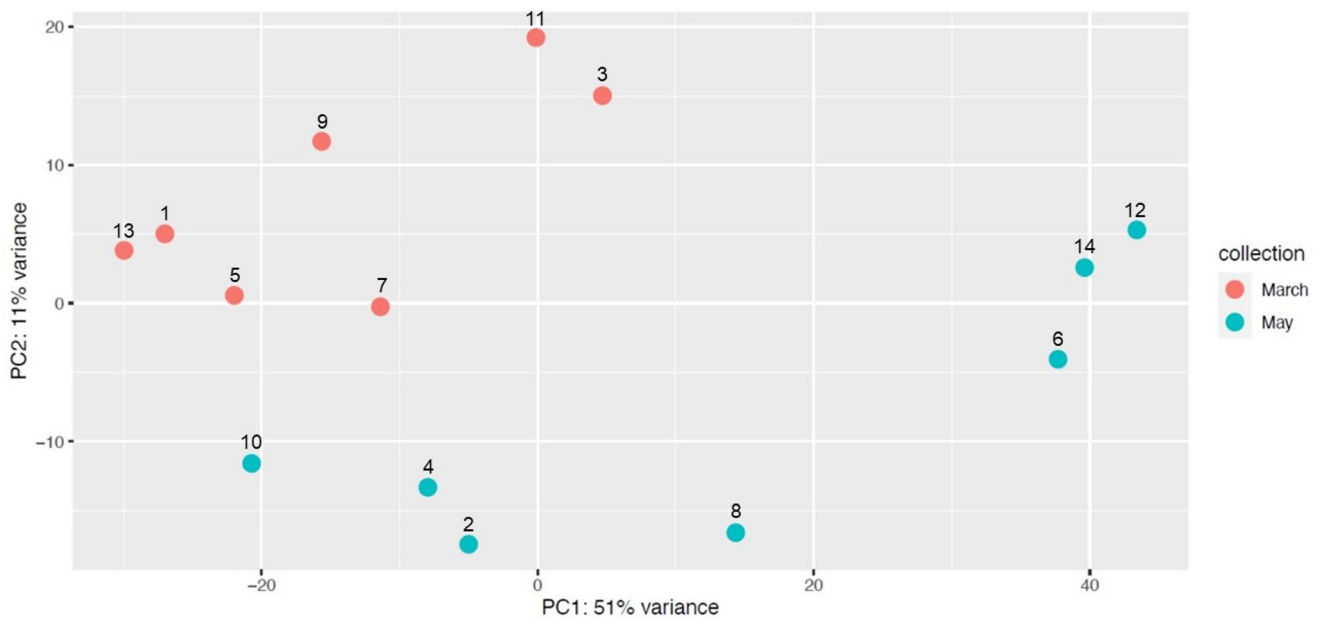


Fig. 4 Principal component analysis (PCA) of normalized gene expression data from *Cynara cardunculus*. Two developmental stages were evaluated (May (stage 5/6) versus March (stage 4)). Colors represent different months' collections. The percentage of variance is indicated on each axis

related to the synthesis of plant secondary metabolites attests to their importance. The metabolic pathways with more DEGs associated are represented in Fig. 5, in which phenylpropanoid biosynthesis and anther/pollen development present some of the top metabolic pathways.

Genes highly up-regulated in the vegetative stage (stage 4)

Due to the high number of genes identified, the $|\log_2FC| \geq 2$ was raised to enhance DEGs with the highest differences between the two developmental stages.

Differentially expressed genes associated with the pathways of flower development and the transition from vegetative to reproductive shoot apical meristems are represented in Fig. 6. Within the DEGs identified in *C. cardunculus*, MADS-box protein EJ2 (*EJ2*), Agamous-like MADS-box AP1 (*API*), Agamous-like MADS-box AGL8 homolog (*TDR4*), MADS-box protein AeAP3-3 (Fragment) (*AP3-2*), Agamous-like MADS-box protein AGL65 (*AGL65*), Agamous-like MADS-box protein AGL29 (*AGL29*), MADS-box transcription factor 50 (*MADS50*), Basic leucine zipper 43 (*BZIP43*), and B-box zinc finger protein 32 (*BBX32*) were identified as up-regulated in stage 4 (vegetative phase).

Regarding the molecular function DNA binding transcription factor activity (GO:0003700) (Fig. 7) several TFs such as transcription factor ORG2 (*ORG2*), Heat stress transcription factor A-4c (*HSA44C*), Heat stress transcription factor A-2c (*HSA22C*), Heat stress transcription factor A-2b (*HSA22B*), and Ethylene-responsive transcription

factor 2 (*ERF2*), Ethylene-responsive transcription factor 1B (*ERF1B*), and WRKY transcription factor SUSIBA2 (*WRKY46*) were found upregulated in stage 4.

Our data revealed also several DEGs up-regulate in stage 4 (vegetative phase) linked to the anther and pollen development pathway (Fig. 8), such as Receptor-like protein 35 (*RLP35*), Receptor-like protein 49 (*RLP49*; codified by two DEGs), Senescence-specific cysteine protease SAG39 (*OsI_14861*), MDIS1-interacting receptor-like kinase 2 (*MIK2*), BURP domain protein RD22 (*RD22*), Receptor-like protein 7 (*RLP7*), Receptor-like protein Cf-9 homolog (*HCR9-0*), Receptor-like protein EIX2 (*EIX2*), Plant intracellular Ras-group-related LRR protein 5 (*IRL5*), Probable leucine-rich repeat receptor-like protein kinase At1g35710 (*At1g35710*), and Probable receptor-like protein kinase At5g59700 (*At5g59700*).

Considering the biological processes, Defense Response (GO:0006952) and abscisic acid-activated signaling pathway (GO:0009738), the Defensin-like protein 1, Anther-specific protein SF18 (Fragment), MLO-like protein 12 (*MLO12*), Kirola, and Root allergen protein (codified by eight DEGs) were found upregulated in stage 4 (vegetative stage) (Fig. 9).

The phenylpropanoid biosynthesis pathway with the higher number of DEGs associates is represented in Fig. 10, in which the Lignin-forming anionic peroxidase, Peroxidase 72 (*PER72*), Peroxidase 12 (*PER12*), Anthocyanidin reductase ((2S)-flavan-3-ol-forming) (*ANR*), Uncharacterized acetyltransferase At3g50280 (*At3g50280*) and two genes Stemmadenine O-acetyltransferase (AT) were identified as upregulated in stage 4 (vegetative stage), playing various

Table 3 The 10% of gene ontology terms per category with a higher number of DEGs associated

Category	GO ID	GO name	# DEGs
BP	GO:0055114	Obsolete oxidation–reduction process	88
	GO:0006355	Regulation of DNA-templated transcription	30
	GO:0006629	Lipid metabolic process	21
	GO:0005975	Carbohydrate metabolic process	15
	GO:0006952	Defense response	13
	GO:0006468	Protein phosphorylation	12
	GO:0055085	Transmembrane transport	10
	GO:0006508	Proteolysis	10
	GO:0016491	Oxidoreductase activity	39
	GO:0020037	Heme binding	34
MF	GO:0005515	Protein binding	31
	GO:0005506	Iron ion binding	30
	GO:0016705	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	28
	GO:0003677	DNA binding	28
	GO:0005524	ATP binding	22
	GO:0003700	DNA-binding transcription factor activity	21
	GO:0003824	Catalytic activity	19
	GO:0016788	Hydrolase activity, acting on ester bonds	16
	GO:0046983	Protein dimerization activity	14
	GO:0004672	Protein kinase activity	12
CC	GO:0016020	Membrane	16
	GO:0005634	Nucleus	9
	GO:0042025	Host cell nucleus	4

Biological process (BP), Molecular Function (MF), and cellular component (CC). #DEGs indicate the number of DEGs annotated with each GO term

roles in plant biology, including lignin formation, flavonoid synthesis, and acetyltransferase activity.

As regards the sesquiterpene lactones biosynthetic pathway (Fig. 11), the Crocetin glucosyltransferase, chloroplastic (*UGT75L6*), Beta-caryophyllene synthase, and (E)-beta-farnesene synthase (codified by three DEGs) were upregulated in stage 4 (vegetative stage).

Genes highly up-regulated in the reproductive stage (stage 5/6)

Within the pathways of the flower developmental and transition from vegetative to reproductive shoot apical meristem (Fig. 6), several genes were upregulated in stages 5/6, such as Agamous-like MADS-box protein AP1 (*API*), Agamous-like MADS-box protein AGL9 homolog (*AGL9*; *FBP2*; codified by two DEGs), and Phytochrome A1 (*PHYA1*).

Regarding the molecular function, DNA binding transcription factor activity (GO:0003700), several TF upregulated in stages 5/6 were identified (Fig. 7), such as Ethylene-responsive transcription factor ERF034 (*ERF034*), Ethylene-responsive transcription factor SHINE 3 (*HN3*), Probable WRKY transcription factor 4, Transcription factor bHLH118, and Dehydration-responsive element-binding protein 1D (*DREB1D*; codified by two DEGs).

The DEGs associated with anther and pollen development pathway, Receptor-like protein 50 (*RLP50*), Receptor-like protein 46 (*RLP46*), Polygalacturonase inhibitor (*pgip*), and Protein YLS3 (*YLS3*) were identified as upregulated in stage 5/6 (reproductive stage) (Fig. 8).

MLP-like protein 28 (*MLP28*), related to the defense response (GO:0006952) (Fig. 9), was identified as up-regulated in stage 5/6 (reproductive stage) in *C. cardunculus*.

From the phenylpropanoid biosynthesis pathway (Fig. 10), Peroxidase N1 (*poxN1*), Beta-glucosidase 12 (*BGLU12*), two genes Beta-glucosidase 24 (*BGLU24*), O-acetylstemmadenine oxidase (*ASO*), Uncharacterized acetyltransferase At3g50280, 4-coumarate-CoA ligase-like 6 (*4CLL6*), Probable cinnamyl alcohol dehydrogenase (*CAD2*), Shikimate O-hydroxycinnamoyltransferase (*HST*), Reticuline oxidase (*BBE1*), Peroxidase 42 (*PER42*), and Gallate 1-beta-glucosyltransferase (*UGT84A13*) were up-regulated in stages 5/6 (reproductive stage).

As regards the sesquiterpene lactones biosynthetic pathway (Fig. 11), the UDP-glycosyltransferase 75C1 (*UGT75C1*), Probable terpene synthase 11 (*TPS11*), Probable 1-deoxy-D-xylulose-5-phosphate synthase 2 chloroplastic (*Os07g0190000*), 3-hydroxy-3-methylglutaryl-coenzyme A reductase, Geranylgeranyl pyrophosphate synthase chloroplastic (*GGPS1*), Germacrene A synthase long form were upregulated in stage 5/6 (reproductive stage).

Phylogenetic relationships of the analyzed proteins

The phylogenetic analysis of the 21 biological relevant DEGs (Table S4) associated with the vegetative-to-reproductive transition revealed a clear clustering according to functional families (Fig. 12). Terpene synthases (*Germacrene A synthase*, *E-β-farnesene synthase* and *probable terpene synthase 11*) formed a highly supported monophyletic clade with bootstrap values ranging from 97 to 99. The peroxidase proteins (*Peroxidase 72*, *Peroxidase N1* and *lignin-forming anionic peroxidase*) clustered together with very strong support (bootstrap=97–100), indicating a conserved evolutionary origin.

The MADS-box transcription factors *AGL8* and *AGL9* grouped in a robust clade (bootstrap=97), while the ethylene-responsive transcription factors *SHINE3* and *ERF2* also clustered with strong support (bootstrap=98). Deeper nodes

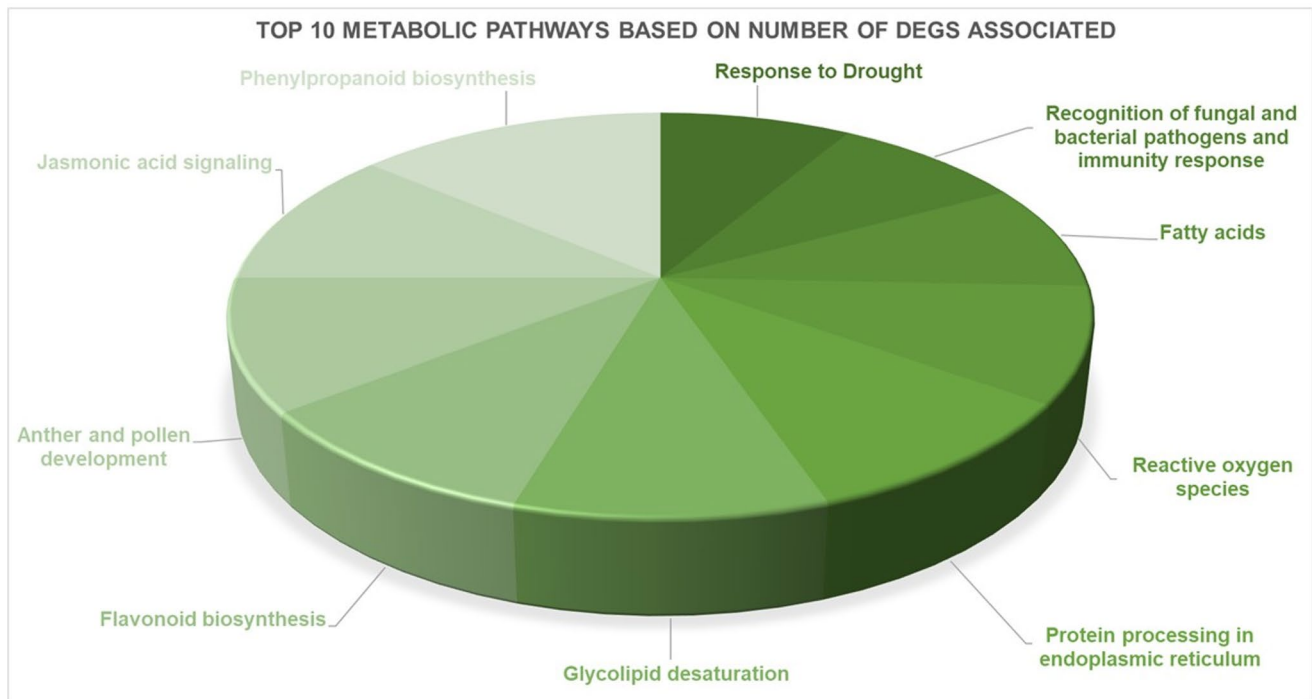


Fig. 5 The top 10 metabolic pathways with a higher number of DEGs associated with reactome and KEGG pathways

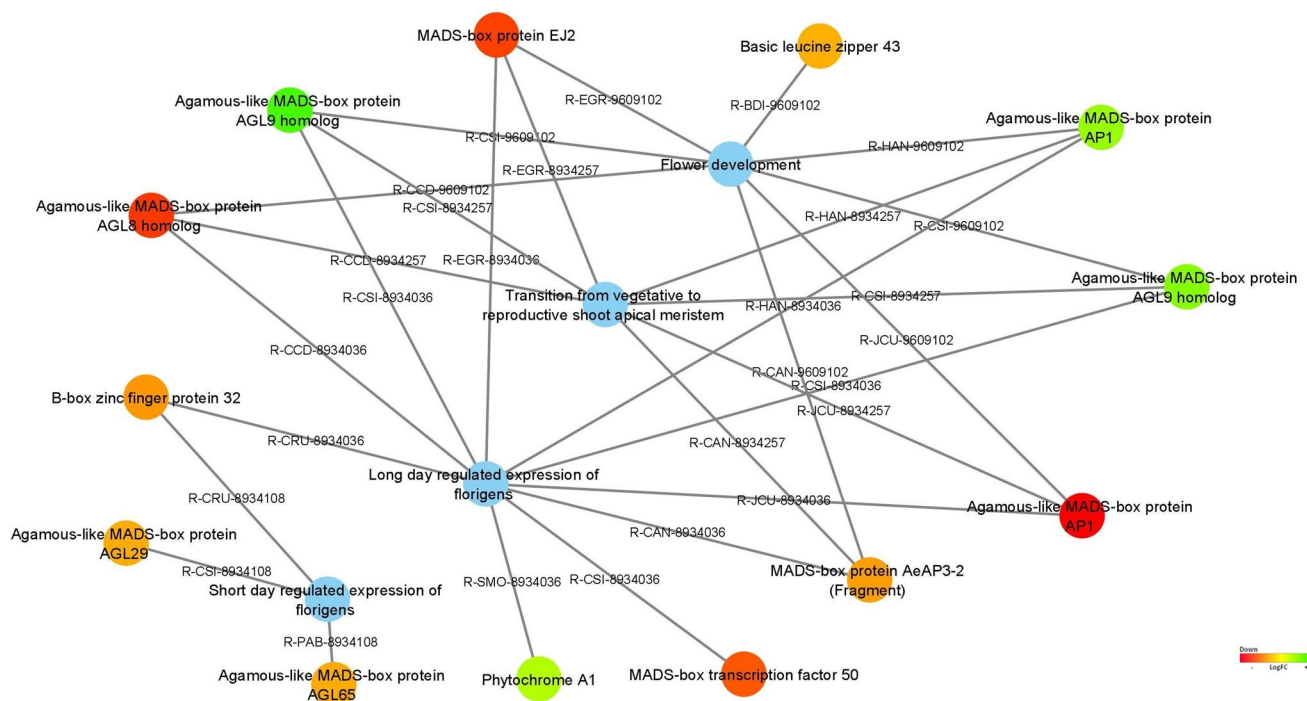
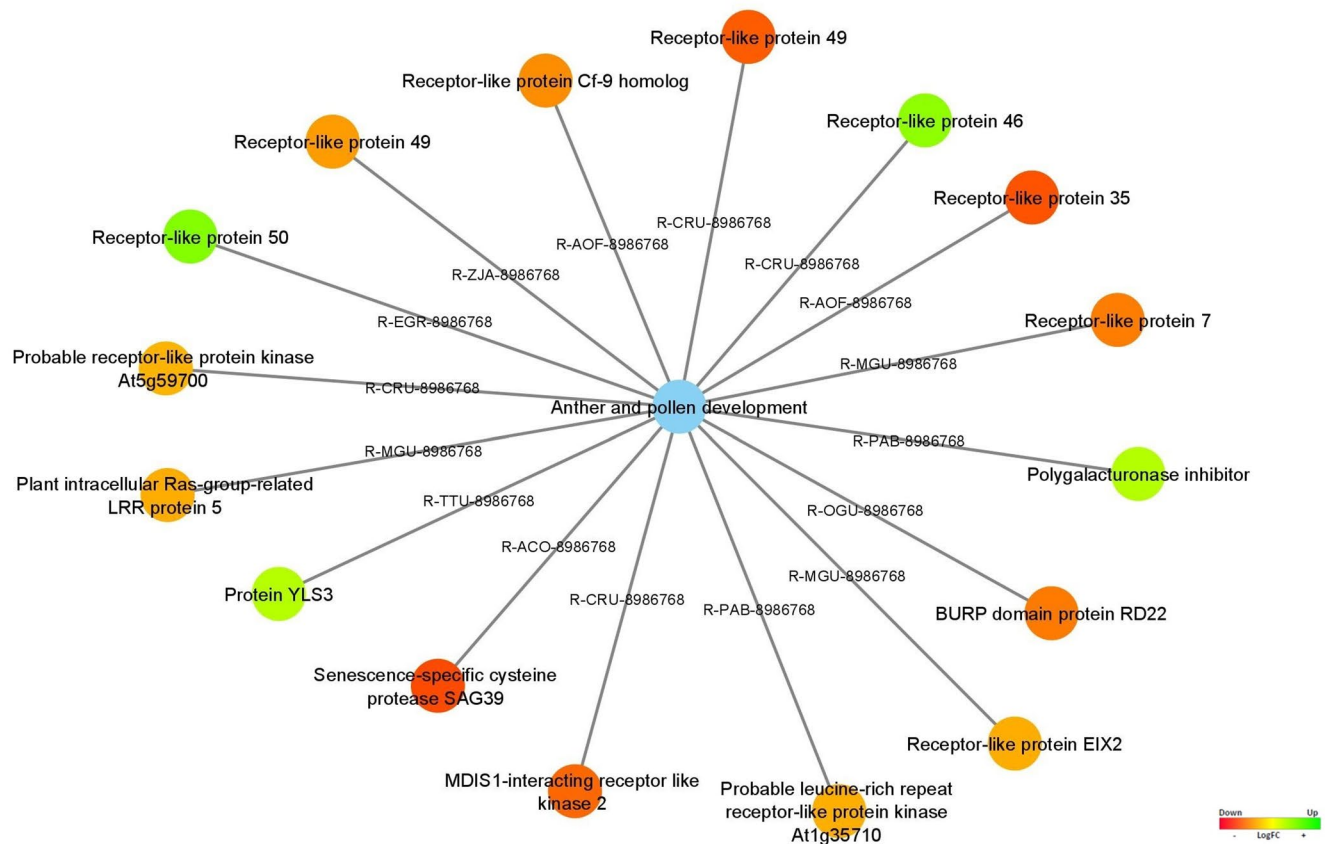


Fig. 6 Differentially expressed genes associated with flower development and the transition from vegetative to reproductive shoot apical meristem pathways are represented as the blue node. The expression

values were colored based on the RGB scale that goes from the most down-regulated (more expressed in stage 4) in red, to the most up-regulated (more expressed in stage 5/6) in green

A network diagram illustrating protein-protein interactions. The central node is "DNA-binding transcription factor activity". It is connected to 20 peripheral nodes, each representing a specific protein or transcription factor. The nodes are color-coded based on their degree of connectivity, as indicated by a legend at the bottom right: red for "Down", yellow for "Neutral", and green for "Up".

- Central Node:** DNA-binding transcription factor activity
- Peripheral Nodes (clockwise from top-left):**
 - WRKY transcription factor SUSIBA2
 - Heat stress transcription factor A-4c
 - Heat stress transcription factor A-2b
 - Heat shock factor protein HSF30
 - Basic leucine zipper 43
 - MADS-box protein AeAP3-2 (Fragment)
 - Dehydration-responsive element-binding protein 1D
 - Ethylene-responsive transcription factor SHINE 3
 - Pathogenesis-related genes transcriptional activator PTI5
 - Ethylene-responsive transcription factor ERF034
 - Ethylene-responsive transcription factor 1B
 - MADS-box protein EJ2
 - Dehydration-responsive element-binding protein 1D
 - Ethylene-responsive transcription factor 2
 - Agamous-like MADS-box protein AGL9 homolog
 - Transcription factor ORG2
 - Transcription factor bHLH118
 - Agamous-like MADS-box protein AP1
 - Agamous-like MADS-box protein AP1
 - Probable WRKY transcription factor 4



most down-regulated (more expressed in stage 4) in red, to the most up-regulated (more expressed in stage 5/6) in green

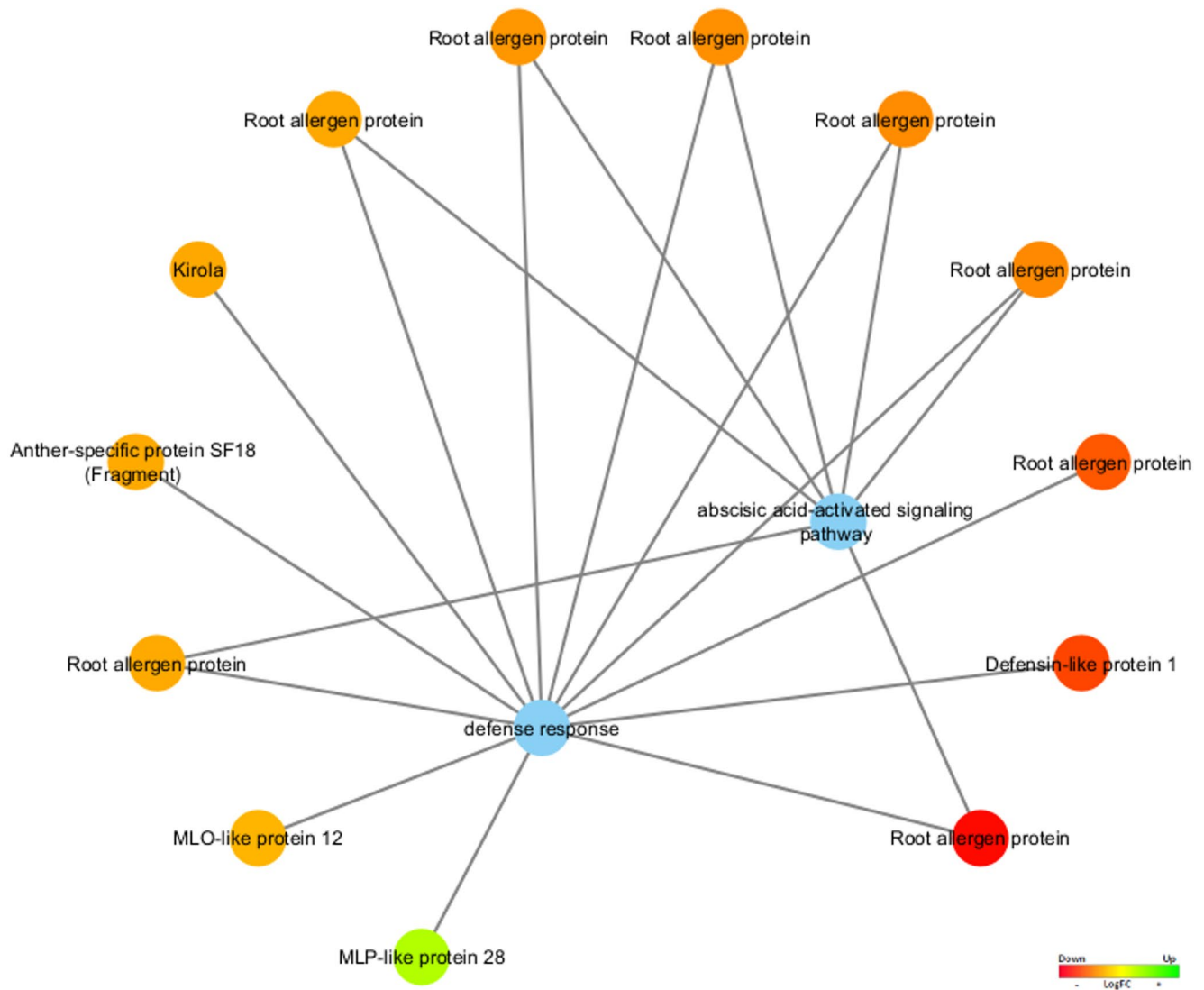


Fig. 9 Functional network of differentially expressed genes associated with the biological processes defense response (GO:0006952) and abscisic acid-activated signaling pathway (GO:0009738) are represented as the blue node. The expression values were colored based

on the RGB scale that goes from the most down-regulated (more expressed in stage 4) in red, to the most up-regulated (more expressed in stage 5/6) in green

connecting the major functional groups showed lower bootstrap values, indicating limited resolution among distantly related protein families across the full set of 21 sequences.

Identification of conserved protein motifs among transcription factor families, along with promoter motif and cis-regulatory element characterization.

MEME analysis identified three conserved protein motifs among the 21 protein sequences analyzed (Fig. S4). The most significant motif (MEME-1) was 49 amino acids long and displayed a very low E-value (5.1×10^{-9}), indicating strong statistical support. This motif was detected in two sequences corresponding to MADS-box transcription factor

genes (AGL8 and AGL9). The motif is highly enriched in basic residues, particularly lysine (K) and arginine (R), which are characteristic of DNA-binding domains. The second most significant motif (MEME-2), with a length of 24 amino acids and an E-value of 6.9×10^{-8} , detected in the three peroxidase sequences, indicating their potential involvement in catalytic activity. In addition, the third motif (MEME-3; E-value = 7.9×10^{-5}) with 46 amino acids was also detected in the two peroxidases (poxN1 and PER72), enriched in alanine (A) and valine (V) residues. The remaining motifs showed weak statistical support and are likely to reflect background sequence variability rather than biologically meaningful features.

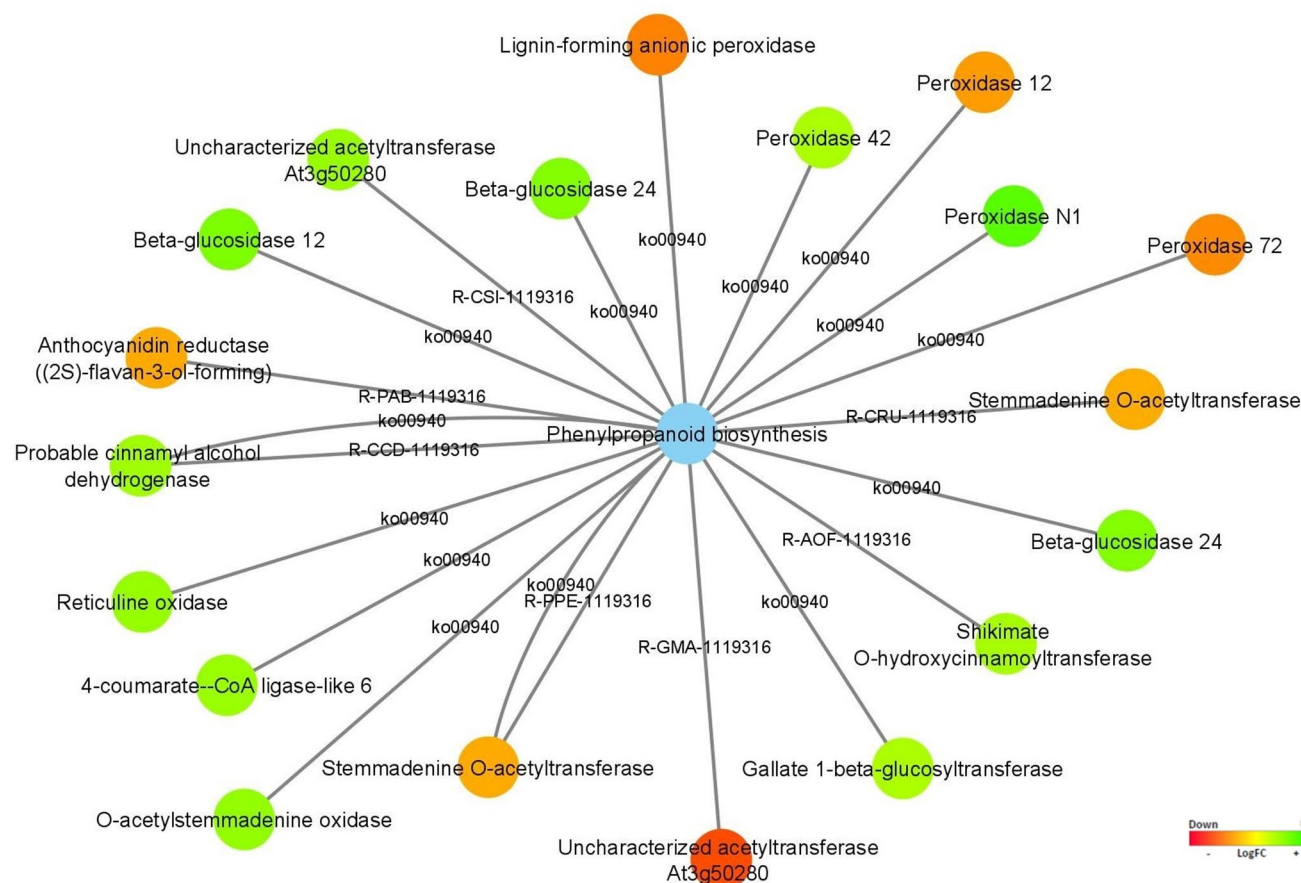


Fig. 10 Differentially expressed genes associated with the phenylpropanoid biosynthesis pathway are represented as the blue node. The expression values were colored based on the RGB scale that goes from

the most down-regulated (more expressed in stage 4) in red, to the most up-regulated (more expressed in stage 5/6) in green

De novo motif analysis of the promoter regions of the 21 genes identified three conserved DNA motifs, each comprising 15 nucleotides. These motifs were characterised by a high GC/C content and strongly conserved nucleotide positions in motifs 1 and 2 (E values of 1.6×10^{-4} and 4.1×10^{-4} , respectively), as well as TA patterns in motif 3 (E value of 6.3×10^{-5}). (Fig. S5).

PlantCARE analysis revealed a diverse set of cis-acting regulatory elements distributed across all seven promoters. Core promoter elements, including TATA-box and CAAT-box, were identified in all sequences, confirming the transcriptional competence of the analyzed regions. In addition, multiple hormone-responsive elements were detected, such as ABRE (abscisic acid responsiveness), AuxRE/TGA-elements (auxin responsiveness), GARE-motif and P-box (gibberellin responsiveness), and ERE (ethylene responsiveness). Several promoters also contained light-responsive elements, including G-box, Box 4, and GT1-motifs, as well as stress-related elements such as ARE, MBS, and TC-rich repeats. Elements associated with developmental regulation, including motifs linked to meristem activity and

floral development, were particularly evident in promoters of MADS-box transcription factors.

Correlation between cynaropicrin accumulation and gene expression

To determine whether cynaropicrin accumulation is accompanied by transcriptional changes during development, we performed a Spearman correlation between metabolite levels and the expression of 28,632 transcripts across individual plants representing stages 4 and 5/6. Five distinct transcripts showed significant correlations with cynaropicrin levels ($|\rho| \geq 0.70$; FDR < 0.05), with correlation coefficients ranging from -0.91 to $+0.87$. Thus, the correlation analysis between cynaropicrin accumulation and gene expression across plants with different metabolite levels and developmental stages identified the following five significantly correlated transcripts: *V2_12g002980* (protein of unknown function), *V2_01g027990* (NLP5), *V2_08g013960* (Transcription repressor OFP8), *V2_01g019160* (Thioredoxin-like 1-1,

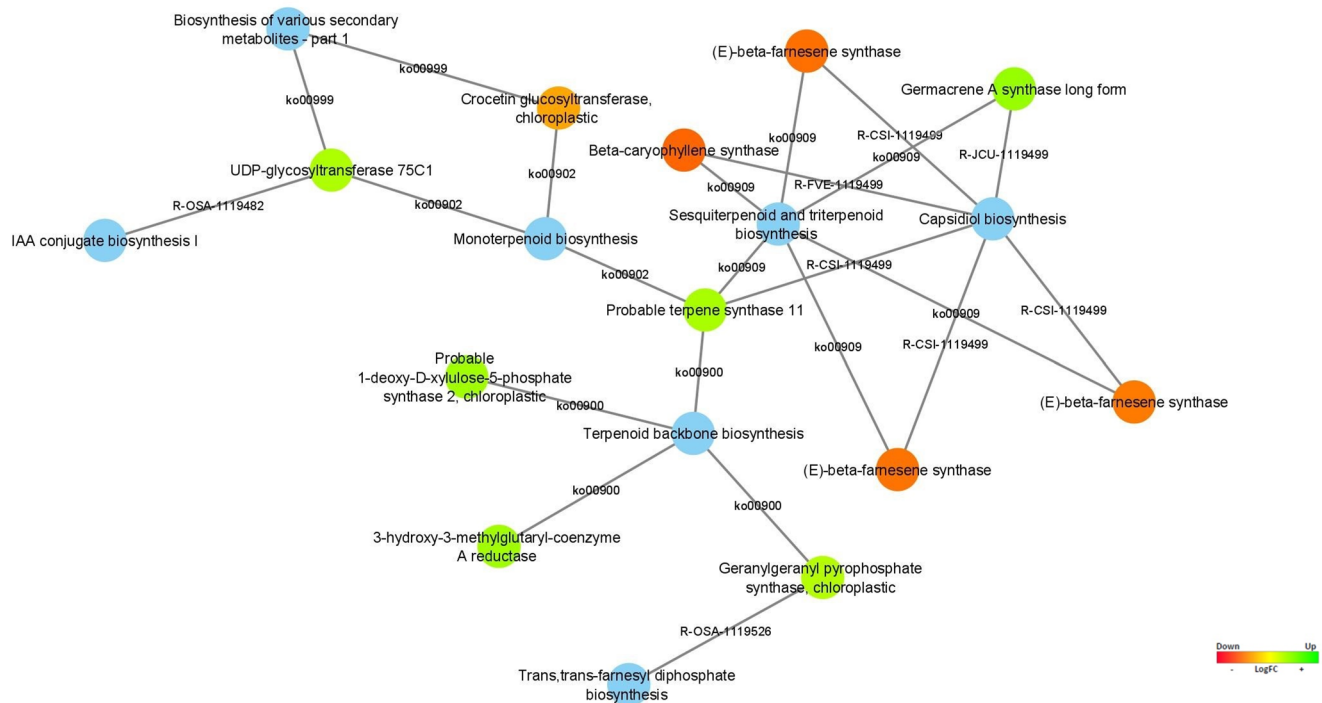


Fig. 11 Differentially expressed genes associated with the sesquiterpene lactones biosynthetic pathway are represented as the blue node. The expression values were colored based on the RGB scale that goes

from the most down-regulated (more expressed in stage 4) in red, to the most up-regulated (more expressed in stage 5/6) in green

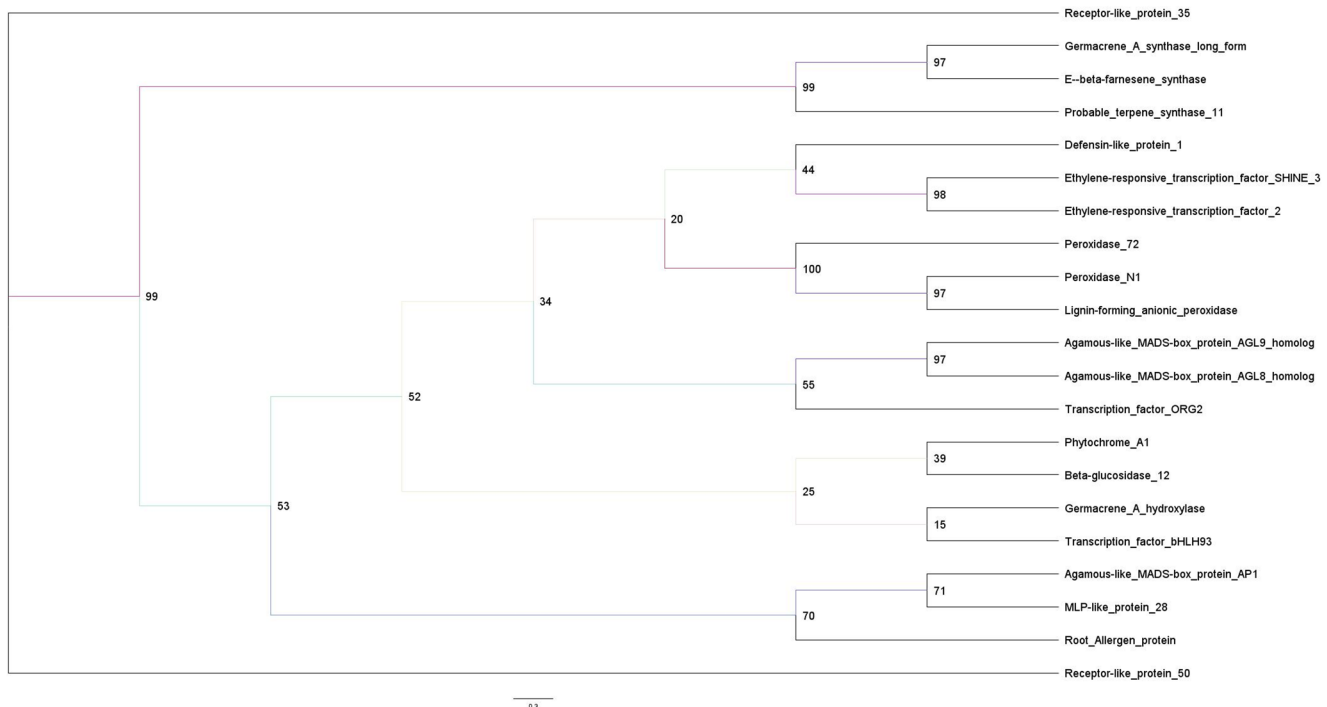
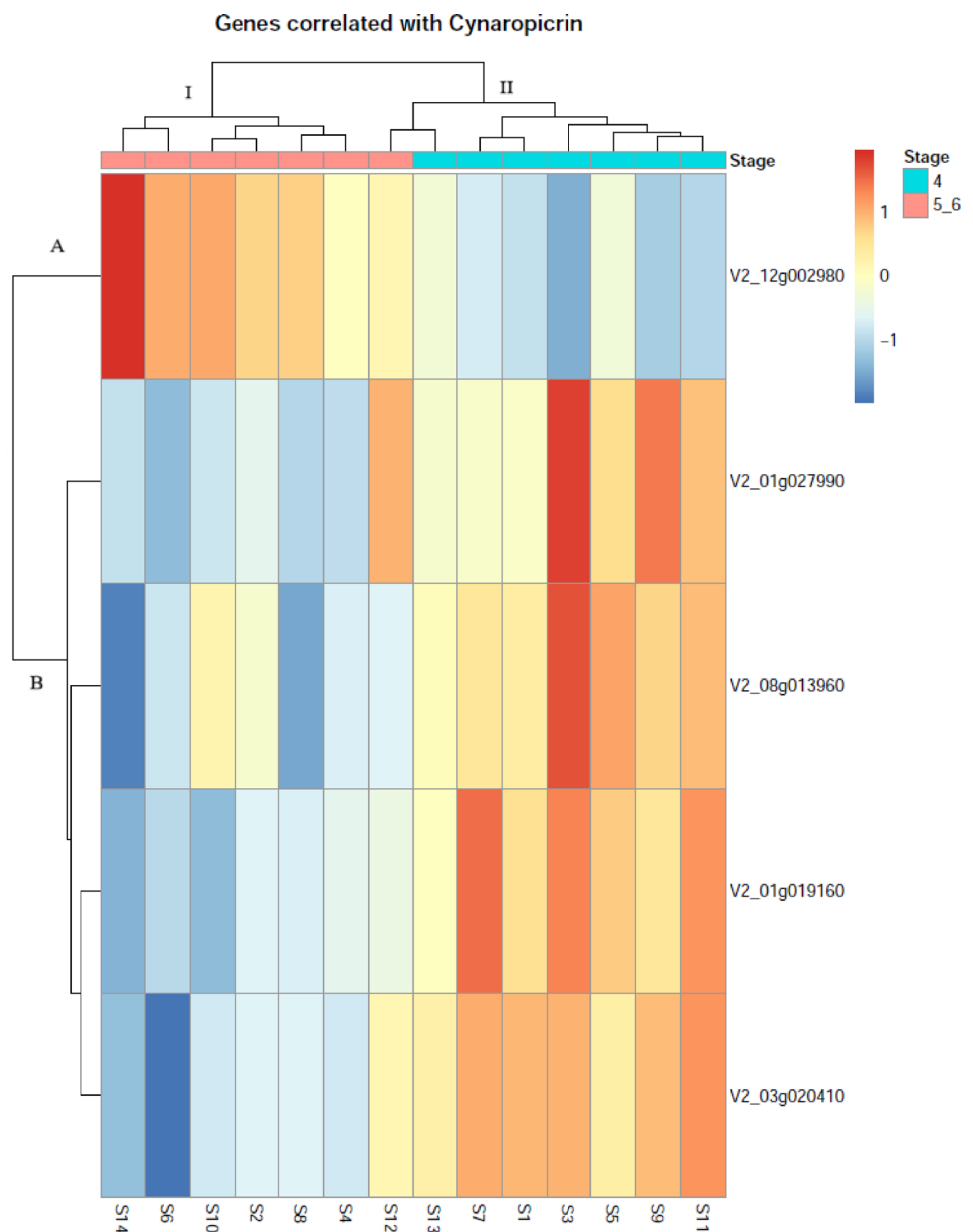


Fig. 12 Maximum likelihood phylogenetic tree of the 21 biologically relevant DEGs associated with the vegetative-to-reproductive transition constructed from the MAFFT alignment using IQ-TREE. Boot-

strap support values (1000 replicates) are indicated at the nodes. The scale bar represents the number of amino acid substitutions per site

Fig. 13 Gene expression patterns associated with cynaropicrin accumulation in *Cynara cardunculus*. Heatmap showing the expression profiles of the five genes significantly correlated with cynaropicrin levels (Spearman's ρ , $FDR < 0.05$, $|\rho| \geq 0.70$) across vegetative (stage 4) and reproductive (stages 5–6) plants. Samples (S1–S14) are ordered according to hierarchical clustering based on gene expression and segregate primarily according to cynaropicrin levels. Expression values are Z-score normalized per gene and represented by a blue to red color scale indicating low to high expression, respectively. The top annotation bar indicates the developmental stage of each sample (vegetative stage 4 and reproductive stages 5–6). Gene hierarchical clustering defines two transcriptional groups: Cluster A, comprising V2_12g002980, which is preferentially expressed in reproductive plants and positively associated with cynaropicrin accumulation; and Cluster B, comprising V2_01g027990, V2_08g013960, V2_01g019160, and V2_03g020410, which show higher expression in vegetative plants and are negatively associated with cynaropicrin levels



chloroplastic), and *V2_03g020410* (protein of unknown function) (Fig. 13).

Hierarchical clustering (Cluster I and II) separated the samples primarily according to cynaropicrin content, with reproductive plants (stage 5/6) grouping together and displaying higher cynaropicrin levels (Cluster I), while vegetative plants (stage 4) formed a distinct cluster characterized by lower cynaropicrin accumulation (Cluster II). Gene expression analysis revealed two contrasting transcriptional patterns associated with cynaropicrin levels.

The *V2_12g002980* (protein of unknown function) was preferentially expressed in reproductive plants and showed a positive correlation with cynaropicrin accumulation. In contrast, *V2_01g027990* (NLP5), *V2_08g013960*

(Transcription repressor OFP8), *V2_01g019160* (Thioredoxin-like 1–1, chloroplastic), and *V2_03g020410* (protein of unknown function) displayed higher expression in vegetative plants and were negatively associated with cynaropicrin levels.

Gene clustering revealed two distinct transcriptional patterns: (i) a gene positively associated with cynaropicrin accumulation and induced during reproductive stage (*V2_12g002980*); and (ii) a group of genes negatively associated with cynaropicrin levels, preferentially expressed during vegetative stage (*V2_01g027990*, *V2_08g013960*, *V2_01g019160*, and *V2_03g020410*).

Discussion

Cardoon, a cross-pollinated perennial species, is characterized by edible inflorescences and serves as a source of floral-based milk coagulants used in traditional cheese production. Additionally, cardoon represents a significant source of secondary metabolites, including phenolics, flavonoids, and sesquiterpene lactones, which contribute to its bioactive properties and biotechnological potential. The present study focused on identifying and interacting with the molecular mechanisms that mediate the vegetative-to-reproductive transition by transcriptomic analysis of two different developmental stages, vegetative and reproductive. RNA-Seq has transformed our ability to examine gene expression dynamics, providing a powerful tool for researchers across many disciplines. Continuous improvements in sequencing technologies and bioinformatics tools are further driving the evolution in this field (Stark et al. 2019).

Using a new sequencing platform, DNBSeg, our study highlighted a high number of DEGs between vegetative and reproductive stages, for the first time in wild cardoon. In our study, a differential gene expression study identified 552 differentially expressed genes (DEGs) ($FC > 2$), with 321 showing higher expression in the vegetative phase and 231 in the reproductive phase. Of these DEGs, 88% were functionally annotated, identifying several DEGs with functions associated with phenylpropanoid biosynthesis, anther/pollen development, flower development, transcription factors (TFs), defence response, and plant growth regulators have been identified. For the growth and development of plants, the temporal regulation of gene expression is essential, and its understanding contributes to revealing the molecular mechanisms underpinning every developmental phase.

Morphological and chemical evaluation

In March, *C. cardunculus* genotypes demonstrated a state of vegetative growth phase 4, characterized by adult leaves' full expansion and maturation, with no signs of inflorescence emergence. According to Archontoulis et al. (2010), the biomass yield during this stage can be very variable depending on the climate conditions.

In May, it was possible to observe the emergence of inflorescences in some of the genotypes sampled. The co-authors also suggest that the inflorescence number in this species may depend on environmental conditions, management, and genetic factors to be triggered (Lee et al. 2023). The plants maintained a constant growth rate during the assessed period. It is also important to note that these analyses were specific to the study's conditions and period, and other variables, such as management and climatic factors, may influence the development of *C. cardunculus* plants.

Cynaropicrin, a sesquiterpene lactone with demonstrated anti-inflammatory, antimicrobial, and anticancer properties (Eljounaidi et al. 2015; Ramos et al. 2017), has been also associated with some physiological functions in plants, including defense against insects and pathogens (Elsebai et al. 2016). The contrasting profiles of cynaropicrin observed from vegetative to reproductive stages in our study suggest a potential involvement of this sesquiterpene lactone in the flowering and capitulum formation in cardoon, offering some insights into the biochemical changes associated with flowering and capitulum formation in *C. cardunculus*.

Novel transcription factor in the regulation of flower development

The transition from vegetative to reproductive shoot apical meristem marks the shift from producing leaves and stems to producing flowers (Blázquez et al. 2001), a process regulated by genetic mechanisms triggered by internal and external signals (Chávez-Hernández et al. 2022). In *C. cardunculus*, floral development is influenced by environmental, genetic, and management factors, resulting in variations in flowering time and the number of inflorescences (Archontoulis et al. 2010; Lee et al. 2023). Transcriptomic studies have been essential to understanding the molecular mechanisms underlying flower development, highlighting the crucial role of transcription factors (TFs) in regulating gene expression.

TFs are proteins that bind to DNA and regulate the transcription of genes involved in plant development (Joshi et al. 2016). Based on the results of Puglia et al. (2020), two-third of the differentially expressed genes (DEGs) identified during capitulum development were TFs. According to our results, 53 DEGs were annotated to the TF category and classified into 8 TF families, with HSF, MADS-box, and bHLH being the most prominent. The MADS-box family (Fig. 7), one of the largest and most studied, plays a central role in floral meristem identity and the determination of floral organs (Schwarz-Sommer et al. 1990; Wang et al. 2019). Overexpression of these genes in the vegetative stage often results in early flowering (Blázquez et al. 2001).

EJ2 gene coding for MADS-box protein EJ2 up-regulated in cardoon vegetative state, has been already described as particularly active during the reproductive phase in *Solanum lycopersicum* and plays a central role in various aspects of this process, which range from the control of meristem identity to the development of the floral organs, the development/ripening, and abscission of fruit and plays a role in the branching process of the inflorescence (Soyk et al. 2017). Other MADS-box genes have been described in several studies. MADS-box transcription factor 50 regulates flowering time and can stimulate the floral transition stage as well as

internode elongation in *A. thaliana* (Tadege et al. 2003; Lee et al. 2004). The Agamous-like MADS-box protein AGL8 homolog (up-regulated in vegetative stage in cardoon) in *A. thaliana*, plays a crucial role in the regulation of floral organ identity and flowering time (Yanofsky et al. 1990; Ferrándiz et al. 2000), while MADS-box protein AeAP3-2 (up-regulated transcript in vegetative stage in cardoon) is involved in flower development in tomatoes and *A. thaliana* (Jack et al. 1992; De Martino et al. 2006). The Agamous-like MADS-box protein AGL65 (*AGL65*) is involved in the regulation of pollen maturation at the late stages of pollen development and pollen tube growth (Adamczyk and Fernandez 2009), while the Agamous-like MADS-box protein AGL29 (*AGL29*) is expressed in pollen and during embryogenesis, expressed in the chalazal endosperm (Kofuji et al. 2003; Bemer et al. 2010).

Additionally, genes codifying for Agamous-like MADS-box protein AP1 (codified by two DEGs, one up and one down-regulated in the reproductive stage), Agamous-like MADS-box protein AGL9 homolog (codified by two DEGs), and Phytochrome A1 were upregulated in the reproductive stage. The Agamous-like MADS-box protein AP1 in *R. chinensis* was also reported as an important gene in the regulation of the development of floral organs, such as petals and stamens, by activating or repressing the expression of downstream genes (Han et al. 2019). The *AGL9* is active in inflorescence development and floral organogenesis (Bonhomme et al. 1997) and Phytochrome A1 is a regulatory photoreceptor that exists in two forms that are reversibly interconvertible by light.

In addition to *MADS-box*, other TF families, such as *ERF*, *bZIP*, *WRKY*, and *bHLH*, are also associated with floral development (Fig. 7). Members of these families were previously described as linked to reproductive development in other plant species, such as *A. thaliana*, *O. sativa*, and *N. tabacum* (Nakano et al. 2006; Li et al. 2018; Hao et al. 2021). The *AP2/ERF* family (33% in our data), for example, is involved in stress responses, development, and hormonal signaling (Feng et al. 2020; Wu et al. 2022). Transcription factors such as *ERF2*, *ERF034*, and *SHINE 3* regulate physiological processes, including fruit ripening, senescence, and responses to biotic and abiotic stresses (Fujimoto et al. 2000; Aharoni et al. 2004; Wu et al. 2022). One of the biggest transcription factor gene families in *A. thaliana* and *C. Cardunculus* var. *altis* is the basic *helix-loop-helix* (*bHLH*) transcription factor family (Puglia et al. 2020; Hao et al. 2021). The *bHLH* family, in turn, is linked to light signaling, regulation of flowering time, and stress responses (Sun et al. 2018), with members like *ORG2* playing roles in multiple biological processes (Heim et al. 2003; Kang et al. 2003; Wang et al. 2007).

Another TF family associated with the molecular function of DNA-binding transcription factor activity is *WRKY*, which is involved in many physiological processes in plants, including growth, development, and responses to both biotic and abiotic stress (Guo et al. 2019). The Probable *WRKY* transcription factor 4 plays a role in plant defense and stress responses. In *A. thaliana*, it is expressed in young, mature, and senescent leaves, and it is induced by biotic and abiotic stresses (Robatzek and Somssich 2001; Yu et al. 2001; Lai et al. 2008; Yamasaki et al. 2012). These TFs and *MADS-box* coordinate a complex signaling network that regulates floral development and plant adaptation to varying environmental conditions.

Regulation patterns of other DEGs directly linked to anther and pollen development

Anther and pollen development are essential processes embedded in the reproductive phase of flowering plants (angiosperms) and involve a series of complex molecular and cellular events, including cell differentiation, division, and maturation (Pearce et al. 2015).

Receptor-like kinases (RLKs) are key proteins in plants, mediating signal transduction and responses to environmental cues like hormones, pathogens, and stress (Mizuno et al. 2007). Notable *RLK* genes differently expressed in our study include homologs to *Atlg35710* and *At5g59700*, and *MIK2* (encodes the MDIS1-interacting receptor-like kinase 2), which regulates growth, development, and stress responses, and is involved in pollen tube perception.

Also, Receptor-like proteins (RLPs), which are vital for plant defense and development, act as molecular sensors to external signals. In this study, several *RLPs*, such as *RLP35*, *RLP49*, *RLP50*, and *RLP46*, were identified as differentially expressed during anther and pollen development (Fig. 8). The observed differential expression of these *RLPs* aligns with their hypothesized roles in cell signaling during anther maturation, pollen development, and possibly self-incompatibility in *C. cardunculus* (Scaglione et al. 2016).

SAG39 (codified for Senescence-specific cysteine protease SAG39) was up-regulated in stage 4, while *YLS3* and *pgip* (codified for Protein YLS3 and Polygalacturonase inhibitor, respectively) were up-regulated in stages 5/6. Senescence-specific cysteine protease SAG39 plays a role in developmental senescence, apoptosis, and stress responses (Liu et al. 2010). Protein YLS3 is expressed in expanding leaves and sepals, triggered by abscisic acid (Yoshida et al. 2001; Kobayashi et al. 2011), while *polygalacturonase inhibitor* (*pgip*) enhances plant resistance to fungal attacks by inhibiting cell wall penetration (Howell and Davis 2005). These results demonstrate the complex molecular processes that underlie stress reactions and plant reproduction.

Defense response

In response to herbivory, plants use a variety of morphological, biochemical, and molecular defense mechanisms to mitigate the impacts of the attack. Both direct and indirect defensive mechanisms facilitate the complex and ever-changing biochemical defense systems against herbivores. The defensive chemicals impact the eating, growth, and survival of herbivores and are either constitutively produced or released in reaction to plant harm. Furthermore, vegetation releases volatile chemical compounds that draw the herbivores' natural predators. These tactics can function separately or in tandem with one another (War et al. 2012).

For example, Defensin-like protein 1, highly represented in stage 4 in our study (Fig. 9), plays an antimicrobial activity by altering fungal membranes (Thevissen et al. 2000). Similarly, root allergen proteins, encoded by eight DEGs in cardoon, can trigger allergic reactions, and their expression varies with environmental and genetic factors (Bonini et al. 1994; Falcon and Caoili 2023). During the reproductive stage, cardoon upregulates *MLP28*, responsible for coding a protein involved in stress responses and disease resistance, which is also linked to the jasmonic acid signaling pathway (Song et al. 2020).

During the reproductive phase, the plant devotes energy to producing flowers and seeds. This energy redistribution can make the plant more sensitive to herbivores, diseases, and other environmental challenges, despite its antibacterial and insecticidal properties (Kant et al. 2015; Quiroz-Pacheco et al. 2020).

Regulation Patterns of DEGs Directly Linked to Phenylpropanoid Biosynthesis

The phenylpropanoid biosynthesis pathway is a complex metabolic pathway responsible for producing various secondary metabolites in plants, such as phenolic compounds including flavonoids, lignans, and other metabolites (Fraser and Chapple 2011). Our study identified several associated DEGs encoding peroxidases from the phenylpropanoid biosynthesis pathway (Fig. 10), where 60% of these transcripts (*Lignin-forming anionic peroxidase*; *PER 72*; *PER 12*) were up-regulated in the plant's vegetative stage (stage 4). One of the many enzymes found in nature that belongs to the class of oxidoreductases is peroxidase. Interaction with hydrogen peroxide and related compounds can accelerate the oxidation of various organic and inorganic substrates. Owing to their broad range of catalytic activity, peroxidases can be used in chemical synthesis, the elimination of phenolic compounds and peroxides, and, considering recent research, the breakdown of mycotoxin (de Oliveira et al. 2021). With potential uses in tobacco, tomato, and sweetgum plants,

anionic peroxidases are particularly important in lignin formation and plant development (Lagrimini 1993), with homologs like *PER 72* and *PER 12* in *Arabidopsis* linked to abiotic and biotic stress responses (Schenk et al. 2000; Fowler and Thomashow 2002). *PER 72* is slightly expressed in roots and induced by cold, while *PER 12* is up-regulated in roots and leaves and expressed in early seed development (Girke et al. 2000). *PER 42*, another peroxidase, may act as a heat shock-like defense protein, with high expression in roots (Schaffer et al. 2001). In tobacco, *peroxidase NI* (*poxNI*) is highly expressed in roots and minimally in lower leaves (Hiraga et al. 2000). However, this study is the first to investigate the presence of these genes in *C. cardunculus*.

Beta-glucosidase transcripts (*BGLU12*, *BGLU24* two isoforms) are highly enriched during the reproductive stage in cardoon, playing roles in phytohormone activation, defense, and cell-wall processes (Cairns and Esen 2010). Research in rice highlights *BGLU24*'s role in seed germination and its interaction with phytohormones like IAA and ABA, suggesting similar functions in cardoon (Ren et al. 2019). Additionally, studies on *C. cardunculus* var. *altilis* reveal that genes like *Shikimate O-hydroxycinnamoyltransferase* (*HCT*) and *4-coumarate-CoA ligase* (*4CL*) are up-regulated during inflorescence development, with *4CLL6* and *HST* showing increased expression in the reproductive stage (Puglia et al. 2020).

The higher expression of these genes during the reproductive stage suggests their involvement in various processes associated with flowering, fruiting, seed development, and putative plant defense mechanisms at this stage of the plant's life cycle.

Metabolite production

In plants, defense responses are crucial to their survival and protection against various biotic and abiotic stresses (Gull et al. 2019). These defense mechanisms are often triggered by the recognition of pathogens, such as bacteria, fungi, viruses, and herbivores (War et al. 2012). The responses involve complex signaling pathways and produce various metabolites that contribute to the plant's defense arsenal. Understanding the growth profile of *C. cardunculus* leaves is crucial for crop production to increase sesquiterpene lactone yield.

Terpenoids, a diverse class of secondary metabolites, play a vital role in plant defense and other biological functions (Singh and Sharma 2015). Their biosynthesis begins with the formation of a terpenoid backbone via the mevalonate or methylerythritol phosphate (MEP) pathways. Key intermediates like trans,trans-farnesyl diphosphate (FPP) serve as precursors for sesquiterpenoids and triterpenoids, which are synthesized by enzymes such as sesquiterpene

synthases and oxidosqualene cyclases (LI et al. 2023a, b). The Probable Terpene Synthase 11 (*TPS11*) is a protein-coding gene that is involved in the biosynthesis of terpenes (Li et al. 2022). *TPS11*, identified in our study as upregulated in the reproductive stage (Fig. 11), has been identified in various plant species, including *A. thaliana* (Aubourg et al. 2002; Tholl and Lee 2011), *Medicago truncatula* (Parker et al. 2014), and *tomato* (Falara et al. 2011). It is primarily expressed in the glandular trichomes of these plants, which are specialized structures that produce and secrete secondary metabolites such as terpenes (Parker et al. 2014). Overall, the *TPS11* protein plays an important role in the biosynthesis of sesquiterpenes in plants, which have a variety of functions, including defense against herbivores and pathogens, as well as attraction of pollinators and seed dispersers.

UDP-glycosyltransferase 75C1 acts on both abscisic acid (ABA) auxin (IAA) biosynthesis and was found highly expressed during fruit ripening and flower development in tomatoes (Sun et al. 2017), supporting the up-regulation of UDP-glycosyltransferase 75C1 in stage 5/6 of cardoon development.

Another important enzyme, beta-caryophyllene synthase which is highly represented in stage 4 of wild cardoon, has previously been identified in *C. cardunculus* L. var. *scolymus* heads (Laghezza Masci et al. 2024) is responsible for producing the sesquiterpene beta-caryophyllene, which is known by its anti-inflammatory and anti-carcinogenic properties. Its expression varies across plant tissues and decreases with seedling age (Gu et al. 2002; Olofsson et al. 2011). Similarly, (E)-beta-farnesene synthase converts FPP into (E)-beta-farnesene, a volatile compound that acts as a signaling molecule during stress or herbivory and influences insect behavior (Bhatia et al. 2015). In wild cardoon, three (*E*)-beta-farnesene synthase genes were found upregulated during the vegetative stage.

Other DEGs responsible for codifying the protein Germacrene A synthase long form were present in our results, which is involved in sesquiterpene lactone biosynthesis, and produces exclusively (+)-germacrene A (GAS). The GAS previously isolated from globe artichoke participates in the first stage of cynaropicrin biosynthesis, where the expression analysis revealed a link between GAS expression level and cynaropicrin concentration (Menin et al. 2012). On the other hand, it was mainly found expressed in roots, head core tissue, and green and etiolated seedlings of chicory, while its expression was lower in green leaves (Bouwmeester et al. 2002). This finding supports our results, in which GAS is upregulated in the reproductive stage, a period with a higher production of cynaropicrin in wild cardoon leaves.

Our results show that the outcomes derived from the various plants during the two months of collection exhibit significant variation in cynaropicrin levels, concerning both

genotype and collection timeframe. This difference suggests that the production of cynaropicrin in plants is activated during the vegetative to reproductive transition in *C. cardunculus*, which can be influenced by environmental, physiological, and genetic factors (Paulino et al. 2023). The elevated presence of cynaropicrin can act as a deterrent against herbivores, functioning as a natural insecticide. According to our genetic results, this gene is upregulated in the reproductive stage, a period with a higher production of cynaropicrin in *C. cardunculus* leaves. Later, germacrene A oxidase (GAO) (upregulated in the vegetative stage in our results; Table S3) and costunolide synthase (COS) genes, also involved in cynaropicrin biosynthesis, were identified and functionally described in *C. cardunculus* (Eljounaidi et al. 2014).

Overall, the regulation of terpenoid biosynthesis in *C. cardunculus* is tightly linked to the vegetative-to-reproductive transition, highlighting the importance of terpenoids in plant defense and their potential applications in agriculture and medicine.

Phylogenetic relationships and conserved protein motifs of differentially expressed genes with cis-regulatory element profiling

Phylogenetic analysis of the 21 DEGs revealed a clear clustering according to major functional families, indicating strong evolutionary conservation and shared ancestry. In particular, terpene synthases, peroxidases, and MADS-box transcription factors formed well-supported clades, reflecting their conserved biological roles across plant lineages. The deep evolutionary conservation and diversification of MADS-box genes across land plants has been extensively documented and underscores their fundamental role in developmental regulation and transcriptional control (Parnicová et al. 2003; Qiu et al. 2023; Zhang et al. 2024).

The close clustering observed among terpene synthases suggests coordinated evolutionary trajectories linked to terpenoid biosynthesis, a key component of plant secondary metabolism and stress adaptation. Similarly, the peroxidase clade exhibited strong phylogenetic support, highlighting the conservation of enzymes involved in oxidative stress responses, cell wall modification, and lignification. Together, these phylogenetic patterns indicate that the selected genes are embedded within conserved functional families, supporting their biological relevance in growth, development, and environmental responses.

The integration of protein motif analysis with phylogenetic relationships revealed a strong correspondence between evolutionary clustering and motif composition. Protein motif analysis identified three conserved motifs across the 21 proteins, mapped specifically to phylogenetically

clusters, indicating preservation of key functional domains likely linked to protein–protein interactions and catalytic activity. The most significant motif, a long and highly conserved region enriched in charged residues, corresponds to the canonical MADS DNA-binding domain, supporting their common evolutionary origin and conserved regulatory function of TF family (Theissen et al. 2000; Smaczniak et al. 2012). This reinforces the importance of the transcriptional regulation during the vegetative-to-reproductive transition. Likewise, motifs enriched in peroxidase proteins were restricted to peroxidase sequences, consistent with conserved residues linked to enzymatic activity.

De novo motif discovery identified conserved sequence patterns shared among promoters, while PlantCARE annotation detected cis-regulatory elements associated with hormonal signaling, light responsiveness, and stress responses. The coexistence of conserved de novo motifs and annotated cis-elements supports coordinated regulation of transcription factor genes by developmental and environmental cues, a regulatory strategy widely reported in plants (Lescot et al. 2002; Bailey et al. 2009; Sheshadri et al. 2016). Notably, the enrichment of hormone and stress responsive cis-elements, suggests that these genes may function as regulatory nodes integrating environmental signaling with secondary metabolic pathways (Toledo-Ortiz et al. 2003; Nakano et al. 2006; Licausi et al. 2013).

Transcriptional control of cynaropicrin biosynthesis during plant development

The integration of gene expression with cynaropicrin accumulation across developmental stages shows that cynaropicrin biosynthesis in *C. cardunculus* is predominantly associated with the reproductive phase (stages 5/6). This pattern is consistent with the developmental regulation of sesquiterpene lactones reported in Asteraceae reported by Liu et al. 2011. In *Cynara cardunculus*, a coordinated increase in sesquiterpene lactones during later developmental stages has also been described, supporting this developmental control in the species (Eljounaidi et al. 2015).

Among the genes correlated with cynaropicrin accumulation, *V2_12g002980* (protein of unknown function) display preferential expression in reproductive plants and showed a positive association with cynaropicrin levels, suggesting a direct involvement in the biosynthetic pathway or in regulatory steps linked to enhanced metabolite production. Such transcriptional induction of pathway related genes during reproductive development has been widely reported for sesquiterpene lactone biosynthesis (Nguyen et al. 2010).

In contrast, *V2_01g027990* (*Nodule inception-like proteins*, *NLP5*), *V2_08g013960* (*OVATE gene*, *OFP8*), *V2_01g019160* (*Thioredoxin-like 1-1*, *chloroplastic*,

TLL1), and *V2_03g020410* (protein of unknown function) were more highly expressed during vegetative stage (stage 4) and exhibited negative correlations with cynaropicrin accumulation (Gaudinier et al. 2018; Hernández-Reyes et al. 2022; Yang et al. 2016). The transcription factor *NLP5*, described as being involved in nitrogen signaling (Gaudinier et al. 2018; Hernández-Reyes et al. 2022), suggests a role in coordinating nitrogen status with metabolic allocation during cardoon vegetative growth. Meanwhile, the transcription repressor *OFP8* (Yang et al. 2016), known to act as a transcriptional repressor, may limit the activation of secondary metabolic pathways during early developmental stages. Moreover, the role of the *thioredoxin-like 1-1*, *chloroplastic*, *TLL1* in the adaptation of the *Arabidopsis thaliana* plant to osmotic stress and in the control of root growth has been described (Lakhssassi et al. 2012).

The expression pattern identified in our study also suggests that these genes may be involved in alternative branches of terpenoid metabolism or in the allocation of metabolic precursors toward other pathways during vegetative growth, rather than directly contributing to cynaropicrin biosynthesis.

Conclusion

Cardoon, a cross-pollinated perennial crop, is a major source of secondary metabolites such as phenylpropanoids, flavonoids, and sesquiterpenes lactones. Its production and regulation can be boosted during inflorescence emergence/development.

The main goal of the current study was to determine the molecular mechanisms involved in the vegetative and reproductive phases using a comparative transcriptome analysis. This work uncovered candidate transcripts not previously associated with flower development or secondary metabolites.

Based on the results presented, significant differences were observed in the morphological growth, cynaropicrin levels, and gene expression of *C. cardunculus* across its different phenological stages. During the vegetative stage (BBCH 4), the plants exhibited considerable leaf area and variable height, but no inflorescences, highlighting a phase of intense vegetative growth. In contrast, at the reproductive stage (BBCH 5/6), inflorescence development was observed, with greater dispersion between plants, suggesting a lack of complete synchronization in flowering. Cynaropicrin quantification showed a significant increase in this compound's content from March to May, which could reflect changes in environmental conditions and plant development. This compound, important for both pharmaceutical and food industries, was found to be more concentrated in May,

indicating that the reproductive stage might be crucial for its accumulation. Transcriptome analysis revealed a substantial number of differentially expressed genes (DEGs), with 321 genes being more expressed in the vegetative stage and 231 in the reproductive stage. DEGs related to secondary metabolite biosynthesis, such as phenylpropanoid pathways, were particularly prominent in both stages, with metabolic pathways related to pollen development and defense responses more active in the vegetative stage. In contrast, genes linked to terpene biosynthesis and flower development were more expressed in the reproductive stage. Novel transcription factors in the regulation of flower development were also identified, with HSF, MADS-box, and bHLH being the most prominent families.

The integrative analysis of the most representative DEGs, revealed coherent regulatory and metabolic modules associated with the floral transition. The coordinated clustering of MADS-box regulators, transcription factors, receptor-like proteins, phenylpropanoid enzymes and sesquiterpene synthases, along with their shared regulatory motifs, indicates that the transition to the reproductive stage is governed by interconnected pathways integrating developmental cues, hormone signaling, and secondary metabolism. This provides a solid molecular framework for understanding the control of flowering and the associated metabolic shifts in Asteraceae.

Overall, this study demonstrates that cynaropicrin accumulation during the transition from stage 4 to stages 5/6 is strongly associated with coordinated transcriptional reprogramming involving both positive and negative regulatory components. The integrative approach adopted here identified a concise and robust set of candidate genes potentially involved in cynaropicrin biosynthesis and regulation, constituting a valuable foundation for future functional and biochemical studies. These findings significantly advance our understanding of sesquiterpene lactone specialization in *C. cardunculus* and open new perspectives for its biotechnological and industrial exploitation.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s1103-025-01679-2>.

Author contributions Conceptualization : Octávio Salgueiro Paulo, Maria F. Duarte, Liliana Marum; Formal analysis : Ana Paulino, Isabel Fernandes, Ana Usié; Funding acquisition : Maria F. Duarte; Investigation : Ana Paulino, Isabel Fernandes, Rita Costa Pires, Ana Usié, Ana Faustino, Jacqueline Santos, Teresa Brás, Daniela Rosa, Octávio Salgueiro Paulo, Maria Fátima Duarte, and Liliana Marum; Methodology : Ana Paulino, Isabel Fernandes, Ana Usié, Octávio Salgueiro Paulo, Maria F. Duarte, Liliana Marum; Project administration : Maria F. Duarte; Resources : Ana Paulino, Rita Costa Pires, Jacqueline Santos, Teresa Brás, Daniela Rosa, Maria Fátima Duarte, and Liliana Marum; Supervision : Octávio Salgueiro Paulo, Maria F. Duarte, Liliana Marum; Writing—original draft : Ana Paulino, Ana Usié, Octávio Salgueiro Paulo, Maria F. Duarte, Liliana Marum; Writing—review

& editing : Ana Paulino, Ana Usié, Octávio Salgueiro Paulo, Maria F. Duarte, Liliana Marum.

Funding Open access funding provided by FCT|FCCN (b-on). This work is supported by Program Alentejo 2020, through the European Fund for Regional Development (FEDER) under the scope of MedCynaraBioTec—Selection of *Cynara cardunculus* genotypes for new biotechnological applications: the value chain improvement of cardoon, a well-adapted Mediterranean crop (ALT20-03-0145-FEDER-039495). Authors also acknowledge FCT for Contrato—Programa to Liliana Marum (CEECINST/00131/2018), and Ana Usié (CEECINST/00100/2021/CP2774/CT0001; <https://doi.org/10.54499/CEECINST/00100/2021/CP2774/CT0001>), Ph.D. grant to Ana Paulino (SFRH/BD/145383/2019; <https://doi.org/10.54499/SFRH/BD/145383/2019>), Ana Faustino (UI/BD/153511/2022; <https://doi.org/10.54499/UI/BD/153511/2022>), and Daniela Rosa (SFRH/BD/143845/2019, <https://doi.org/10.54499/SFRH/BD/143845/2019>; COVID/BD/153537/2024, <https://doi.org/10.54499/COVID/BD/153537/2024>), Project UIDB/05183/2025 (<https://doi.org/10.54499/UIDB/05183/2025>) to Mediterranean Institute for Agriculture, Environment and Development (MED), and Project LA/P/0121/2020 (<https://doi.org/10.54499/LA/P/0121/2020>) to CHANGE—Global Change and Sustainability Institute.

Data availability The data used in this study are part of the BioProject accession number PRJNA975359, available through the NCBI database. The specific BioSample IDs corresponding to each sample are listed in Table 2.

Competing interests The authors declare no competing interests.

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/>.

References

- Acquadro A, Portis E, Valentino D, Barchi L, Lanteri S (2020) Mind the gap: Hi-c technology boosts contiguity of the globe artichoke genome in low-recombination regions. *G3 Genes, Genomes, Genetics* 10:3557–3564. <https://doi.org/10.1534/g3.120.401446>
- Adamczyk BJ, Fernandez DE (2009) MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in *Arabidopsis* [W][OA]. *Plant Physiol* 149:1713–1723. <https://doi.org/10.1104/pp.109.135806>
- Aharoni A, Dixit S, Jetter R, Thoenes E, Van Arkel G, Pereira A (2004) The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in *Arabidopsis* W inside a box sign. *Plant Cell* 16:2463–2480. <https://doi.org/10.1105/tpc.104.022897>

- Anders S, Pyl PT, Huber W (2015) HTSeq-A python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Andrews S (2010) FastQC: a Quality Control Tool for High Throughput Sequence Data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Archontoulis SV, Struik PC, Vos J, Danalatos NG (2010) Phenological growth stages of *Cynara cardunculus*: codification and description according to the BBCH scale. *Ann Appl Biol* 156:253–270. <https://doi.org/10.1111/j.1744-7348.2009.00384.x>
- Aubourg S, Lechamy A, Bohlmann J (2002) Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Mol Genet Genomics* 267:730–745. <https://doi.org/10.1007/s00438-002-0709-y>
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37:W202–W208. <https://doi.org/10.1093/nar/gkp335>
- Bemer M, Heijmans K, Airolidi C, Davies B, Angenent GC (2010) An atlas of type I MADS box gene expression during female gametophyte and seed development in *Arabidopsis*. *Plant Physiol* 154:287–300. <https://doi.org/10.1104/pp.110.160770>
- Benjamini Y, Hochberg Y (2000) On the adaptive control of the false discovery rate in multiple testing with independent statistics
- Bhatia V, Maisnam J, Jain A, Sharma KK, Bhattacharya R (2015) Aphid-repellent pheromone E- β -farnesene is generated in transgenic *Arabidopsis thaliana* over-expressing farnesyl diphosphate synthase2. *Ann Bot* 115:581–591. <https://doi.org/10.1093/aob/mcu250>
- Blázquez M, Koornneef M, Putterill J (2001) Flowering on time: genes that regulate the floral transition workshop on the molecular basis of flowering time control. *EMBO Rep* 2(12):1078–1082. <https://doi.org/10.1093/embo-reports/kve254>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bonhomme F, Sommer H, Bernier G, Jacquard A (1997) Characterization of SaMADS D from *Sinapis alba* suggests a dual function of the gene: in inflorescence development and floral organogenesis. Kluwer Academic Publishers
- Bonini S, Magrini L, Rotiroti G, Ronchetti MP, Onorati P (1994) Genetic and environmental factors in the changing incidence of allergy. *Allergy* 49:6–14
- Bouwmeester HJ, Kodde J, Verstappen FWA, Altug IG, De Kraker JW, Wallaart TE (2002) Isolation and characterization of two germacrene A synthase cDNA clones from chicory. *Plant Physiol* 129:134–144. <https://doi.org/10.1104/pp.001024>
- Brás T, Guerreiro O, Duarte MF, Neves LA (2015) Impact of extraction parameters and concentration by nanofiltration on the recovery of phenolic compounds from *Cynara cardunculus* var. *altalis*: assessment of antioxidant activity. *Ind Crops Prod* 67:137–142. <https://doi.org/10.1016/j.indcrop.2015.01.005>
- Brás T, Paulino AFC, Neves LA, Crespo JG, Duarte MF (2020) Ultrasound assisted extraction of cynaropicrin from *Cynara cardunculus* leaves: optimization using the response surface methodology and the effect of pulse mode. *Ind Crops Prod* 150:112395. <https://doi.org/10.1016/j.indcrop.2020.112395>
- Cairns JRK, Esen A (2010) β -Glucosidases. *Cell Mol Life Sci* 67:3389–3405
- Castro MM, Rosa D, Ferro AM, Faustino A, Paulino A, Brás T, Machado E, Cruz CP, Belo ADF, Nozes P, Portugal J, Ramoa S, Mendonça D, Simões F, Duarte MF, Marum L (2021) Genetic diversity and population structure of *Cynara cardunculus* L. in southern Portugal. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0252792>
- Chávez-Hernández EC, Quiroz S, García-Ponce B, Álvarez-Buylla ER (2022) The flowering transition pathways converge into a complex gene regulatory network that underlies the phase changes of the shoot apical meristem in *Arabidopsis thaliana*. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2022.852047>
- De Martino G, Pan I, Emmanuel E, Levy A, Irish VF (2006) Functional analyses of two tomato *APETALA3* genes demonstrate diversification in their roles in regulating floral development. *Plant Cell* 18:1833–1845. <https://doi.org/10.1105/tpc.106.042978>
- de Oliveira FK, Santos LO, Buffon JG (2021) Mechanism of action, sources, and application of peroxidases. *Food Res Int* 143:110266. <https://doi.org/10.1016/j.foodres.2021.110266>
- Divekar PA, Narayana S, Divekar BA, Kumar R, Gadratagi BG, Ray A, Singh AK, Rani V, Singh V, Singh AK, Kumar A, Singh RP, Meena RS, Behera TK (2022) Plant secondary metabolites as defense tools against herbivores for sustainable crop protection. *Int J Mol Sci*. <https://doi.org/10.3390/ijms23052690>
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Eljounaidi K, Cankar K, Comino C, Moglia A, Hehn A, Bourgaud F, Bouwmeester H, Menin B, Lanteri S, Beekwilder J (2014) Cytochrome P450s from *Cynara cardunculus* L. CYP71AV9 and CYP71BL5, catalyze distinct hydroxylations in the sesquiterpene lactone biosynthetic pathway. *Plant Sci* 223:59–68. <https://doi.org/10.1016/j.plantsci.2014.03.007>
- Eljounaidi K, Comino C, Moglia A, Cankar K, Genre A, Hehn A, Bourgaud F, Beekwilder J, Lanteri S (2015) Accumulation of cynaropicrin in globe artichoke and localization of enzymes involved in its biosynthesis. *Plant Sci* 239:128–136. <https://doi.org/10.1016/j.plantsci.2015.07.020>
- Elsebai MF, Mocan A, Atanasov AG (2016) Cynaropicrin: a comprehensive research review and therapeutic potential as an anti-hepatitis C virus agent. *Front Pharmacol*. <https://doi.org/10.3389/fphar.2016.00472>
- Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
- Falcon RMG, Caoili SEC (2023) Immunologic, genetic, and ecological interplay of factors involved in allergic diseases. *Front Allergy* 4:1215616. <https://doi.org/10.3389/falgy.2023.1215616>
- Feng K, Hou XL, Xing GM, Liu JX, Duan AQ, Xu ZS, Li MY, Zhuang J, Xiong AS (2020) Advances in AP2/ERF super-family transcription factors in plant. *Crit Rev Biotechnol* 40(6):750–776. <https://doi.org/10.1080/07388551.2020.1768509>
- Ferrándiz C, Gu Q, Martienssen R, Yanofsky MF (2000) Redundant regulation of meristem identity and plant architecture by *fruitfull*, *Apetalal* and *cauliflower*. *Development* 127:725–734
- Fowler S, Thomashow MF (2002) *Arabidopsis* transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell* 14:1675–1690. <https://doi.org/10.1105/tpc.003483>
- Fraser CM, Chapple C (2011) The phenylpropanoid pathway in *Arabidopsis*. *Arabidopsis Book* 9:e0152. <https://doi.org/10.1199/tab.0152>
- Fujimoto SY, Ohta M, Usui A, Shinshi H, Ohme-Takagi M (2000) *Arabidopsis* ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant Cell* 12(3):393–404. <https://doi.org/10.1105/tpc.12.3.393>
- Gatto A, De Paola D, Bagnoli F, Vendramin GG, Sonnante G (2013) Population structure of *Cynara cardunculus* complex and the origin of the conspecific crops artichoke and cardoon. *Ann Bot* 112:855–865. <https://doi.org/10.1093/aob/mct150>

- Gaudinier A, Rodriguez-Medina J, Zhang L, Olson A, Liseron-Monfils C, Bågman AM, Foret J, Abbitt S, Tang M, Li B, Runcie DE, Kliebenstein DJ, Shen B, Frank MJ, Ware D, Brady SM (2018) Transcriptional regulation of nitrogen-associated metabolism and growth. *Nature* 563(7730):259–264. <https://doi.org/10.1038/s41586-018-0656-3>
- Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J (2000) Microarray analysis of developing Arabidopsis seeds. *Plant Physiol* 124(4):1570–1581. <https://doi.org/10.1104/pp.124.4.1570>
- Gominho J, Lourenço a, Palma P, Lourenço ME, Curt MD, Fernández J, Pereira H (2011) Large scale cultivation of *Cynara cardunculus* L. for biomass production—a case study. *Ind Crops Prod* 33:1–6. <https://doi.org/10.1016/j.indcrop.2010.09.011>
- Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36:3420–3435. <https://doi.org/10.1093/nar/gkn176>
- Gu YQ, Wildermuth MC, Chakravarthy S, Loh YT, Yang C, He X, Han Y, Martin GB (2002) Tomato transcription factors *Pti4*, *Pti5*, and *Pti6* activate defense responses when expressed in Arabidopsis. *Plant Cell* 14:817–831. <https://doi.org/10.1105/tpc.000794>
- Gull A, Ahmad Lone A, Ul Islam Wani N (2019) Biotic and Abiotic Stresses in Plants. In: Abiotic and biotic stress in plants. IntechOpen <https://doi.org/10.5772/intechopen.85832>
- Guo H, Zhang Y, Wang Z, Lin L, Cui M, Long Y, Xing Z (2019) Genome-wide identification of *WRKY* transcription factors in the asteranae. *Plants* 8:1–19. <https://doi.org/10.3390/plants8100393>
- Han Y, Tang A, Yu J, Cheng T, Wang J, Yang W, Pan H, Zhang Q (2019) *RcAP1*, a Homolog of *APETALA1*, is associated with flower bud differentiation and floral organ morphogenesis in *Rosa chinensis*. *Int J Mol Sci*. <https://doi.org/10.3390/ijms20143557>
- Hao Y, Zong X, Ren P, Qian Y, Fu A (2021) Basic helix-loop-helix (*Bhlh*) transcription factors regulate a wide range of functions in Arabidopsis. *Int J Mol Sci*. <https://doi.org/10.3390/ijms22137152>
- Harder LD, Prusinkiewicz P (2013) The interplay between inflorescence development and function as the crucible of architectural diversity. *Ann Bot* 112:1477–1493
- Heim MA, Jakoby M, Werber M, Martin C, Weissshaar B, Bailey PC (2003) The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Mol Biol Evol* 20:735–747. <https://doi.org/10.1093/molbev/msg088>
- Hernández-Reyes C, Lichtenberg E, Keller J, Delaux PM, Ott T, Schenk ST (2022) NIN-like proteins: interesting players in rhizobia-induced nitrate signaling response during interaction with non-legume host *Arabidopsis thaliana*. *Mol Plant Microbe Interact* 35(3):230–243. <https://doi.org/10.1094/MPMI-10-21-0261-R>
- Hiraga S, Ito H, Sasaki K, Yamakawa H, Mitsuhara I, Toshima H, Matsui H, Honnia' M, Ohashr Y (2000) Wound-induced expression of a tobacco peroxidase is not enhanced by ethephon and suppressed by methyl jasmonate and coronatine. *Plant Cell Physiol* 41(2):165–170. <https://doi.org/10.1093/pcp/41.2.165>
- Howell JT, Davis MR (2005) Emerging technologies / technologies naissantes plant defense mechanisms against fungal pathogens: polygalacturonase inhibitor proteins. *Can J Plant Pathol* 27:5–15
- Jack T, Brockman LL, Meyerowitz EM (1992) The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell* 68(21):683–697. [https://doi.org/10.1016/0092-8674\(92\)90144-2](https://doi.org/10.1016/0092-8674(92)90144-2)
- Joshi R, Wani SH, Singh B, Bohra A, Dar ZA, Lone AA, Pareek A, Singla-Pareek SL (2016) Transcription factors and plants response to drought stress: current understanding and future directions. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2016.01029>
- Kang HG, Foley RC, Oñate-Sánchez L, Lin C, Singh KB (2003) Target genes for OBP3, a Dof transcription factor, include novel basic helix-loop-helix domain proteins inducible by salicylic acid. *Plant J* 35:362–372. <https://doi.org/10.1046/j.1365-3113X.2003.01812.x>
- Kant MR, Jonckheere W, Knecht B, Lemos F, Liu J, Schimmel BCJ, Villarroel CA, Ataíde LMS, Dermauw W, Glas JJ, Egas M, Janssen A, Van Leeuwen T, Schuurink RC, Sabelis MW, Alba JM (2015) Mechanisms and ecological consequences of plant defence induction and suppression in herbivore communities. *Ann Bot* 115:1015–1051
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780
- Kobayashi Y, Motose H, Iwamoto K, Fukuda H (2011) Expression and genome-wide analysis of the xylogen-type gene family. *Plant Cell Physiol* 52:1095–1106. <https://doi.org/10.1093/pcp/pcr060>
- Kofuji R, Sumikawa N, Yamasaki M, Kondo K, Ueda K, Ito M, Hasebe M (2003) Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. *Mol Biol Evol* 20:1963–1977. <https://doi.org/10.1093/molbev/msg216>
- Krizek BA, Fletcher JC (2005) Molecular mechanisms of flower development: an armchair guide. *Nat Rev Genet* 6:688–698
- Laghezza Masci V, Alicandri E, Antonelli C, Paolacci AR, Marabottini R, Tomassi W, Scarascia Mugnozza G, Tiezzi A, Garzoli S, Vinciguerra V, Vettrano AM, Ovidi E, Ciaffai M (2024) *Cynara cardunculus* L. var. *scolymus* L. landrace “Carciofo Ortano” as a source of bioactive compounds. *Plants*. <https://doi.org/10.3390/plants13060761>
- Lagrimini L (1993) Molecular characterization of the lignin-forming peroxidase: Role in growth, development and response to stress. Progress summary report, April 1, 1992–March 31. <https://doi.org/10.2172/10130082>
- Lai Z, Vinod K, Zheng Z, Fan B, Chen Z (2008) Roles of Arabidopsis *WRKY3* and *WRKY4* transcription factors in plant responses to pathogens. *BMC Plant Biol*. <https://doi.org/10.1186/1471-2229-8-68>
- Lakhssassi N, Doblas VG, Rosado A, del Valle AE, Posé D, Jimenez AJ, Castillo AG, Valpuesta V, Borsani O, Botella MA (2012) The Arabidopsis tetratricopeptide thioredoxin-like gene family is required for osmotic stress tolerance and male sporogenesis. *Plant Physiol* 158(3):1252–1266. <https://doi.org/10.1104/pp.111.188920>
- Lattanzio V, Kroon PA, Linsalata V, Cardinali A (2009) Globe artichoke: a functional food and source of nutraceutical ingredients. *J Funct Foods* 1:131–144. <https://doi.org/10.1016/j.jff.2009.01.002>
- Lee S, Kim J, Han JJ, Han MJ, An G (2004) Functional analyses of the flowering time gene *OsMADS50*, the putative suppressor of overexpression of CO 1/Agamous-Like 20 (*SOC1/AGL20*) ortholog in rice. *Plant J* 38:754–764. <https://doi.org/10.1111/j.1365-3113X.2004.02082.x>
- Lee Z, Kim S, Choi SJ, Joung E, Kwon M, Park HJ, Shim JS (2023) Regulation of flowering time by environmental factors in plants. *Plants* 12(21):3680. <https://doi.org/10.3390/plants12213680>
- Lescot M, Déhais P, Moreau Y, De Moor B, Rouzé P, Rombauts S (2002) Plantcare: a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res Database Issue* 30(1):325–327. <https://doi.org/10.1093/nar/30.1.325>
- Li C, Zha W, Li W, Wang J, You A (2023a) Advances in the biosynthesis of terpenoids and their ecological functions in plant resistance. *Int J Mol Sci* 24(14):11561. <https://doi.org/10.3390/ijms241411561>
- Li L, Jia X, Qin L (2023b) Advances in the study of key genes and transcription factors regulating the mevalonate synthesis pathway

- in edible and medicinal plants. Food Sci Technol. <https://doi.org/10.1590/fst.127922>
- Li W, Li X, Chao J, Zhang Z, Wang W, Guo Y (2018) NAC family transcription factors in tobacco and their potential role in regulating leaf senescence. Front Plant Sci. <https://doi.org/10.3389/fpls.2018.01900>
- Li X, Hu Y, Su M, Zhang M, Du J, Zhou H, Zhang X, Ye Z (2022) Genome-wide analysis of terpene synthase gene family to explore candidate genes related to disease resistance in *Prunus persica*. Front Plant Sci. <https://doi.org/10.3389/fpls.2022.1032838>
- Licausi F, Ohme-Takagi M, Perata P (2013) APETALA2/ethylene responsive factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. New Phytol 199:639–649. <https://doi.org/10.1111/nph.12291>
- Liu L, Zhou Y, Szczerba MW, Li X, Lin Y (2010) Identification and application of a rice senescence-associated promoter. Plant Physiol 153:1239–1249. <https://doi.org/10.1104/pp.110.157123>
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. <https://doi.org/10.1186/s13059-014-0550-8>
- Menin B, Comino C, Portis E, Moglia A, Cankar K, Bouwmeester HJ, Lanteri S, Beekwilder J (2012) Genetic mapping and characterization of the globe artichoke (+)-germacrene A synthase gene, encoding the first dedicated enzyme for biosynthesis of the bitter sesquiterpene lactone cynaropicrin. Plant Sci 190:1–8. <https://doi.org/10.1016/j.plantsci.2012.03.006>
- Mizuno S, Osakabe Y, Maruyama K, Ito T, Osakabe K, Sato T, Shinozaki K, Yamaguchi-Shinozaki K (2007) Receptor-like protein kinase 2 (RPK 2) is a novel factor controlling anther development in *Arabidopsis thaliana*. Plant J 50:751–766. <https://doi.org/10.1111/j.1365-3113X.2007.03083.x>
- Nakano T, Suzuki K, Fujimura T, Shinshi H (2006) Genome-wide analysis of the *ERF* gene family in arabidopsis and rice. Plant Physiol 140:411–432. <https://doi.org/10.1104/pp.105.073783>
- Nguyen DT, Göpfert JC, Ikezawa N, Macnevin G, Kathiresan M, Conrad J, Spring O, Ro DK (2010) Biochemical conservation and evolution of germacrene A oxidase in asteraceae. J Biol Chem 285(22):16588–16598. <https://doi.org/10.1074/jbc.M110.111757>
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32(1):268–274. <https://doi.org/10.1093/molbev/msu300>
- Olofsson L, Engström A, Lundgren A, Brodelius PE (2011) Relative expression of genes of terpene metabolism in different tissues of *Artemisia annua* L. BMC Plant Biol. <https://doi.org/10.1186/1471-2229-11-45>
- OmicsBox—Bioinformatics Made Easy, BioBam Bioinformatics (2019). <https://www.biobam.com>
- Parenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, Colombo L (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. Plant Cell 15(7):1538–1551. <https://doi.org/10.1105/tpc.011544>
- Parker MT, Zhong Y, Dai X, Wang S, Zhao P (2014) Comparative genomic and transcriptomic analysis of terpene synthases in Arabidopsis and Medicago. IET Syst Biol 8:146–153. <https://doi.org/10.1049/iet-syb.2013.0032>
- Paulino A, Pires RC, Fernandes I, Santos J, Brás T, Rosa D, Paulo OS, Duarte MF, Marum L (2023) Transcriptome analysis of *Cynara cardunculus* in southern Portugal. In: Acta horticulturae. International Society for Horticultural Science, pp 351–358
- Pearce S, Ferguson A, King J, Wilson ZA (2015) Flowermet: a gene expression correlation network for anther and pollen development. Plant Physiol 167:1717–1730. <https://doi.org/10.1104/pp.114.253807>
- Petropoulos S, Fernandes Â, Pereira C, Tzortzakos N, Vaz J, Soković M, Barros L, Ferreira ICFR (2019) Bioactivities, chemical composition and nutritional value of *Cynara cardunculus* L. seeds. Food Chem 289:404–412. <https://doi.org/10.1016/j.foodchem.2019.03.066>
- Portis E, Acquadro A, Lanteri S (2019) The globe artichoke genome, 1st edn. Springer Cham, Switzerland
- Portis E, Barchi L, Acquadro A (2005) Genetic diversity assessment in cultivated cardoon by AFLP (amplified fragment length polymorphism) and microsatellite markers
- Puglia GD, Prjibelski AD, Vitale D, Bushmanova E, Schmid KJ, Raccuia SA (2020) Hybrid transcriptome sequencing approach improved assembly and gene annotation in *Cynara cardunculus* (L.). BMC Genomics. <https://doi.org/10.1186/s12864-020-6670-5>
- Qiu Y, Li Z, Walther D, Köhler C (2023) Updated phylogeny and protein structure predictions revise the hypothesis on the origin of MADS-box transcription factors in land plants. Mol Biol Evol 40(1):msad194. <https://doi.org/10.1093/molbev/msad194>
- Quiroz-Pacheco EN, Mora F, Boege K, Dominguez CA, del-Val E (2020) Effects of herbivory and its timing on reproductive success of a tropical deciduous tree. Ann Bot 126:957–969. <https://doi.org/10.1093/AOB/MCAA117>
- R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing
- Raccuia S, Melilli MG (2007) Biomass and grain oil yields in *Cynara cardunculus* L. genotypes grown in a Mediterranean environment. Field Crops Res 101:187–197. <https://doi.org/10.1016/j.fcr.2006.11.006>
- Rambaut, A. (2018). FigTree v1.4.4
- Ramos P, Guerra ÂR, Guerreiro O, Freire CSR, Silva AMS, Duarte MF, Silvestre AJD (2013) Lipophilic extracts of *Cynara cardunculus* L. var. altilis (DC): a source of valuable bioactive terpenic compounds. J Agric Food Chem 61:8420–8429. <https://doi.org/10.1021/jf402253a>
- Ramos PAB, Ferro AM, Oliveira MM, Gonçalves S, Freire CSR, Silvestre AJD, Duarte MF (2019) Biosynthesis and bioactivity of *Cynara cardunculus* L. guaianolides and hydroxycinnamic acids: a genomic, biochemical and health-promoting perspective
- Ramos PAB, Guerra ÂR, Guerreiro O, Santos SAO, Oliveira H, Freire CSR, Silvestre AJD, Duarte MF (2017) Antiproliferative effects of *Cynara cardunculus* L. Var. altilis (DC) lipophilic extracts. Int J Mol Sci. <https://doi.org/10.3390/ijms18010063>
- Rau D, Rodriguez M, Rapposelli E, Murgia ML, Papa R, Brown AHD, Attene G (2016) Spatial genetic structure in wild cardoon, the ancestor of cultivated globe artichoke: limited gene flow, fragmentation and population history. Plant Sci 253:194–205. <https://doi.org/10.1016/j.plantsci.2016.09.013>
- Ren R, Li D, Zhen C, Chen D, Chen X (2019) Specific roles of Os4B-Glu10, Os6B-Glu24, and Os9B-Glu33 in seed germination, root elongation, and drought tolerance in rice. Planta. <https://doi.org/10.1007/s00425-019-03125-2>
- Robatzek S, Somssich IE (2001) A new member of the Arabidopsis WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defence-related processes. Plant J 28:123–133. <https://doi.org/10.1046/j.1365-3113X.2001.01131.x>
- Rottenberg A, Zohary D (1996) The wild ancestry of cultivated artichoke. Genet Resour Crop Evol 43:53–58. <https://doi.org/10.1007/BF00126940>
- Scaglione D, Reyes-Chin-Wo S, Acquadro A, Froenicke L, Portis E, Beitel C, Tirone M, Mauro R, Lo Monaco A, Mauromicale G, Faccioli P, Cattivelli L, Rieseberg L, Michelmoro R, Lanteri S (2016) The genome sequence of the outbreeding globe artichoke constructed de novo incorporating a phase-aware low-pass sequencing strategy of F1 progeny. Sci Rep 6:1–17. <https://doi.org/10.1038/srep19427>

- Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M, Wisman E (2001) Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*
- Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM (2000) Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Plant Biol*
- Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H (1990) Genetic Control of Flower Development by Homeotic Genes in *Antirrhinum majus*
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. <https://doi.org/10.1101/gr.1239303>
- Sheshadri SA, Nishanth MJ, Simon B (2016) Stress-mediated *cis*-element transcription factor interactions interconnecting primary and specialized metabolism in *planta*. *Front Plant Sci* 7:1725. <http://doi.org/10.3389/fpls.2016.01725>
- Singh B, Sharma RA (2015) Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. 3 *Biotech* 5:129–151. <https://doi.org/10.1007/s13205-014-0220-2>
- Smaczniak C, Immink RG, Angenent GC, Kaufmann K (2012) Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development* 139(17):3081–3098. <https://doi.org/10.1242/dev.074674>
- Song L, Wang J, Jia H, Kamran A, Qin Y, Liu Y, Hao K, Han F, Zhang C, Li B, Li Y, Shen L, Wang F, Wu Y, Yang J (2020) Identification and functional characterization of *NbMLP28*, a novel MLP-like protein 28 enhancing potato virus y resistance in *Nicotiana benthamiana*. *BMC Microbiol*. <https://doi.org/10.1186/s12866-020-01725-7>
- Sonnante G, Carluccio AV, De Paolis A, Pignone D (2008) Identification of artichoke SSR markers: molecular variation and patterns of diversity in genetically cohesive taxa and wild allies. *Genet Resour Crop Evol* 55:1029–1046. <https://doi.org/10.1007/s10722-008-9310-5>
- Sonnante G, Carluccio AV, Vilatersana R, Pignone D (2007) On the origin of artichoke and cardoon from the *Cynara* gene pool as revealed by rDNA sequence variation. *Genet Resour Crop Evol* 54:483–495. <https://doi.org/10.1007/s10722-006-9199-9>
- Soyk S, Lemmon ZH, Oved M, Fisher J, Liberatore KL, Park SJ, Goren A, Jiang K, Ramos A, van der Knaap E, Van Eck J, Zamir D, Eshed Y, Lippman ZB (2017) Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Cell* 169:1142–1155.e12. <https://doi.org/10.1016/j.cell.2017.04.032>
- Stark R, Grzelak M, Hadfield J (2019) RNA sequencing: the teenage years. *Nat Rev Genet* 20:631–656
- Sun X, Wang Y, Sui N (2018) Transcriptional regulation of *bHLH* during plant response to stress. *Biochem Biophys Res Commun* 503:397–401
- Sun Y, Ji K, Liang B, Du Y, Jiang L, Wang J, Kai W, Zhang Y, Zhai X, Chen P, Wang H, Leng P (2017) Suppressing ABA uridine diphosphate glucosyltransferase (*SIUGT75C1*) alters fruit ripening and the stress response in tomato. *Plant J* 91:574–589. <https://doi.org/10.1111/tpj.13588>
- Tadege M, Sheldon CC, Helliwell CA, Upadhyaya NM, Dennis ES, Peacock WJ (2003) Reciprocal control of flowering time by *OsSOC1* in transgenic *Arabidopsis* and by FLC in transgenic rice. *Plant Biotechnol J* 1:361–369. <https://doi.org/10.1046/j.1467-7652.2003.00034.x>
- Thevissen K, Osborn RW, Acland DP, Broekaert WF (2000) Specific binding sites for an antifungal plant defensin from dahlia (*Dahlia merckii*) on fungal cells are required for antifungal activity
- Theissen G, Becker A, Di Rosa A et al (2000) A short history of MADS-box genes in plants. *Plant Mol Biol* 42:115–149. <https://doi.org/10.1023/A:1006332105728>
- Tholl D, Lee S (2011) Terpene specialized metabolism in *Arabidopsis thaliana*, 0143 edn. American Society of Plant Biologists
- Toledo-Ortiz G, Huq E, Quail PH (2003) The *Arabidopsis* basic/helix-loop-helix transcription factor family. *Plant Cell* 15(8):1749–1770. <https://doi.org/10.1105/tpc.013839>
- Velez Z, Campinho M, Guerra Á, García L, Ramos P, Guerreiro O, Felício L, Schmitt F, Duarte M (2012) Biological characterization of *Cynara cardunculus* L. methanolic extracts: antioxidant, anti-proliferative, anti-migratory and anti-angiogenic activities. *Agriculture* 2:472–492. <https://doi.org/10.3390/agriculture2040472>
- Wang HY, Klatte M, Jakoby M, Bäumlein H, Weisshaar B, Bauer P (2007) Iron deficiency-mediated stress regulation of four subgroup *Ib BHLH* genes in *Arabidopsis thaliana*. *Planta* 226:897–908. <https://doi.org/10.1007/s00425-007-0535-x>
- Wang Y, Zhang J, Hu Z, Guo X, Tian S, Chen G (2019) Genome-wide analysis of the MADS-box transcription factor family in *Solanum lycopersicum*. *Int J Mol Sci*. <https://doi.org/10.3390/ijms20122961>
- War AR, Paulraj MG, Ahmad T, Buhroo AA, Hussain B, Ignacimuthu S, Sharma HC (2012) Mechanisms of plant defense against insect herbivores. *Plant Signal Behav* 7:1306–1320
- Wiklund A (1992) The genus *Cymra* L. (Asteraceae-Cardueae). *Genus* 109:75–123
- Wingett SW, Andrews S (2018) Fastq screen: A tool for multi-genome mapping and quality control [version 1; referees: 3 approved, 1 approved with reservations]. *F1000Res* 7. <https://doi.org/10.12688/f1000research.15931.1>
- Wu Y, Li X, Zhang J, Zhao H, Tan S, Xu W, Pan J, Yang F, Pi E (2022) ERF subfamily transcription factors and their function in plant responses to abiotic stresses. *Front Plant Sci* 13:1042084
- Yamasaki K, Kigawa T, Watanabe S, Inoue M, Yamasaki T, Seki M, Shinozaki K, Yokoyama S (2012) Structural basis for sequence-specific DNA recognition by an *Arabidopsis* WRKY transcription factor. *J Biol Chem* 287:7683–7691. <https://doi.org/10.1074/jbc.M111.279844>
- Yang C, Shen W, He Y, Tian Z, Li J (2016) OVATE family protein 8 positively mediates brassinosteroid signaling through interacting with the GSK3-like kinase in Rice. *PLoS Genet* 12(6):e1006118. <https://doi.org/10.1371/journal.pgen.1006118>
- Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM (1990) The protein encoded by the *arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature* 346:35–39
- Yoshida S, Ito M, Nishida I, Watanabe A (2001) Isolation and RNA gel blot analysis of genes that could serve as potential molecular markers for leaf senescence in *Arabidopsis thaliana*. *Plant Cell Physiol*. <https://doi.org/10.1093/pcp/pce021>
- Yu D, Chen C, Chen Z (2001) Evidence for an important role of WRKY DNA binding proteins in the regulation of *NPRI* gene expression. *Plant Cell*. <https://doi.org/10.2307/3871384>
- Zhang Z, Li X, Wang J et al (2024) Evolution and function of MADS-box transcription factors in plants. *Int J Mol Sci* 25:13278. <https://doi.org/10.3390/ijms252413278>
- Falara V, Akhtar TA, Nguyen TT, Spyropoulou EA, Bleeker PM, Schauvinhold I, Matsuba Y, Bonini ME, Schillmiller AL, Last RL, Schuurink RC, Pichersky E. (2011) The tomato terpene synthase gene family. *Plant Physiol*. Oct;157(2):770-89. doi: 10.1104/pp.111.179648
- Liu Q, Majdi M, Cankar K, Goedbloed M, Charnikhova T, Verstappen FW, de Vos RC, Beekwilder J, van der Krol S, Bouwmeester HJ (2011). Reconstitution of the costunolide biosynthetic pathway in yeast and *Nicotiana benthamiana*. *PLoS One*. 6(8):e23255. doi: 10.1371/journal.pone.0023255