



## OPEN ACCESS

## EDITED BY

Antonio J. Castro,  
Spanish National Research Council (CSIC),  
Spain

## REVIEWED BY

Gianni Barcaccia,  
University of Padua, Italy  
Shuli Fan,  
Chinese Academy of Agricultural Sciences,  
China

## \*CORRESPONDENCE

Tom Ruttink  
✉ tom.ruttink@ilvo.vlaanderen.be

RECEIVED 07 March 2023

ACCEPTED 02 May 2023

PUBLISHED 13 June 2023

## CITATION

Waegneer E, Rombauts S, Baert J,  
Dauchot N, De Keyser A, Eeckhaut T,  
Haegeman A, Liu C, Maudoux O, Notté C,  
Staelens A, Van der Veken J, Van Laere K  
and Ruttink T (2023) Industrial chicory  
genome gives insights into the molecular  
timetable of anther development and male  
sterility.

*Front. Plant Sci.* 14:1181529.  
doi: 10.3389/fpls.2023.1181529

## COPYRIGHT

© 2023 Waegneer, Rombauts, Baert,  
Dauchot, De Keyser, Eeckhaut, Haegeman,  
Liu, Maudoux, Notté, Staelens, Van der  
Veken, Van Laere and Ruttink. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that  
the original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Industrial chicory genome gives insights into the molecular timetable of anther development and male sterility

Evelien Waegneer<sup>1,2</sup>, Stephane Rombauts<sup>3,4</sup>, Joost Baert<sup>1</sup>,  
Nicolas Dauchot<sup>5</sup>, Annick De Keyser<sup>3,4</sup>, Tom Eeckhaut<sup>1</sup>,  
Annelies Haegeman<sup>1</sup>, Chang Liu<sup>6</sup>, Olivier Maudoux<sup>7</sup>,  
Christine Notté<sup>7</sup>, Ariane Staelens<sup>1</sup>, Jeroen Van der Veken<sup>1</sup>,  
Katrijn Van Laere<sup>1</sup> and Tom Ruttink<sup>1,3\*</sup>

<sup>1</sup>Plant Sciences Unit, Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Melle, Belgium, <sup>2</sup>Laboratory for Plant Genetics and Crop Improvement, Division of Crop Biotechnics, Department of Biosystems, Katholieke Universiteit Leuven, Leuven, Belgium, <sup>3</sup>Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium, <sup>4</sup>Center for Plant Systems Biology, VIB, Ghent, Belgium, <sup>5</sup>Unit of Cellular and Molecular Plant Biology, UNamur, Namur, Belgium, <sup>6</sup>Department of Epigenetics, Institute of Biology, University of Hohenheim, Stuttgart, Germany, <sup>7</sup>Chicoline, A division of Cosucra Groupe Warcoing S.A., Warcoing, Belgium

Industrial chicory (*Cichorium intybus* var. *sativum*) is a biannual crop mostly cultivated for extraction of inulin, a fructose polymer used as a dietary fiber. F1 hybrid breeding is a promising breeding strategy in chicory but relies on stable male sterile lines to prevent self-pollination. Here, we report the assembly and annotation of a new industrial chicory reference genome. Additionally, we performed RNA-Seq on subsequent stages of flower bud development of a fertile line and two cytoplasmic male sterile (CMS) clones. Comparison of fertile and CMS flower bud transcriptomes combined with morphological microscopic analysis of anthers, provided a molecular understanding of anther development and identified key genes in a range of underlying processes, including tapetum development, sink establishment, pollen wall development and anther dehiscence. We also described the role of phytohormones in the regulation of these processes under normal fertile flower bud development. In parallel, we evaluated which processes are disturbed in CMS clones and could contribute to the male sterile phenotype. Taken together, this study provides a state-of-the-art industrial chicory reference genome, an annotated and curated candidate gene set related to anther development and male sterility as well as a detailed molecular timetable of flower bud development in fertile and CMS lines.

## KEYWORDS

industrial chicory, genome assembly and annotation, anther development, cytoplasmic male sterility (CMS), transcriptome profiling (RNA-seq), pollen development, regulatory pathway

## 1 Introduction

Chicory (*Cichorium intybus* L.) is a biannual crop belonging to the Asteraceae family. The species is cultivated as a leafy vegetable (e.g., witloof) or as a root crop (e.g., industrial chicory) because of its high content of inulin, a fructose polymer that is used as a prebiotic soluble dietary fiber (Shoaib et al., 2016; Aldahak et al., 2021). One of the breeding strategies used in chicory is hybrid breeding, where two compatible homozygous lines are crossed to obtain high yielding heterotic F1 offspring. Mass production of hybrid seed requires male sterile lines to prevent self-pollination, as male sterility inherently eliminates the need for manual emasculation or chemical treatment to inhibit pollen growth. Cytoplasmic male sterility (CMS) has been described in many crops (reviewed by Bohra et al. (2016)). The male sterility phenotype in CMS is maternally inheritable and often caused by altered mitochondrial genes. The CMS phenotype can be suppressed by nuclear Restorer-of-fertility genes, indicating that defective communication between mitochondria and nucleus might be at the basis. CMS can occur either spontaneously, or can be induced *via* intra- or interspecific protoplast fusion or mutagenesis. Elucidation of the mechanisms causing CMS requires knowledge on two aspects. First, a comprehensive understanding of normal flower bud development leading to fertile and mature pollen, including a detailed description of the underlying morphological, physiological, and molecular processes; and second, an inventory of the processes that are affected in CMS lines to delineate the possible causes and downstream effects that prevent pollen development and maturation.

Driven by the importance of CMS for hybrid breeding in agricultural crops, considerable efforts to elucidate the underlying mechanisms have unveiled an array of causative genes and morphological symptoms across various species (reviewed by Hanson and Bentolila (2004); Chase (2007); Chen and Liu (2014) and Kim and Zhang (2018)). Four models for possible CMS mechanisms are supported by currently available evidence; the cytotoxicity model, the energy deficiency model, the aberrant programmed cell death (PCD) model and the retrograde regulation model (Chen and Liu, 2014). The cytotoxicity model assumes that a CMS-causing protein is cytotoxic for the cell, thereby killing sporophytic and gametophytic tissues in the anther. The energy deficiency model proposes that male gametophytic development requires a high amount of energy and CMS mitochondria fail to meet these energy requirements in stamen, resulting in abortion of pollen. The aberrant PCD model predicts that male sterility is the result of disturbed timing of PCD of the tapetum, a tissue in the anther wall with an important role in feeding developing pollen. Lastly, the retrograde regulation model suggests that mitochondrial proteins alter nuclear gene expression *via* retrograde signals, which in turn cause the defects that result in male sterility. Previous research has shown that CMS involves complex processes, and therefore it is likely that most CMS phenotypes can be caused by combinations of these four models. In chicory, CMS studies have been limited to a morphological study of line CMS524 (Habarugira et al., 2015) and identification of the causative mitochondrial defect in a CMS cybrid (Rambaud et al., 1997; Varotto et al., 2001).

While most studies on chicory fertile flower bud development have focused on morphology, molecular studies are limited to flower initiation (Perilleux et al., 2013; Mathieu et al., 2020), and a detailed temporal description of molecular events during flower bud development is lacking. Starting from flower initiation, flower bud development to anthesis takes approximately 15 days in chicory, and consists of a differentiation phase and a pollen maturation phase (Habarugira et al., 2015). The differentiation phase consists of the formation of floral organs and differentiation of the anther walls into epidermis, endothecium, middle layer, and tapetum cells, surrounding the developing pollen (Habarugira et al., 2015). The subsequent pollen maturation phase is marked by degeneration of the middle layer, tapetum, septum, and stomium to facilitate pollen release.

Currently, a detailed description of the molecular events driving flower bud development in chicory is lacking, yet is critical to understand the molecular mechanisms leading to CMS. Here, we studied two chicory CMS clones; CMS36 and CMS30. These CMS clones were developed by crossing a wild chicory genotype with an inbred line of industrial chicory (Van der Veken et al., 2018). In subsequent backcrosses with industrial chicory progressively fewer pollen was produced until no pollen was present in the third and fourth backcross generation. While the cause and mechanism of male sterility is unknown in these clones, CMS may result from the incompatibility between wild chicory cytoplasm and industrial chicory nuclei. However, in both CMS clones fertility could be restored at elevated temperature (Van der Veken et al., 2018). The response time between heat shock and fertility restoration was three weeks, suggesting that early stage flower buds (size between 2 mm and 4 mm) are most responsive to fertility restoration. CMS30 male sterility was more robust as only few pollen were produced after a heat shock, whereas in CMS36 full fertility restoration was observed (Van der Veken et al., 2018).

With the aim to gain insight in the molecular regulation of flower bud development in fertile chicory and to uncover transcriptome changes associated with CMS, we first created a chromosome-scale reference genome sequence for *C. intybus* var. *sativum* (industrial chicory) and identified a functionally annotated gene set. Using that gene set as reference, we performed RNA-Seq transcriptome sequencing at eight subsequent stages of flower bud development in a fertile line and two CMS clones, CMS36 and CMS30. The temporal expression profiles of thousands of differentially expressed genes were then clustered to delineate co-expressed gene sets that, in turn, were used to dissect the temporal sequence of events during flower bud development using the gene functional annotations per co-expression cluster. Transient activation of regulatory networks, hormonal signaling pathways, and biochemical or metabolic pathways captured by the temporal expression clusters were anchored to cellular events by morphological analysis *via* light microscopy of developing flower buds. Thus, we give a comprehensive overview of the molecular processes during anther development, including ethylene, jasmonate, auxin, brassinosteroid and gibberellin signaling and response, regulation of sporopollenin biosynthesis, energy, phenylpropanoid and isoprenoid metabolism, tapetum

development and degradation, pollen development and maturation, and anther dehiscence. Additionally, we analyzed which of these developmental, biochemical, metabolic, and hormonal regulatory pathways were disturbed in clones CMS36 and CMS30, and evaluate those molecular signatures in the context of the four CMS models known from other species. Taken together, these studies provide a reference genome sequence with functionally annotated gene sets as state-of-the-art genomics resource for chicory, validated by manual curation of hundreds of candidate genes assigned to a broad range of genetic networks, thus paving the way for future structural genome biology and comparative genomics studies. They further provide a detailed timetable of molecular events that can be used as reference framework to profile the molecular phenotypes of a range of additional CMS clones to investigate which other modes of action may cause CMS in chicory, and identified key regulatory components that may be targeted to modulate CMS phenotypes.

## 2 Materials and methods

### 2.1 Plant material

For genome sequence assembly, an eighth generation inbred line of industrial chicory (*C. intybus* var. *sativum*, L8001; derived from family H79 from the ILVO chicory breeding program), was used for genomic DNA extraction (see below). For transcriptome and morphological analysis of flower buds, an industrial chicory clone (K1337) and an inbred line (L8018) were used as fertile control lines with normal pollen production (referred to as ‘fertile’), in addition to CMS clones 12-712-30 (CMS30) and 12-712-36 (CMS36). Both CMS clones can become fertile under favorable environmental circumstances, such as elevated temperatures during anther development, but CMS30 shows more stable male sterility, while CMS36 shows full fertility restoration (Van der Veken et al., 2018). All plants were kept on the field protected by a plastic cover during winter for vernalization and grown on a container field until bolting. Upon bolting, plants were transferred to a controlled growth chamber at 16h/8h day/night with temperature of  $15 \pm 2^\circ\text{C}$  and SONT lights (MASTER GreenPower 600W EL Plus, Philips, The Netherlands) with a photosynthetic photon flux density (PPFD) of  $210 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$  measured at 1 m above the floor using a LI-250A light meter (LI-COR, United States). Various tissues were sampled from line L10003 (two further generations of selfing after L8001), grown under similar conditions, to create the small tissue panel.

### 2.2 DNA extraction for genome assembly

For PacBio and Illumina whole genome shotgun (WGS) sequencing, high molecular weight (HMW) genomic DNA was extracted with a modified CTAB protocol (Murray and Thompson, 1980). In short, 0.5 g of chicory root powder was incubated in 2% CTAB buffer at pH 5.0, supplemented with 0.2%  $\beta$ -mercaptoethanol and 1% PVP-40, for 1 h at  $60^\circ\text{C}$  with regular

inversion. The mixture was extracted twice with chloroform: isoamylalcohol (24:1), and precipitated by adding 0.5 volume of 5 M NaCl, 0.05 volume 7.5 M ammonium acetate and 0.6 final volume ice-cold isopropanol. After centrifugation (15 min,  $4^\circ\text{C}$ , 4000 rpm), the pellet was washed three times with 70% ethanol. The pellet was resuspended in 10 mM Tris pH 8.0 and treated with RNase A, followed by phenol:chloroform:isoamylalcohol (25:24:1) extraction, chloroform:isoamylalcohol extraction and ammonium acetate/ethanol precipitation. The pellet was finally resuspended in 10 mM Tris pH 8.0.

Several approaches for extracting HMW nuclear genomic DNA were tried, but the best performing protocol for Oxford Nanopore Technologies (ONT) sequencing was similar to the BioNano Prep High Polysaccharides Plant Tissue DNA Isolation protocol (document number 30218, revision C, [www.bionanogenomics.com](http://www.bionanogenomics.com)). In short, etiolated fresh young leaves of industrial chicory line L8001 were chopped into a fine mesh with a razor blade. This material was filtered to recover nuclei, which were embedded into CleanCut Low Melting agarose plugs (Bio-Rad). Five plugs per falcon tube were incubated in BioNano Prep Lysis Buffer (BioNano Prep Plant DNA isolation kit; part # 80003), with proteinase-K (Qiagen) at  $50^\circ\text{C}$  overnight with intermittent mild shaking. The next day, the proteinase K solution was refreshed and incubated for another 2 h at  $50^\circ\text{C}$ , followed by RNase A (Qiagen) treatment at  $37^\circ\text{C}$  for 1 h. Subsequently, the agarose plugs were washed (5 times 15 min) and then melted at  $70^\circ\text{C}$  for 2 min in an 1.5 ml Eppendorf tube (1 plug per tube) placed at  $43^\circ\text{C}$ , and agarase (ThermoFisher Scientific) and RNase A were added and incubated at  $43^\circ\text{C}$  for 1 h. The material from the melted and digested plug was then dialyzed over a  $0.1 \mu\text{m}$  dialysis membrane (Millipore) floating on a TE wash buffer for 2 h. The remaining drop on the membrane was used as the starting material for library construction for ONT sequencing (HMW genomic DNA).

For HiC sequencing, nuclei were first extracted from chicory leaf protoplasts. Protoplasts were obtained after 3 h incubation in digestion buffer (1.5% cellulose R10, 0.4% macerozyme, 0.5 M mannitol, 20 mM KCl, 20 mM MES pH 5.5, 10 mM  $\text{CaCl}_2$ ). Debris was removed by filtering the digestion mixture using a  $100 \mu\text{m}$  filter, followed by centrifugation (100 g, 10 min,  $20^\circ\text{C}$ ). Protoplast were further purified using a sucrose gradient. Protoplast were resuspended in nuclei isolation solution pH 7.0 (0.1 mM spermidine, 10 mM phosphate buffer pH 7.0, 2.5 mM EDTA, 10 mM NaCl, 10 mM KCl, 0.2 M sucrose, 0.15% Triton X-100, 2.5 mM DTT), and formaldehyde was added (1% final concentration), followed by 15 min incubation at RT. Glycine was added (0.2 M final concentration), followed by 20 min incubation at RT and subsequently 7 min on ice. Protoplasts were transferred to large syringes, and passed 4 times through a 25G 5/8 gauge needle. The lysate was subsequently filtered through a  $10 \mu\text{m}$  filter, followed by centrifugation (400 g, 10 min). The pellet was resuspended in another nuclei isolation buffer (0.1mM spermidine, 10 mM MES-KOH pH 5.5, 2.5 mM EDTA, 10 mM NaCl, 10 mM KCl, 0.2 M sucrose, 0.15% Triton X-100, 2.5 mM DTT). The lysate was again filtered ( $10 \mu\text{m}$  filter), centrifuged (400 g, 10 min) and stored in 40% glycerol. Hi-C libraries using DpnII were generated according to Liu et al. (2017).

## 2.3 Sequencing and assembly

The PacBio data was generated on an RS II system, using P6 chemistry (Wageningen University, NL). The Illumina data was obtained with Illumina HiSeq2500 (WGS and Hi-C), MiSeq, and NextSeq500 (KULeuven Genomics Core facility, Belgium). ONT sequencing was done using the MinION Mk1B, where the Ligation Sequencing Kit 1D (SQK-LSK108, ONT) and Rapid Sequencing Kit (SQK-RAD0004, Life Technologies) were used for library preparation and tested on a FLO-MIN106 flow cell. Finally, a library was built from the obtained genomic DNA using the ligation protocol (SQK-LSK108) and sequenced (72 h run) on a fresh FLO-MIN106 flow cell. All the results from the sequencing tests and the final full run were combined for assembly. Base calling of the ONT raw data was done with guppy v3.4.1. The data were primarily assembled using the longer ONT reads combined with PacBio reads, while the Illumina reads contributed to the polishing of the genome. The ONT reads were assembled using the Flye software (v2.5), with *-min-overlap* 3000 to increase stringency at the initial overlay step, *-ont-raw* to accommodate the ONT error rate, and default parameters including nine rounds of polishing through consensus (Kolmogorov et al., 2019). Long-Read scaffolding was attempted with the PacBio reads. This assembly was further polished using Pilon (Walker et al., 2014), with Illumina short reads. The assembly was purged from haplotigs and repeat-containing contigs using *purge\_haplotigs*. The Hi-C reads were used to scaffold the 13,490 contigs into 10,515 Hi-C scaffolds using Juicer (Durand et al., 2016) and the 3D-DNA (v180114) pipeline in diploid mode, with two rounds of mis-join correction. Contigs <15 kb were not considered at this stage to reduce ambiguous mapping of the Hi-C read pairs. At this stage, the total assembly of 948 Mb contained 17 large scaffolds (N90 = 40 Mb) roughly representing chromosome arms, together with a large set of shorter contigs. These Hi-C scaffolds were further ordered and linked into the nine chromosomes using a genetic map. The resulting assembly yielded an anchored genome of 877 Mb. A final polish was applied using Medaka (<https://github.com/nanoporetech/medaka>) (v0.11.2, ONT data) and Racon (Vaser et al., 2017) (PacBio data) aiming at filling gaps resulting from the Hi-C and genetic map scaffolding. BUSCO genome completeness analysis was performed using the eudicots\_odb10 reference set of 2326 genes (Manni et al., 2021). Whole genome triplication events were analyzed with i-ADHoRe software (Proost et al., 2012).

## 2.4 Gene prediction

Protein coding genes were predicted using Augustus (Stanke et al., 2008) on the polished and frozen reference genome sequence, including hints from RNA-Seq (flower development and tissue panel) and alignments with proteins from *Arabidopsis thaliana* (Araport11, 2016), *Solanum lycopersicum* (Heinz1706, annotation ITAG3.2, solgenomics) and *Lactuca sativa* (NCBI taxonID 4236,

annotation release 100). To reduce the degree of gene model over-prediction due to repeated elements (transposable elements and simple sequence repeats), high abundant repeats were *de novo* predicted with RepeatModeler (<https://github.com/Dfam-consortium/RepeatModeler>). The obtained repeat library for chicory was used to soft-mask the genome sequence with RepeatMasker. Further curation of the predicted gene models was done *via* the ORCAE interface (Sterck et al., 2012). BUSCO gene set completeness analysis was performed using the eudicots\_odb10 reference set of 2326 genes.

## 2.5 Sampling of flower bud developmental stages

For transcriptome analysis, flower buds of the fertile line K1337 and CMS clones CMS30 and CMS36 were sampled in the growth chambers between 9:30 and 11:30 AM. Buds were divided into eight size categories using a digital caliper to measure bud size: 3-4 mm (FB4), 4-5 mm (FB5), 5-6 mm (FB6), 6-7 mm (FB7), 7-8 mm (FB8), 8-9 mm (FB9), 9-10 mm (FB10) and 10-12 mm (FB12). At least five flower buds of three different plants of the same clone were pooled. For the samples containing the smallest bud sizes, up to 20 flower buds were harvested to obtain sufficient material for RNA extraction. Samples were immediately frozen in liquid nitrogen and stored at -80°C until RNA extraction. Two biological replicate samples were taken per clone per size category. For morphological analysis, samples of fertile line L8018, CMS30 and CMS36 were harvested in identical conditions and size categories representing flower bud developmental stages (FB4-FB12).

## 2.6 RNA extraction and sequencing

Pooled samples were ground and 100 mg tissue was used for RNA extraction. Total RNA was extracted using a modified CTAB protocol (Luypaert et al., 2017), including DNase treatment (DNA-free, Invitrogen). RNA quantity and integrity were measured using Nanodrop ND-1000 (Thermo Fisher) and Experion (Biorad) instruments, respectively. For all samples, 3 µg (50 ng.mL<sup>-1</sup>) of RNA was used for TruSeq stranded RNA-Seq library preparation, quality control with Agilent Technologies 2200 TapeStation and paired-end (2x100 bp) sequencing on an Illumina NovaSeq6000 instrument (Macrogen, Korea).

## 2.7 RNA-Seq data analysis

RNA-Seq reads were trimmed, de-replicated and mapped against the CDS sequence of 54,020 predicted genes of the novel *C. intybus* var. *sativum* reference genome sequence (<https://bioinformatics.psb.ugent.be/orcae/overview/Cicin>), using BWA-mem (Li and Durbin, 2009). Data analysis was done using R v.4.0.5 with the package edgeR (McCarthy et al., 2012). Genes

with expression in less than two samples and with less than four Reads Per Kilobase of transcript per Million mapped reads (RPKM) were discarded. Normalization to scale the raw library sizes was done using the trimmed mean of M-values (to the reference, selected using default settings) (Robinson and Oshlack, 2010). A principal component analysis (PCA) was performed with log<sub>2</sub> transformed counts of all samples using the DESeq2 package in R v4.0.5 (Love et al., 2014). To identify differentially expressed genes (DEGs) of the fertile line (K1337) between the eight developmental stages FB4-FB12, we fitted a negative binomial generalized log-linear model to the RPKM values for each gene. We identified lowest and highest expression level within the fertile line developmental series and subsequently performed likelihood ratio tests (LRT). Thresholds for further filtering of the DEGs were set at minimal 4-fold-change in expression (log<sub>2</sub>-fold-change of 2, here called the FC4 gene set) and an adjusted p-value (Benjamini-Hochberg multiple testing correction) cutoff of 0.05 (Benjamini and Hochberg, 1995).

Differentially expressed genes were assigned to co-expression clusters based on the developmental stage of their maximum expression level. Clustering was performed using the Mfuzz package in R v4.0.5 (Kumar and E Futschik, 2007), followed by manual curation. Genes with different expression profile in one or both CMS clones compared to the fertile line were manually selected by examining the expression profiles.

## 2.8 Gene orthology analysis of FC4 gene set

To assign genes in the FC4 gene set to certain biochemical pathways, *Arabidopsis* gene identifiers were extracted from KEGG (Kanehisa and Goto, 2000; Kanehisa, 2019; Kanehisa et al., 2021), while *Arabidopsis* genes with a transcriptional response to a given phytohormone were extracted from Nemhauser et al. (2006). These coding DNA sequences (CDS) were used as queries to identify the *C. intybus* homologs using a tBLASTn search (e-value <1e-50), against the protein sequences of all predicted *C. intybus* gene models. Orthologs that belonged to the FC4 gene set were extracted and their (co-)expression pattern was further analyzed in the fertile line and CMS clones. For the hormone response genes, multiple response genes were not unique for one hormone, but an overview of genes involved in these processes is given in Supplemental Data Set 2. ORCAE was used for manual curation of the structural gene annotation of these genes (Sterck et al., 2012). Each predicted gene model was manually curated, using available supporting data in ORCAE (such as RNA-Seq read alignments, splicing sites, gapped alignments of *de novo* assembled transcripts, and blast hits of orthologues from closely related species).

## 2.9 Phylogenetic analysis of genes in biochemical pathways

To identify the correct ortholog for genes involved in biochemical and genetic pathways (genes described in section

“Early tapetum development”, “Phenylpropanoid metabolism”, “Energy metabolism”, “Isoprenoid metabolism”, “Anther dehiscence” and rate-limiting hormone biosynthesis genes in “Plant hormones”) phylogenetic analysis was performed. First, per gene the corresponding homology groups (HOMgroups) with all known orthologues of multiple species (*Arabidopsis thaliana*, *Lactuca sativa*, *Daucus carota* and *Helianthus annuus*) were extracted from the PLAZA dicots 5.0 database (<https://bioinformatics.psb.ugent.be/plaza>) (Van Bel et al., 2021). Next, we performed multiple sequence alignment at the protein sequence level per HOMgroup, followed by construction of a phylogenetic tree (Neighbor-Joining method, Jukes-Cantor model, bootstrap replicates set to 100) in CLC Genomics Main Workbench. Gene functions were assigned according to the closest *Arabidopsis* ortholog. For the HOMgroup including extracellular peroxidases, elucidating the true gene function was challenging due to the size of the gene family and high level of protein sequence similarity between different gene family members (Lüthje and Martinez-Cortes, 2018). Therefore, it should be noted that we cannot be sure all described extracellular peroxidases are truly involved in lignin polymerization.

## 2.10 Morphological analysis

Morphological analysis was performed based on the protocol reported by (Habarugira et al., 2015), with modifications. Harvested flower buds were vacuum fixated for 20 min in formalin acetic acid-alcohol (FAA) solution (10:7:2:1 ethanol 99%:dH<sub>2</sub>O:formalin 37%:glacial acetic acid) and then further incubated in FAA for 24 h at room temperature. Fixed samples were stored in 50% ethanol at -20°C until further use. Samples were dehydrated by changing the solvent in two steps (70% and 85% ethanol) to 100% ethanol. Dehydrated samples were prepared for embedding by infiltration with glycol methacrylate (Technovit 7100, Heraeus Kulzer, Germany) in 1.5 mL Eppendorf tubes. Large flower bud samples (>4 mm) were dissected before transfer to the infiltration solution. On initial infiltration with a 1:1 ethanol:Technovit + Technovit with hardener 1 solution, another 20 min vacuum step was performed to enhance infiltration. Afterwards, samples remained for 2 h at 4°C in the 1:1 infiltration solution. Samples were transferred into 100% Technovit solution and incubated overnight. For embedding, 1 mL Technovit hardener 2 was added to 15 mL Technovit solution and samples were embedded. Samples were kept for 1 h at room temperature and were later incubated at 37°C overnight. Next, blocks were sectioned at 5 µm using a HM360 Microtome (Thermo Scientific). Slides were stained for 10 min with 1% Toluidine Blue O (Acros Organics, Geel) in demineralized water and rinsed twice in demineralized water. Slides were subsequently covered and sealed with DPX new mounting medium (Merck KGaA, Darmstadt, Germany) and incubated for 3 days prior to microscopic analysis. Slides were analyzed using an inverted microscope with a magnification of 400x and 1000x (Leica DMi8, Wetzlar, Germany) equipped with a camera (Leica DFC450, Wetzlar, Germany) and processed using LAS software (Leica Wetzlar, Germany).

## 2.11 Accession numbers

All sequencing data was submitted to NCBI SRA under the BioProject number: PRJNA899436; Illumina Short read data: SRR22243842-SRR22243848; Hi-C: SRR22542269, SRR22542268 (restriction site: DpnII); PacBio: SRR22404031; ONT: SRR22704481-SRR22704487). Raw sequence reads of the flower bud development series and the small tissue panel were submitted to NCBI SRA under BioProject PRJNA898887, SRR22213225-SRR22213284.

The genome is publicly available on ORCAE (<https://bioinformatics.psb.ugent.be/orcae/overview/Cicin>).

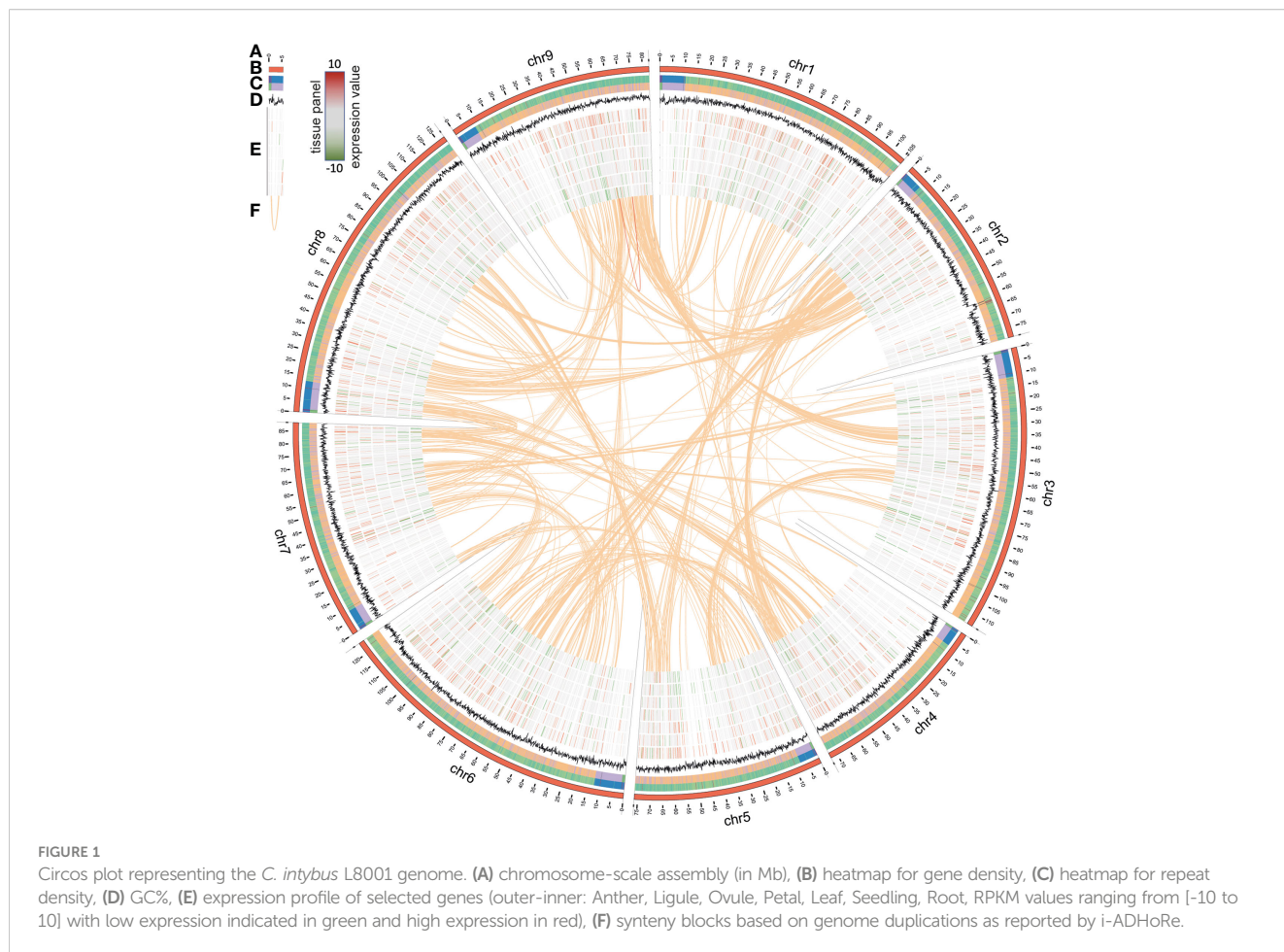
## 3 Results

### 3.1 *Cichorium intybus* L8001\_v1 genome assembly and gene annotation

An eighth generation inbred line (L8001) of *C. intybus* var. *sativum* (industrial chicory) was used for Whole-Genome Sequencing (WGS) and assembly. Genomic read data were created with Illumina HiSeq2500 (WGS, PE-100, 488M read-pairs, coverage 40x) NextSeq (WGS, PE-100, 394M read-pairs, 33x), MiSeq (WGS, PE-295, 24M read-pairs, 6x), PacBio RS II (1.38 M reads; range 1001 bp to 38,4 kb; 7.8 Gb; read-N50 8.3 kb;

genome coverage ~7x) and Oxford Nanopore Technologies (ONT) MinION (2.99 M reads; range 100 bp to 694 kb; 16.8 Gb; read-N50 10.7 kb; genome coverage ~15x). In short, the primary assembly was generated with Flye (Kolmogorov et al., 2019), using the ONT reads. Long read polishing was performed with Racon using the ONT and PacBio reads combined, and further polishing was performed with Pilon using Illumina WGS short reads. Hi-C proximity ligation libraries (PE-120, 46M reads) were used to construct a contact probability map, and scaffolds were ordered and oriented with 3D-DNA/Juicer. The Hi-C data allowed for the scaffolding into chromosome arms, but a linkage map was needed to confirm, join, and order these into the final 9 chromosomes with a total chromosome-anchored length of 877 Mb (leaving about 49 Mb unanchored) (Figures 1A–D, F; Supplemental Table 1) (Cadalen et al., 2010). Similarly to previous publications in lettuce and chicory, we observe triplication patterns across the genome (Figure 1F) (Reyes-Chin-Wo et al., 2017; Fan et al., 2022).

Augustus was trained prior to gene prediction using the eudicots gene set from BUSCO. *De novo* repeat detection was performed on the final assembly and the genome was repeat-masked prior to gene prediction (Figure 1C). Augustus predicted 53,386 genes (Figure 1B) using hints from RNA-Seq of the flower bud developmental series and a small tissue panel (Figure 1E) and proteins from reference proteomes (*Arabidopsis*, tomato, lettuce) as biological evidence (Supplemental Table 1). BUSCO genome completeness analysis



using the eudicots reference set (n=2326 genes) resulted in C:92.7% [S:88.9%, D:3.8%], F:1.2%, M:6.1%, thus only missing 140 genes from the whole BUSCO reference set. Manual curation of predicted genes putatively involved in hormonal, metabolic, and regulatory pathways was done *via* ORCAE, combined with orthology identification *via* detailed gene family analyses, protein alignment, phylogenetic analyses, and clade classification (Supplemental Data Set 2 lists all genes, functional description, assignment per pathway as described in Figures 2–4, and normalized expression values during flower bud development). The L8001\_v1 reference genome sequence, all

annotated gene models, functional description, and expression data in a small tissue panel is publicly available *via* ORCAE (<https://bioinformatics.psb.ugent.be/orcae/overview/Cicin>). Recently, another *C. intybus* genome was published by Fan et al. (2022), which we compared to our genome assembly (Supplemental Table 2; Supplemental Figures 1, 2). Genome assembly statistics and BUSCO analysis were mostly similar, except for average contig size and genome size and most of both assemblies were collinear, except for three chromosomes where the two assemblies deviate in the orientation and/or order of the chromosome arms.

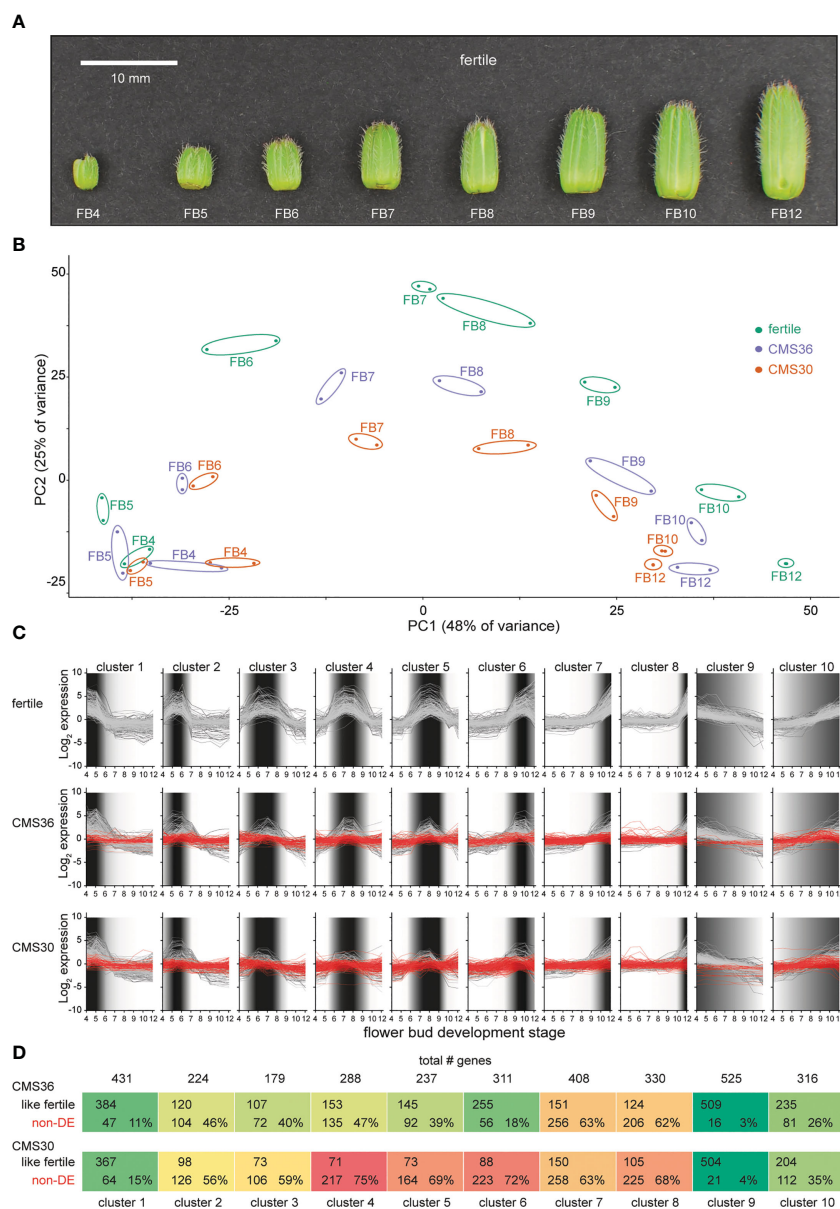
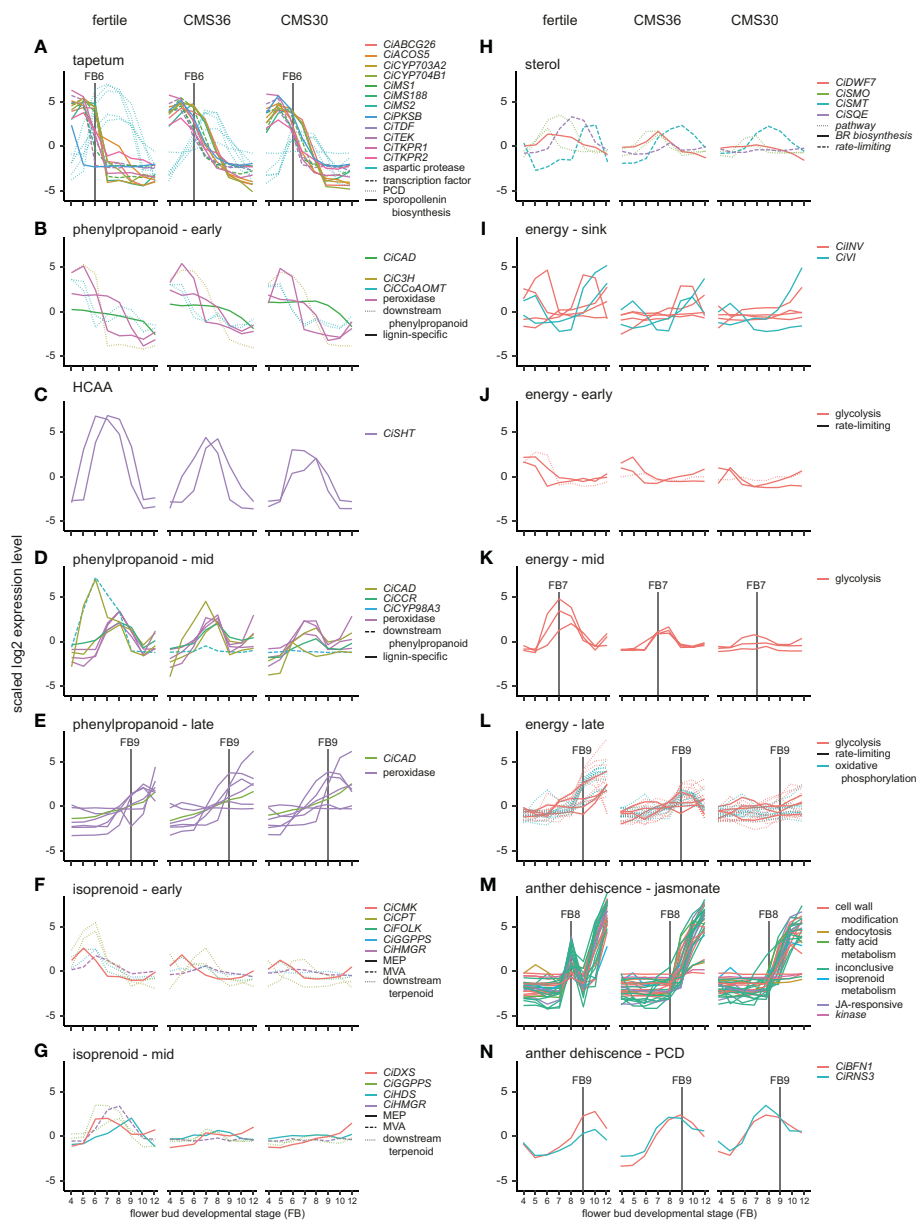


FIGURE 2

Global transcriptome analysis of flower bud development in fertile line K1337 and male sterile CMS clones CMS36 and CMS30 of industrial chicory (*Cichorium intybus* var. *sativum*). (A) Developmental series of flower buds (FB) at increasing size (4–12 mm) sampled for microscopy and RNA-Seq. (B) Principal component analysis on 30971 transcript expression profiles. (C) Ten clusters of differential expression (DE) profiles identified within the > 4-fold-change gene set of the fertile line. Dark shaded areas indicate the periods of high relative expression level. (D) Comparison of expression profiles between fertile line and CMS lines, either having a similar expression profile to the fertile line (like fertile), or DE in fertile line but not DE across the developmental series in the corresponding CMS line (non-DE).



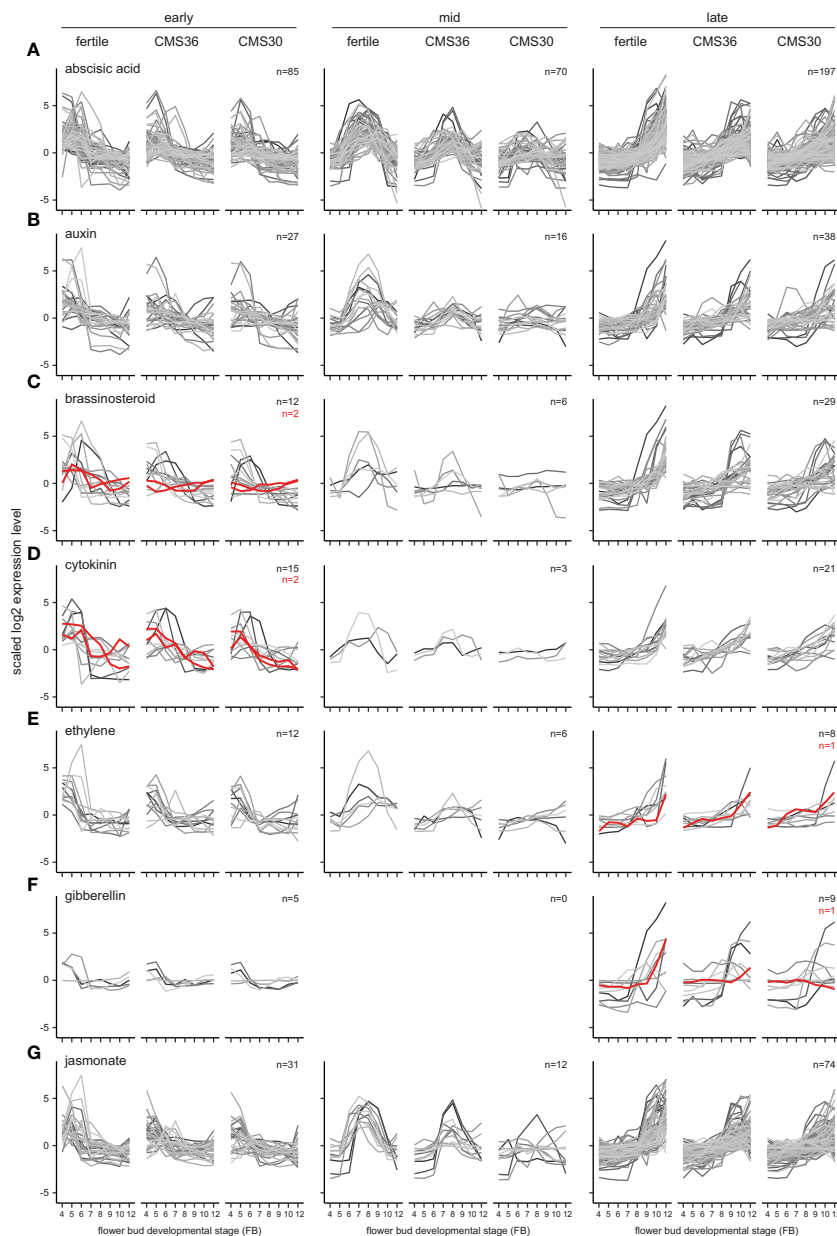
**FIGURE 3**  
 Expression profiles of genes involved in specific pathways important for anther processes of fertile line K1337 and male sterile CMS clones CMS36 and CMS30 of industrial chicory (*Cichorium intybus* var. *sativum*). x-axis shows flower bud developmental stages, ranging from 4mm (FB4) to 12mm (FB12). y-axis shows scaled log<sub>2</sub> Reads Per Kilobase transcript per Million reads mapped (RPKM) values. Individual genes are shown with different colors, while pathway grouping is shown by different line types. Gene identifiers per subgroup are listed in [Supplemental Data Set 2](#). For information on ortholog assignment, see Materials and Methods. The following (regulatory) pathways are shown: (A) tapetum development, (B) phenylpropanoid metabolism in early FB stages, (C) hydroxycinnamic acid amides (HCAA) pathway, (D) phenylpropanoid metabolism in mid FB stages, (E) phenylpropanoid metabolism in late FB stages, (F) isoprenoid metabolism in early FB stages, (G) isoprenoid metabolism in mid FB stages, (H) sterol metabolism, (I) genes related to sink strength of energy metabolism, (J) energy metabolism in early FB stages, (K) energy metabolism in mid FB stages, (L) energy metabolism in late FB stages, (M) jasmonate regulated gene cluster linked to anther dehiscence, (N) programmed cell death (PCD) genes linked to anther dehiscence.

### 3.2 Temporal transcriptome profiling during flower bud development

The chicory flower bud developmental process was reconstructed by sampling a series of flower buds of increasing sizes (4-12 mm, see Materials and Methods), reflecting eight progressive stages of flower bud development (referred to as FB4 – FB12), in a fertile line and the two CMS clones, CMS30 and

CMS36 (Figure 2A). These developmental sample series were subjected to light microscopy to map morphological development, and to transcriptome profiling to yield temporal expression profiles for 30,971 out of 53,386 predicted genes (58%), with detectable expression levels across all samples (see Materials and Methods). Principal component analysis (PCA) on the 30,971 transcript expression profiles (Figure 2B) revealed that subsequent stages of flower bud development (FB4-12) of all three lines are resolved



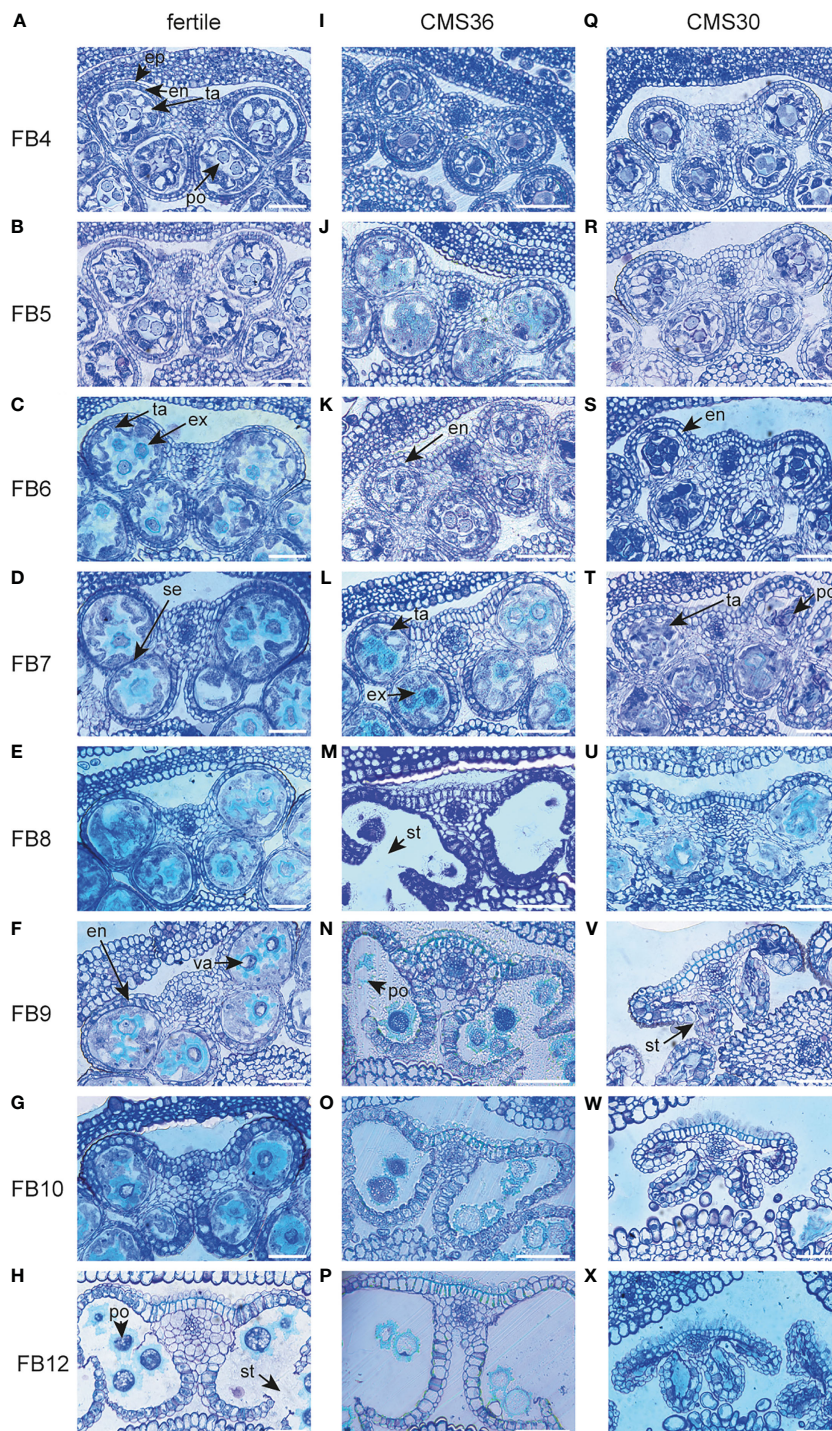


**FIGURE 4**  
 Expression profiles of phytohormone biosynthesis pathway genes (red) and response genes (greyscale) of fertile line K1337 and male sterile CMS clones CMS36 and CMS30 of industrial chicory (*Cichorium intybus* var. *sativum*). x-axis shows flower bud developmental stages, ranging from 4mm (FB4) to 12mm (FB12). y-axis shows scaled log<sub>2</sub> Reads Per Kilobase transcript per Million mapped (RPKM) values. Gene identifiers per subgroup are listed in [Supplemental Data Set 2](#). For information on ortholog assignment, see Materials and Methods. The following phytohormone pathways are shown: **(A)** abscisic acid, **(B)** auxin, **(C)** brassinosteroid, **(D)** cytokinin, **(E)** ethylene, **(F)** gibberellin, **(G)** jasmonate.

along the first principal component axis, suggesting that the majority of the expression variance in the dataset (PC1, 48%) reflects temporal changes in the transcriptome throughout flower bud development. Differences between transcriptomes of fertile and CMS clones are resolved on the second principal component axis (PC2, 25%) (Figure 2B). The PCA showed that in early stages (FB4–5) samples from fertile buds cluster closely together with samples from CMS buds, suggesting little initial transcriptional differences between fertile and CMS flower buds. From FB6 onwards, PCA analysis showed that the transcriptional difference between fertile and CMS buds progressively increased (Figure 2B). Accordingly,

microscopic analysis of morphological development revealed that CMS anther morphology initially does not markedly differ from fertile anther morphology, except for slight differences in tapetum degradation and pollen morphology in FB4–7 (CMS36) and FB4–6 (CMS30) (Figures 5A–C, I–K, Q–S). In later stages (FB8/FB9), CMS clones are characterized by early endothecium lignification (FB6) and anther dehiscence in FB8/FB9 (CMS36/CMS30), while CMS30 additionally displays inwards collapse of the anther and early pollen degradation (Figures 5F, H, K, M, S, V).

Next, we performed likelihood ratio tests (LRT) on the fertile line gene expression profiles, and identified 21,247 statistically



**FIGURE 5**

Morphological analysis and comparison of anthers of fertile line L8018 and male sterile CMS clones CMS36 and CMS30 of industrial chicory (*Cichorium intybus* var. *sativum*). (A) fertile 4mm flower bud (FB) stage with epidermis (ep), endothecium (en), tapetum (ta) and pollen (po). (B) fertile 5mm FB stage. (C) fertile 6mm FB stage with degraded tapetum cell walls (ta) and developing exine walls of pollen (ex). (D) fertile 7mm FB stage with the start of septum degradation (se). (E) fertile 8mm FB stage. (F) fertile 9mm FB stage with the start of endothecium expansion and lignification (en) and pollen vacuolization (va). (G) fertile 10mm FB stage. (H) fertile 12mm FB stage with stomium opening (st) and pollen starch bodies (po). (I) CMS36 4mm FB stage. (J) CMS36 5mm FB stage. (K) CMS36 6mm FB stage with the start of endothecium expansion and lignification (en). (L) CMS36 7mm FB stage with degraded tapetum cell walls (ta) and start of pollen exine wall formation (po). (M) CMS36 8mm FB stage with stomium opening (st). (N) CMS36 9mm FB stage with start of pollen degradation (po). (O) CMS36 10mm FB stage. (P) CMS36 12mm FB stage. (Q) CMS30 4mm FB stage. (R) CMS30 5mm FB stage. (S) CMS30 6mm FB stage with start of endothecium expansion and lignification (en). (T) CMS30 7mm FB stage with degraded tapetum cell walls (ta) and start of pollen degradation (po). (U) CMS30 8mm FB stage. (V) CMS30 9mm FB stage with stomium opening (st) and inwards collapse of the anther. (W) CMS30 10mm FB stage. (X) CMS30 12mm FB stage. Scale bar indicates 50  $\mu$ m.

significant ( $p < 0.05$ ) differentially expressed genes (DEG) within the fertile flower bud developmental series. This indicates that the vast majority of genes expressed in flower buds (69% of 30,971 expressed genes) showed dynamic expression during flower bud development. Of these, 8039 genes (25%) displayed at least a 2-fold-change in expression, while 4067 genes displayed a >4-fold-change in expression (13%, from hereon called the FC4 gene set). Next, we looked at expression levels of the FC4 gene set in the tissue panel (anther, ligule, ovule, petal, leaf, root and seedling) to assess flower bud development specific expression (Figure 1E; Supplemental Data Set 1). As expected, higher expression of these genes was observed in flower tissues (anther, ligule, ovule and petal), whereas in leaf, seedling and root tissues these genes were less expressed (Figure 1E).

The expression profiles of the FC4 gene set within the fertile line time series were then clustered to investigate the temporal sequence of events, and to delineate subsequent biological processes during flower bud development. Of the 4067 FC4 genes, 3279 genes (80%) were divided into 10 clusters; one cluster with 316 gradually up-regulated genes and one cluster with 525 gradually downregulated genes, and eight clusters containing genes transiently expressed in specific FB stages, thus revealing overall very high levels of co-expression (Figure 2C; Supplemental Data Set 1). Per cluster, genes with similar transcriptional upregulation in the fertile line and both CMS clones were separated from genes with little or no transcriptional upregulation in CMS36 and/or CMS30 (Figures 2C, D). Consistent with the global PCA analysis, a relatively small fraction of genes in clusters transiently expressed during early stages was not upregulated in CMS clones. The fraction of genes without transcriptional upregulation in CMS clones became more pronounced in clusters expressed during later stages of flower bud development. The fraction of genes without transcriptional upregulation in CMS30 was also consistently higher than that fraction of genes in CMS36 (Figure 2D), which could correspond to the severity of morphological differences observed by microscopy (Figure 5), or the different stability of sterility between the two CMS clones. This means that in general, processes that normally occur in fertile flowers are suppressed or absent in the CMS clones, and the associated gene expression in those cell types or organs is also not activated normally at the appropriate time. Thus, the comparative analysis of fertile versus CMS clones allows to uncouple the developmental processes that occur simultaneously in whole flower buds and this, in turn, adds resolution to delineate co-expressed genes and disentangle molecular processes.

### 3.3 Dissecting molecular processes during flower bud development

In the following, we performed a targeted analysis based on the functional description of genes, assigned genes to known regulatory pathways or metabolic processes based on their functional annotation, combined with their temporal expression dynamics, to describe the activity of a range of biological processes during

subsequent stages of chicory flower bud development. To further strengthen the high quality of the reference genome sequence and gene annotation, we manually curated the gene model structure of all candidate genes described in detail below (for further details, see Supplemental Data Set 2) and eliminated pseudogenes to ensure accurate expression profiling and functional annotation, and pave the way for future functional studies on individual genes.

#### 3.3.1 Early tapetum development

The tapetum is a cell layer in the anther that mainly functions as feeding tissue for developing pollen. An important function of the tapetum layer is synthesis of sporopollenin, a pollen wall precursor (Wang et al., 2018). The developmental timing of tapetum development and degradation is important to generate viable pollen and tapetum defects lead to male sterility. To investigate whether male sterility in the clones CMS36 and CMS30 involves defects in tapetum development, functioning, or degradation, we analyzed expression profiles of important regulators of tapetum development, as well as genes involved in tapetum function. For instance, in a genetic pathway proposed by Li et al. (2017), *DYTI* and *TDF1* are transcription factors regulating tapetum development, while *AMS* and *MS188* regulate sporopollenin synthesis, tapetum degradation and pollen wall development. In *Arabidopsis*, *AMS* regulates *TEK* expression, important for sexine and nexine formation in the two layers of the outer pollen wall (exine) (Li et al., 2017). *MS1* is a transcription factor involved in late tapetum development and pollen wall formation in *Arabidopsis* (Li et al., 2017; Wang et al., 2018). Here, light microscopic analysis revealed that degradation of tapetum cell walls, an early hallmark of tapetum degradation, takes place around FB6 in the fertile line (Figure 5C). Accordingly, microscopic analysis revealed that CMS36 and CMS30 display a global pattern of tapetum degradation that is similar to the fertile line, except that the cell walls are still intact at FB6 in CMS36 and CMS30 and degradation only starts at FB7 (Figures 5K, L, S, T), suggesting a slight delay of tapetum developmental timing. As expected, key transcription factors regulating tapetum development (*CiTDF1*, *CiMS188*, and *CiTEK*) were highly expressed just prior to tapetum degradation (FB4-FB5) in the fertile line (Figure 3A). Two putative *CiMS1* genes were highly expressed during early stages (FB4, FB5 and FB6) (Figure 3A). Furthermore, another eight genes involved in sporopollenin biosynthesis were all highly expressed at stages FB4 and FB5, and were strongly down-regulated at FB7, consistent with degradation of the tapetum layer (Figure 3A). In both CMS clones, *CiTDF1*, *CiMS188*, *CiTEK*, *CiMS1*, and sporopollenin biosynthesis genes showed expression profiles similar to the fertile line. However, in both CMS clones, high expression of these genes was generally maintained until FB7 instead of FB6 (Figure 3A), consistent with a slight delay of tapetum developmental timing (Figures 5K, L, S, T). Additionally, we searched for core marker genes for developmental programmed cell death (dPCD) and identified five genes encoding aspartic protease (Olvera-Carrillo et al., 2015). These genes were highly upregulated from FB6 until FB9 in fertile lines, coinciding with tapetum cell wall degradation (Figure 3A). In CMS36, upregulation of four out of five genes was delayed until FB7 and maintained until FB8, while in CMS30, upregulation of these four

was similar to the fertile line but less strong (Figure 3A). The fifth aspartic protease did not show any upregulation in CMS36 or CMS30 (Figure 3A). Taken together, this confirms that tapetum development and degradation is delayed in CMS clones, but not severely disturbed.

### 3.3.2 Phenylpropanoid metabolism

The phenylpropanoid pathway is part of secondary metabolism and produces a variety of phenolic compounds in plants including precursors for lignin, flavonoids and suberin, a component of the pollen wall (Ma and Constabel, 2019). These compounds have many functions during flower bud development, like providing building blocks for the pollen wall and several lignification processes (Battat et al., 2019). Lignin biosynthesis and breakdown plays a key role during anther development, most notably during endothecium lignification and anther dehiscence. In the fertile line, genes involved in phenylpropanoid metabolism and monolignol biosynthesis show three different expression profiles during flower bud development (Figures 3B–E). For instance, during FB4–6, the upregulated genes encode enzymes such as *CINNAMYL ALCOHOL DEHYDROGENASE (CiCAD)* and *CAFFEYOYL-COA O-METHYLTRANSFERASE (CiCCoAOMT)* acting in the phenylpropanoid pathway that produces monolignols (Figure 3B). In FB7–9, the majority of upregulated genes encode monolignol biosynthesis enzymes (*CiCAD* and *CINNAMOYL-COA REDUCTASE (CiCCR)*) or extracellular peroxidases (Figure 3D). In FB9–12, the vast majority of upregulated genes encode extracellular peroxidases (Figure 3E), putatively involved in monolignol polymerization at the cell wall (Xie et al., 2018). This suggests that biosynthesis of precursors for flavonols and monolignols takes place until FB9, whereas during stages FB9–12 lignin polymerization becomes important. Accordingly, microscopic analysis revealed that endothecium lignification started at FB9 in the fertile line. In CMS clones, genes upregulated during FB4–5 and FB6–9 show expression dynamics similar to the fertile line (Figures 3B–D). In late stages (FB9–12), one peroxidase is upregulated earlier in CMS clones (FB7) compared to the fertile line (FB9), which is consistent with earlier endothecium lignification in CMS clones. Two other peroxidases are not upregulated in CMS clones in FB12, while the fertile line does show upregulation in FB12 (Figure 3E). Taken together, the changed expression patterns of mostly peroxidases could reflect the earlier endothecium lignification, and earlier anther dehiscence in CMS clones.

The phenylpropanoid pathway also results in biosynthesis of other phenolic compounds such as hydroxycinnamic acid amides (HCAAs) (Figure 3C). Specific conjugates from HCAAs, i.e., hydroxycinnamoyl spermidines, are produced in the tapetum and accumulate in the *Arabidopsis* pollen coat (Fellenberg et al., 2009), and play a key role in pollen development. In sunflower, two *SPERMIDINE HYDROXYCINNAMOYL TRANSFERASE (SHT)* genes are hub genes regulating the flux from the phenylpropanoid pathway into the HCAA pathway during pollen development (Li et al., 2021). The chicory orthologs of these genes (*CiSHT*) show peak expression at FB6–8 in fertile flower buds (Figure 3C). In CMS36 and CMS30, *CiSHT* is reduced, with stronger reduction in

CMS30 compared to CMS36 (Figure 3C). Together these results indicate that HCAA biosynthesis, important pollen wall components, take place during FB6–8, and is decreased in CMS clones, but not severely disturbed.

### 3.3.3 Isoprenoid metabolism

Isoprenoid metabolism produces precursors for a variety of compounds with numerous functions, including plant hormones such as gibberellin, brassinosteroids, cytokinins and abscisic acid (Vranova et al., 2013). Additionally, isoprenoids are involved in photosynthetic processes, and many sterols in plants are produced *via* the isoprenoid metabolic pathway (Vranova et al., 2013). All isoprenoids are derived from the same five-carbon building block: isopentenyl diphosphate (IPP), which can be produced *via* two independent pathways in plants (Vranova et al., 2013; Tholl, 2015). Some key differences between these two pathways distinguish their effect on male sterility. The mevalonate (MVA) pathway occurs in the cytosol and leads to the production of triterpenoids and sterols, whereas the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway takes place in plastids and is mostly important for production of carotenoids and chlorophyll side chains (Suzuki et al., 2004; Vranova et al., 2013). Mutations in enzymes involved in the MVA pathway, but not the MEP pathway, have been linked to male sterility (Suzuki et al., 2004; Okada et al., 2008). It is thought that male sterility may result from disturbed sterol composition in anthers, as an important sterol precursor, squalene, could rescue the male sterile phenotype in MVA pathway mutants (Suzuki et al., 2004; Okada et al., 2008; Vranova et al., 2013).

We identified genes involved in both the MVA and MEP pathway and acting downstream in the terpenoid biosynthesis pathway. In the MVA pathway, we found two *HYDROXYMETHYLGLUTARYL-COA REDUCTASE (CiHMGGR)* genes, encoding a rate-limiting enzyme in the biosynthesis pathway. One gene was upregulated in FB6, and the other in FB7–8 in the fertile line, while no upregulation was observed in CMS clones (Figures 3F, G). In the MEP pathway, we found *1-DEOXY-DXYLULOSE 5-PHOSPHATE SYNTHASE (CiDXS)*, a rate-limiting enzyme, *4-(CYTIDINE 5'-DIPHOSPHO)-2-C-METHYL-D-ERYTHRITOL KINASE (CiCMK)* and *4-HYDROXY-3-METHYLBUT-2-ENYL DIPHOSPHATE SYNTHASE (CiHDS)*. *CiCMK* was upregulated at FB5 in the fertile line and both CMS clones (Figure 3F). *CiDXS* was upregulated at FB6–7 in the fertile line, while *CiHDS* was upregulated at FB9 (Figure 3G). In contrast, no upregulation of these genes was observed in the CMS clones (Figures 3F, G). Genes acting downstream in the terpenoid pathway included *CIS-PRENYLTRANSFERASE (CiCPT)*, *FARNESOL KINASE (CiFOLK)* and *GERANYLGERANYL PYROPHOSPHATE SYNTHASE (CiGGPPS)*. *CiCPT*, *CiFOLK*, and *CiGGPPS* were upregulated at FB5–6 in the fertile line, but in CMS clones, these genes were either not upregulated or upregulated at different stages (Figures 3F, G). Two other *CiGGPPS* genes were upregulated at FB6–8 in the fertile line, but no upregulation was observed in the CMS clones (Figure 3G). Taken together, these results suggest that both the MVA and MEP pathways, as well as the downstream terpenoid pathway are disrupted in CMS clones.

We identified four genes in the FC4 gene set involved in the sterol biosynthesis pathway: the rate-limiting enzyme SQUALENE

EPOXIDASE (*CiSQE*), *C-24 STEROL METHYLTRANSFERASE* (*CiSMT*) and two 4 *ALPHA-METHYL OXIDASEs* (*CiSMOs*). CMS clones displayed differences in gene expression of these genes compared to the fertile line, although these differences were less pronounced in CMS36 than in CMS30 (Figure 5H). Most noteworthy is the lack of upregulation in both CMS clones of the rate-limiting *CiSQE*, acting upstream in the sterol biosynthesis pathway (Shi et al., 1996; Carland et al., 2010; Doblas et al., 2013), suggesting compromised sterol biosynthesis in both CMS clones (Figure 3H). In addition, we found a brassinosteroid biosynthesis gene (*CiDWF7*) acting downstream in the sterol biosynthesis pathway, which was slightly upregulated at FB6-8 in the fertile line, and had similar expression in CMS36, but not in CMS30 (Figure 3H). Disturbed sterol composition in anthers could also be a contributing factor to the male sterility phenotype, similar to previous observations in other plant species (Suzuki et al., 2004; Okada et al., 2008; Vranova et al., 2013).

### 3.3.4 Energy metabolism

Anther and pollen development are highly energy-demanding processes, and disturbances in carbohydrate allocation or energy metabolism can lead to male sterility (Chen and Liu, 2014; Geng et al., 2018). An important factor controlling anther development is the increase of sink strength *via* cell wall invertase (INV) activity, which cleaves sucrose in glucose and fructose, compounds that can subsequently be transported into sink tissues (Zhang et al., 2010; Bihmidine et al., 2013; Morey et al., 2018; Chen et al., 2019; Yan et al., 2019). Therefore, we compared the expression profiles of genes involved in sink strength and glucose metabolism in the fertile line and CMS clones (Figures 3I–L).

In the fertile line, out of four *CiINV* genes, two genes were highly expressed in FB4-6, one gene in FB9-10, and one gene gradually increased expression during subsequent stages of flower bud development (Figure 3I), suggesting that the flower bud indeed acts as a strong sink during these stages. In addition to cell wall invertases, we also found two vacuolar invertases (*CiVI*) that were highly upregulated in both FB4-6 and FB9-12 (Figure 3I). Late stage upregulation of *CiVI* coincides with vacuolization in chicory pollen (Figures 5F, G, 6). Such *CiVI* have been found to be highly upregulated during etiolated leaf growth in chicory and this is indicative of a growing sink tissue (Van den Ende et al., 2002). In addition, as VI produces large amounts of fructose and glucose, an osmotic effect is generated that could be the driving force behind pollen vacuolization (Van den Ende et al., 2002). Furthermore, genes involved in glucose metabolism (glycolysis and oxidative phosphorylation) (Kanehisa and Goto, 2000; Kanehisa, 2019; Kanehisa et al., 2021) showed three types of transient expression dynamics; early upregulation (FB4-6), intermediate upregulation (FB6-8) and late upregulation (FB9-12) (Figures 3J–L). In FB4-6, both CMS36 and CMS30 did not show *CiINV* upregulation, whereas *CiVI* was upregulated together with genes involved in glycolysis in CMS36 and CMS30 (Figure 3J). The lack of *CiINV* upregulation in FB4-6 may suggest that less sucrose is available and perhaps there is insufficient sucrose to support proper functioning of anther tissues and early pollen development in CMS clones. In mid stages (FB6-8), genes involved in glycolysis showed weak upregulation in CMS36,

and no upregulation in CMS30 (Figure 3J). In late stages (FB9-12), CMS36 displayed weak upregulation of *CiINV* and *CiVI*, while CMS30 displayed no upregulation (Figure 3I). Accordingly, microscopic analysis revealed that CMS36 pollen are degrading and CMS30 contained no pollen at FB9-12, suggesting that *CiINV* and *CiVI* expression was needed for carbohydrate accumulation during late stages of pollen development (Figures 5N–P, V, X).

### 3.3.5 Anther dehiscence

In the final stage of anther development, anther dehiscence releases mature pollen from the anther. This is a complex process that involves coordination of pollen maturation and degradation of specialized cell types in the anther, and defects in timing or failure of dehiscence results in male sterility (as reviewed by Wilson et al. (2011)). Anther dehiscence requires three events: expansion and lignification of the endothecium; septum degradation to form a bi-locular anther; and finally stomium degradation for anther opening (Goldberg et al., 1993). Microscopic analysis of fertile flower buds revealed that septum degradation started at FB7 and finished at FB12, endothecium lignification took place at FB10, and stomium degradation resulting in anther opening occurs at FB12 (Figures 5D–H). We identified a specific cluster with about 20 co-expressed genes that showed a peak at FB8, followed by temporal suppression at FB9, and subsequent upregulation in FB10 and FB12 (Figure 3M). Two lines of evidence suggest that this cluster may reflect a functional module that regulates anther dehiscence. First, the presence of jasmonate response genes in the cluster suggests that expression of these genes might be regulated by jasmonate, a known regulator of anther dehiscence. Second, the cluster also contained genes involved in cell wall degradation and modification such as polygalacturonases, expansin, cellulase, pectinase and invertase. Such cell wall degrading enzymes are involved in septum and stomium degradation (Wilson et al., 2011). Likewise, a class III peroxidase was previously associated to lignification of the endothecium cell wall (Shigeto et al., 2015).

Microscopic analysis of CMS flower buds revealed that anther opening occurs around FB8 (CMS36) and FB9 (CMS30), thus earlier compared to the fertile line (FB12, Figures 5H, M, V). This was accompanied by a temporal shift in endothecium lignification, which occurred at FB6 in both CMS clones, earlier compared to the fertile line, where endothecium lignification occurs at FB9 (Figures 5F, K, S). Start of septum degradation was not observed in CMS clones, but the septum was fully degraded by the time of anther opening (Figures 5M, V). We also found two marker genes for dPCD, which were upregulated at FB9-10 in the fertile line; *BIFUNCTIONAL NUCLEASE1* (*CiBFN1*) and *RIBONUCLEASE3* (*CiRNS3*) (Olvera-Carrillo et al., 2015). In CMS clones, upregulation of these genes started at FB7, indicating that dPCD processes also drive the earlier timing of anther dehiscence in CMS clones (Figure 3N).

Notably, the jasmonate-regulated gene cluster showed different expression dynamics in both CMS clones compared to the fertile line, as most genes were progressively upregulated starting at FB9 until FB12, lacking the transient FB8 peak in the fertile line (Figure 3M). In CMS30, anther dehiscence was also characterized by an inwards collapse of the anther (Figure 3V). Endothecium lignification, septum degradation, anther dehydration, and pollen swelling are all necessary for correct anther opening, and especially

dehydration of endothecium and epidermal cells are required for outward bending of the anther (Wilson et al., 2011). In both barley and rice, pollen swelling is partially responsible for the force for rupture of the partially degraded septum (Matsui et al., 2000a; Matsui et al., 2000b). Additionally, pollen swelling might result from water relocation because of an increase in potassium ions in the pollen, as demonstrated in multiple species (Bashe and Mascarenhas, 1984; Heslop-Harrison et al., 1987; Matsui et al., 2000b). It is possible that the increase in potassium ions in the pollen attracts water from the surrounding tissues and might contribute to the dehydration of the endothecium and epidermis (Rehman and Yun, 2006; Wilson et al., 2011). Conversely, the absence of intact pollen in CMS30 could thus explain the inward collapse of the anther (Figures 5T, U).

## 3.4 Plant hormones

To investigate the regulatory role of plant hormones during flower bud development, we analyzed the expression profiles of genes involved in biosynthesis and response pathways of abscisic acid, auxin, brassinosteroids, cytokinins, ethylene, gibberellin and jasmonate (Kanehisa and Goto, 2000; Nemhauser et al., 2006; Kanehisa, 2019; Kanehisa et al., 2021) (see Materials and Methods, Figure 4 and Supplemental Data Set 2). We found putative response genes for most hormones upregulated in early, mid and/or late stages (Figure 4). Here, we discuss the most noteworthy expression profiles based on co-expression of rate-limiting phytohormone biosynthesis enzymes and response genes, contrast a fertile line with CMS clones, and place them in the frame of the known role of hormones in anther development as previously described in literature.

### 3.4.1 Early FB stages

In early stages of flower bud development (FB4-6) two rate-limiting brassinosteroid genes (*BRASSINOSTEROID-6-OXIDASE*) are upregulated. In CMS clones, these genes are not upregulated, similar to observations of *CiINV* in early stages (Figures 3J, 4C). In rice, brassinosteroid signaling is important for establishing sink strength of the developing anther and pollen grains, and defects in brassinosteroid synthesis/signaling led to reduced starch accumulation in pollen grains (Zhu et al., 2015), indicating that brassinosteroids could be linked to the energy metabolism in chicory.

### 3.4.2 Intermediate FB stages

In the intermediate stages of flower bud development (FB6-9) in the fertile line, auxin response genes are highly upregulated, suggesting transient activity of the auxin-dependent pathway (Figure 4B, mid). In contrast to the fertile line, transient auxin response gene expression is partially suppressed in CMS36, and almost completely suppressed in CMS30 (Figure 4B, mid). Furthermore, microscopic analysis showed that CMS36 anthers contain intact pollen, while CMS30 anthers contain degrading pollen at stages FB6-8 (Figures 5K-M, S-U). As developing pollen accumulate auxin (Cecchetti et al., 2008; Salinas-Grenet et al., 2018), our observations suggest that absence of normal pollen development explains, at least in part, the lack of

upregulation of genes involved in pollen development in CMS clones, as well as auxin-dependent gene expression. In addition, a limited number of ethylene response genes were upregulated during FB6-9, and showed a similar expression profile to auxin response genes in all three genotypes (Figure 4E, mid). Co-expression of genes involved in auxin, ethylene, and brassinosteroid metabolism, response, and/or signaling thus potentially facilitates hormonal crosstalk in the developmental stage after tapetum degradation and before endothecium lignification (FB6-9).

### 3.4.3 Late FB stages

In late stages (FB9-12) of flower bud development in the fertile line and CMS clones, jasmonate response genes are upregulated together with ethylene and gibberellin biosynthesis and response genes, suggesting transient activity (Figures 4E-G, late). Late jasmonate response genes include genes involved in programmed cell death, cell wall modification, anther dehiscence, carbohydrate transport, and auxin and gibberellin response. Both ethylene and gibberellin response genes include cell wall modification, and gibberellin response genes also include jasmonate biosynthesis and pollen development genes. These results suggest that these hormones regulate anther dehiscence, as cell wall modification and programmed cell death are key processes for anther dehiscence (Wilson et al., 2011). Genes linked to auxin response in late stages were also involved in cell wall modification, gibberellin and ethylene pathways. Taken together, these results suggest that concerted action of jasmonate, gibberellin, ethylene, and auxin regulate anther dehiscence in chicory.

## 4 Discussion

### 4.1 Genomic resource to build a comprehensive overview of gene regulatory networks controlling flower bud development

In this study, we report a new genome assembly for industrial chicory (*C. intybus* var. *sativum*), complete with structural annotations of genes. Furthermore, we identified genes involved in flower bud development and CMS, which were assigned to biochemical pathways and were manually curated. Comparison of the L8001\_v1 genome assembly and the genome assembly recently published by Fan et al. (2022) revealed global collinearity and some inversions and translocations of large chromosomal segments. Both biological and technical causes could explain these structural differences. First, the observed inversions and translocations could reflect actual biological differences in genome composition as industrial chicory was used in this study, while forage chicory was used by Fan et al. (2022). Second, different assembly strategies were used; in this study contigs were assembled using ONT reads, followed by polishing with short read data, and subsequent assembly into pseudochromosomes using HiC and final anchoring to a genetic map, whereas Fan et al. (2022) used long range PacBio HiFi sequencing. The L8001\_v1 genome, as well as all

predicted and curated genes are available on ORCAE as a public resource and can be used for further genetic studies, comparative genomics and analysis of chromosome structural elements.

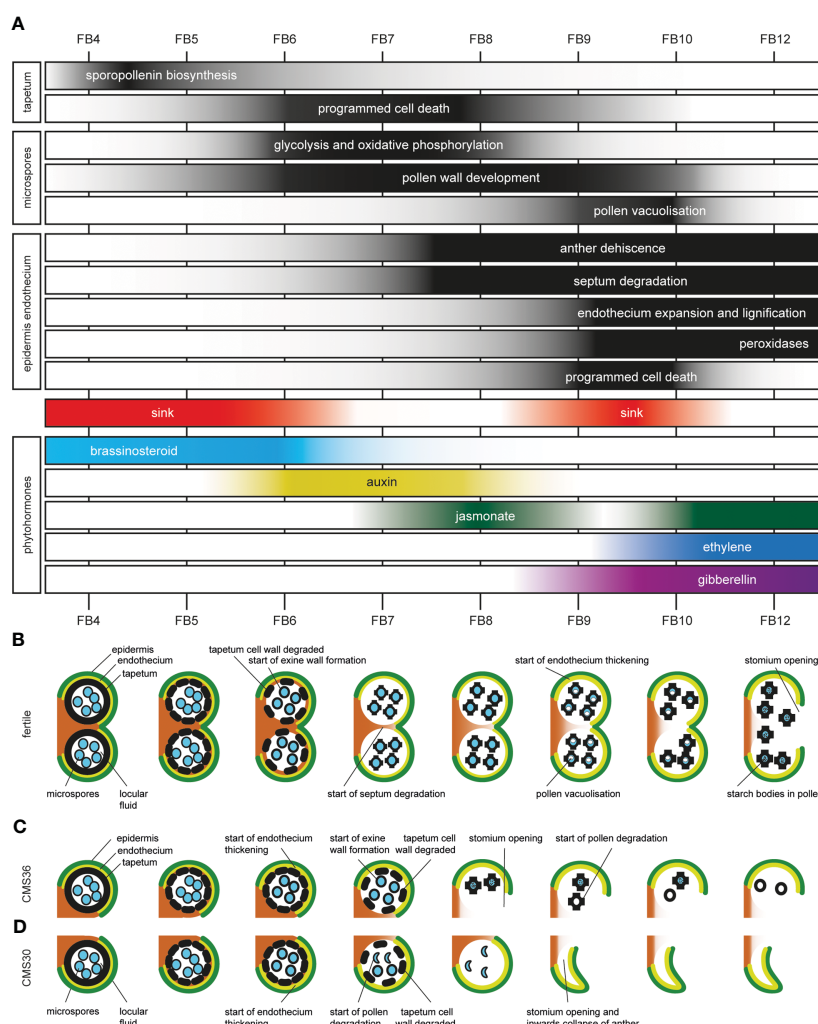
### 4.2 Fertile anther development is strictly timed and regulated

Based on the timing of morphological processes observed *via* light microscopy, combined with the temporal expression profiles of genes and their functional annotation, we were able to reconstruct the temporal sequence of morphological, cellular, and metabolic events, and link those to hormonal action and regulatory genetic pathways. Using the contrast in gene expression profiles between the fertile line and male sterile CMS clones, we could assign some of these pathways to anther and pollen development, and reconstruct a timetable of molecular events (Figure 6). In this discussion, the processes described in the results will be integrated and placed in their respective tissue context during fertile flower bud development: the

tapetum as feeding tissue supporting pollen development; pollen development; and the role of epidermis, endothecium and phytohormones in anther opening (Figure 6).

#### 4.2.1 Tapetum as feeding tissue

In the fertile line, the tapetum is present during FB4-5. During these stages, sporopollenin biosynthesis genes, as well as important transcriptional regulators (including *CiTDF1*, *CiMS188*, *CiTEK*, and *CiMS1*) are highly expressed (Figure 6A). In turn, expression of *CiVI* and *CiINV* facilitate that the anther acts as a strong sink tissue early during flower bud development (FB4-6) and help fulfill the high metabolic needs of the tapetum (Figure 6A). During these stages, brassinosteroid biosynthesis genes were upregulated early during normal pollen development but not in CMS clones, and are potentially linked to the sink strength of the anther (Figure 6A) (Zhu et al., 2015). From FB6 until FB9, dPCD marker genes (aspartic proteases) are upregulated and indicate breakdown of



**FIGURE 6** Timetable of molecular, cellular, metabolic and morphological events during anther development in fertile and CMS clones of industrial chicory (*Cichorium intybus* var. *sativum*). (A) Overview of transient activity of processes in the tapetum, developing pollen, epidermis, endothecium, and activation of sink strength and phytohormones. (B) Schematic overview of fertile anther development in chicory. (C) Schematic overview of anther development in male sterile chicory clone CMS36. (D) Schematic overview of anther development in male sterile chicory clone CMS30.

the tapetum to release pollen wall compounds and sugars for the developing pollen in the locule (Figure 6A).

#### 4.2.2 Pollen development

At early stages (FB4-5), tapetum cell walls are intact and developing pollen do not have a pollen wall yet, which starts to develop around the time of tapetum cell wall degradation (FB6) (Figure 6B). The pollen wall develops further until after the onset of anther dehiscence (Figure 6B). Exine wall formation during FB6-8 also coincides with HCAA biosynthesis, which are important components of the pollen wall. During FB6-9, genes involved in glycolysis are upregulated, which could be a signature of pollen metabolism (consistent with absence of upregulation of these genes in CMS clones that lack active pollen) (Figure 6A). In addition, both the MVA and MEP pathway for isoprenoid synthesis were active in early FB stages. The isoprenoid biosynthesis pathway produces the precursors of many important compounds, such as certain plant hormones, but also sterols and terpenes (Vranova et al., 2013; Tholl, 2015). Two hormones may be associated with the development and metabolic activity of pollen. Auxin response genes showed co-expression with energy metabolism genes in both the fertile line and CMS clones, coinciding with early stages of pollen development, and consistent with previous reports that developing pollen accumulate auxin (Cecchetti et al., 2008; Salinas-Grenet et al., 2018) (Figure 6A). In FB9-10 fertile pollen start to vacuolize, which coincides with a second period of upregulated *CiVI* and *CiINV* expression, suggesting that the anther again acts as a sink, and pollen accumulate sucrose to establish a carbohydrate pool (Figures 6A, B). In addition, the upregulation of *CiVI* also suggests that pollen further convert sucrose to glucose and fructose, which creates an osmotic effect needed for pollen vacuolization.

#### 4.2.3 Role of epidermis, endothecium and phytohormones in anther opening

During the last stages of anther development, morphological changes in endothecium and epidermis drive anther dehiscence. In the fertile line, this starts at FB7 with the onset of septum degradation (Figure 6B). At FB9, the endothecium starts expanding and endothecium cell walls become lignified, consistent with subsequent upregulation of genes involved in monolignol biosynthesis (FB6-9) and monolignol polymerization by peroxidases (FB10-12) (Figures 6A, B). Finally, stomium degradation results in anther opening and release of mature pollen.

Several lines of evidence suggest that auxin, jasmonate, ethylene, and gibberellin interact to regulate the timing of anther dehiscence events. As described above, in FB6 auxin response genes are transiently expressed in developing pollen. In *Arabidopsis*, auxin accumulation has been linked to synchronization of pollen development and anther dehiscence, such as the timing of the lignification process that marks the start of anther dehiscence (Cecchetti et al., 2008; Salinas-Grenet et al., 2018).

The transient increase in auxin response genes is followed by a sharp and transient peak in jasmonate response genes at FB8 (Figure 6A) during fertile flower bud development. After a marked transient downregulation of jasmonate-responsive gene expression, the jasmonate-dependent genes are again upregulated in FB10-12 (Figure 6A). As previously described, jasmonate is

known to be involved in anther dehiscence in many species, as evidenced by indehiscent anthers of jasmonate biosynthesis mutants (Wilson et al., 2011), and exogenously applied high jasmonate concentrations induce precocious anther dehiscence (Cecchetti et al., 2013). In addition, auxin can negatively regulate jasmonate biosynthesis and such hormonal crosstalk may regulate the timing of anther dehiscence (Cecchetti et al., 2013). Therefore, we postulate that the transient peak of jasmonate response genes at FB8, potentially in combination with negative regulation by auxin, has an important regulatory role in the timing of anther dehiscence.

Together with upregulation of jasmonate responsive genes in FB10-12, ethylene-dependent gene expression in FB10-12 was observed, while gibberellin response genes are upregulated during F9-12, suggesting hormonal crosstalk between these three phytohormones (Figure 6A). Gibberellin is important for stamen development, and is involved in filament elongation, tapetum development, pollen formation, and anther dehiscence in *Arabidopsis* (Wilson et al., 2011). Both gibberellin and jasmonate are needed for anther dehiscence in *Arabidopsis* and crosstalk between gibberellin and jasmonate-dependent pathways is required for stamen development (Peng, 2009; Marciniak and Przedniczek, 2019). A role for ethylene during anther dehiscence has been shown in petunia and tobacco (Rieu et al., 2003; Wang and Kumar, 2007). During FB10-12, also extracellular peroxidases are upregulated, known to be important for monolignol polymerization (Figure 6A).

### 4.3 Energy deficiency and retrograde signaling models may explain male sterility in CMS36 and CMS30

Global analysis of the FC4 gene set revealed that gradually more processes become uncoupled in CMS compared to the fertile line, as evidenced by PCA analysis, and the fraction of divergent gene expression profiles in CMS clones compared to the fertile line (Figure 2). Here, we provide an overview of processes that are affected in CMS clones, and evaluate the four possible mechanisms underlying CMS.

In the tapetum of CMS36 and CMS30, present during stages FB4-6, both the genetic regulators and sporopollenin genes were upregulated in early stages, but display slightly delayed developmental timing, which was consistent with microscopic observations of delayed tapetum degradation (Figures 3A, 6B-D).

The pollen in CMS36 develop normally during FB4-8 but start to degrade after anther opening, while the pollen in CMS30 initially develop normally, but start to degrade at FB7 (Figures 6B-D). Glycolysis, oxidative phosphorylation, and auxin-response genes are less (CMS36) or not (CMS30) upregulated during stages FB6-9, indicating that auxin accumulation is not taking place in CMS clones. Likewise, brassinosteroid genes are not upregulated in CMS clones. The lack of upregulation of brassinosteroid-associated genes, in turn, could be explained by the suppressed expression of isoprenoid and sterol biosynthesis pathways in CMS clones, which normally produce precursors for brassinosteroid biosynthesis.

Anther opening is premature in CMS36 (FB8) and CMS30 (FB9) compared to the fertile line (FB12), which could be due to disturbed phytohormone regulation (Figures 6B-D). Auxin accumulation is



important for the timing of anther dehiscence processes, but is decreased in developing CMS pollen. It is possible that the altered auxin levels at stages FB6-9 further lead to the disturbed jasmonate signaling in later stages, and subsequently earlier timing of dehiscence. These gene expression dynamics suggest that deregulated jasmonate induced processes, together with lack of auxin-jasmonate crosstalk, could explain the earlier timing of anther opening in CMS36 and CMS30.

Based on these results, we can evaluate which CMS model fits our observations best. As tapetum degradation seems to occur similarly in the fertile line and CMS clones, despite a slight delay in timing, evidenced by microscopy and expression data from dPCD markers, the aberrant PCD model is an unlikely mechanism. The cytotoxicity model could explain the degradation of developing pollen in CMS30 and for CMS36 in later stages. In line with the energy deficiency model, a lower sink strength in early and late stages (as evidenced by gene expression of *CiINV*), may suggest that tapetum energy requirements to support pollen development may not be met. As both CMS36 and CMS30 were established by crossing a wild chicory genotype (as mother) with an industrial chicory inbred line (as father), followed by subsequent backcrosses, it is likely that the male sterility phenotype is the result of incompatibility between mitochondria of the wild chicory and nuclear genome of the industrial chicory, compliant with the retrograde regulation model. In summary, our results are mostly in line with the energy deficiency model, and/or disturbance of retrograde signaling from mitochondrion to nucleus, or a combination of multiple mechanisms (Chen and Liu, 2014).

## 5 Conclusion

In conclusion, we provided a new industrial chicory reference genome assembly with an annotated and curated gene set related to flower bud development and male sterility that can be used for further genetic studies. The genome can also be used as a resource for genetic studies related to processes not discussed in this paper. Additionally, we provided a comprehensive molecular and morphological overview of processes related to chicory anther development and processes which are disturbed in CMS. This overview can be the basis of further studies on anther development and male sterility in chicory.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material. All data was submitted to NCBI SRA under BioProject numbers PRJNA899436 and PRJNA898887.

## Author contributions

EW, SR, OM and TR designed the research. EW, SR, JB, AK, TE, AH, CL, JV, KL and TR performed the research. JB, ND, and

OM created and provided the inbred line L8001. AH contributed new analytic/computational tools. EW, SR, TE, AH, CL, KL and TR analyzed data. EW, SR and TR wrote the paper with input from all other authors. All authors contributed to the article and approved the submitted version.

## Funding

Initial sequencing was supported in the framework of a DGA project (Grant D31-1221) with the Walloon region (DGARNE-Belgium). EW and JV were supported by a scholarship from Chicoline, a division of Cosucra Groupe Warcoing S.A., Belgium. All authors declare no other competing interests.

## Acknowledgments

We thank Ellen De Keyser and Laurence Desmet for the sampling and RNA extractions of chicory flower buds, Magali Losschaert and Sophie Carbonelle for preparing and staining microscopy sections and subsequent light microscopy and the ILVO field teams for maintenance of plant material. We also thank Bertrand Vandoorne for constructive feedback.

## Conflict of interest

Authors OM and CN are employed by Chicoline, a division of Cosucra Groupe Warcoing.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

This study received funding from Cosucra Groupe Warcoing S.A. Belgium as scholarship to EW and JV. The funder had the following involvement in the study: Cosucra Groupe Warcoing S.A. Belgium was partner in the DGA project Grant D31-1221 with the Walloon region DGARNE Belgium that initiated genome sequencing. ND and OM, affiliated with Cosucra Groupe Warcoing, created and provided the inbred line L8001.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1181529/full#supplementary-material>

## References

- Aldahak, L., Salem, K. F. M., Al-Salim, S. H. F., and Al-Khayri, J. M. (2021). "Advances in chicory (*Cichorium intybus* L.) breeding strategies," in *Advances in plant breeding strategies: vegetable crops: volume 10: leaves, flowerheads, green pods, mushrooms and truffles*. Eds. J. M. Al-Khayri, S. M. Jain and D. V. Johnson (Cham: Springer International Publishing).
- Bashe, D., and Mascarenhas, J. P. (1984). Changes in potassium ion concentrations during pollen dehydration and germination in relation to protein synthesis. *Plant Sci. Lett.* 35, 55–60. doi: 10.1016/0304-4211(84)90157-3
- Battat, M., Eitan, A., Rogachev, I., Hanhineva, K., Fernie, A., Tohge, T., et al. (2019). A MYB triad controls primary and phenylpropanoid metabolites for pollen coat patterning. *Plant Physiol.* 180, 87–108. doi: 10.1104/pp.19.00009
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bihmidine, S., Hunter, C. T., Johns, C. E., Koch, K. E., and Braun, D. M. (2013). Regulation of assimilate import into sink organs: update on molecular drivers of sink strength. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00177
- Bohra, A., Jha, U. C., Adhimoalam, P., Bisht, D., and Singh, N. P. (2016). Cytoplasmic male sterility (CMS) in hybrid breeding in field crops. *Plant Cell Rep.* 35, 967–993. doi: 10.1007/s00299-016-1949-3
- Cadalen, T., Mörchen, M., Blassiau, C., Clabaut, A., Scheer, I., Hilbert, J. L., et al. (2010). Development of SSR markers and construction of a consensus genetic map for chicory (*Cichorium intybus* L.). *Mol. Breed.* 25, 699–722. doi: 10.1007/s11032-009-9369-5
- Carland, F., Fujioka, S., and Nelson, T. (2010). The sterol methyltransferases SMT1, SMT2, and SMT3 influence arabidopsis development through nonbrassinosteroid products. *Plant Physiol.* 153, 741–756. doi: 10.1104/pp.109.152587
- Cecchetti, V., Altamura, M. M., Brunetti, P., Petrocelli, V., Falasca, G., Ljung, K., et al. (2013). Auxin controls arabidopsis anther dehiscence by regulating endothecium lignification and jasmonic acid biosynthesis. *Plant J.* 74, 411–422. doi: 10.1111/tpj.12130
- Cecchetti, V., Altamura, M. M., Falasca, G., Costantino, P., and Cardarelli, M. (2008). Auxin regulates arabidopsis anther dehiscence, pollen maturation, and filament elongation. *Plant Cell* 20, 1760–1774. doi: 10.1105/tpc.107.057570
- Chase, C. D. (2007). Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. *Trends Genet.* 23, 81–90. doi: 10.1016/j.tig.2006.12.004
- Chen, L., and Liu, Y. G. (2014). Male Sterility and fertility restoration in crops. *Annu. Rev. Plant Biol.* 65, 579–606. doi: 10.1146/annurev-arplant-050213-040119
- Chen, Z., Qin, C., Wang, M., Liao, F., Liao, Q., Liu, X., et al. (2019). Ethylene-mediated improvement in sucrose accumulation in ripening sugarcane involves increased sink strength. *BMC Plant Biol.* 19, 285. doi: 10.1186/s12870-019-1882-z
- Doblas, V. G., Amorim-Silva, V., Pose, D., Rosado, A., Esteban, A., Arro, M., et al. (2013). The SUD1 gene encodes a putative E3 ubiquitin ligase and is a positive regulator of 3-hydroxy-3-methylglutaryl coenzyme a reductase activity in arabidopsis. *Plant Cell* 25, 728–743. doi: 10.1105/tpc.112.108696
- Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., et al. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 3, 95–98. doi: 10.1016/j.cels.2016.07.002
- Fan, W., Wang, S., Wang, H., Wang, A., Jiang, F., Liu, H., et al. (2022). The genomes of chicory, endive, great burdock and yacon provide insights into asteraceae palaeopolyploidization history and plant inulin production. *Mol. Ecol. Resour.* 22, 3124–3140. doi: 10.1111/1755-0998.13675
- Fellenberg, C., Bottcher, C., and Vogt, T. (2009). Phenylpropanoid polyamine conjugate biosynthesis in arabidopsis thaliana flower buds. *Phytochemistry* 70, 1392–1400. doi: 10.1016/j.phytochem.2009.08.010
- Geng, X., Ye, J., Yang, X., Li, S., Zhang, L., and Song, X. (2018). Identification of proteins involved in carbohydrate metabolism and energy metabolism pathways and their regulation of cytoplasmic Male sterility in wheat. *Int. J. Mol. Sci.* 19, 324. doi: 10.3390/ijms19020324
- Goldberg, R. B., Beals, T. P., and Sanders, P. M. (1993). Anther development: basic principles and practical applications. *Plant Cell* 5, 1217–1229. doi: 10.1105/tpc.5.10.1217
- Habarugira, I., Hendriks, T., Quillet, M. C., Hilbert, J. L., and Rambaud, C. (2015). Effects of nuclear genomes on anther development in cytoplasmic male sterile chicories (*Cichorium intybus* L.): morphological analysis. *Sci. World J.* 2015, 529521. doi: 10.1155/2015/529521
- Hanson, M. R., and Bentolila, S. (2004). Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* 16 Suppl, S154–S169. doi: 10.1105/tpc.015966
- Heslop-Harrison, J. S., Heslop-Harrison, Y., and Reger, B. J. (1987). Anther-filament extension in liliaceae: potassium ion movement and some anatomical features. *Ann. Bot.* 59, 505–515. doi: 10.1093/oxfordjournals.aob.a087344
- Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* 28, 1947–1951. doi: 10.1002/pro.3715
- Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 49, D545–D551. doi: 10.1093/nar/gkaa970
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kim, Y. J., and Zhang, D. (2018). Molecular control of Male fertility for crop hybrid breeding. *Trends Plant Sci.* 23, 53–65. doi: 10.1016/j.tplants.2017.10.001
- Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8
- Kumar, L., and E Futschik, M. (2007). Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* 2, 5–7. doi: 10.6026/97320630002005
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, D. D., Xue, J. S., Zhu, J., and Yang, Z. N. (2017). Gene regulatory network for tapetum development in arabidopsis thaliana. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01559
- Li, S., Yaermainaiti, S., Tian, X.-M., Wang, Z.-W., Xu, W.-J., Luo, J., et al. (2021). Dynamic metabolic and transcriptomic profiling reveals the biosynthetic characteristics of hydroxycinnamic acid amides (HCAs) in sunflower pollen. *Food Res. Int.* 149, 110678. doi: 10.1016/j.foodres.2021.110678
- Liu, C., Cheng, Y.-J., Wang, J.-W., and Weigel, D. (2017). Prominent topologically associated domains differentiate global chromatin packing in rice from arabidopsis. *Nat. Plants* 3, 742–748. doi: 10.1038/s41477-017-0005-9
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8
- Lüthje, S., and Martinez-Cortes, T. (2018). Membrane-bound class III peroxidases: unexpected enzymes with exciting functions. *Int. J. Mol. Sci.* 19, 2876. doi: 10.3390/ijms19102876
- Luypaert, G., Witters, J., Van Huylenbroeck, J., De Clercq, P., De Riek, J., and De Keyser, E. (2017). Induced expression of selected plant defence related genes in pot azalea, rhododendron simsii hybrid. *Euphytica* 213, 227. doi: 10.1007/s10681-017-2010-5
- Ma, D., and Constabel, C. P. (2019). MYB repressors as regulators of phenylpropanoid metabolism in plants. *Trends Plant Sci.* 24, 275–289. doi: 10.1016/j.tplants.2018.12.003
- Manni, M., Berkeley, M. R., Seppey, M., and Zdobnov, E. M. (2021). BUSCO: assessing genomic data quality and beyond. *Curr. Protoc.* 1, e323. doi: 10.1002/cpz1.323
- Marciniak, K., and Przedniczek, K. (2019). Comprehensive insight into gibberellin- and jasmonate-mediated stamen development. *Genes* 10, 811. doi: 10.3390/genes1008811
- Mathieu, A. S., Perilleux, C., Jacquemin, G., Renard, M. E., Lutts, S., and Quinet, M. (2020). Impact of vernalization and heat on flowering induction, development and fertility in root chicory (*Cichorium intybus* L. var. sativum). *J. Plant Physiol.* 254, 153272. doi: 10.1016/j.jplph.2020.153272
- Matsui, T., Omasa, K., and Horie, T. (2000a). High temperature at flowering inhibits swelling of pollen grains, a driving force for thecae dehiscence in rice (*Oryza sativa* L.). *Plant Prod. Sci.* 3, 430–434. doi: 10.1626/pps.3.430
- Matsui, T., Omasa, K., and Horie, T. (2000b). Mechanism of septum opening in anthers of two-rowed barley (*Hordeum vulgare* L.). *Ann. Bot.* 86, 47–51. doi: 10.1006/anbo.2000.1158
- McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. *Nucleic Acids Res.* 40, 4288–4297. doi: 10.1093/nar/gks042
- Morey, S. R., Hirose, T., Hashida, Y., Miyao, A., Hirochika, H., Ohsugi, R., et al. (2018). Genetic evidence for the role of a rice vacuolar invertase as a molecular sink strength determinant. *Rice (N Y)* 11, 6. doi: 10.1186/s12284-018-0201-x
- Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4326. doi: 10.1093/nar/8.19.4321
- Nemhauser, J. L., Hong, F., and Chory, J. (2006). Different plant hormones regulate similar processes through largely nonoverlapping transcriptional responses. *Cell* 126, 467–475. doi: 10.1016/j.cell.2006.05.050
- Okada, K., Kasahara, H., Yamaguchi, S., Kawaide, H., Kamiya, Y., Nojiri, H., et al. (2008). Genetic evidence for the role of isopentenyl diphosphate isomerases in the mevalonate pathway and plant development in arabidopsis. *Plant Cell Physiol.* 49, 604–616. doi: 10.1093/pcp/pcn032
- Olvera-Carrillo, Y., Van Bel, M., Van Hautegeem, T., Fendrych, M., Huysmans, M., Simaskova, M., et al. (2015). A conserved core of programmed cell death indicator genes discriminates developmentally and environmentally induced programmed cell death in plants. *Plant Physiol.* 169, 2684–2699. doi: 10.1104/pp.15.00769
- Peng, J. (2009). Gibberellin and jasmonate crosstalk during stamen development. *J. Integr. Plant Biol.* 51, 1064–1070. doi: 10.1111/j.1744-7909.2009.00881.x

- Perilleux, C., Pieltain, A., Jacquemin, G., Bouche, F., Detry, N., D'aloia, M., et al. (2013). A root chicory MADS box sequence and the arabidopsis flowering repressor FLC share common features that suggest conserved function in vernalization and de-vernization responses. *Plant J.* 75, 390–402. doi: 10.1111/tj.12208
- Proost, S., Fostier, J., De Witte, D., Dhoedt, B., Demeester, P., Van De Peer, Y., et al. (2012). I-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res.* 40, e11. doi: 10.1093/nar/gkr955
- Rambaud, C., Bellamy, A., Dubreucq, A., Bourquin, J. C., and Vasseur, J. (1997). Molecular analysis of the fourth progeny of plants derived from a cytoplasmic male sterile chicory cybrid. *Plant Breed.* 116, 481–486. doi: 10.1111/j.1439-0523.1997.tb01035.x
- Rehman, S., and Yun, S. J. (2006). Developmental regulation of K accumulation in pollen, anthers, and papillae: are anther dehiscence, papillae hydration, and pollen swelling leading to pollination and fertilization in barley (*Hordeum vulgare* L.) regulated by changes in K concentration? *J. Exp. Bot.* 57, 1315–1321. doi: 10.1093/jxb/erj106
- Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikiti, S., Song, C., et al. (2017). Genome assembly with *in vitro* proximity ligation data and whole-genome triplication in lettuce. *Nat. Commun.* 8, 14953. doi: 10.1038/ncomms14953
- Rieu, I., Wolters-Arts, M., Derksen, J., Mariani, C., and Weterings, K. (2003). Ethylene regulates the timing of anther dehiscence in tobacco. *Planta* 217, 131–137. doi: 10.1007/s00425-003-0976-9
- Robinson, M. D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11, R25. doi: 10.1186/gb-2010-11-3-r25
- Salinas-Grenet, H., Herrera-Vasquez, A., Parra, S., Cortez, A., Gutierrez, L., Pollmann, S., et al. (2018). Modulation of auxin levels in pollen grains affects stamen development and anther dehiscence in arabidopsis. *Int. J. Mol. Sci.* 19, 2480. doi: 10.3390/ijms19092480
- Shi, J., Gonzales, R. A., and Bhattacharyya, M. K. (1996). Identification and characterization of an s-adenosyl-L-methionine: delta 24-sterol-C-methyltransferase cDNA from soybean. *J. Biol. Chem.* 271, 9384–9389. doi: 10.1074/jbc.271.16.9384
- Shigeto, J., Itoh, Y., Hirao, S., Ohira, K., Fujita, K., and Tsutsumi, Y. (2015). Simultaneously disrupting AtPrx2, AtPrx25 and AtPrx71 alters lignin content and structure in arabidopsis stem. *J. Integr. Plant Biol.* 57, 349–356. doi: 10.1111/jipb.12334
- Shoib, M., Shehzad, A., Omar, M., Rakha, A., Raza, H., Sharif, H. R., et al. (2016). Inulin: properties, health benefits and food applications. *Carbohydr. Polym.* 147, 444–454. doi: 10.1016/j.carbpol.2016.04.020
- Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntentically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24, 637–644. doi: 10.1093/bioinformatics/btn013
- Sterck, L., Billiau, K., Abeel, T., Rouze, P., and Van De Peer, Y. (2012). ORCAE: online resource for community annotation of eukaryotes. *Nat. Methods* 9, 1041. doi: 10.1038/nmeth.2242
- Suzuki, M., Kamide, Y., Nagata, N., Seki, H., Ohyama, K., Kato, H., et al. (2004). Loss of function of 3-hydroxy-3-methylglutaryl coenzyme a reductase 1 (HMG1) in arabidopsis leads to dwarfing, early senescence and male sterility, and reduced sterol levels. *Plant J.* 37, 750–761. doi: 10.1111/j.1365-313x.2004.02003.x
- Tholl, D. (2015). Biosynthesis and biological functions of terpenoids in plants. *Adv. Biochem. Eng. Biotechnol.* 148, 63–106. doi: 10.1007/10\_2014\_295
- Van Bel, M., Silvestri, F., Weitz, E. M., Krefit, L., Botzki, A., Coppens, F., et al. (2021). PLAZA 5.0: extending the scope and power of comparative and functional genomics in plants. *Nucleic Acids Res.* 50, D1468–D1474. doi: 10.1093/nar/gkab1024
- Van den Ende, W., Michiels, A., Le Roy, K., and Van Laere, A. (2002). Cloning of a vacuolar invertase from Belgian endive leaves (*Cichorium intybus*). *Physiologia Plantarum* 115, 504–512. doi: 10.1034/j.1399-3054.2002.1150404.x
- Van der Veken, J., Vandermoere, S., Maudoux, O., Notté, C., Eeckhaut, T., Baert, J., et al. (2018). Temperature related instability of a naturally induced cytoplasmic male sterility in industrial chicory (*Cichorium intybus* L.). *Plant Breed.* 137, 423–432. doi: 10.1111/pbr.12582
- Varotto, S., Nenz, E., Lucchin, M., and Parrini, P. (2001). Production of asymmetric somatic hybrid plants between *cichorium intybus* L. and *helianthus annuus* L. *Theor. Appl. Genet.* 102, 950–956. doi: 10.1007/s001220000485
- Vaser, R., Sovic, I., Nagarajan, N., and Sikic, M. (2017). Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116
- Vranova, E., Coman, D., and Grussem, W. (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu. Rev. Plant Biol.* 64, 665–700. doi: 10.1146/annurev-arplant-050312-120116
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9, e112963. doi: 10.1371/journal.pone.0112963
- Wang, K., Guo, Z. L., Zhou, W. T., Zhang, C., Zhang, Z. Y., Lou, Y., et al. (2018). The regulation of sporopollenin biosynthesis genes for rapid pollen wall formation. *Plant Physiol.* 178, 283–294. doi: 10.1104/pp.18.00219
- Wang, Y., and Kumar, P. P. (2007). Characterization of two ethylene receptors PhERS1 and PhETR2 from petunia: PhETR2 regulates timing of anther dehiscence. *J. Exp. Bot.* 58, 533–544. doi: 10.1093/jxb/erl229
- Wilson, Z. A., Song, J., Taylor, B., and Yang, C. (2011). The final split: the regulation of anther dehiscence. *J. Exp. Bot.* 62, 1633–1649. doi: 10.1093/jxb/err014
- Xie, M., Zhang, J., Tschaplinski, T. J., Tuskan, G. A., Chen, J. G., and Muchero, W. (2018). Regulation of lignin biosynthesis and its role in growth-defense tradeoffs. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01427
- Yan, W., Wu, X., Li, Y., Liu, G., Cui, Z., Jiang, T., et al. (2019). Cell wall invertase 3 affects cassava productivity via regulating sugar allocation from source to sink. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00541
- Zhang, H., Liang, W., Yang, X., Luo, X., Jiang, N., Ma, H., et al. (2010). Carbon starved anther encodes a MYB domain protein that regulates sugar partitioning required for rice pollen development. *Plant Cell* 22, 672–689. doi: 10.1105/tpc.109.073668
- Zhu, X., Liang, W., Cui, X., Chen, M., Yin, C., Luo, Z., et al. (2015). Brassinosteroids promote development of rice pollen grains and seeds by triggering expression of carbon starved anther, a MYB domain protein. *Plant J.* 82, 570–581. doi: 10.1111/tj.12820