



OPEN Dissecting flowering time and flower color in *Carum carvi* utilizing a long-read draft genome and a GBS-based QTL mapping

Daniel von Maydell¹✉, Fang-Shiang Lim², Martin Junghanns³, Yvonne Poeschl⁴, Holger Budahn¹, Frank Marthe¹, Jens Keilwagen^{1,2} & Thomas Schmutzer^{1,4}

Caraway (*Carum carvi* L.) is a major essential oil crop with biennial and annual flowering types. As basic research resource, we developed a draft genome assembly using long-read ONT sequencing and provide a structural and functional gene annotation for an annual caraway inbred line. To elucidate the genetic control of flowering, a genotyping-by-sequencing (GBS) was conducted for an F_2 population ($N = 187$) generating 731 strictly filtered SNPs. A linkage map was constructed spanning 663 cM across 10 linkage groups with in total 634 (full map) or 259 (thinned map) SNPs. Contrary to its dominance in F_1 , annual flowering occurred in only 36% of F_2 plants under late sowing conditions. Furthermore, the annual F_2 plants exhibited delayed flowering compared to the annual parent. QTL analysis identified five significant QTLs for (adjusted) flowering time (LG02, LG03, LG05, LG08 and LG10) explaining 6.1% to 10.5% (in total 42.7%) of phenotypic variance. The results support a polygenic predominately additive model for flowering induction in caraway. In addition, two QTLs for flower color (LG01, LG10) were detected explaining 10.9% and 26.0% of phenotypic variance, respectively. This study provides a comprehensive genomic resource for caraway, bridging the gap between traditional breeding and molecular improvement.

Keywords Flowering regulation, Genetic map, Linkage, Quantitative trait locus, Umbelliferae, Vernalization

Caraway (*Carum carvi* L., $2n=2x=20$) is a key essential oil crop of the Apiaceae family, with an estimated genome size of 2.14 pg/1C¹. It occurs in two distinct flowering types: biennial and annual². Biennials require a cold stimulus (vernalization) once they have reached a juvenile stage of at least six leaves and a rootstock diameter of 5 mm before winter to initiate flowering³. Depending on growing area, biennials need to be sown from May to end of June to reach that size³, meaning they occupy the field for two growing seasons. In contrast, annuals complete their life cycle within a single year. While annuals are typically sown in spring, they can also be grown as "winter-annuals" (autumn sowing) to achieve higher grain yields by utilizing early spring moisture and a prolonged vegetation period. However, this cultivation system requires a high degree of winter-hardiness, a trait currently found in only one commercial annual variety ('Aprim', Agritec Ltd.).

In general, genetic diversity is limited in the annual gene pool, which is thought to have originated from North Africa⁴. In contrast, the biennial gene pool is widespread across Eurasia and offers a vast reservoir of genetic variation for agronomic traits, quality, and environmental adaptation. To enrich the annual gene pool, introgression from biennials is a promising breeding strategy. While crossbreeding is possible through synchronized flowering by artificial vernalization, fixation of an annual flowering habit and early flowering, which is associated with higher yields⁵, remains challenging. Flowering type in caraway is reported to be controlled by a major locus, with annual flowering habit being dominant^{3,6}. This dominance allows the recessive biennial allele to persist undetected throughout the breeding process. Further on, Nemeth and Pluhar⁶ hypothesized that several minor loci of biennials may decelerate flowering time.

In a screening of various annual and biennial genotypes, few exceptionally rose-flowering biennial genotypes were detected⁷. Although not a primary trait for field production, rose-flowering could increase the economic

¹Institute for Breeding Research On Horticultural Crops, Julius Kühn Institute, Quedlinburg, Germany. ²Institute for Biosafety in Plant Biotechnology, Julius Kühn Institute, Quedlinburg, Germany. ³Dr. Junghanns GmbH, OT Groß-Schierstedt, Aschersleben, Germany. ⁴Institute of Agricultural and Nutritional Sciences, Chair of Plant Breeding, Martin Luther University Halle-Wittenberg, Halle, Germany. ✉email: daniel.maydell@julius-kuehn.de

potential of caraway as an ornamental plant. Since these genotypes are small wild types, an introgression of the rose-flowering trait into cultivated material is necessary for economic use.

Identifying diagnostic genetic markers for flowering type, flowering time, and flower color could aid in efficiently screening early breeding generations to fix annual, early-flowering, or rose-flowering habits. So far, single nucleotide polymorphism (SNP)-based diagnostic (PACE/KASP) markers have been developed from a diversity panel of annual and biennial flowering types⁸. These markers have been proven useful for checking crossing success in various crosses between annuals and biennials, but cannot reliably predict the flowering habit. Bocianowski and Seidler-Lozykowska⁹ reported significant associations for several quantitative traits using randomly amplified polymorphic DNA (RAPD) markers, but did not include flowering time. In general, the genetic dissection of traits in caraway is hampered by a lack of physical sequence information (a genome assembly) and a genetic map.

In this study, we aimed to address these limitations by:

1. Detecting quantitative trait loci (QTL) affecting flowering type, flowering time, and flower color based on a biparental F_2 mapping population utilizing genotyping-by-sequencing (GBS) data. Meanwhile, developing a genetic map to bring genetic markers into an order.
2. Introducing an annotated draft genome assembly of caraway based on long-read sequencing data and using this assembly for efficient mapping of GBS data as well as for the genetic dissection of detected QTL regions to identify candidate genes.

Material and methods

Construction of a draft genome assembly & annotation

DNA extraction and purification

In total, 2 g of leaf material was taken from the annual S_7 inbred line ‘WS49’ (JKI). This line was chosen because of increased homozygosity by inbreeding and its key role within summer-annual and winter-annual breeding programs, notably carrying genetic determinants for winter-hardiness. Thus, ‘WS49’ provides a robust reference suitable for broad genetic applications beyond the scope of this study. For high molecular weight (HMW) DNA extraction the gravity column based NucleoBond HMW DNA kit from Macherey–Nagel was used. We then enriched long DNA fragments by applying the Short Read Eliminator (SRE) XS kit from PacBio. For quality check, DNA quantification was done using the Qubit dsDNA BR Assay-Kit on the Qubit fluorometer. Concentration after HMW DNA extraction was measured at 556 ng/μl and after SRE a concentration of 137 ng/μl was achieved.

Genome sequencing

The sequencing library was constructed using the SQK-LSK114 kit from Oxford Nanopore Technologies. Purified HMW DNA was used and 2 μl with 20 fmol were loaded onto a PromethION flow cell (FLO-PRO114M). Sequencing was performed on a PromethION24 system using in total three flow cells running for 72 h each (Figure S1 shows overall flow cell performance over run time). Base calling was run in SUP accuracy mode with 400 bps using Guppy (version 6.4.6) and a minimal length of 200 nucleotides was required. The three flow cells reached raw read N50 values (23.47 kb, 25.48 kb and 25.35 kb) and generated a total of 321.97 Gb (67.43 Gb, 110.73 Gb and 143.81 Gb). With 298.63 Gb, 92.8% of all reads passed a min Q score of 10. Descriptive statistical reports were generated with toulligQC (<https://github.com/GenomiqueENS/toulligQC>) which stated high-quality of the sequencing process.

Genome assembly and performance statistics

To construct the draft genome of caraway the Flye assembler from Kolmogorov, et al¹⁰ was used in version 2.9.1. The genome size of caraway was estimated at 2.1 Gbp based on flow cytometric data of a caraway germplasm collection as previously reported¹. This value was used as the genome size parameter for the Flye assembly. Before use, ONT reads were screened for contamination, adapters were removed and reads were required to have an average high quality ($Q > 20$). In the genome assembly process standard parameters were used except ‘min_overlap’ set to 7,000 bp, the ‘min_read_length’ set to 21.5 kb and the coverage of the initial disjointing process was set to ‘40’. The constructed genome graph was inspected using GfaViz¹¹. For a comprehensive assessment of the genome assembly, we measure the three relevance levels contiguity, completeness and correctness using an ensemble of bioinformatic methods.

The taxonomic classification and profiling of the constructed assembly were performed using KRAKEN2 in the version 2.1.3¹². The analysis was based on a k-mer concept using a subset of the nt database for bacteria, fungi, human, plant, and virus genomes. To assess the completeness of the assembly, the Benchmarking Universal Single-Copy Orthologs (BUSCO, version 5.8) software was utilized by screening the conserved eudicotyledons_odb12 gene set, which comprises of 2,805 orthologous genes from 76 genomes^{13,14}. Quantitative measures were provided for the classes: complete, duplicated, fragmented, and missing. Furthermore, we utilized the long read-based Inspector software in version 1.3.1¹⁵ to evaluate the assembly and to identify potential regions of assembly errors.

Annotation

Structural gene annotation was predicted using a combination of ab initio and homology-based gene prediction. Ab initio gene prediction was performed with Tiberius v1.1.8 using the plant-trained model eudicotyledons.yaml¹⁶. Homology-based gene prediction was then performed with GeMoMa v1.9^{17–19}, using annotations from the following reference species: *Arabidopsis thaliana* TAIR10 (GCF_000001735.4)²⁰, *Daucus carota* v3.0 (GCF_001625215.2)²¹, and *Coriandrum sativum* (<http://cgdb.bio2db.com/>; accessed 17 January 2026)^{22,23}. The

Tiberius gene predictions were supplied to GeMoMa as external predictions. Default parameters were used, except that the GAF filter expression was replaced with $\text{start} = \text{'M'}$ and $\text{stop} = \text{'*'}^1$ and ($\text{evidence} > 1$ or (!isNaN(pAA) and $\text{pAA} > 0.7$)) to retain high-confidence gene models. Similar to the assessment of genome annotation quality was subsequently assessed with BUSCO using the eudicotyledons_odb12 lineage^{13,14} and OMark using the Viridiplantae database²⁴. The resulting gene models were functionally annotated using InterProScan v5.59–91.0 with the parameters -f TSV -applications TIGRFAM, FunFam, SFLD, PANTHER, Gene3D, Hamap, PrositeProfiles, SMART, CDD, PRINTS, PIRSR, PrositePatterns, AntiFam, Pfam, PIRSF -pathways -goterms²⁵.

Biparental mapping population

One plant of the biennial wild accession CARUM 63 from the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK, Seeland OT Gatersleben, Germany) was crossed with one plant of the annual inbred line “L32” of the JKI of the 5th inbred generation. This inbred line displays the standard white flower color of caraway, while the biennial accession shows a prominent rose flower color (see Fig. 1). Both, homogeneous biennial flowering and rose flower color has been determined in a two-year field trial (accession ID here: Cc42_RUS_wd)⁷. The taxonomic identity of CARUM 63 was verified by the curators of the IPK genebank, and reference herbarium specimens for the IPK caraway collection are maintained at the Herbarium Gatersleben (GAT). The biennial accession was once donated by the botanical garden in Halle (Saale), Germany, but originates from Russia (according to passport data from IPK). This accession was also chosen because it was late flowering among various biennial accessions, which might allow the detection of additional QTLs for flowering time⁷. One F_1 plant from the cross was self-fertilized to produce a segregating F_2 mapping population. Because the biennial parent retained residual heterozygosity, the use of a single F_1 founder ensured that the F_2 population segregated from a defined set of parental haplotypes. While the use of multiple independent F_1 individuals could potentially capture additional haplotype combinations and recombination events, it would also increase genetic heterogeneity within the mapping population. The present design therefore provides a genetically well-defined biparental population suitable for linkage and QTL mapping.

Phenotyping

Seeds of F_2 and controls were sown and seedlings were pre-grown in the greenhouse in May 2023. In June, pre-grown plants were transplanted to a field trial consisting of five rows, with 10 cm spacing between plants and 100 cm between rows. As controls, 90 annual plants and 90 biennial plants were planted in rows 1 and 5, respectively. Since the biennial parent did not produce sufficient seeds by self-fertilization, the cultivar ‘Rekord’ was used as phenotypic biennial control during the field trial. In total 187 F_2 plants were grown in the intermediate rows 2 to 4 (see Fig. 2).

For each F_2 plant, flowering time was recorded as the number of days from sowing until the beginning of flowering, defined as the emergence of the first anthers (BOF1). To account for the gap between annual and biennial flowering, a second measure was calculated (BOF2) by subtracting a fixed period of 145 days without new flowering individuals (“winter period”) from the values of biennial plants. The adjustment value of 145 days used to calculate BOF2 was derived directly from the observed vegetative winter gap in the field. The final annual individual flowered 161 days after sowing (November 08, 2023), whereas the earliest biennial individual initiated flowering 307 days after sowing (April 02, 2024), revealing a 146-day interval entirely devoid of flowering events. This reduced the bimodal distribution of the data to improve the robustness of the used parametric models. In addition, BOF2 accounts for the environmental truncation under field conditions, acknowledging that several

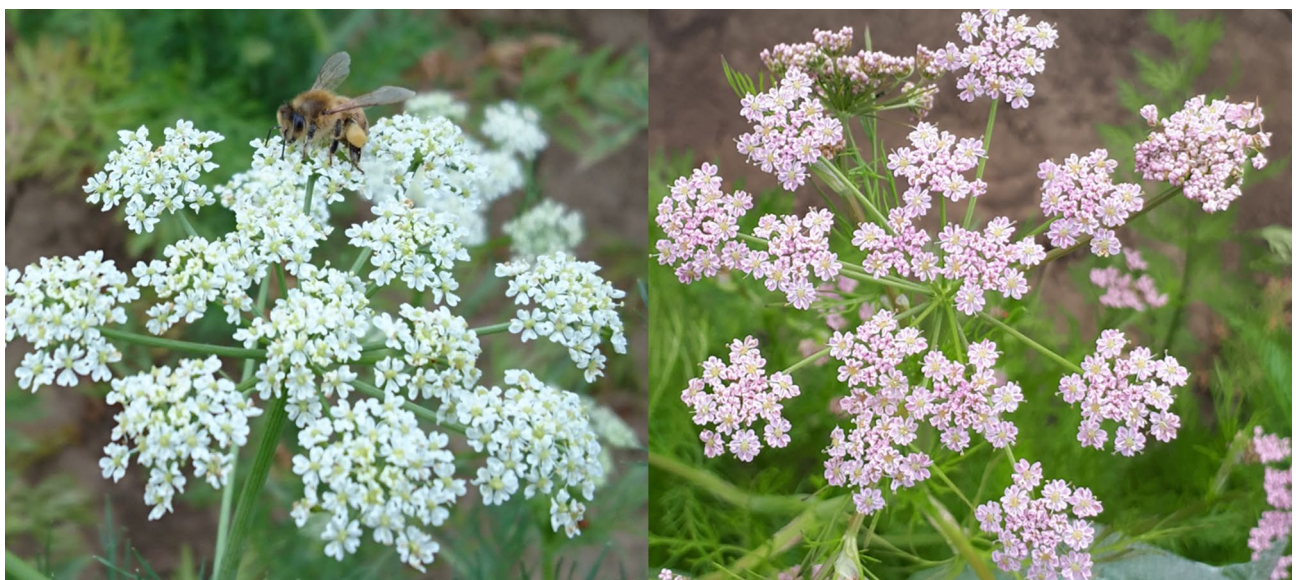


Fig. 1. Phenotypic differences between parental plants. The annual inbred line “L32” (left) shows white flowers, while the biennial accession CARUM 63 (right) displays rose-colored flowers.

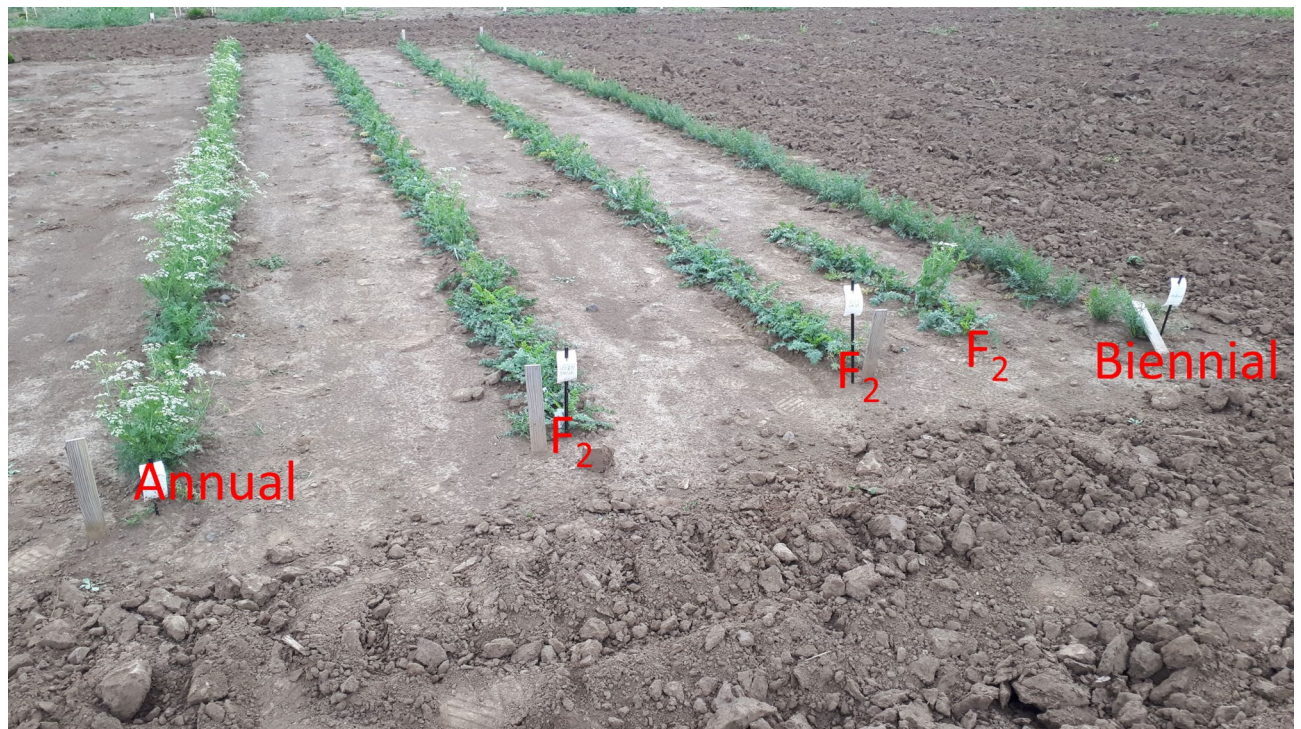


Fig. 2. Field trial in September 2023. Row 1 (from the left) shows the annual control with white flowers in full bloom. Rows 2–4 represent the still sparsely flowering F_2 mapping population, and row 5 contains the non-flowering biennial control.

genotypes with a potential annual flowering habit were physically prevented from flowering by the onset of winter due to the late sowing date, which is one reasonable explanation for the observed segregation ratio in F_2 (see [discussion](#)). To search for flowering time loci independent from flowering type, a standardized beginning of flowering (BOF3) was calculated by Z-transforming BOF1 separately for annual and biennial flowering types. Through this transformation, the resulting BOF3 value no longer reflects the absolute time to flowering, but rather how early or late an individual plant flowers relative to its flowering type. This effectively scales down the massive genetic effect of the annual/biennial switch to zero, enabling the downstream mapping analysis to detect genetic loci that modulate the timing of flowering within both groups. As pivotal advantage to two separate MQM analyses, a moderate sample size is retained. In addition, flowering time was simplified recorded as flowering type, which differentiated between annual (i.e. flowering before winter) and biennial flowering types (i.e. flowering after winter; FT1). In contrast to BOF1–3, FT1 also includes plants, which died during winter. These plants, which showed a healthy condition before winter but were predominately damaged by mouse were regarded as biennial plants to avoid selection bias against biennial genotypes.

We conducted a chi-square test for goodness-of-fit using the function `chisq.test` in R²⁶.

Flower color was scored (rose or white) at an early flowering stage, when at least one umbel had fully developed petals. All plants showing a rose coloration (disregarding color intensity) were defined as rose-flowering. Although strong variation in rose color intensity was observed, preliminary field observations indicated that pigmentation intensity was strongly influenced by environmental conditions, particularly pigment bleaching under high light exposure. As a consequence, quantitative scoring of color intensity under field conditions was considered insufficiently robust and reproducible. The objective of this study was therefore not to quantify pigmentation intensity but to distinguish between the presence and absence of anthocyanin-derived flower coloration. Collapsing the phenotype into a binary classification reduced environmentally induced variation and enabled consistent scoring across all individuals for subsequent genetic analyses.

An overview of all trait descriptors is presented in Table 1 (result section).

Genotyping by sequencing, SNP discovery and filtering

DNA was extracted from young developing leaves using DNeasy Plant Mini Kit (Qiagen, Venlo, Netherlands). DNA quality and quantity were tested using Nanodrop 8000 and Qubit Flex Fluorometer using dsDNA Assay BR (all Thermo Fisher Scientific, Waltham, USA) and gel electrophoresis. Library preparation for GBS was carried out at the Institute of Integrative Biology and Systems (IBIS) of University Laval (Quebec, Canada). Genomic DNA was normalized to 10 ng/μl and digested with the restriction enzymes PstI and MspI²⁷. Libraries were sequenced using an Illumina NovaSeq 6000 platform at Génome Québec (Montréal, Canada), generating 150-bp paired-end reads.

Raw sequences were processed and SNP calling was performed using the internal Galaxy server (version 25.0) within JKI²⁸. First, GBS raw reads were trimmed using Trim Galore v0.6.7; non-default parameters: Phred

Abbr	Definition	Values	N
BOF1	Beginning of flowering (first anthers visible) in days after sowing. [One outlier (F133: BOF1 = 178; Z > 3 within annuals) during mapping and for calculating BOF2/BOF3 removed.]	106–356	163
BOF2	Beginning of flowering (first anthers visible) in adjusted days after sowing; adjusted by subtracting 145 days for biennial plants (winter period without new flowering plants) from BOF1	106–234	163
BOF3	Standardized beginning of flowering (Z-score); calculated by Z-transforming BOF1 separately for annual and biennial flowering types	–1.55 – 2.68	163
FT1	Flowering type, differentiating between annual and biennial flowering (including plants died over winter as biennial flowering types)	0 = annual 2 = biennial	180
COL	Flower color	0 = white 2 = rose	164

Table 1. Description of phenotypic traits and scoring methods for flowering time and flower color. Abbr. = abbreviation; N = sample size.

quality score threshold for low-quality ends: 30; discard reads shorter than 50 nucleotides²⁹. Trimmed reads were mapped against a draft assembly (2171 contigs, see below) using BWA MEM v0.7.18^{30,31}. SNP calling was carried out using bcftools mpileup and bcftools call v1.15.1 (non-default-parameters: output-tags DP, DPR)³².

SNPs were filtered in three steps: (i) quality filters (minimum quality score ≥ 40, minimum read depth per SNP = 4) to eliminate low-confidence base calls while minimizing the loss of informative loci; (ii) selection of SNPs homozygous and polymorphic between parental genotypes; (iii) additional filters (≤ 10% missing data, 35–65% heterozygosity, and 15–35% representation of both homozygous genotypes in mapping population). These stringent downstream thresholds were applied to rigorously purge technical noise, thereby ensuring a structurally sound framework for de novo linkage map construction.

Construction of a linkage map and QTL mapping

A linkage map was constructed using the software JoinMap 5.0 using default settings³³. Markers were grouped using a LOD threshold of 4 (no changes were observed up to a LOD of 10), and regression mapping with Kosambi’s mapping function was applied. Loci with mean χ^2 contributions > 5 were removed. To ensure model stability and minimize multicollinearity during QTL mapping, the genetic map was thinned to a maximum density of one marker per 1 cM. In clusters of tightly linked markers, the marker with the lowest proportion of missing data was chosen to represent the bin.

QTL analysis was performed using MapQTL 5³⁴. Genome-wide LOD significance thresholds ($\alpha = 0.05$) for each trait were determined by 1,000 permutation tests. Initially, interval mapping (IM) was used to identify candidate QTL regions. To refine QTL positions and to detect loci potentially masked by major-effect QTL, multiple QTL mapping (MQM) was subsequently applied. To ensure model parsimony, only cofactors improving the explained variance at significant QTLs were retained in the final MQM model (final cofactor settings per trait can be found in Table S1F). Final QTL positions, effects, and explained variance were estimated from the MQM model, whereas 2-LOD- and 1-LOD support intervals were derived from a restricted MQM (rMQM). Markers immediately outside the intervals were defined as flanking markers. To validate QTLs for the non-parametrically distributed traits (COL, FT1), a Kruskal–Wallis rank sum test was performed.

Although BLINK was originally developed for association mapping, the Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) model³⁵ was used as implemented in the R package GAPIT³⁶ to cross-validate MQM peaks. Unlike single-locus models (MLM/GLM), BLINK’s multi-locus approach effectively eliminated redundant signals caused by high local LD. BLINK was performed using the entire set of 731 filtered SNPs. This allowed for a higher-resolution scan of the genome, leveraging the full allelic information available in our GBS data. To accommodate the high number of underlying contigs within the software’s architecture, markers were assigned to a single continuous linkage group and numbered consecutively.

Results

Draft genome assembly & annotation basic statistics

Draft genome assembly

The constructed draft assembly of caraway reached a total size of 1.87 Gbp consisting of 2,171 fragments. The assembly exhibited high contiguity with a contig N50 of 9.5 Mb, and 820 scaffolds were found to be larger than 100 kb in length. Taxonomic profiling via KRAKEN2 revealed that the vast majority of the constructed contigs were associated with plants, indicating low contamination rates (Table S2). Only a minor fraction of three contigs (0.14%) remained unclassified. The BUSCO assessment confirmed a high level of genomic completeness (97.9%, Table S3). Detailed contiguity and completeness metrics are further summarized in the snail plot (Fig. 3). Additionally, the final evaluation with Inspector confirmed the quality of the assembly by an overall mapping rate of 99.74% and a depth in large contigs of 158.92.

Annotation

Structural gene prediction yielded 41,835 genes and 51,645 transcripts. Quality assessment of the structural gene annotation using BUSCO indicated comparable results for Carum carvi, Daucus carota and Arabidopsis thaliana, and much better results than Coriandrum sativum (Table S3). Furthermore, BUSCO results for genome and annotation were highly similar, with slightly better results for the annotation of 98.6% completeness. Additionally, quality assessment with Omarm indicated that structural gene prediction was of high quality, with the majority of predictions being highly consistent and only a few being missing, inconsistent or partial (Fig. 4).

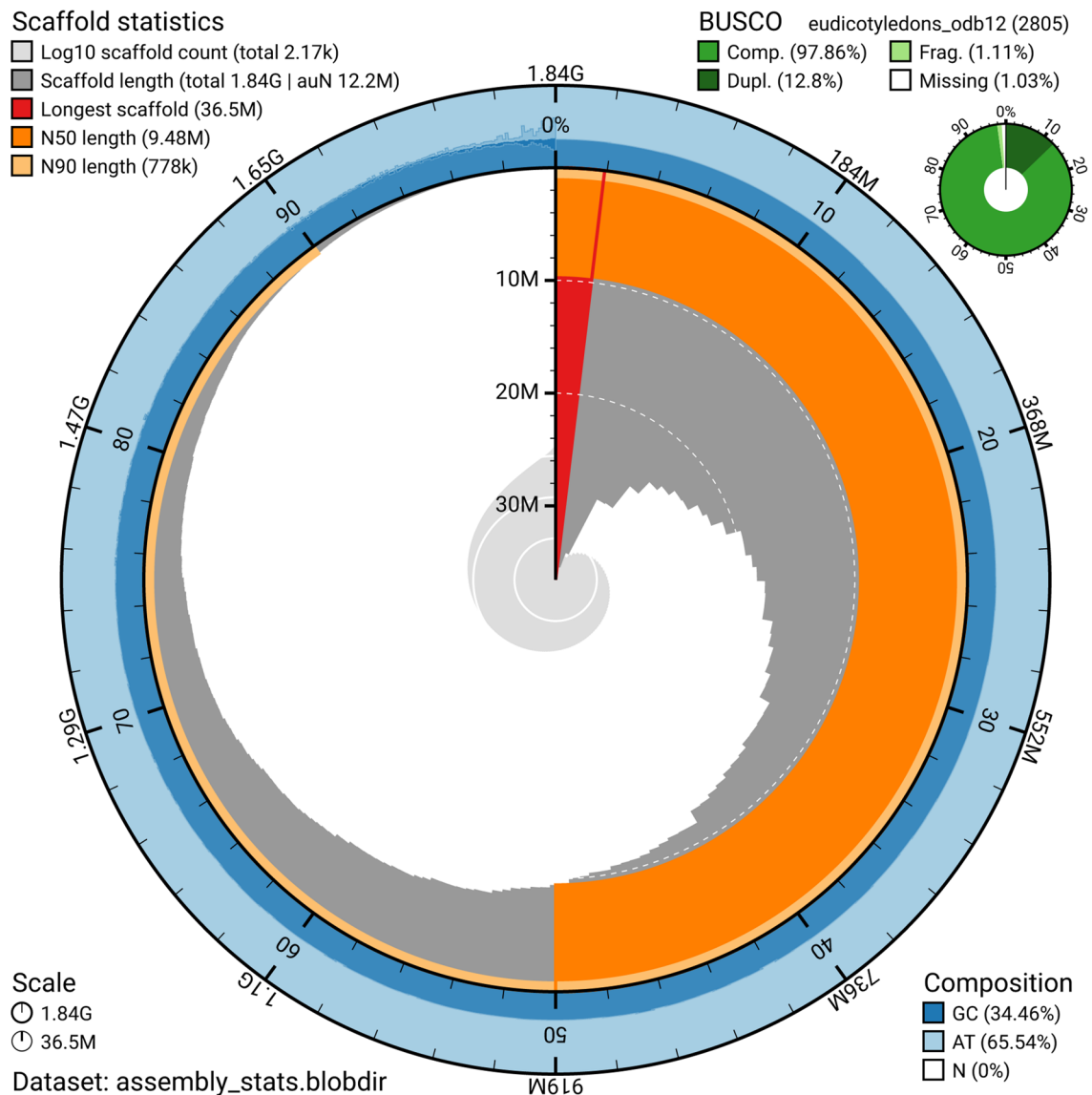


Fig. 3. Snail plot. Summarizes the main assembly metrics for the 1.87 Gbp draft genome.

Phenotypic data

Flowering time and flowering type

F₁: Prior to the establishment of the F₂ mapping population, phenotypic attributes of the F₁ generation were assessed under controlled greenhouse conditions. While finally a single F₁ plant was selected to produce the F₂ seeds utilized in this study, notably, all cultivated F₁ individuals initiated flowering without requiring a vernalization period, indicating that the annual flowering habit is qualitatively dominant over the biennial trait. However, a considerable developmental delay in flowering time was observed in the F₁ plants compared to annual parental genotypes grown simultaneously under identical conditions.

F₂ and controls (field trial): All 90 plants of the annual control had initiated flowering by mid-September 2023, approximately 106 days after sowing. In contrast, none of the biennial control plants flowered during the first year. Among the 187 cultivated F₂ plants, 23 plants died before initiating flowering. Seven and 16 of these plants died before or during winter, respectively. These 16 plants were regarded as biennial flowering types (FT1), but were disregarded for quantitative trait descriptors (BOF1, BOF2, BOF3). Consequently, flowering type was assessed for 180 plants, while other traits descriptors were obtained for 164 plants (Table 1).

Overall, 36% (64 plants) were classified as annual flowering types and 64% (116 plants) as biennial flowering types (Fig. 5A). For annual plants, the beginning of flowering (BOF1) ranged from 106 days after sowing (September 14, 2023) to 161 days after sowing (November 08, 2023). One outlier (178 days after sowing; November 25, 2023) was removed. For biennial plants, the beginning of flowering ranged from 307 days after sowing (April 02, 2024) to 356 days after sowing (May 21, 2024) (Fig. 5B). BOF1 was used to compute BOF2 and BOF3 (see material & methods and Table 1) for mapping purposes. All phenotypic data are deposited in Table S1A.

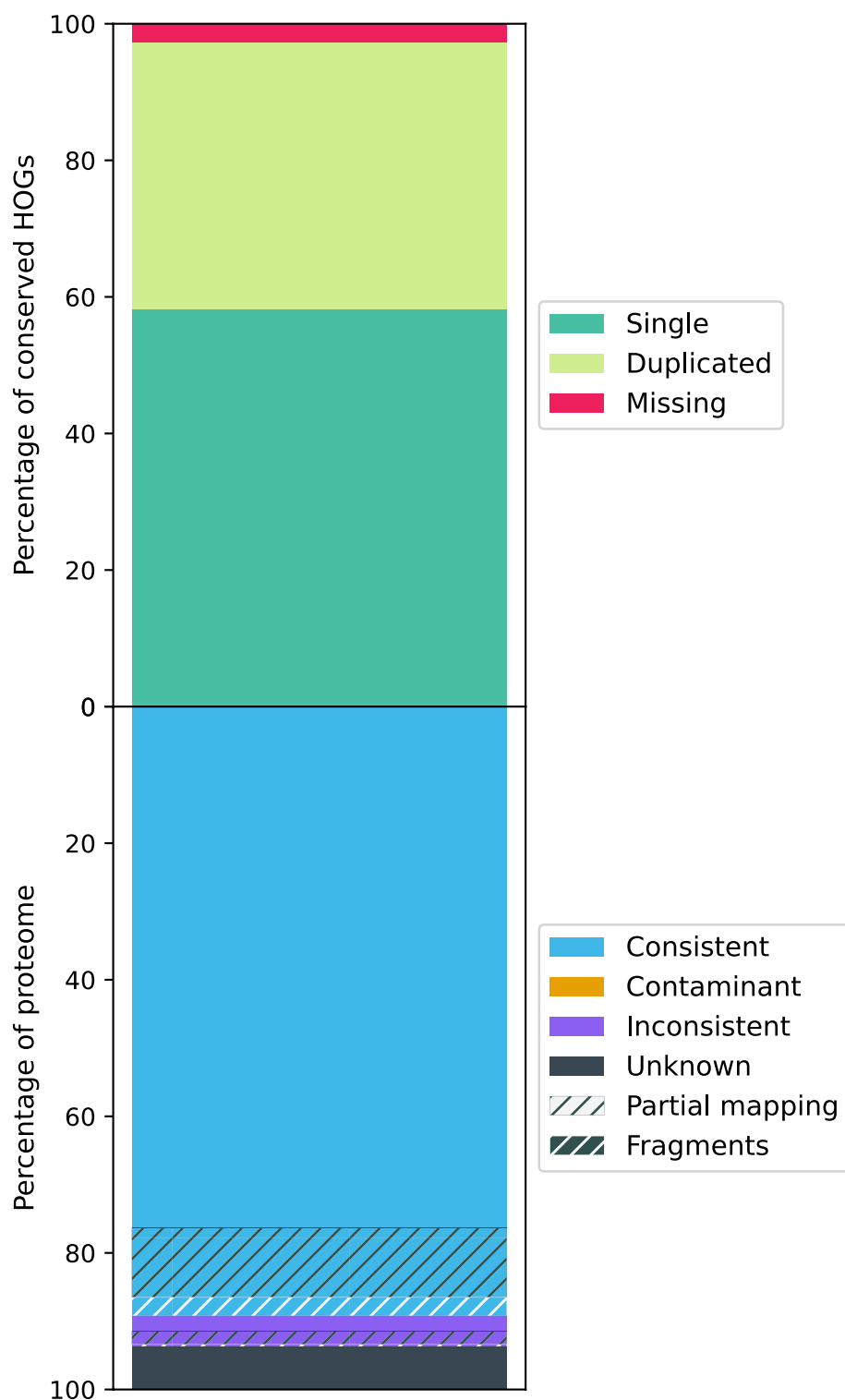


Fig. 4. Proteome completeness and consistency. Results from the OMARK assessment for the *Carum carvi* draft assembly.

We formally tested the observed segregation against a 3:1 ratio (annual:biennial according to Nemeth and Pluhar⁶) using chi-square test for goodness-of-fit, which was rejected: $\chi^2 = 149.36$, $p < 0.001$.

Flower color

Under greenhouse conditions during crossings and F_2 production, we observed that the biennial parent exhibit a strongly bleached rose coloration. In the field trial, we observed a wide range of flower colors in F_2 plants, ranging from white and light rose to deep rose (as in Fig. 1). Due to the continuous range and noticeable bleaching on

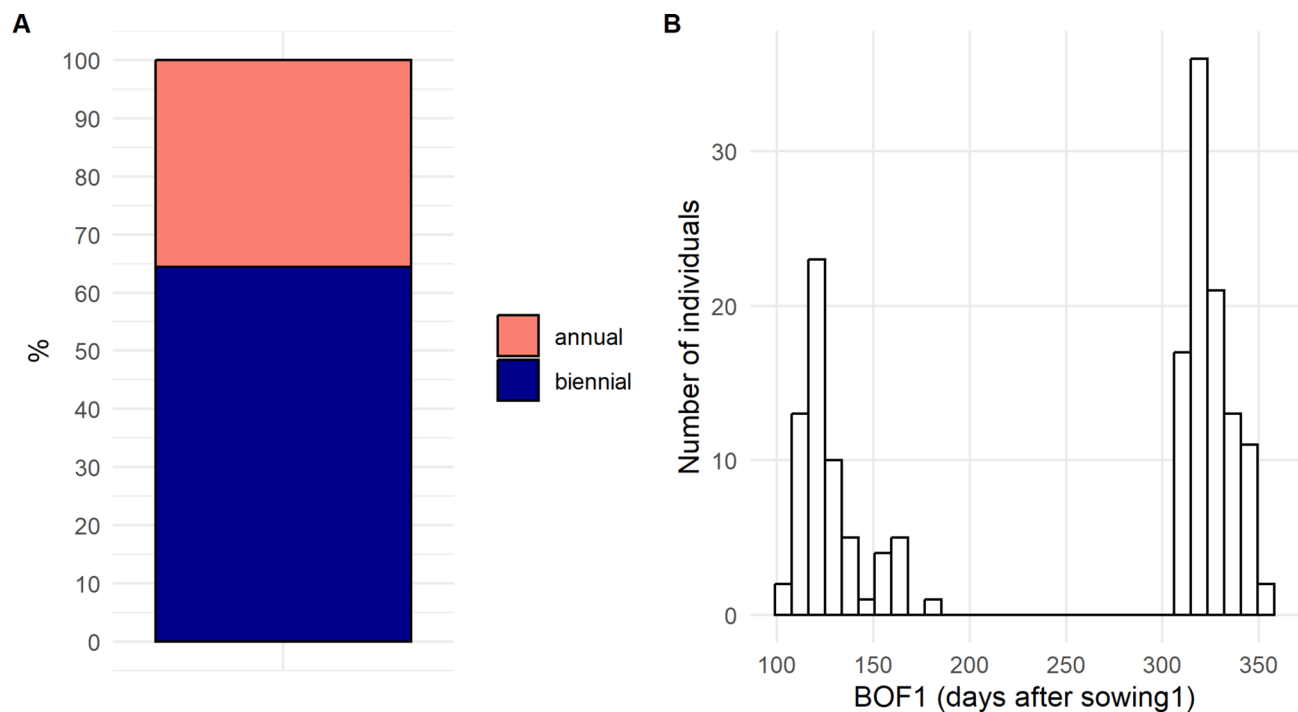


Fig. 5. Flowering time. **(A)** Percentage of annual (red) and biennial (blue) flowering types among 180 F₂ plants. **(B)** Variation in beginning of flowering (BOF1) in days after sowing, recorded for 164 F₂ plants.

COL	Count			Percentage		
	Annual	Biennial	All	Annual	Biennial	All
Rose	28	69	97	43.8	69.0	59.1
White	36	31	67	56.2	31.0	40.9
Sum	64	100	164			

Table 2. Distribution of flower color phenotypes (rose vs. white) among annual and biennial flowering types in the mapping population (n = 164). Frequencies are provided as absolute counts and relative percentages (%).

sunny days, we did not define intermediate flower colors. Thus, all plants showing a rose coloration were defined as rose-flowering. In total, we detected 59.1% rose-flowering plants, while biennial flowering types had a higher percentage of 69.0% (Table 2, Fig. 6A). In more detail, for early-flowering annuals and late-flowering biennials, a higher proportion of white-flowering plants was observed in the F₂ generation (Fig. 6B).

Genotyping by sequencing, mapping and filtering

In total 2,249 million sequences were gained by GBS for all 187 F₂ plants and parental plants after trimming. The number of sequences per sample ranged from 0.8 to 32.8 million sequences and averaged 11.9 million. On average 95.6% of the sequences were mapped to the draft genome assembly. 97.5% and 99.4% of the sequences of the annual and of the biennial parent were mapped to the draft genome, respectively. After a first quality filter 9,787 SNPs were gained by GBS. While the biennial parent showed a heterozygosity of 24.81%, the annual parent had a heterozygosity of 5.92%. After filtering for homozygous SNPs between parents and the segregation ratio, 731 SNPs remained (Table S1B), which were used for linkage mapping in Joinmap. The full filtering pipeline and the effect of each filter (including intermediate steps) on SNP number is reported in Table S1B.

Linkage map

Markers were grouped into 10 linkage groups. From the 731 initial SNPs, 634 SNPs were mapped by regression mapping (full map in Table S1C). A thinning to one marker per cM resulted in a final map of 259 markers spanning a total of 663 cM across 10 LGs (thinned map in Table 3, Table S1D). The number of SNPs per LG ranged from 11 to 46 SNPs and averaged 26 SNPs. The length of the LGs ranged from 29.9 cM to 89.2 cM and averaged 66.3 cM. The average inter-marker distance was 2.9 cM. The largest gap was located on LG09 with 23.6 cM (Table 3). An overview of the genetic map including marker density is displayed in Fig. 7.

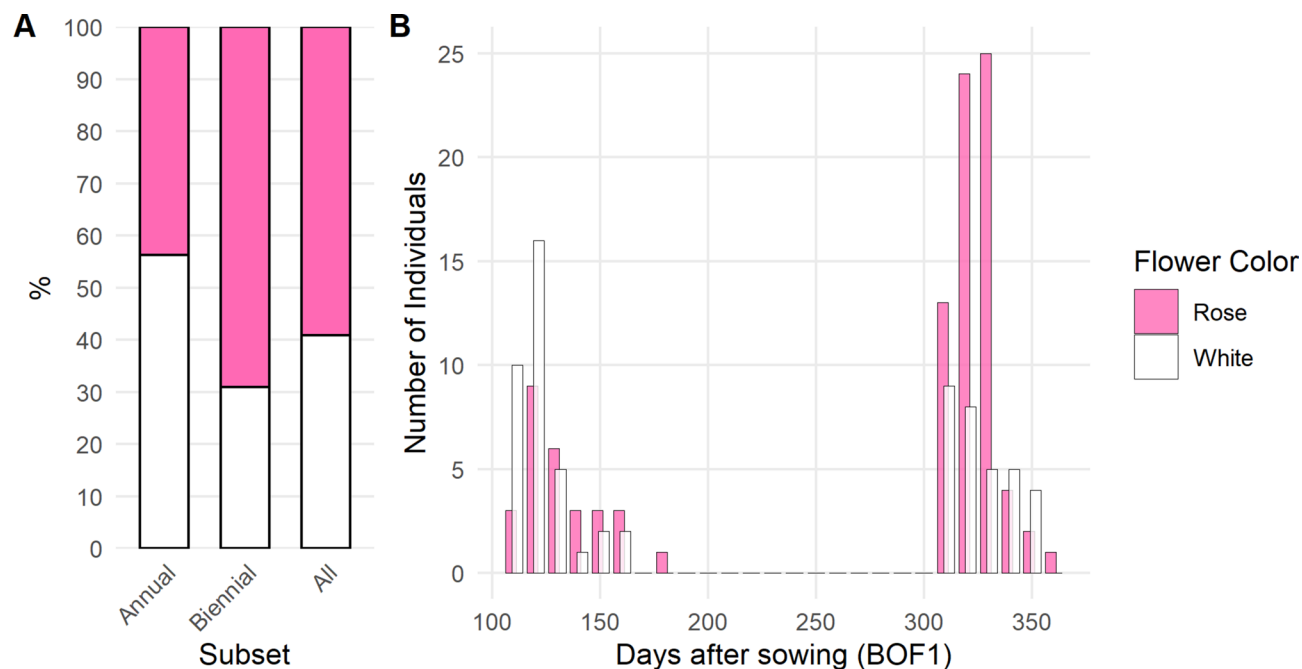


Fig. 6. Flower color. **(A)** Relative frequency (%) of rose and white flowering phenotypes, stratified by annual ($n=64$) and biennial ($n=100$) flowering types. **(B)** Variation in the beginning of flowering (BOF1, days after sowing) for rose and white flowering individuals.

Group	Number of SNPs	Length (cM)	Largest gap (cM)	Average gap (cM)
LG01	25	86.6	22.1	3.6
LG02	46	89.2	9.8	2
LG03	25	79.3	13.6	3.3
LG04	30	72	17.5	2.5
LG05	30	71.2	13.9	2.5
LG06	30	65.9	8.3	2.3
LG07	31	67	8.4	2.2
LG08	20	47.6	12.6	2.5
LG09	11	54.2	23.6	5.4
LG10	11	29.9	8.6	3
Average	25.9	66.3	-	2.9
Total	259	662.9	-	-

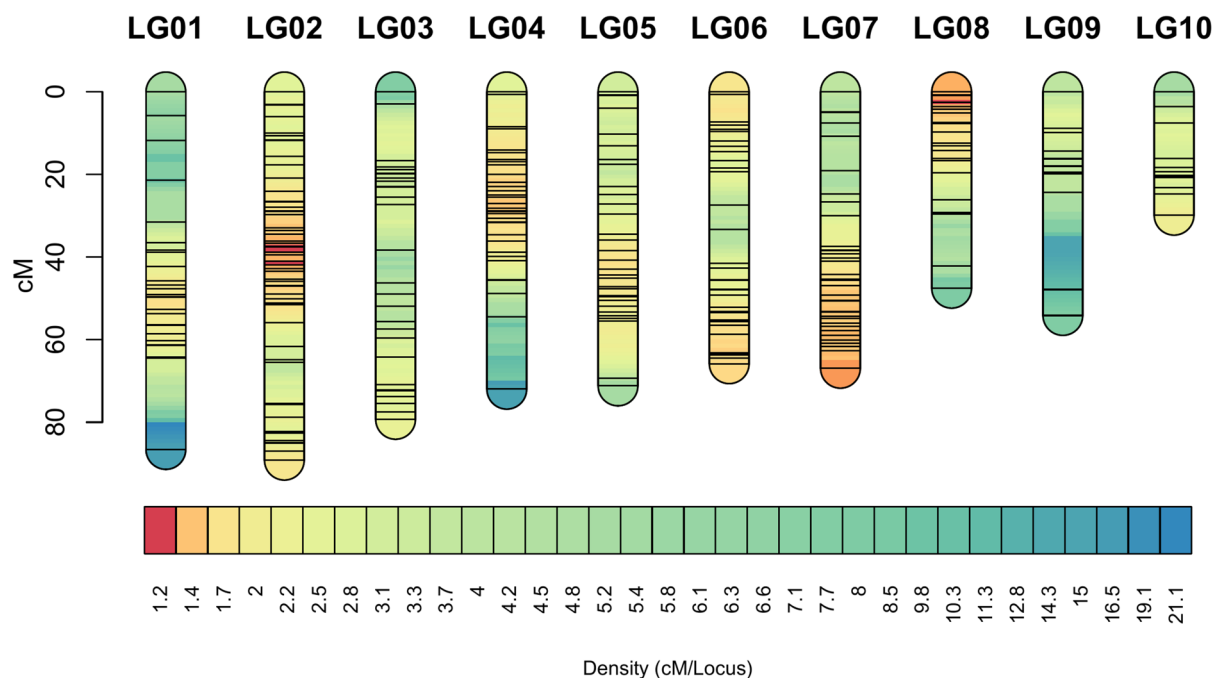
Table 3. Summary of the genetic linkage map developed via regression mapping thinned to one marker per cM. The map comprises 259 SNP markers distributed across 10 linkage groups (LG).

QTL mapping (IM, MQM, Kruskal–Wallis), BLINK model & candidate genes

Flowering time and flowering type

We conducted IM (Figure S2, raw results Table S1E) to identify candidate QTL regions followed by MQM (raw results in Table S1F and Table S1G for restricted MQM) to refine QTL positions and to detect potentially masked QTLs at subthreshold-maxima of IM. The genome wide LOD threshold (for $\alpha=0.05$) as calculated by the permutation test was 3.5 for BOF2 and 3.6 for BOF1, BOF3 and FT1, respectively. The final MQM for BOF2 revealed five significant QTLs on LG02, LG03, LG05, LG08 and LG10 (Fig. 8, Table 4). Interestingly, the QTL on LG05 had a subthreshold maximum in IM, but became a prominent QTL in MQM. Without mitigating strong difference between annuals and biennials (BOF1, FT1) only the QTLs on LG02 and LG05 surpassed the significance threshold (Fig. 8, Table S1H). In contrast, after removing the difference between annuals and biennials by separate Z-transformation (BOF3) the QTL on LG10 became most prominent (LOD = 6.73, explained variance = 14.7%, Fig. 8, Table S1H).

Results from classical QTL mapping were compared with results from a BLINK model (raw results in Table S1K). The FDR threshold of the BLINK model (for $\alpha=0.05$) was about 3.7. In general, BLINK corroborated the results of MQM (Table 4) with two deviations: Additionally, significant QTLs were detected for BOF1 on LG08 (as already detected for BOF2) as well as for FT1 on LG06 (Table S1K). Detected peak markers of BLINK are not



Rendered by LinkageMapView

Fig. 7. Thinned linkage map. Density distribution of SNP markers across the 10 linkage groups (LG) of the thinned genetic map comprising 259 loci.

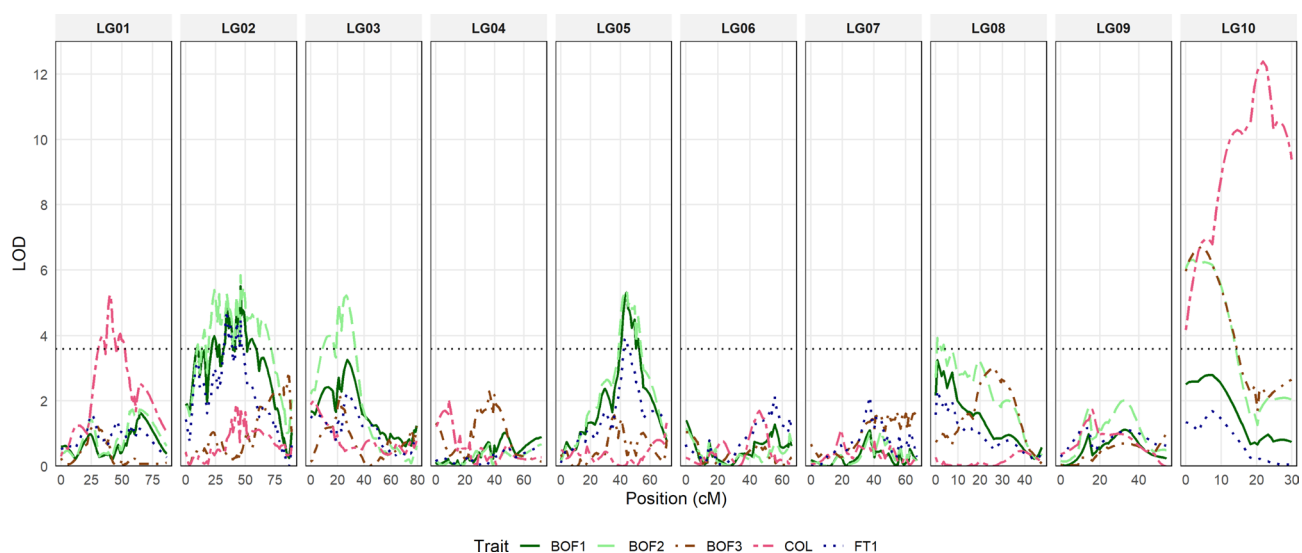


Fig. 8. LOD curves rMQM. Results (LOD values) of rMQM for flowering time descriptors (BOF1 = darkgreen solid lines, BOF2 = lightgreen longdash lines, BOF3 = brown dashed-dotted line, FT1 = blue dotted lines) and flower color (COL = rose twodash lines). Dotted black line = Genome-wide significance threshold for BOF1, BOF3, FT1 and COL (LOD = 3.6), threshold for BOF2 (LOD = 3.5) not shown.

identical but co-segregated with peak markers of MQM. The Kruskal Wallis test (Figure S3, Table S1L) for FT1 corroborated QTLs on LG02, LG03, LG05 and LG10, but also showed significant peaks ($K^* > 8$) on LG06, LG07 and LG08. All significant QTLs from BLINK and the Kruskal–Wallis test not reaching the threshold in MQM are regarded as suggestive QTLs requiring additional evidence in subsequent studies. While following statistics focused on the five QTLs detected for BOF2, detailed statistics for the other trait descriptors and suggestive QTLs (including loci LOD > 2 from MQM, which might become of interest in future studies) are added to the

QTL	Trait descriptor	Group	Position cM	1-LOD cM	LOD MQM (IM) [BLINK]	Expl. Var %	Add MQM [BLINK]	Dom	Peak marker
Cc-FT02.1	BOF2	LG02	45.9	41.9–51.2	5.84 (6.39) [7.47]	9.2 [6.11]	13.17 [13.89]	2.66	contig_1125_12302196
Cc-FT03.1	BOF2	LG03	26.5	18.8–32.3	5.23 (3.14) [5.65]	8.3 [4.18]	11.61 [11.17]	–3.36	contig_1045_5362953
Cc-FT05.1	BOF2	LG05	44.0	39.5–52.9	5.39 (2.47) [6.69]	8.6 [4.82]	12.95 [12.68]	–2.53	contig_1514_1706860
Cc-FT08.1	BOF2	LG08	0.8	0.0–9.6	3.93 (2.61) [4.93]	6.1 [2.84]	10.11 [10.03]	–0.26	contig_1688_9875227
Cc-FT10.1	BOF2	LG10	2.0	0.0–11.5	6.32 (5.61) [7.00]	10.5 [6.90]	14.40 [13.05]	4.13	contig_880_18056394
Cc-COL01.1	COL	LG01	39.8	37.6–42.3	5.25 (5.33) [6.15]	10.9	–0.49 [–0.5]	0.11	contig_460_5388281
Cc-COL10.1	COL	LG10	21.8	18.3–24.2	12.39 (13.11) [12.74]	26.0	0.73 [0.7]	0.29	contig_1950_3018163

Table 4. Characteristics of significant QTL for flower color and flowering time (BOF2). Corresponding values for the IM and BLINK models are provided in parentheses () and square brackets [], respectively. 1-LOD intervals were extracted from the restricted MQM model. Expl. Var. indicates the percentage of explained phenotypic variance (%); Add. and Dom. represent additive and dominance effects in relation to the biennial allele, respectively. Significant scores (based on permutation tests for MQM/IM or FDR for BLINK) are indicated in bold.

supplements (Table S1H, Figure S4). Table S1H also includes an extended version of Table 4 for significant QTLs, providing 2-LOD intervals as well as flanking markers useful for fine mapping.

The identified five QTLs for BOF2 explained 6.1% (Cc-FT08.1) to 10.5% (Cc-FT10.1) of the phenotypic variance, respectively (in total 42.7%). In relation to the biennial allele, these loci exhibited similar positive additive effects from 10.11 (Cc-FT08.1) to 14.40 (Cc-FT10.1) adjusted days after sowing. In comparison, dominance effects were low (–3.36 to 4.13 adjusted days after sowing; Table 4, Fig. 9). The 1-LOD interval ranged from 9.6 to 13.5 cM (Table 4). We found a medium correlation between the number of biennial alleles across the five QTLs and flowering time (BOF2, Spearman $r = 0.61$, Fig. 10). In detail, most plants with seven or more biennial alleles flowered in the second year (after 161 adjusted days after sowing), while plants with four to six biennial alleles exhibited high plasticity (Fig. 10). In addition, we scrutinized the effect of the QTLs on BOF2 within annuals and biennials, separately. Most conspicuously, Cc-FT10.1 showed additive effects within both groups (Figure S5). Thus, Cc-FT10.1 contributed to flowering time in general but less to the strong difference between annuals and biennials, explaining why this QTL was masked in BOF1 without the adjustments in BOF2 or BOF3.

To identify potential regulators of flowering time, we exemplarily screened for candidate genes within the 1-LOD intervals of Cc-FT02.1, which is most likely associated with vernalization requirement. We identified 1,571 genes (2,002 transcripts) within this interval (Table S4). Notably, a Histone-lysine N-methyltransferase (containing a SET domain and a RING/FYVE/PHD-type zinc finger) was identified in very close proximity to the peak of Cc-FT02.1 (32,536 bp; ~0 cM).

Flower color

The genome wide LOD threshold (for $\alpha = 0.05$) as calculated by the permutation test was 3.6 for COL. Based on this threshold, initial interval mapping revealed two QTLs for COL on LG01 and LG10, which were corroborated by the final MQM, BLINK (Table 4, Fig. 9) and the Kruskal–Wallis test (Figure S3). At the major QTL on LG10 (explaining 26.0% of the variance, Table 4), most homozygous biennial (a) genotypes exhibited a rose flower, whereas most homozygous annual (b) genotypes exhibited a white flower (Fig. 9). Interestingly, the QTL on LG01 showed an opposing effect; specifically, most homozygous annual (b) genotypes showed a rose flower. For both QTLs, heterozygous genotypes were either white- or rose-flowering, with a slightly higher percentage of rose-flowering plants (Fig. 9). Analyzing the interaction of both QTLs at homozygous stages, we found that the biennial allele of Cc-COL10.1 was obligatory for rose flowering, but the annual allele at Cc-COL01.1 had a stabilizing effect (Fig. 11).

It must be noted as a statistical limitation that standard quantitative parameters, including additive effects, dominance effects, and explained phenotypic variance, were extracted from the parametric mixture models for the binary trait COL (Table 4, as also true for FT1, Table S1). Because these traits are nominal rather than continuously distributed, these values do not represent absolute quantitative physical metrics, but serve as

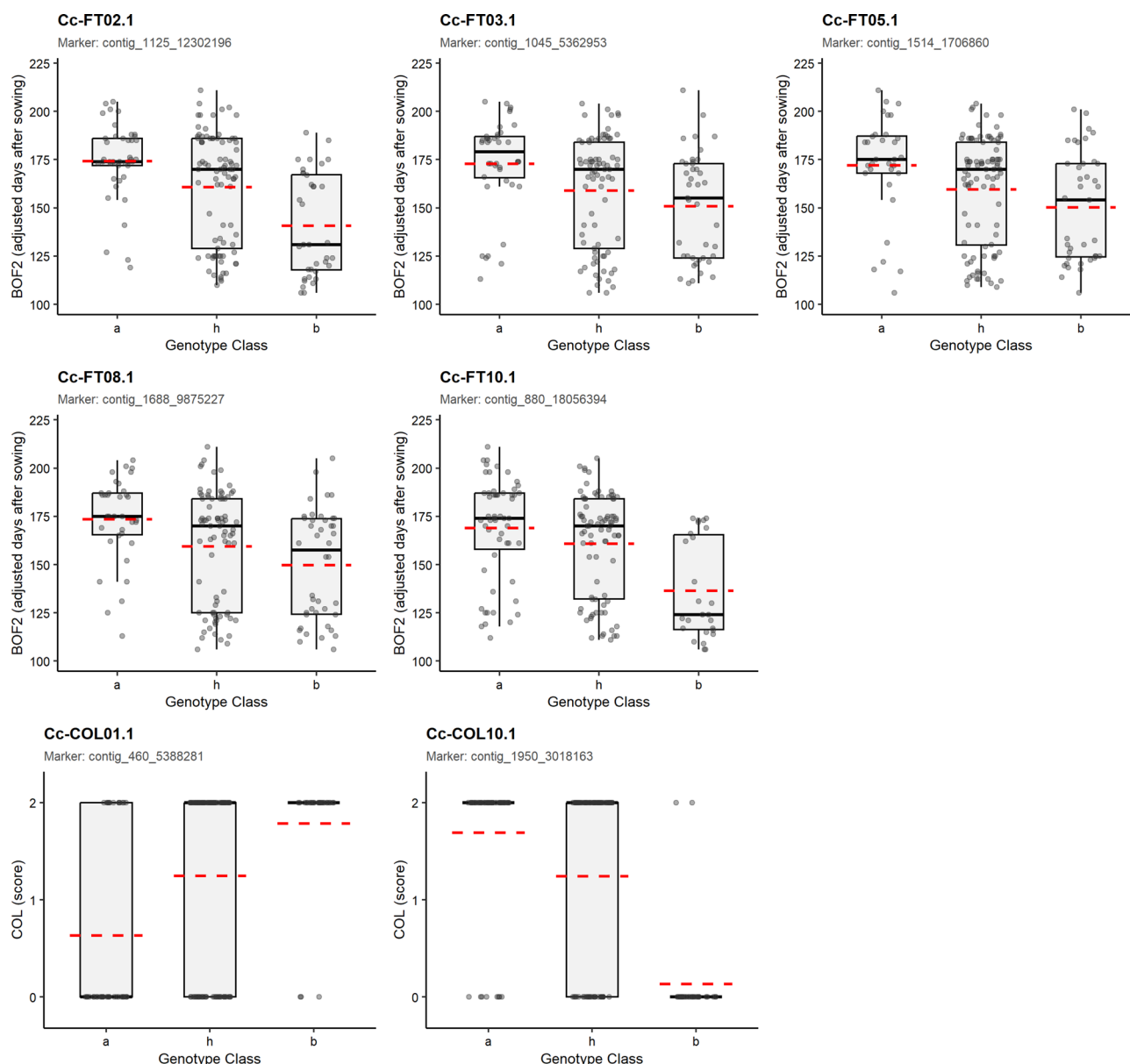


Fig. 9. Allelic effects of identified QTL for flowering time (5, BOF2) and flower color (2, COL). Genotypes at peak markers: (a) homozygous biennial, (h) heterozygous, (b) homozygous annual. BOF2 is shown in adjusted days after sowing; COL is shown as scoring (0 = white, 2 = rose). Solid lines indicate medians; red dashed lines indicate means. Individual data points are shown as gray jitters.

mathematical approximations to indicate the relative genetic power and the phenotypic direction driven by the respective parental alleles.

We screened for candidate genes within the 1-LOD interval of Cc-COL10.1 and identified 320 genes (367 transcripts) within this interval (Table S4). Notably, three myb8 related transcription factors were located in a distance of 650,685 to 818,677 bp (approx. 1–2 cM) of the peak marker for Cc-COL10.1.

Figure 12 summarizes the QTL positions on the genetic map for flower color and flowering time (BOF2).

Discussion

Annotated draft genome assembly and genetic map as research resource

The draft genome assembly presented here represents a significant milestone for caraway research, providing the first high-quality genomic backbone for this species. With a total size of 1.87 Gbp and a contig N50 of 9.5 Mb, the assembly demonstrates exceptional contiguity for a draft version. The high BUSCO completeness (97.9% for the genome, 98.6% for the annotation), combined with the low contamination rates, underscores the reliability of this resource. As a genetic resource, this assembly bridges the gap between traditional breeding and modern molecular genetics in caraway. As demonstrated in this study it allowed us to efficiently map GBS data compared to prior approaches using ‘mock references’¹ and to identify candidate genes for detected QTLs. In future it may

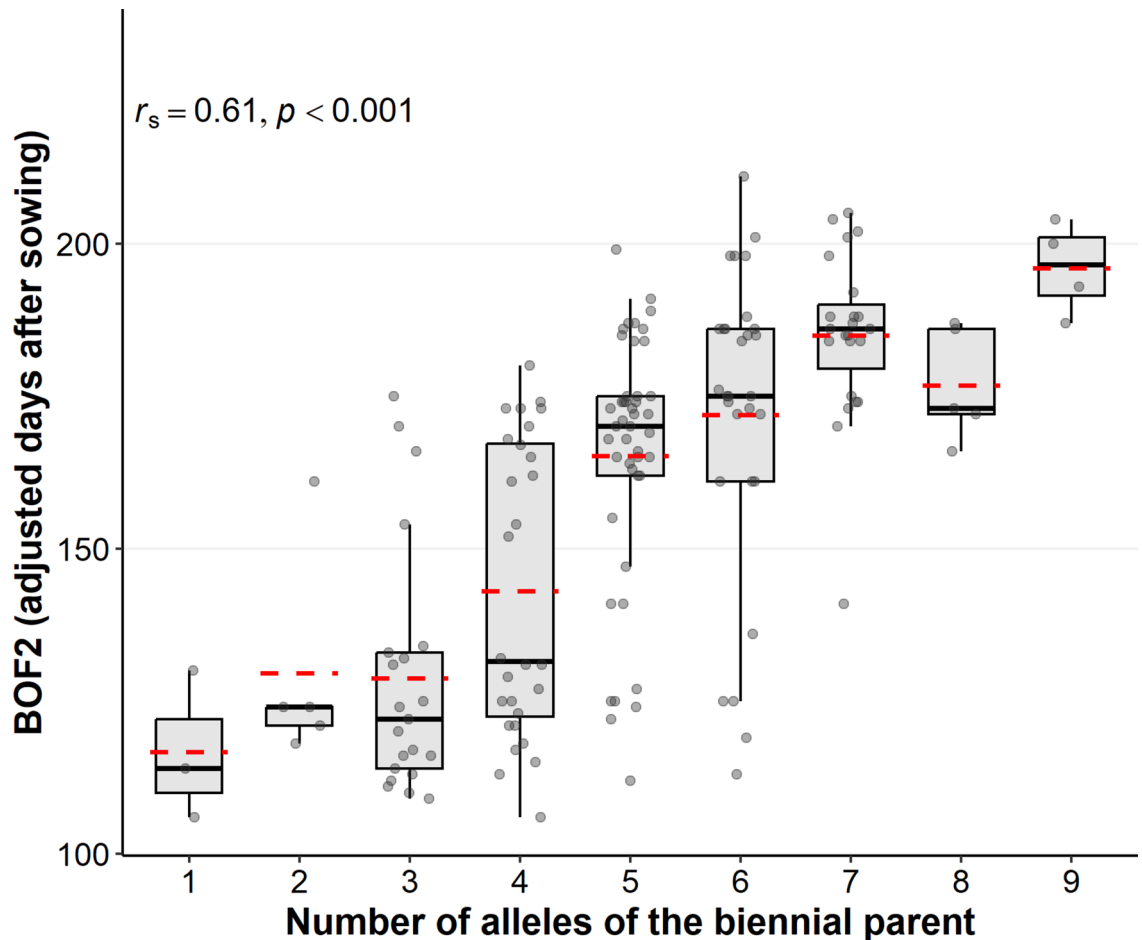


Fig. 10. Phenotypic distribution of flowering time (BOF2). Based on the total number of biennial parent alleles across five identified QTL (Cc-FT02.1, Cc-FT03.1, Cc-FT05.1, Cc-FT08.1, Cc-FT10.1). Solid lines indicate medians; red dashed lines indicate means. Individual data points are shown as gray jitters.

facilitate comparative genomics within the Apiaceae using the increasing number of available genomes in this family^{23,37–40} and support the development of tools for marker-assisted selection.

The caraway assembly (1.8 Gb) is smaller than the estimated cytometric genome size of 2.1 Gb. We attribute this ~14% difference to two main factors. First, cytometric genome-size estimates are relative measurements that are prone to overestimation due to the interference of cytosolic secondary metabolites with DNA-intercalating fluorochromes⁴¹. Second, high-copy repetitive sequences—such as centromeric and sub-telomeric stretches of satellite sequences, ribosomal DNA, and long, near-identical LTR-retrotransposons—cannot be fully resolved even with long reads; these regions tend to collapse during the assembly process, resulting in reduced copy numbers. Crucially, this underrepresented fraction occurs within the repeat space rather than the gene space, which is consistent with the 98.6% complete BUSCOs and the high read-mapping rate achieved here. For comparison, the barley MorexV3 pseudomolecules account for a total assembled genome size of 4.2 Gb, whereas its cytometric estimate was calculated at ~5 Gb, a gap equivalent to an entire chromosome that known missing repeats cannot fully explain⁴². An identical difference recurs across the 76 long-read genomes of the barley pangenome⁴³. We therefore consider the present assembly a comprehensive representation of the caraway gene space, with the residual difference reflecting collapsed repeats and the upper-bound nature of the cytometric estimate.

Despite its high quality, the current assembly remains a draft at the scaffold level. While the contig N50 is high, the 2,171 fragments are not yet anchored into definitive chromosomes. Here, the given genetic map offers first possibilities to resolve the physical orientation of the scaffolds, but a high number of scaffolds is not represented by SNPs. Future efforts should employ Hi-C sequencing⁴⁴ to bridge the remaining gaps and achieve a true chromosome-level assembly. While structural annotation is excellent, transcriptomics (RNA-Seq) could further refine the gene models, particularly for isoforms and non-coding RNAs.

The wild biennial parent utilized in this study exhibited a high baseline heterozygosity (24.81%, Table CCC), which could be expected for a predominant outbreeding species⁴⁵. Including these heterozygous loci would have increased the number of SNPs for mapping. However, this would have introduced structurally divergent segregation patterns (resembling a pseudo-testcross) and potentially GBS-related technical artifacts like pseudo-heterozygosity from paralogous regions⁴⁶. Indeed, this is supported by the fact that the annual inbred

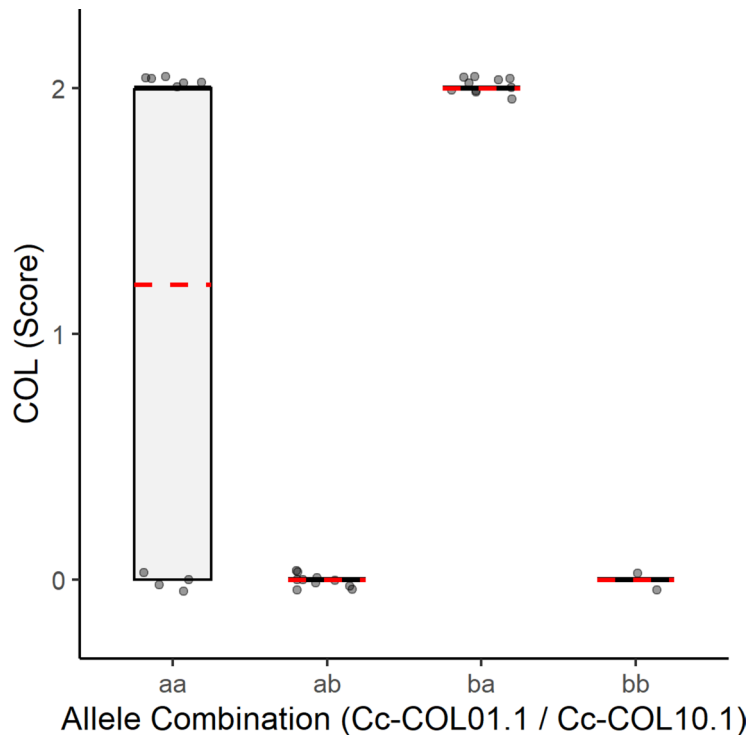


Fig. 11. Phenotypic distribution of flower color for combinations of homozygous genotypes of two identified QTL (Cc-COL01.1 and Cc-COL10.1). Genotypes at peak markers: (a) homozygous biennial, (b) homozygous annual. COL is shown as scoring (0 = white, 2 = rose). Solid lines indicate medians; red dashed lines indicate means. Individual data points are shown as gray jitters.

parent of the fifth inbred generation retained an observed heterozygosity of 5.92%, which likely overestimates true biological heterozygosity due to such technical noise. To prevent severe mapping errors in the absence of a chromosome-level reference map for validation, we deliberately chose to exclude parental heterozygous loci. In general, we applied strict, conservative filter settings and focused on a lower but sufficient number of homozygous loci to construct a first robust map.

Flowering time and flowering type

Genetic architecture of flowering time and flowering type in caraway – beyond the 3:1 mendelian ratio

In the following discussion it is crucial to differentiate between flowering time as a quantitative trait and flowering type (annual vs. biennial) as a qualitative trait. In this study, ‘flowering time’ usually included the prevalent quantitative difference between flowering types (BOF1 and BOF2) so that we predominately searched for loci controlling flowering type using these trait descriptors. However, we also searched for flowering time loci independent from flowering type using Z-standardization (BOF3). While flowering type is primarily associated with vernalization requirement in caraway³, other mechanisms may contribute, such as strong general flowering time regulators shifting flowering induction beyond the first vegetation period. Therefore, it can become challenging to disentangle general flowering time loci from specific loci governing vernalization response.

We will focus on the five-QTL model supported by the final MQM model and corroborated by BLINK for BOF2 using adjusted flowering time. The mitigation of the bimodal distribution of data (compared to BOF1 and FT1) clearly improved the signal for QTL detection, both for loci rather related to vernalization response and to flowering time in general (see below). Additional loci detected only by Kruskal–Wallis (for FT1) or BLINK as well as sub-threshold signals in MQM might represent minor-effect QTL and should be interpreted cautiously. Overall, the integration of MQM, IM, BLINK, and Kruskal–Wallis analyses (for nominal scaled data) allowed us to distinguish high-confidence QTLs from suggestive loci, providing a conservative yet comprehensive characterization of the genetic architecture of flowering traits in this F_2 population. To emphasize the explorative character of this study, all suggestive loci are presented in the supplements. As a side note, BLINK although designed for GWAS, proved to be a fast and efficient tool to detect QTLs in F_2 populations. However, without a chromosome-level assembly, this model cannot replace the complex mapping pipeline.

Our observations in F_1 generation confirm a dominant annual flowering type when assessed as a qualitative trait, which initially aligns with earlier literature^{3,6}. However, when treated as a quantitative trait, we observed a considerable delay in flowering time in F_1 plants compared to annual parents. This suggests that while ‘annuality’ per se may appear dominant (qualitative view), the underlying flowering time alleles exhibit incomplete dominance or a strong additive component (quantitative view).

This was further corroborated by the F_2 field trial: While all plants of the annual parent reached anthesis, only a few F_2 plants flowered within the same timeframe. The observed higher proportion of biennials in F_2 can

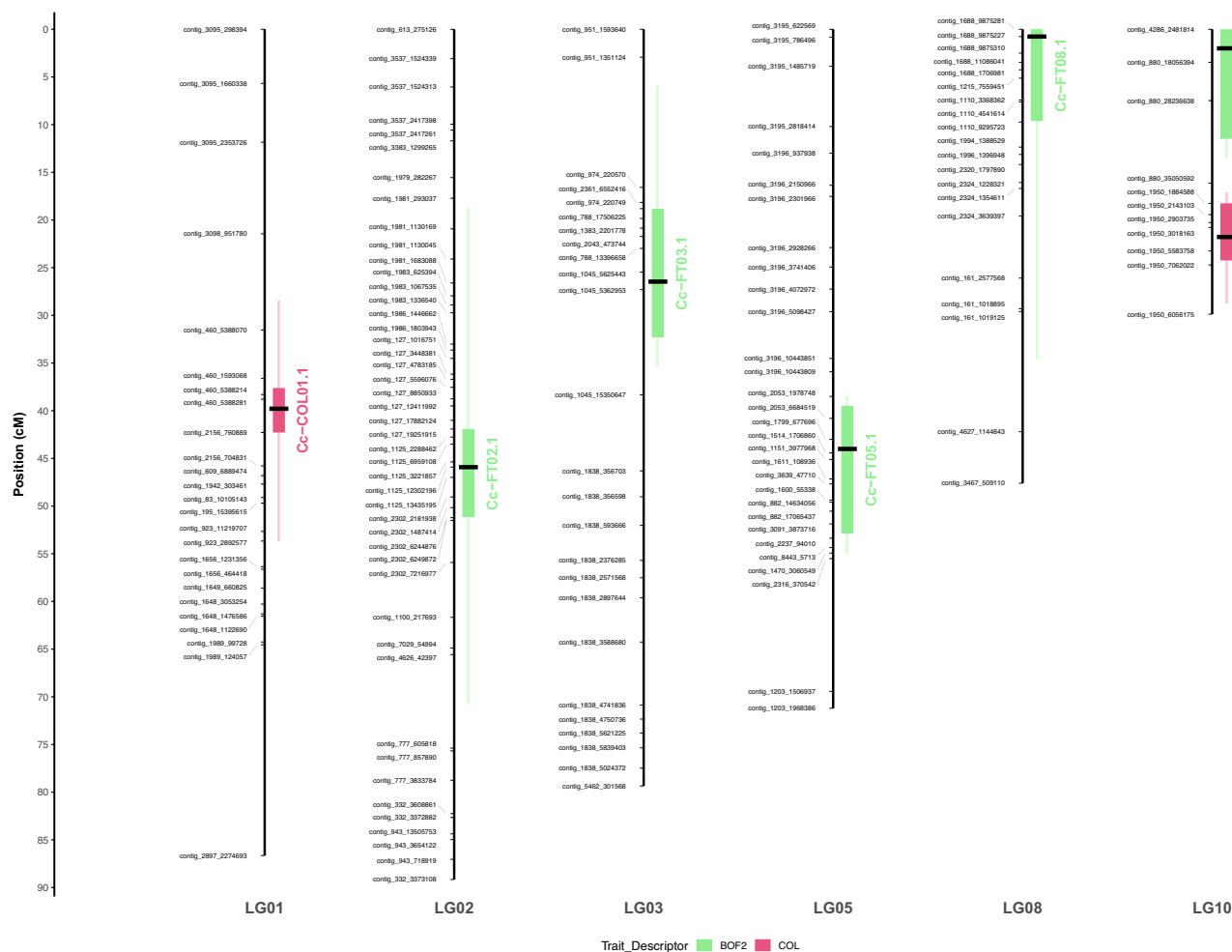


Fig. 12. QTL positions on the genetic map (reduced to LGs covering a QTL). Black horizontal bars highlight the position of the peak from MQM. Thick and thin colored vertical bars show the 1-LOD and 2-LOD intervals, respectively.

only be explained on a quantitative scale in combination with the late sowing date, which may have masked the ‘annual’ phenotype of late-flowering genotypes. This biological truncation further justifies the use of the adjusted BOF2 trait descriptor for flowering time, which better reflects the true genetic potential of these individuals. Indeed, the five detected QTLs for BOF2 predominantly showed additive effects with only marginal dominance. No QTL showed an obligatory biennial habit for the homozygous biennial genotype (Fig. 9), so that it is unlikely that a single vernalization switch operates in this specific population. This supports the hypothesis that flowering time in this cross is not governed by a single ‘on/off’ switch, but by several minor-to-medium effect QTL (6.1 to 10.5% explained variance, Table 4). Therefore, our data do not support a major QTL affecting flowering type with a classic 3:1 ratio (annual vs. biennial) associated with vernalization requirement as proposed by Nemeth and Pluhar⁶. This discrepancy might stem from the distinct genetic background of our Russian wild biennial parent compared to the cultivar ‘Maud’ used by Nemeth and Pluhar⁶.

Nevertheless, since flowering of biennials can be induced by cold, the vernalization pathway must play a crucial role in flowering time regulation in caraway. Although we cannot fully disentangle whether the detected QTLs modulate vernalization requirement or flowering time in general or both, comparing the result of the different flowering time descriptors (BOF1, BOF2, BOF3 and FT1) and the effects within flowering types can provide first hints, which, however, need additional evidence: Since, the effect of the QTLs Cc-FT02.1, Cc-FT03.1, Cc-FT05.1, and Cc-FT08.1 is predominantly explained by the differentiation between annuals and biennials (no significance for BOF3) these QTLs are likely involved in the vernalization pathway. In contrast, QTL Cc-FT10.1 exhibited considerable effects within both flowering types (high significance for BOF3). This suggests that Cc-FT10.1 acts as a general flowering time regulator, potentially explaining why the Russian wild genotype shows delayed flowering even among other biennial genotypes⁷. As a simplified working thesis, we finally propose a polygenic model where Cc-FT02.1, Cc-FT03.1, Cc-FT05.1, and Cc-FT08.1 primarily but not exclusively modulate vernalization response, whereas Cc-FT10.1 acts as a general flowering time regulator.

Comparison to other Apiaceae and candidate genes

In Apiaceae vegetables and medicinal plants grown for their root structures, early/annual flowering (or bolting) has been analyzed as detrimental behavior reducing quality and yield^{47,48}, whereas genetic studies on flowering time in seed drugs seem to be scarce. Similar to caraway, in multiple Apiaceae species annual flowering habit was found to be dominant. In carrot one vernalization locus has been mapped to chromosome 2⁴⁹, yet the monogenic model was later extended to a multiallelic digenic model⁵⁰. In celery one major locus was identified for controlling vernalization requirement⁵¹. On first glance, these studies appear to contrast with our proposed polygenic model for the investigated F₂ in caraway. However, it should be recognized that each study merely represents a snapshot of the genetic diversity within a species or family. Looking into details, we see difficulties to distinguish between a qualitative (“annual”) and quantitative (“early”) description of flowering in other species, too. E.g. in carrot even annuals need a cold stimulus⁵². Moreover, various reports of different rates of early/annual flowering in diverse Apiaceae species depending on environmental conditions and cultivation systems, including sowing date⁴⁸, demonstrate some complexity of the genetic control of flowering in Apiaceae.

The transition from vegetative to reproductive growth in plants is governed by a complex integrated network of pathways. While transcriptomic studies in *Angelica sinensis* and *Daucus carota* have identified a wide array of orthologs^{53–55}, pinning down genes behind certain flowering time QTLs remains challenging even in well-studied Apiaceae like carrot. Genes involved in the cold-induced suppression of floral repressors (e.g., *FLC*-like) or the activation of vernalization response genes (e.g., *VRN*-like) represent the most promising candidates for QTLs predominantly associated with flowering type, such as Cc-FT02.1.

Indeed, we identified a Histone-lysine N-methyltransferase (containing a SET domain and a RING/FYVE/PHD-type zinc finger) in close proximity to the peak marker of Cc-FT02.1. Based on its predicted domain architecture, this gene represents a plausible candidate for involvement in chromatin-mediated vernalization responses. In particular, the combination of SET and PHD domains is reminiscent of functions associated with Polycomb-mediated regulation and cold-responsive chromatin remodeling described for vernalization-related factors such as *VRN2* or *VIN3*^{56,57}. However, this functional relationship remains hypothetical and cannot be inferred from positional information alone. The inherent uncertainty of the peak position and the high gene density within the 1-LOD interval must be considered, especially since this interval may not represent a conservative confidence threshold in a medium-sized F₂ population. Therefore, fine-mapping and subsequent functional validation remain indispensable to confirm the causal genes underlying Cc-FT02.1 and other identified QTLs.

Flower color

Genetic architecture of flower color – environmental sensitivity, an expected major “biennial” QTL & a surprising modifying ‘annual’ QTL

Whereas the biennial parent showed a deep rose flowering habit in our field trials in 2019 and 2020⁷, the substantial bleaching of flower color in both the parent and the F₁ generation under greenhouse conditions indicated that flower color in caraway is a plastic trait, highly sensitive to environmental factors. Observations from various field trials further show that most caraway genotypes, including annuals, can produce rose/reddish pigments under stress conditions (though not reaching the deep rose intensity shown in Fig. 1), which adds an additional layer of complexity to phenotyping.

Under field conditions, the rose coloration reappeared in the F₂ generation with varying intensities; however, environmental effects (bleaching) were again confirmed on sunny days. Since a quantitative scoring of color intensity was considered insufficiently robust and reproducible due to these environmental effects, in this study we focused on presence and absence of flower coloration. Disregarding color intensity, the major QTL detected on LG10 showed incomplete dominance of the biennial rose-flowering allele (Table 4). Given that flower color could not be assessed on a quantitative scale and that heterozygotes exhibited high plasticity (Fig. 9), this might suggest an intermediate inheritance pattern.

In addition to the major locus on LG10, a second QTL was identified on LG01. Interestingly, the annual allele at this locus contributed positively to rose coloration. It might be that we observe an epistatic effect between both QTLs in the F₂ population that was latent in the parental lines, as the combination of homozygous annual alleles on LG01 and homozygous biennial alleles on LG10 resulted in a consistent rose phenotype (Fig. 11). This indicates that while LG10 may act as a primary switch, the LG01 locus functions as a genetic modifier. Furthermore, the observation that this pigmentation extends beyond the petals to the internodes, peduncles, and ripening seeds suggests a systemic regulation of pigment biosynthesis rather than an organ-specific floral trait.

Comparison to other Apiaceae and candidate genes

Likely, flower color is associated with anthocyanin synthesis. Purple carrots, which “accumulate large quantities of anthocyanins in their root tissues, as well as in other plant parts”⁵⁸ might be the best characterized model within Apiaceae. Here, two QTLs (P₁ and P₃) have been mapped to chromosome 3^{59,60}. It has further been shown that the MYB transcription factor DcMYB7 (within P₃), rather than structural genes, controls purple pigmentation in carrot roots^{58,61}. In caraway, it is plausible that similar to carrot rather regulatory genes may be responsible for the difference between rose- and white-flowering genotypes. The fact that white-flowering genotypes can develop rose pigmentation under stress conditions (based on own observations) is consistent with the hypothesis that the structural biosynthetic pathway is intact and only lacks a regulatory trigger. Therefore, the detected three *myb8* related transcription factors nearby the peak for Cc-COL10.1 represent plausible candidate genes. However, again fine-mapping and subsequent functional validation remain indispensable to confirm the causal genes.

Limitations and future perspectives

In general, all findings of QTL mapping are restricted to the evaluated biparental population and one specific environment. A validation across different populations and environments remains necessary. Without such a validation, all detected QTLs should be considered preliminary. Moreover, we used a moderate sample size ($N = \text{up to } 187$) for QTL mapping, which limits the detection of small-effect QTLs. Therefore, considering the polygenic nature of flowering induction observed in our population, in future validations the sample size should be increased. Although we detected plausible candidate genes, at the current state these findings are highly speculative. QTL confidence intervals are still very large and include an extensive list of annotated genes. Further studies, such as fine-mapping and gene expression analysis are necessary to narrow down QTL intervals and the list of potential candidate genes.

For future research on flowering time, the inclusion of cultivated biennial material could be of particular interest. This would also allow to validate the observations of Nemeth and Pluhar⁶, who proposed one major QTL affecting flowering time and flowering type in the cultivar 'Maud'. Since we found that a relatively late sowing potentially masked the "true" flowering type of F_2 plants, it seems recommendable to use earlier sowing dates in subsequent studies.

Although the environmental sensitivity of flower color in caraway complicates its practical application, it remains a trait of interest for the development of biennial or winter-annual cultivars. Such varieties, which typically flower under milder spring conditions, are more likely to consistently display the rose phenotype, offering opportunities for specialized branding and market differentiation. Beyond agriculture, the aesthetic combination of rose flowers, subtly tinted seeds, and pigmented stalks suggests a potential for ornamental or floristic use.

Conclusion

The QTL mapping in this study provides novel insights into the genetic architecture of flowering time and flower color in caraway establishing a crucial foundation for future functional studies. Together, the linkage map and the annotated draft genome assembly represent a milestone resource for caraway, which will promote the molecular elucidation of important traits that have remained uncharacterized until now.

Data availability

Phenotypic data, filtering pipeline, SNP matrix, genetic maps, results from QTL mapping, functional annotation and list of candidate genes are added to the supplements. The GBS data (FASTQ files) generated for this study have been deposited in the European Nucleotide Archive (ENA) under the project accession number PRJEB108909. Individual sample accessions are listed in Supplementary Table S1M. The long-read whole genome sequencing data collected in this study, the genome assembly and annotation have been deposited at the European Nucleotide Archive (ENA) under the project accession number PRJEB111921.

Received: 4 May 2026; Accepted: 7 July 2026

Published online: 14 July 2026

References

1. von Maydell, D. et al. On genetic diversity in caraway: Genotyping of a large germplasm collection. *PLoS ONE* **15**, e0244666. <https://doi.org/10.1371/journal.pone.0244666> (2021).
2. Pank, F. in *Handbuch des Arznei- und Gewürzpflanzenanbaus*. 4 (ed B. Hoppe) (Verein für Arznei- und Gewürzpflanzen SALUPLANTA e.V.) 728–762 (2012).
3. Németh, É. in *Caraway The Genus Carum Vol. 7 Medicinal and Aromatic Plants - Industrial Profiles* (ed E. Nemeth) Ch. (harwood academic publishers) 5, 69–81 (1998).
4. Németh, É. in *Caraway The Genus Carum Vol. 7 Medicinal and Aromatic Plants - Industrial Profiles* (ed E. Nemeth) Ch. (harwood academic publishers) 5, 1–8 (1998).
5. von Maydell, D. et al. On heterosis, inbreeding depression and general combining ability in annual caraway (*Carum carvi*). *Euphytica* **217**, 163. <https://doi.org/10.1007/s10681-021-02893-x> (2021).
6. Nemeth, E. & Pluhar, Z. Preliminary observations on the inheritance of caraway flowering. *Beiträge zur Züchtungsforschung* **2**, Jahrgang, 116–118 (1996).
7. von Maydell, D. et al. Genetic variation of annual and biennial caraway (*Carum carvi*) germplasm offers diverse opportunities for breeding. *Ind. Crops Prod.* **208**, 117798. <https://doi.org/10.1016/j.indcrop.2023.117798> (2024).
8. von Maydell, D., Keilwagen, J., Lehnert, H., Berner, T. & Marthe, F. On the search for the vernalization locus in caraway (*Carum carvi*) using genotyping by sequencing data. *J. Kulturpflanzen* **74**, 263–270. <https://doi.org/10.5073/jfk.2022.11-12.06> (2022).
9. Bocianowski, J. & Seidler-Lozykowska, K. The relationship between RAPD markers and quantitative traits of caraway (*Carum carvi* L.). *Ind Crops Prod.* **36**, 135–139 (2012).
10. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546. <https://doi.org/10.1038/s41587-019-0072-8> (2019).
11. Gonnella, G., Niehus, N. & Kurtz, S. GfaViz: Flexible and interactive visualization of GFA sequence graphs. *Bioinformatics* **35**, 2853–2855. <https://doi.org/10.1093/bioinformatics/bty1046> (2019).
12. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257. <https://doi.org/10.1186/s13059-019-1891-0> (2019).
13. Seppely, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness. In *Gene Prediction: Methods and Protocols* (ed. Kollmar, M.) 227–245 (Springer New York, 2019).
14. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351> (2015).
15. Chen, Y., Zhang, Y., Wang, A. Y., Gao, M. & Chong, Z. Accurate long-read *de novo* assembly evaluation with Inspector. *Genome Biol.* **22**, 312. <https://doi.org/10.1186/s13059-021-02527-4> (2021).
16. Gabriel, L., Becker, F., Hoff, K. J. & Stanke, M. Tiberius: End-to-end deep learning with an HMM for gene prediction. *Bioinformatics* **40**, btac685. <https://doi.org/10.1093/bioinformatics/btac685> (2024).

17. Keilwagen, J. et al. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89–e89. <https://doi.org/10.1093/nar/gkw092> (2016).
18. Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O. & Grau, J. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics* **19**, 189. <https://doi.org/10.1186/s12859-018-2203-5> (2018).
19. Keilwagen, J., Hartung, F. & Grau, J. GeMoMa: Homology-based gene prediction utilizing intron position conservation and RNA-seq data. In *Gene Prediction: Methods and Protocols* (ed. Kollmar, M.) 161–177 (Springer New York, 2019).
20. Lamesch, P. et al. The *Arabidopsis* information resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210. <https://doi.org/10.1093/nar/gkr1090> (2012).
21. Coe, K. et al. Population genomics identifies genetic signatures of carrot domestication and improvement and uncovers the origin of high-carotenoid orange carrots. *Nat. Plants* **9**, 1643–1658. <https://doi.org/10.1038/s41477-023-01526-6> (2023).
22. Song, X. et al. Coriander genomics database: A genomic, transcriptomic, and metabolic database for coriander. *Hortic. Res.* **7**, 55. <https://doi.org/10.1038/s41438-020-0261-0> (2020).
23. Song, X. et al. Deciphering the high-quality genome sequence of coriander that causes controversial feelings. *Plant Biotechnol. J.* <https://doi.org/10.1111/pbi.13310> (2019).
24. Nevers, Y. et al. Quality assessment of gene repertoire annotations with OMArk. *Nat. Biotechnol.* **43**, 124–133. <https://doi.org/10.1038/s41587-024-02147-w> (2025).
25. Jones, P. et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031> (2014).
26. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. (2025).
27. Bastien, M. et al. Genotyping-by-Sequencing in Potato. 283–296. https://doi.org/10.1007/978-3-319-66135-3_15 (2017).
28. The Galaxy, C. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* **50**, W345–W351. <https://doi.org/10.1093/nar/gkac247> (2022).
29. Trim Galore. In GitHub repository. GitHub. href="https://github.com/FelixKrueger/TrimGalore.com/fenderglass/Flye"><https://github.com/FelixKrueger/TrimGalore.com/fenderglass/Flye> (2021).
30. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. <<https://ui.adsabs.harvard.edu/abs/2013arXiv1303.3997L>>. (2013).
31. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595. <https://doi.org/10.1093/bioinformatics/btp698> (2010).
32. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> (2009).
33. van Ooijen, J. W. JoinMap® 5, Software for the calculation of genetic linkage maps in experimental populations. (2018).
34. van Ooijen, J. W. MapQTL 5, Software for the Mapping of Quantitative Trait Loci in Experimental Populations. *Plant Research International* (2004).
35. Huang, M., Liu, X. L., Zhou, Y., Summers, R. M. & Zhang, Z. W. BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* <https://doi.org/10.1093/gigascience/giy154> (2019).
36. Lipka, A. E. et al. GAPIT: Genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–2399. <https://doi.org/10.1093/bioinformatics/bts444> (2012).
37. Iorizzo, M. et al. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat. Genet.* **48**, 657–666. <https://doi.org/10.1038/ng.3565> (2016).
38. Palumbo, F., Galla, G., Vitulo, N. & Barcaccia, G. First draft genome sequencing of fennel (*Foeniculum vulgare* Mill.): Identification of simple sequence repeats and their application in marker-assisted breeding. *Mol. Breed.* <https://doi.org/10.1007/s11032-018-0884-0> (2018).
39. Liu, H. et al. The parsley genome assembly and DNA methylome shed light on apigenin biosynthesis in the Apiaceae. *Plant Physiol.* **197**, kiaf077. <https://doi.org/10.1093/plphys/kiaf077> (2025).
40. Liu, H. et al. The high-quality genome of *Cryptotaenia japonica* and comparative genomics analysis reveals anthocyanin biosynthesis in Apiaceae. *Plant J.* **118**, 717–730. <https://doi.org/10.1111/tjp.16628> (2024).
41. Dolezel, J., Greilhuber, J. & Suda, J. Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* **2**, 2233–2244. <https://doi.org/10.1038/nprot.2007.310> (2007).
42. Navrátilová, P. et al. Prospects of telomere-to-telomere assembly in barley: Analysis of sequence gaps in the MorexV3 reference genome. *Plant Biotechnol. J.* **20**, 1373–1386. <https://doi.org/10.1111/pbi.13816> (2022).
43. Jayakodi, M. et al. Structural variation in the pangenome of wild and domesticated barley. *Nature* **636**, 654–662. <https://doi.org/10.1038/s41586-024-08187-1> (2024).
44. Belton, J. M. et al. Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276. <https://doi.org/10.1016/j.ymeth.2012.05.001> (2012).
45. von Maydell, D., Brandes, J., Lehnert, H., Junghanns, W. & Marthe, F. Breeding synthetic varieties in annual caraway: Observations on the outcrossing rate in a polycross using a high-throughput genotyping system. *Euphytica* **217**, 1. <https://doi.org/10.1007/s10681-020-02732-5> (2020).
46. McKinney, G. J., Waples, R. K., Seeb, L. W. & Seeb, J. E. Paralogues are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Mol. Ecol. Resour.* **17**, 656–669. <https://doi.org/10.1111/1755-0998.12613> (2017).
47. Wang, X.-J. et al. Origin, evolution, breeding, and omics of Apiaceae: A family of vegetables and medicinal plants. *Hortic. Res.* **9**, uhac076. <https://doi.org/10.1093/hr/uhac076> (2022).
48. Li, M., Li, M., Wang, L., Li, M. & Wei, J. Apiaceae medicinal plants in China: A review of traditional uses, phytochemistry, bolting and flowering (BF), and BF control methods. *Molecules* **28**, 4384 (2023).
49. Alessandro, M. S., Galmarini, C. R., Iorizzo, M. & Simon, P. W. Molecular mapping of vernalization requirement and fertility restoration genes in carrot. *Theor. Appl. Genet.* **126**, 415–423. <https://doi.org/10.1007/s00122-012-1989-1> (2013).
50. Wohlfeiler, J., Alessandro, M. S., Cavagnaro, P. F. & Galmarini, C. R. Multiallelic digenic control of vernalization requirement in carrot (*Daucus carota* L.). *Euphytica* **215**, 37. <https://doi.org/10.1007/s10681-019-2360-2> (2019).
51. Quiros, C. F., Douches, D. & D'Antonio, V. Inheritance of annual habit in celery: Cosegregation with isozyme and anthocyanin markers. *Theor. Appl. Genet.* **74**, 203–208. <https://doi.org/10.1007/bf00289969> (1987).
52. Wohlfeiler, J., Alessandro, M. S., Morales, A., Cavagnaro, P. F. & Galmarini, C. R. Vernalization requirement, but not post-vernalization day length, conditions flowering in carrot (*Daucus carota* L.). *Plants* **11**, 1075 (2022).
53. Luo, M. et al. Regulatory networks of flowering genes in *Angelica sinensis* during vernalization. *Plants* **11**, 1355 (2022).
54. Cholin, S. S. et al. Deciphering carotenoid and flowering pathway gene variations in eastern and western carrots (*Daucus carota* L.). *Genes* **15**, 1462 (2024).
55. Zhan, Z. et al. Molecular cloning, expression analysis, and subcellular localization of FLOWERING LOCUS T (FT) in carrot (*Daucus carota* L.). *Mol. Breed.* **37**, 149. <https://doi.org/10.1007/s11032-017-0749-y> (2017).
56. Gendall, A. R., Levy, Y. Y., Wilson, A. & Dean, C. The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in *Arabidopsis*. *Cell* **107**, 525–535. [https://doi.org/10.1016/S0092-8674\(01\)00573-6](https://doi.org/10.1016/S0092-8674(01)00573-6) (2001).
57. Labandera, A.-M. et al. The PRT6 N-degron pathway restricts VERNALIZATION 2 to endogenous hypoxic niches to modulate plant development. *New Phytol.* **229**, 126–139. <https://doi.org/10.1111/nph.16477> (2021).

58. Bannoud, F. et al. Dissecting the genetic control of root and leaf tissue-specific anthocyanin pigmentation in carrot (*Daucus carota* L.). *Theor. Appl. Genet.* **132**, 2485–2507. <https://doi.org/10.1007/s00122-019-03366-5> (2019).
59. Yildiz, M. et al. Expression and mapping of anthocyanin biosynthesis genes in carrot. *Theor. Appl. Genet.* **126**, 1689–1702. <https://doi.org/10.1007/s00122-013-2084-y> (2013).
60. Cavagnaro, P. F. et al. A gene-derived SNP-based high resolution linkage map of carrot including the location of QTL conditioning root and leaf anthocyanin pigmentation. *BMC Genomics* **15**, 1118. <https://doi.org/10.1186/1471-2164-15-1118> (2014).
61. Xu, Z. S., Yang, Q. Q., Feng, K. & Xiong, A. S. Changing carrot color: Insertions in DcMYB7 alter the regulation of anthocyanin biosynthesis and modification. *Plant Physiol.* **181**, 195–207. <https://doi.org/10.1104/pp.19.00523> (2019).

Acknowledgements

We thank Brian Boyle and Sharen Roland and their teams in Canada for excellent library preparation and sequencing. For sample preparation and technical assistance on the PromethION system we want to thank Jana Müglitz and Samuel Streicher from the Breeding Informatics group at MLU and Beate Kamm (JKI) for high quality DNA extraction for GBS. Not at least we thank the team VQ, Annette Benecke, Thomas Heck and Konstantin Bradke (JKI) for supporting planting and sampling of the plants in the field.

Author contributions

DvM wrote the original draft, designed the trial, conducted phenotypic analysis and genetic mapping, participated in funding acquisition and project administration. FSL contributed to the original draft, conducted annotation and candidate gene identification. MJ advised DvM and contributed to funding acquisition and project administration. YPG contributed to long-read sequencing and generation of the draft assembly. HB supervised genetic mapping. FM supervised the trial and contributed to funding acquisition and project administration. JK contributed to the original draft, supervised annotation and candidate gene identification. TS contributed to the original draft, conducted long-read sequencing and generated the draft assembly. All authors read and approved the final version of the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This project was funded by the Fachagentur Nachwachsende Rohstoffe (FNR) on behalf of the Federal Ministry of Agriculture, Food and Regional Identity (BMLEH), Funding ID: 2220NR103A. A fund for open access publication was granted by the Julius Kühn Institute (JKI).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-61767-1>.

Correspondence and requests for materials should be addressed to D.M.

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

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

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

© The Author(s) 2026