

# QTL analysis of flowering time and flower color in diverse crosses of biennial and annual caraway (*Carum carvi*) under early sowing conditions

Daniel Maydell

[daniel.maydell@julius-kuehn.de](mailto:daniel.maydell@julius-kuehn.de)

Julius Kühn Institute

**Beate Kamm**

Julius Kühn Institute

**Akhil Pinnuri**

Dr. Junghanns GmbH

**Fang-Shiang Lim**

Julius Kühn Institute

**Jolien Zimmer**

Julius Kühn Institute

**Martin Junghanns**

Dr. Junghanns GmbH

**Frank Marthe**

Julius Kühn Institute

---

## Research Article

**Keywords:** Flower color, genetic map, linkage, quantitative trait locus, Umbelliferae, vernalization

**Posted Date:** September 28th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-10973509/v1>

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

[Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

**QTL analysis of flowering time and flower color in diverse crosses of biennial and annual caraway (*Carum carvi*) under early sowing conditions**

Daniel von Maydell<sup>\*1</sup>, Beate Kamm<sup>1</sup>, Akhil Pinnuri<sup>1,3</sup>, Fang-Shiang Lim<sup>2</sup>, Jolien Zimmer<sup>1</sup>, Martin Junghanns<sup>3</sup>, Frank Marthe<sup>1</sup>

<sup>1</sup> Julius Kühn Institute, Institute for Breeding Research on Horticultural Crops, Quedlinburg, Germany

<sup>2</sup> Julius Kühn Institute, Institute for Biosafety in Plant Biotechnology, Quedlinburg, Germany

<sup>3</sup> Dr. Junghanns GmbH, Aschersleben, OT Groß-Schierstedt

**\* Correspondence:**

Daniel von Maydell

daniel.maydell@julius-kuehn.de

ORCIDs

Daniel von Maydell (ORCID: 0000-0001-7674-1559)

## Abstract

Background: Caraway (*Carum carvi* L.,  $2n = 2x = 20$ ) is a major essential oil crop with biennial and annual flowering. Combining both genepools offers a promising strategy to enhance genetic diversity and climate resilience, requiring a clear understanding of flowering time and vernalization genetics. Despite the increasing availability of genome assemblies in Apiaceae, genetic mapping studies beyond carrot remain scarce. Building on previous findings of polygenic control in a wild background, this study investigated two expanded  $F_2$  mapping populations ( $N = 288$  each) comparing a cultivated (P1) and a wild (P2) biennial parent crossed with the same annual parent. In addition, phenotypic data of 15 other  $F_2$  populations were analyzed. Phenotyping under early sowing conditions was combined with genotyping-by-sequencing (GBS).

Results: GBS identified 1,245 (P1) and 1,560 (P2) segregating SNPs, which were used to construct linkage maps and a consensus physical map comprising 1.34 Gb (72% of the draft assembly). Early sowing resulted in a strong dominance of annual flowering (P1: 97.9%; P2: 99.7%), contrasting with previous observations under late sowing. QTL mapping revealed contrasting genetic architectures between the cultivated and wild backgrounds. In cultivated P1, two novel major QTLs likely associated with vernalization response were identified on LG02 and LG06, jointly explaining 30.7% of phenotypic variance, suggesting an oligogenic architecture. Diagnostic PACE markers were developed and validated for future fine-mapping of these QTLs, achieving a 96.3–98.8% GBS conversion rate. Conversely, wild P2 confirmed a cumulative effect of four previously identified vernalization associated QTLs, explaining only 7.5% of phenotypic variance and supporting a polygenic architecture. Across populations, a potentially common QTL on LG10 was detected as a candidate general flowering time regulator. Additionally, two previously reported QTLs on LG01 and LG10 for rose flower color were re-detected.

40 Conclusions: While we observe an increasing number of genome assemblies in Apiaceae, QTL studies  
41 are still rare in this family. This comprehensive study on flowering time in caraway sheds new light on  
42 the potential diversity and complexity of flowering time regulation in Apiaceae, demonstrating a strong  
43 environmental impact and a contrasting genetic architecture between a wild and a cultivated biennial  
44 background.

45

## 46 **Keywords**

47 Flower color, genetic map, linkage, quantitative trait locus, Umbelliferae, vernalization

48

49

50

51

52

## Introduction

Caraway (*Carum carvi* L.,  $2n = 2x = 20$ ) is a major essential oil crop of the Apiaceae family used for various pharmaceutical, culinary and industrial purposes (Mahboubi 2019). It is characterized by two distinct flowering types: biennial and annual (Németh 1998; Pank 2012), which is associated with the prevalent population structure of caraway genetic resources (von Maydell et al. 2021). While biennials require a winter vernalization period and occupy the field for two growing seasons, annuals complete their life cycle within a single year. To enrich the limited genetic diversity of the annual genepool, introgression from widespread biennial populations is a promising breeding strategy. Particularly for breeding of “climate-resilient” winter annual caraway (autumn sowing), the attributes of biennials like winterhardiness can be very valuable (von Maydell et al. 2026a). However, the genetic dissection of agronomic traits in caraway has long been hampered by a lack of genomic resources (Misginna et al. 2026).

Recently, von Maydell et al. (2026b) established the first basic genomic framework for caraway by developing a draft genome assembly and a structural and functional gene annotation. Moreover, a genetic map consisting of 10 linkage groups (LGs) was established by analyzing an  $F_2$  mapping population from a cross of the wild-type biennial accession ‘CARUM 63’ and the annual inbred line ‘L32’. They identified five quantitative trait loci (QTLs) for flowering time on LGs LG02, LG03, LG05, LG08, and LG10, which collectively explained 42.7% of the phenotypic variance, supporting a polygenic, predominantly additive model for flowering time regulation of this cross. Furthermore, it was hypothesized that the QTL on LG10 might be a general flowering time regulator, whereas the other QTLs might be predominately but not exclusively involved in vernalization response. Additionally, von Maydell et al. (2026b) analyzed flower color, since the biennial parent ‘CARUM 63’ showed a rare rose flower color. They identified one major QTL for flower color on LG10 (26% explained phenotypic variance), with the biennial allele showing a rose coloration and a second QTL

77 on LG01 (10.9% explained phenotypic variance), interestingly with the annual allele (from the white  
78 parent) seeming to stabilize the rose coloration.

79 Despite these recent insights, several critical questions and methodological limitations remain to be  
80 addressed. First, the initial genetic map was restricted by marker density and population size ( $N = 187$ ).  
81 Second, environmental factors, specifically sowing date, might strongly influence flowering  
82 phenology. Under the late sowing conditions (LSC) of the previous study, only 36% of the  $F_2$  plants  
83 exhibited an annual habit. Taking into account that annual flowering should be dominant (100% annual  
84 flowering in  $F_1$ ), this suggested that the "true" genetic flowering type of many individuals may have  
85 been masked by late cultivation. This issue affected the discrimination between QTLs involved in  
86 vernalization response and QTLs generally regulating flowering time. Third, the polygenic model of  
87 von Maydell et al. (2026b) using a "wild" biennial parent contrasted with the observations from  
88 Nemeth and Pluhar (1996), who suggested one major QTL for vernalization requirement and additional  
89 minor QTLs affecting flowering time based on a segregation analysis using a cultivated biennial parent  
90 ('Maud'). To address these gaps, this study strived to achieve the following objectives:

91 **1. Map improvement, cross-validation & integration of physical and genetic information:** We  
92 aimed to refine and cross-validate the caraway genetic map by increasing both sequencing depth  
93 (marker density) and population size (up to 288 plants per population) across two distinct  $F_2$  mapping  
94 populations: Population P1 (biennial cultivar 'Rekord' x 'L32') and Population P2 (CARUM 63 x annual  
95 inbred line 'L32'). To allow direct comparison of QTL intervals across different mapping populations,  
96 a consensus physical framework was constructed using the draft genome assembly.

97 **2. Environmental impact on QTL detection in wild biennial background (P2):** To test the  
98 hypothesis that early sowing unmasks the "true" genetic flowering habit, we conducted a comparative  
99 QTL mapping under early sowing conditions (ESC) using the expanded  $F_2$  P2, allowing a direct  
100 environment-comparison with the identical cross from von Maydell et al. (2026b).

101 **3. Comparison of QTL detection with cultivated biennial background (P1) & Diversity of**  
102 **flowering time regulation:** To investigate the contrasting findings between cultivated and wild  
103 material, we performed a second QTL mapping using the newly developed F<sub>2</sub> Population P1 dissecting  
104 the genetic architecture of a standard commercial biennial background ('Rekord'). In case of common  
105 QTLs, this would also allow a second validation of QTLs.

106 **4. Diversity of flowering time regulation** Additionally, we expanded the genetic landscape by  
107 analyzing phenotypic data from crosses involving 15 different biennial maternal genotypes and an  
108 annual paternal genotype (N = 48 to 92 per each F<sub>2</sub>).

109 **5. Re-detection of QTLs for flower color (in P2):** In addition, we tested whether the detected QTLs  
110 for flower color from von Maydell et al. (2026b) could be re-detected in P2 under ESC.

111 **6. Diagnostic markers for fine mapping:** Finally, to bridge the gap to fine mapping aiming towards  
112 marker-assisted selection (MAS), we designed PCR-based PACE markers targeting the peak and  
113 flanking regions of the most promising identified QTLs for flowering time. We validated functionality  
114 and evaluated their conversion rate from GBS data.

115

# 1 Material and methods

## 1.1 Biparental QTL mapping of 2 F<sub>2</sub> populations

### 1.1.1 Biparental mapping populations

We crossed two different biennial genotypes with the annual inbred line ‘L32’ of the Julius Kühn Institute (JKI, Quedlinburg, Germany) of the 5th inbred generation. One F<sub>1</sub> plant from each cross was self-fertilized to produce two segregating F<sub>2</sub> mapping populations. As biennial parent for population P1 we used the Czech cultivar ‘Rekord’ (originally purchased from Agritec Plant Research Ltd., Šumperk, Czech Republic), which is a frequently cultivated variety. As biennial parent for population P2 we used the biennial wild-type accession ‘CARUM 63’ from the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK, Seeland OT Gatersleben, Germany), which had shown late flowering among biennial accessions and a rare rose flower color (von Maydell et al. 2024). According to IPK passport data, this accession originates from Russia and was originally provided as a donation by the Botanical Garden of Halle (Saale), Germany. The biological classification of ‘CARUM 63’ was confirmed by IPK genebank curators, with official reference specimens archived in the Gatersleben Herbarium (GAT). P2 represents the same cross as used in a recent mapping under LSC (von Maydell et al. 2026b). The biennial parents ‘Rekord’ and ‘Carum 63’ were also chosen because of a high genetic distance (von Maydell et al. 2021) increasing the possibility to identify diverse QTLs.

### 1.1.2 Genotyping by sequencing, SNP discovery and filtering

Genomic DNA was extracted from young leaves using DNeasy Plant Mini Kit, Qiagen (for P1) or NucleoSpin Plant II Mini Kit, Marchery-Nagel (for P2 and parents). GBS library preparation followed the protocols described in detail in von Maydell et al. (2026b). Briefly, DNA was digested with *PstI* and *MspI* (Bastien et al. 2017; Poland et al. 2012), and sequenced on an Illumina NovaSeq 6000 platform (150-bp paired-end).

Bioinformatic processing was conducted using the Galaxy platform according to the pipeline established by von Maydell et al. (2026b). Briefly, raw reads were trimmed using Trim Galore (Krueger 2021), mapped against the caraway draft assembly (von Maydell et al. 2026b) using BWA-mem (Li 2013), followed by SNP calling using bcftools (Li et al. 2009). The resulting SNPs were filtered for quality (score  $\geq 40$ , depth  $\geq 4$ ,  $\leq 10\%$  missing data), homozygous polymorphisms between parental genotypes, and segregation patterns (15–35% for both homozygous genotypes, 35–65% heterozygosity). We used this segregation pattern filter as potentially less conservative pre-filter than a chi square test, which could be strict at high sample sizes. Deviating from von Maydell et al. (2026b), we performed a redundancy reduction prior to map construction: SNPs showing identical genotypes (except for missing values) across all F<sub>2</sub> individuals were grouped into bins, with only the most informative marker (lowest amount of missing data) per bin being retained for the mapping process. Altogether, these filters are meant as preliminary quality filters to gain a reasonable number of high-quality SNPs reducing computational time in the downstream mapping process, which includes additional filters.

### 1.1.3 Construction of linkage maps

Linkage groups (LGs) were established using JoinMap 5.0 (van Ooijen 2018) with a minimum LOD score threshold of 7 (no more changes until LOD = 10). Loci with mean  $\chi^2$  contributions  $>5$  and loci causing spurious linkage between LGs were removed (merely one SNP per population). LGs were ordered according to the map from von Maydell et al. (2026b) using common contigs. The regression mapping algorithm with the Kosambi mapping function was applied to construct genetic maps. To ensure stable models and low multicollinearity during QTL mapping, the genetic map was thinned to a maximum density of one marker per 1 cM (the marker with the lowest proportion of missing data was chosen to represent the bin).

#### **1.1.4 Construction of a consensus physical map and map integration**

To compare QTL intervals across different mapping populations, a consensus physical framework was constructed using the draft genome assembly. Instead of calculating a joint genetic consensus map in centimorgans (cM), which is prone to distortions due to population-specific recombination rates, markers from both individual genetic maps (P1,  $n = 870$ ; P2,  $n = 735$ ) were projected onto the physical genomic sequences. Contig ordering within each linkage group (LG) was established based on the median genetic cM position of all anchored markers per contig, using the denser P1 map as the primary reference skeleton. To seamlessly integrate contigs exclusively mapped in P2 without scale-induced positional shifts, P2 genetic positions were harmonized onto the P1 reference scale via linear interpolation using shared anchor contigs prior to final order assignment. In the same way we integrated the map of von Maydell et al. (2026b), which contained a few contigs not included in the maps of this study. Contig orientation (Forward/Reverse) was mathematically determined using Spearman's rank correlation coefficient ( $\rho$ ) between the physical bp coordinates and genetic cM positions for contigs containing  $\geq 2$  markers. Orientations with a correlation  $r > 0.3$  were classified accordingly, while contigs with single markers or conflicting signals were conservatively designated as "Unknown". To visualize collinearity and recombination landscapes between the populations, Marey maps were generated by plotting genetic distances against the cumulative pseudo-physical chromosome lengths.

#### **1.1.5 Phenotyping**

Seeds of  $F_2$  and parents were sown in the greenhouse at March 17, 2025. To represent parents as controls, self-fertilized seeds from the parental plant were used. The pre-grown plants were acclimatized under protected outdoor conditions for two weeks at the end of April. In early May, plants were transplanted to the field in rows of 24 plants, with 50 cm spacing between plants and 50 cm between rows (100 cm after every second row as a passage). During this early outdoor-period (4 weeks) the lowest (night) temperature ranged from 2 to 12 °C and averaged about 5°C; the highest (day)

temperature ranged from 10 to 28°C. In total, 288 plants per each F<sub>2</sub> and 92 plants per each parent were planted (Figure 1).

### 1.1.6 QTL Mapping

QTL analysis followed the approach described by von Maydell et al. (2026b). Briefly, analysis was performed in MapQTL 5 (van Ooijen 2004), with genome-wide LOD thresholds ( $\alpha = 0.05$ ) determined via 1,000 permutation tests. Initial interval mapping (IM) was followed by multiple QTL mapping (MQM), final co-factor settings per trait in Table S1. Final QTL parameters such as position of the peak and explained variance were estimated from the MQM model, while 1- and 2-LOD support intervals and flanking markers were derived from restricted MQM (rMQM). For the binary trait flower color, we compared MQM results with the non-parametric Kruskal-Wallis test.

For cross-validation of MQM, we used the Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) model (Huang et al. 2019), which is implemented in the R package GAPIT (Lipka et al. 2012). Here, we used the full set of filtered SNPs. The effectiveness of this model, originally designed for genome-wide association studies (GWAS), for cross-validation had been shown in von Maydell et al. (2026b). Notably, BLINK effectively excludes loci correlated with peak markers in contrast to standard models like GLM or MLM, which is a highly effective filter against redundancy in absence of a chromosome-level assembly.

## 1.2 Phenotypic analysis of 15 additional F<sub>2</sub> populations

To screen more diversity, we evaluated well-documented, unpublished data of the JKI working group from the years 1999 and 2000 for 15 different F<sub>2</sub> populations resulted from crosses between biennial and annual caraway. As maternal material the biennial cultivars ‘Cesky’, ‘Hollandi’, ‘Kami’,

‘Konczewicki’, ‘Mansholts’, ‘Maud’, ‘Niederdeutscher’, ‘Plewiski’, ‘Prochan’, ‘Sylvia’, ‘Trojica’, ‘Volhouden’ as well as three biennial accessions from the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK, Seeland OT Gatersleben, Germany) were used. The accessions CARUM 10, CARUM 13 and CARUM 16 origin from Romania, Czech Republic and Italy, respectively and can be defined as wild-type or landrace. As paternal parent an annual selection of the JKI was used.

In mid of March 1999 or 2000, F<sub>2</sub> populations and the annual parent were directly sown into the field in plots/rows with 45 cm distance between plots at the JKI location in Quedlinburg. After germination plants were singled to a distance of 40 cm between plants. A randomized complete block design with 4 repetitions and 21 plants per genotype and repetition was used. The number of annual flowering plants per genotype was collected for all 84 plants, however, merely documented for F<sub>2</sub> populations grown in 1999. Begin of flowering was measured for 12 randomly chosen annual flowering plants per plot, resulting in a sample size of 48 plants per genotype. Deviating from this in 1999, the total sample size of the annual parent was 96.

### **1.3 Diagnostic markers for selected flanking and peak markers**

We built and evaluated PACE markers (3cr Bioscience) for the flanks and the peaks of the two major QTLs detected in P1 on LG02 and LG06 as listed in Table 1. We aimed to select peak markers at maximum 2.0 cM in distance to the peak and flanking markers at the margins of the 1.5 LOD interval (Dupuis and Siegmund 1999) to balance mapping precision and positional certainty. Primers were designed by 3cr Bioscience and produced by Life Technologies (Thermo Fisher Scientific). Sequence information (50 bases up and down of the SNP of interest) and primer sequences are deposited in Table S4. PACE assays were conducted following the protocol from von Maydell et al. (2020) using the CFX96 real-time system and the CFX Maestro software (Bio-Rad). In general, 40 PCR cycles (excluding touchdown) at 57°C annealing temperature were applied. For a comparison with GBS

results, PACE assays were conducted for 88 F<sub>2</sub> plants per locus, including parental DNA and extraction buffer as controls. We used the same DNA samples as used for GBS, without normalizing DNA concentration to simulate the conditions of an applied high-throughput screening. Based on Qubit results, DNA concentration ranged from 20 to 372 ng/μl.

Table 1 List of PACE markers used in this study and basic details of corresponding SNPs. Extended informations in Table S4.

| Marker_ID | LG   | cM   | SNP_position         | Purpose     |
|-----------|------|------|----------------------|-------------|
| M178      | LG02 | 37.2 | contig_127_1178505   | Left flank  |
| M183      | LG02 | 41.3 | contig_127_11822745  | Peak        |
| M175      | LG02 | 47.6 | contig_1125_6569290  | Right flank |
| M198      | LG06 | 33.5 | contig_2317_10137733 | Left flank  |
| M192      | LG06 | 43.1 | contig_2029_245099   | Peak        |
| M204      | LG06 | 47.0 | contig_907_3309145   | Right flank |

## 1.4 Statistics and Visualizations

245 Basic test statistics such as ANOVA, Fishers's exact test or chi-square goodness-of-fit test were  
246 conducted with base R (R Core Team 2025). For various filtering steps frequently the R packages tidyr  
247 (Wickham et al. 2019) and dplyr were employed. Figures were created using ggplot2 (Wickham 2016).

248

249

## 2 Results

### 2.1 Biparental QTL mapping of 2 F<sub>2</sub> populations

#### 2.1.1 Genotyping by sequencing

In total, we generated about 3,850 M reads (average 13.4 M) and 3,400 M reads (average 11.8 M) for P1 and P2, respectively. We gained a sufficient number of sequences (approx. > 2 M reads) for 280 and 275 F<sub>2</sub> individuals for P1 and P2, respectively. To ensure a low number of missing values for parents, we generated a high number of reads for parents (about 192.9 M reads for ‘Rekord’, 626.3 M reads for ‘CARUM 63’ and 397.4 M reads for ‘L32’. After a first quality filter, 13,093 and 12,169 SNPs were gained by GBS for P1 and P2, respectively. Filters for homozygous SNPs between parents, the segregation ratio and redundancy resulted in a final set of 1,245 and 1,560 high-quality SNPs for subsequent mapping for P1 and P2, respectively. Based on these datasets the parents ‘Rekord’, ‘CARUM 63’ and ‘L32’ had a heterozygosity of 17.4%, 31.8% and 9.2/9.6% (P1/P2). We tested whether the integration of SNPs, which were heterozygous in the maternal genotypes and showing a 1:2:1 segregation-like pattern in F<sub>2</sub> could be used to increase marker density. However, we found that the number of markers was merely marginally increased after redundancy reduction and the map construction was not improved. The SNP matrices and effect of all filters are shown in Table S1 C and D.

#### 2.1.2 Linkage maps & Consensus physical map

For both populations, markers were grouped into 10 linkage groups (LGs) after excluding one marker per each population, which caused spurious linkage between LG9 and LG10 (P1) or LG01 and LG03 (P2). All LGs could be aligned to the reference map from Paper 1 von Maydell et al. (2026b) based on common contigs without contradictory findings. Table 2 and Figure 2 compare the thinned linkage

273 maps of both populations. We chose the thinned maps for comparison because these are not biased by  
274 regions with high marker density. Full and thinned maps are provided in Table S2.

275 Map P1 comprised a total of 392 SNP markers covering a total genetic distance of 796.7 cM. Marker  
276 numbers per linkage group (LG) fluctuated between 14 (LG09) and 60 (LG02), averaging 39.2 markers  
277 per LG. The mean inter-marker distance (average gap) was 2.4 cM. In comparison, Map P2 contained  
278 337 SNP markers spanning a total genetic length of 780.1 cM. The number of markers per LG ranged  
279 from 21 (LG07) to 57 (LG02), with an average of 33.7 markers per LG. This resulted in a mean average  
280 gap of 2.6 cM. While the average distance between adjacent markers remained relatively consistent  
281 between both maps, distinct differences were observed regarding the maximum interval sizes (largest  
282 gaps) within individual linkage groups. Map P1 achieved a substantially higher resolution for LG07  
283 and LG08. Conversely, Map P1 exhibited large marker-deficient intervals on LG09 and LG10.  
284 Notably, the largest gap on LG10 increased from 8.6 cM (Map P2) to 36.1 cM (Map P1).

285

286 Table 2: Summary of the genetic linkage maps (Map P1/Map P2) developed via regression mapping  
 287 thinned to one marker per cM. Summary of the consensus physical map.

|                | Linkage Maps P1/P2 |               |                     |                     | Consensus Physical Map |                   |                |
|----------------|--------------------|---------------|---------------------|---------------------|------------------------|-------------------|----------------|
| Group          | Number<br>SNPs     | Length (cM)   | Largest<br>gap (cM) | Average<br>gap (cM) | Number<br>SNPs         | Number<br>Contigs | Length<br>(Mb) |
| <b>LG01</b>    | 41 / 53            | 94.8 / 106.7  | 9.1 / 8.8           | 2.4 / 2.1           | 296                    | 30                | 189.71         |
| <b>LG02</b>    | 60 / 57            | 85.2 / 85     | 6.6 / 9.3           | 1.4 / 1.5           | 363                    | 31                | 152.48         |
| <b>LG03</b>    | 37 / 29            | 81.5 / 69.9   | 11.9 / 10.1         | 2.3 / 2.5           | 196                    | 23                | 177.53         |
| <b>LG04</b>    | 30 / 30            | 89.9 / 69.6   | 16.3 / 6.8          | 3.1 / 2.4           | 214                    | 29                | 136.37         |
| <b>LG05</b>    | 43 / 33            | 68.5 / 73.4   | 5.6 / 10.8          | 1.6 / 2.3           | 269                    | 31                | 146.41         |
| <b>LG06</b>    | 48 / 22            | 77 / 57.7     | 11.4 / 14.1         | 1.6 / 2.7           | 226                    | 27                | 139.89         |
| <b>LG07</b>    | 44 / 21            | 64.7 / 68.9   | 7.4 / 15.2          | 1.5 / 3.4           | 242                    | 22                | 161.25         |
| <b>LG08</b>    | 43 / 34            | 73.4 / 82     | 5.3 / 18.9          | 1.7 / 2.5           | 228                    | 35                | 121.2          |
| <b>LG09</b>    | 14 / 22            | 67.3 / 81.6   | 32.7 / 22.6         | 5.2 / 3.9           | 71                     | 7                 | 26.92          |
| <b>LG10</b>    | 32 / 36            | 94.4 / 85.3   | 36.1 / 8.6          | 3 / 2.4             | 134                    | 12                | 128.62         |
| <b>Average</b> | 39.2 / 33.7        | 79.7 / 78     | - / -               | 2.4 / 2.6           | 223.9                  | 24.7              | 138.04         |
| <b>Total</b>   | 392 / 337          | 796.7 / 780.1 | - / -               | - / -               | 2239                   | 247               | 1380.38        |

288  
 289 Across the 10 linkage groups (LG01–LG10), a total of 2,239 SNPs from all three available full maps  
 290 were placed into the consensus physical map across 247 contigs, covering a total physical length of  
 291 1.38 Gb. On average, each linkage group comprised 223.9 SNPs, 24.7 contigs, and an average physical  
 292 length of 138.04 Mbp (Table 2, Figure 3, Table S2 E, Figure S1). Notably, LG09 was the smallest  
 293 linkage group, exhibiting significantly lower values than average with only 71 SNPs, 7 contigs, and a  
 294 physical span of 26.92 Mbp (Table 2). Remarkably, the genetic map from von Maydell et al. (2026b)

295 contained a severely different order of contigs in LG08 and merely a fragmented part from LG10  
296 (Figure 3).

297

### 298 **2.1.3 Phenotypic data**

299 The raw phenotypic data for P1 and P2 are placed in Table S1 A and B, respectively.

#### 300 **2.1.3.1 Flowering time**

301 The annual parent ‘L32’ exhibited a highly synchronized onset of flowering, compressing its  
302 phenotypic range to between 87 and 116 days after sowing (DAS, Figure 4). It reached a distinct single  
303 peak at 92.5 DAS, with 100% of the population flowering within first year (92 out of 92 plants).

304 The F<sub>2</sub> population P1 showed a distinct bimodal distribution stretching over a wide phenotypic range  
305 (79 to 224 DAS), with the vast majority of individuals segregating as annuals (97.91%, 281 out of 287  
306 plants, Figure 4). Kernel density estimation identified a major flowering peak at 103.0 DAS (10.5 DAS  
307 later than the annual parent), followed by a secondary, smaller peak at 118.6 DAS. Notably, the  
308 corresponding biennial parent ‘Rekord’ exhibited a remarkably high annual flowering fraction of  
309 25.00% (24 out of 96 plants) despite its classified biennial habit. However, the first flowering plant  
310 occurred at 130 DAS. Merely, 19 of 287 plants of P1 fall within the range of annual flowering plants  
311 of ‘Rekord’.

312 The F<sub>2</sub> population P2 displayed a strongly left-skewed distribution, with an almost complete shift  
313 towards the annual habit (99.65% annual fraction, 283 out of 284 plants, Figure 4). The earliest P2  
314 individuals initiated flowering at 87 DAS, matching the earliest annual parents, and reached a single  
315 major peak at 98.9 DAS (6.4 DAS later than the annual parent, but earlier than P1). In contrast to  
316 ‘Rekord’, the biennial parent ‘CARUM 63’ maintained its biennial character far more strictly, exhibiting  
317 a minimal annual flowering fraction of just 4.88% (4 out of 82 plants). These plants flowered within a

late ten-day window from 131 to 141 DAS, overlapping with the late-flowering tail of the P2 population (6 plants, extending up to 158 DAS).

In general, plants of ‘CARUM 63’ showed a typical biennial morphotype with remaining in a leaf rosette stage, developing a low number of leaves, which lying flat on the soil (Figure 1 D). 34 out of 82 plants survived the winter and flowered in the second year. In contrast, plants of ‘Rekord’ soon developed a bush of many erected leaves (Figure 1 E) and all flowering plants (25%) merely developed weak inflorescences, which scarcely produced any viable seeds. All plants of ‘Rekord’ died during winter. Notably, in both biennial parents a high number of plants died not because of frosts, but was killed by mice. In general, F<sub>2</sub> plants were visually much more vigorous regarding overall number and size of leaves and inflorescences than the inbred line ‘L32’ (Figure 1). Interestingly, we observed that many plants of P1 were strongly affected by diseases during ripening stage (in particular powdery mildew) independent from flowering time, whereas plants of P2 were usually healthy and produced healthy seeds.

#### **2.1.3.2 Flower color**

In general, the determination of flower color was difficult due to strong bleaching of the rose coloration under ESC with flowering predominately in summer with high radiation in 2025. We merely detected purely white flowering plants (212, 73.9%) and plants with a slight rose coloration (75, 26.1%). For mapping white flowering plants were noted as “0” and rose flowering plants as “2”.

#### **2.1.4 QTL mapping**

Detailed results from mapping (IM, MQM and rMQM) with MapQTL5 including final cofactor settings are placed in Table S1 E. LOD profiles from the rMQM are presented in Figure 5. Kruskal-Wallis and BLINK results are provided in Table S1 F-H. In general, MQM, IM, BLINK and Kruskal-

Wallis showed highly congruent results. In the following sections we will focus on the description of the MQM results. All QTL statistics with additional information are presented in Table S1 H and corresponding phenotypic data per QTL are shown in Figure S2. Table 3 shows a selection of the most interesting detected QTLs. Each QTL identified in this study has a unique identifier across studies and populations, because currently evidence is insufficient to clearly determine identical loci across studies and populations. However, we will discuss the likelihood of identical loci below.

#### **2.1.4.1 Flowering time**

##### **Major QTLs on LG02/LG06 of P1**

Most conspicuously, we detected two major QTLs for BOF in P1 on LG02 (peak locus contig\_127\_8357699 at 40.3 cM) and on LG06 (peak locus contig\_2317\_2291222 at 41.8 cM). Cc-FT02.1 mapped with a LOD score of 10.32 and explained 12.1% of the phenotypic variance with an additive effect of -5.6 days in relation to the annual allele, whereas Cc-FT06.1 exhibited an even stronger effect with a LOD score of 15.22, explaining 18.6% of the variance and an additive effect of -6.6 days (Table 3, Figure 5). Notably, these effects and explained variances were reached without inclusion of non-flowering or very late (>160 DAS) flowering plants. The confidence interval of Cc-FT02.2 slightly overlaps at its boundary with the previously identified QTL Cc-FT02.1 (von Maydell et al. 2026b), but shows a distinct peak position (Figure 7). On the consensus map, the 1-LOD intervals of the QTLs on LG02 and LG06 encompassed 12.4 Mb (457/572 annotated genes/transcripts) and approximately 6 Mb (258/335 annotated genes/transcripts), respectively. In close proximity (<0.2 Mb) of both peak markers, Zinc finger RING/FYVE/PHD-type coding genes were detected. A list of all genes and transcripts within 1-LOD intervals is provided in Table S1 I.

To further scrutinize the two major QTLs, we investigated the phenotypic distribution for the four homozygous combinations. Genotypes with a double homozygous annual (aa) combination on average

flowered earliest about 98 DAS, whereas genotypes with a double homozygous biennial (bb) combination flowered latest about 128 DAS (Figure 6). In an ANOVA, among ‘normally’ flowering homozygous individuals (80–160 days), both loci displayed strong, highly significant additive main effects ( $p < 0.001$ ), jointly explaining 56.01% of the phenotypic variance without significant linear interaction ( $p = 0.208$ ). However, a distinct epistatic threshold effect emerged when examining non-flowering (NF) and very late flowering (VLF) phenotypes (>160 days): these extreme phenotypes occurred exclusively in double homozygous biennial (‘bb’) individuals (6 out of 19 plants; 31.6%), whereas all aa, ab, and ba genotypes flowered ‘normally’ (100%). This enrichment of non-flowering phenotypes in the bb class was highly significant (Fisher's exact test,  $p = <0.001$ ).

#### **Potentially common QTL on LG10 of P1 and P2**

On LG10, a consistent QTL (Cc-FT10.2) was identified in both populations, with peak marker contig\_1019\_2561338 in P1 (22.8 cM, LOD = 4.12,  $R^2 = 8.9\%$ , Add = -4.2) and contig\_875\_3986579 in P2 (29.4 cM, LOD = 7.56,  $R^2 = 9.8\%$ , Add = -3.6). In the consensus physical map, Contig 875 directly precedes Contig 1019, placing the peak markers ~2.72 Mb apart plus an assembly gap of unknown size. Due to this tight consensus alignment and matching additive directions both signals may anchor to the identical locus. However, it must be noted that there is a large gap without markers in P1, which was even more severe in P2 2023. Since a large part of >59 Mb including contigs 875 and 1019 is completely missing in von Maydell et al. (2026b) a comparison with the QTL Cc-FT10.1 (P2 2023) is scarcely possible. The shown 1-LOD interval (Figure 7) can likely be limited by the fragmentation of LG10 so that 1-LOD intervals can potentially overlap between all three observations.

## Other QTLs

Four additional minor QTLs associated with flowering time exceeded genome-wide significance thresholds (LOD 4.28–5.13), accounting for 5.4%–7.2% of phenotypic variance. A single QTL was mapped in P1 (LG05 at 13.9 cM), while three QTLs were identified in P2 (LG03 at 1.1 cM, LG05 at 66.1 cM, and LG09 at 27.7 cM). Additive effects were negative across LG03 and LG05, whereas the LG09 locus displayed a positive additive effect in relation to the annual allele (Table 3). Potentially, on LG05 several minor QTLs might be, because it was difficult to determine a peak position in MQM, also resulting in extreme broad 1-LOD intervals (Figure 7).

## Missing vernalization-responsive QTLs (on LG02, LG04, LG05, LG08) in P2 under ESC?

None of the four minor to medium QTLs for flowering time potentially associated with vernalization response detected under LSC (von Maydell et al. 2026b) was explicitly re-detected under ESC of this study. In case of the QTLs on LG05 this might be not fully clear due to the large overlapping 1-LOD interval (Figure 7). Interestingly, the medium QTL detected on LG02 for P2 in 2023 resided on the neighboring contig of the strong QTL detected in P1 (Figure 7).

We tested whether the cumulative additive effect of those four genetic regions (von Maydell et al. 2026b) might still be detectable by analyzing the phenotypic distribution of the cumulative number of biennial ‘b’ alleles in P2. For this purpose we used the peak markers from von Maydell et al. (2026b) if applicable or a marker in close physical proximity to the peak marker. Linear regression analysis revealed a highly significant cumulative effect across these loci ( $N = 260$ ,  $p = 7.01 \times 10^{-6}$ ). To account for potential selection bias and genomic background noise, we performed 1,000 empirical permutations sampling four random loci across the genome. The cumulative QTL effect remained statistically significant ( $p = 0.025$ ). On average, each additional biennial ‘b’ allele delayed flowering by 1.85 days ( $R^2 = 7.54\%$ ). This indicated a significant dosage-dependent polygenic influence of the four QTL

regions on flowering time, even though the individual loci did not reach the significance threshold for QTL detection under ESC.

#### **Tests for segregation ratio of flowering type**

To test a two-locus dominant epistatic model (expected 15:1 ratio) for flowering type in P1 while accounting for incomplete penetrance, the expected phenotypic frequencies were adjusted for the observed 25% annual-flowering rate in the biennial parent under ESC of this study. Assuming 25% of genetically biennial individuals express an annual phenotype, the adjusted expected phenotypic ratio shifts to 61:3 (95.31% annual: 4.69% biennial). In an  $F_2$  population of  $N = 287$  plants (observed: 281 annual, 6 biennial), a chi-square goodness-of-fit test revealed a slight but significant deviation from this model (Chi-square = 4.33,  $df = 1$ ,  $p = 0.037$ ). For comparison, the simple two-locus model (15:1 ratio) without penetrance adjustment showed a much stronger deviation from the observed data (Chi-square = 8.47,  $df = 1$ ,  $p = 0.0036$ ).

Further, we formally tested whether segregation in P2 was consistent with the four-locus model proposed by Maydell et al. (2026). Under this model, the biennial phenotype is expected only in plants homozygous for the biennial allele at all four loci. Accounting for the incomplete penetrance of the biennial parent under early sowing (4.88% annual flowering), the expected frequency of biennial  $F_2$  plants was 0.372% (1.06 of 284 plants). Because the expected number of biennial individuals under the four-locus model was very low, segregation was evaluated using an exact binomial test rather than a chi-square test. One biennial  $F_2$  plant was observed, which was fully consistent with this expectation (exact binomial test,  $P = 1.00$ ). However, notably this individual did not harbor all eight alleles from the four-locus model (Figure 8s).

437 **Table 3:** Characteristics of selected significant QTLs associated with flower color (COL) and flowering time (BOF). Corresponding results  
438 from the IM and BLINK analyses are shown in parentheses () and square brackets [], respectively. The 1-LOD confidence intervals were  
439 obtained from the restricted MQM model. Expl. Var. denotes the proportion of phenotypic variance explained (%), while Add. and Dom.  
440 represent the additive and dominance effects of the annual allele, respectively. Significant values, as determined by permutation tests for  
441 MQM/IM or by false discovery rate (FDR) correction for BLINK, are highlighted in bold.

| QTL_ID     | Population | Trait | LG   | Position<br>cM | 1-LOD<br>cM   | LOD<br>MQM<br>(IM)<br>[BLINK]                    | Expl.<br>Var.<br>% | Add<br>MQM<br>[BLINK] | Dom   | Peak Locus          |
|------------|------------|-------|------|----------------|---------------|--------------------------------------------------|--------------------|-----------------------|-------|---------------------|
| Cc-FT02.2  | P1         | BOF   | LG02 | 40.3           | 39.2-<br>43.2 | <b>10.32</b><br><b>(6.54)</b><br><b>[21.97]</b>  | 12.1<br>[8.1]      | -5.6                  | -0.1  | contig_127_8357699  |
| Cc-FT06.1  | P1         | BOF   | LG06 | 41.8           | 39.8-<br>43.1 | <b>15.22</b><br><b>(10.84)</b><br><b>[14.15]</b> | 18.6<br>[14.9]     | -6.6                  | -3.0  | contig_2317_2291222 |
| Cc-FT10.2  | P1         | BOF   | LG10 | 22.8           | 9.8-<br>44.9  | <b>4.12</b><br>(2.65)<br><b>[4.95]</b>           | 8.9<br>[2.4]       | -4.2                  | -3.2  | contig_1019_2561338 |
| Cc-FT10.3  | P2         | BOF   | LG10 | 29.4           | 26.0-<br>32.4 | <b>7.56</b><br><b>(4.45)</b><br><b>[9.71]</b>    | 9.8<br>[5.4]       | -3.6                  | -0.2  | contig_875_3986579  |
| Cc-COL01.2 | P2         | COL   | LG01 | 53.5           | 53.1-<br>55.6 | <b>8.34</b><br><b>(6.33)</b><br><b>[5.00]</b>    | 11.9<br>[4.71]     | 0.42                  | -0.21 | contig_83_28558179  |
| Cc-COL10.2 | P2         | COL   | LG10 | 64.3           | 62.8-<br>73.4 | <b>6.16</b><br><b>(4.13)</b><br><b>[8.14]</b>    | 8.6<br>[6.34]      | -0.36                 | 0.025 | contig_1950_3482482 |

#### **2.1.4.2 Flower color**

QTL mapping revealed two significant loci associated with flower color. One QTL was located on LG01 at 53.53 cM and showed an overlapping 1-LOD interval with the corresponding locus identified by von Maydell et al. (2026b) (Figure 7). This locus showed the strongest association (LOD = 8.34) and explained 11.9% of the phenotypic variance (Table 3). It exhibited a positive additive effect in relation to the annual allele. The other QTL was located on LG10 at 64.30 cM with the peak marker in close physical proximity (<0.5 Mb) to the peak marker of the corresponding locus in von Maydell et al. (2026b) (Figure 7). This locus was detected with a LOD score of 6.16, explaining 8.6% of the phenotypic variance (Table 3). In contrast to the QTL on LG01, it displayed a negative additive effect in relation to the annual allele. Comparing the phenotypes for homozygous combinations of both loci, the combination 'ab' (annual LG01, biennial LG10) had the highest proportion of rose flowering plants (72.0%), whereas the reverse combination had no rose flowering plants (Figure 9).

## 2.2 Phenotypic analysis of 15 additional F<sub>2</sub> populations

Across the F<sub>2</sub> populations from 1999, 95.2% (maternal biennial cultivar ‘Volhouden’) to 98.8% (maternal biennial cultivars ‘Plewiski’, ‘Kami’, and ‘Prochan’) of plants (N = 84) showed an annual flowering behavior (details Table S3). For F<sub>2</sub> populations analyzed in 2000, the proportion of annual flowering plants was evenly high, but exact quantification results are not available. On a quantitative scale F<sub>2</sub> populations (N = 42), flowered on average 9.29 and 13.78 days later than the annual parent in 1999 and 2000, respectively (Figure 10). For each comparison between the annual parent and an F<sub>2</sub> population the difference was significant according to the Tukey HSD test accounting for multiple testing. The strongest difference showed the F<sub>2</sub> from ‘CARUM 16’ with a later flowering of 21.10 days in 2000 (Figure 10). In general, the phenotypic range per genotype was broader in 2000. Notably, the F<sub>2</sub> populations of ‘Niederdeutscher’ and ‘Konczewicki’ were tested in both years and showed larger average differences from the annual parent in 2000 indicating a year effect. Raw data and detailed results are deposited in Table S3.

## 2.3 Diagnostic markers for selected flanking and peak markers

We tested PACE marker designs for several SNPs in proximity of flanking and peak marker positions of QTLs on LG02 and LG06 detected for P1. We selected the SNP and corresponding PACE marker showing best discrimination (Figure 11) between all three genotypic constitutions of the SNP per position for this study. Therefore, the used loci for PACE marker design can deviate from the exact positions of the peak/flanks as given in Table 3. The selected markers representing the peak are at maximum 2.0 cM in distance to the peak of the QTL and the selected flanking markers are approximately at the margins of the 1.5 LOD. All designed markers showed a high conversion rate with GBS data ranging from 96.30 to 98.77% (Table 4).

Table 4: Comparison of GBS and PACE results for selected flanking and peak markers.

| Marker | Sample Size | PACE Failed | GBS Failed | Valid Comparisons | Matches | Conversion Rate |
|--------|-------------|-------------|------------|-------------------|---------|-----------------|
| M175   | 88          | 2           | 5          | 81                | 80      | 98.77           |
| M178   | 88          | 2           | 5          | 81                | 80      | 98.77           |
| M183   | 88          | 1           | 5          | 82                | 81      | 98.78           |
| M192   | 88          | 2           | 5          | 81                | 80      | 98.77           |
| M198   | 88          | 2           | 5          | 81                | 78      | 96.30           |
| M204   | 88          | 3           | 5          | 80                | 78      | 97.50           |

## 3 Discussion

### 3.1 Map improvement, cross-validation & integration of physical and genetic information

#### 3.1.1 General considerations regarding mapping strategy

We consider that classical F<sub>2</sub> mapping with filtering for homozygous contrasts between parents is the most effective mapping strategy for caraway compared to realistic alternative strategies such as pseudo-testcross or BC1 (developing of RILs or DH is not manageable for an orphan crop like caraway). First, the biological and experimental setup of caraway highly favors an F<sub>2</sub> approach. Producing an F<sub>1</sub> generation is straightforward, and a single F<sub>1</sub> individual can easily be selfed to yield a large mapping population of potentially more than 2000 F<sub>2</sub> plants, providing excellent genetic resolution also for subsequent fine mapping without complex crossing schemes. Second, while integrating markers that are heterozygous in the parents might theoretically increase marker density, our data demonstrates that this is practically ineffective. Although the maternal lines showed considerable heterozygosity (17.4% to 31.8%), utilizing SNPs that were heterozygous in the maternal genotypes and showed a 1:2:1 segregation-like pattern in the F<sub>2</sub> population merely marginally increased the number of unique markers after redundancy reduction and failed to improve map construction. Furthermore, we observed unexpected levels of heterozygosity in the paternal line ‘L32’ (>9%), which actually represent a fifth inbreeding generation. In such an advanced inbreeding stage, high heterozygosity is highly biologically implausible. This strongly indicates a high risk of pseudo-heterozygosity typical for GBS (McKinney et al. 2017). Attempting to integrate these markers could introduce severe noise. Consequently, focusing strictly on the high-confidence, classical F<sub>2</sub> configuration remains the most robust and reliable strategy for high-quality linkage map assembly in caraway.

### 3.1.2 Genetic maps & consensus physical map

In the previous study (von Maydell et al. 2026b), a substantial number of SNPs was lost during filtering due to missing values in both the parents and the F<sub>2</sub> population. By substantially increasing the sequencing depth, particularly for the parental lines, we successfully mitigated the impact of missing-value filtering. Consequently, substantially more high-quality SNPs were retained throughout the filtering pipeline prior to linkage mapping (1,245 for P1 and 1,560 for P2 versus 731 SNPs in von Maydell et al. (2026b)). Additionally, the overall map length expanded from the previous map of 662.9 cM to 780.1 cM (P2), which is consistent with the larger F<sub>2</sub> population size (288 vs. 187 plants, notably originating from the same cross) allowing for a higher resolution of recombination events. Although the final number of markers integrated into the thinned linkage maps for QTL mapping (392 and 337 SNPs for P1 and P2, respectively) remains relatively modest, it still marks an improvement over the 259 SNPs utilized in von Maydell et al. (2026b). Crucially, for subsequent fine-mapping efforts, the larger pool of markers available in the full genetic maps provides a valuable resource for marker saturation within specific regions of interest.

However, the resolution of our genetic maps remains constrained by larger gaps. Whereas we could not solve this issue for mapping, the constructed consensus physical map filled some of these gaps differentiating between populations. It still appears that approximately 0.52 Mbp (28%) of the draft assembly remains unplaced in the consensus physical map and that the current ordering and orientation are prone to errors in regions with low marker density or contradictory signals between populations. In particular, the small physical length of LG09 indicates a high degree of missing physical information. The published karyogram of caraway does not support the existence of such a fragmented chromosome (Pank 2012). The restriction-enzyme-based genome reduction inherent to GBS can result in uneven genomic representation potentially contributing to the remaining marker-deficient intervals (Davey et al. 2013). In addition, segregation distortion can affect linkage-map construction and marker ordering, and the exclusion of distorted markers may reduce genome coverage (Zhu et al. 2007; Zuo et al. 2019).

However, less stringent filtering resulted in severe computational problems during linkage-map construction in JoinMap in our dataset. Therefore, stringent filtering was retained as a practical measure to obtain a stable and interpretable linkage map for caraway in absence of a chromosome-resolved assembly.

Despite these limitations, the consensus physical framework currently provides the most practical coordinate system for systematically comparing overlapping QTL intervals across independent mapping populations. Furthermore, this pseudo-chromosomal map may provide an empirical resource to validate a future chromosome-level genome assembly, e.g. generated based on Hi-C (Belton et al. 2012; Chen et al. 2020; Zhou et al. 2020).

### **3.2 Environmental impact on QTL detection in wild biennial background (P2)**

By observing 99.7% annual flowering in P2 under ESC, we confirmed that the “true” phenotype was masked in P2 under LSC with merely 36% annual flowering F<sub>2</sub> plants (von Maydell et al. 2026b). The biennial parent of P2 (‘CARUM 63’) merely showed a low proportion of annual flowering individuals (4.88%), indicating that vernalization requirement was largely functional and was not fully overcome under ESC in the biennial genetic background. Therefore, we cannot explain the high proportion of annual flowering F<sub>2</sub> plants solely by environmental effects, such as a complete fulfillment of vernalization requirement by mild temperatures in the first outdoor period in April or the complete overcoming of vernalization requirement by heat or day-length during summer. Instead, the higher proportion of annual flowering under early sowing is likely driven by advanced plant development compared to late sowing, interacting with the genetic background. Thus, we need to explain the observations from ESC and LSC by an adequate genetic model.

The observed phenotypic ratio (one biennial plant among 284 F<sub>2</sub> plants) is consistent with a four-locus model in which homozygosity for the biennial allele at all four loci is required to re-establish the biennial phenotype. For four independently segregating loci, such a model would predict a frequency of 1/256 for the fully biennial genotype, which is close to the observed frequency of 1/284. von Maydell et al. (2026b) identified four minor to medium significant loci associated with flowering type on LG02, LG03, LG05 and LG08 jointly explaining 32.2% of the phenotypic variance, which would fit to a four-loci (+ x, i.e., additional minor loci) model (The fifth QTL on LG10 is discussed below). Vernalization potentially works via an allele-dosage dependent threshold model causing predominantly a qualitative difference between annuals and biennials. Thus, vernalization responsive loci do not necessarily need to exert a quantitative effect on flowering time among annual flowering F<sub>2</sub> plants. While the loci detected under LSC contributed in different combinations to the extreme difference between annuals and biennials (Additive effects of 31 to 47 DAS per locus, for uncorrected BOF), this was not possible under ESC in the absence of a sufficient number of biennial F<sub>2</sub> plants. This can explain why these four loci were not individually re-detected under ESC despite the considerably increased sample size (187 to 284). Yet, the detected significant cumulative effect indicates that a low additive effect remained.

This is also supported by the fact that the F<sub>2</sub> population shows later flowering than the annual parent. this difference could also be explained by loci independent of vernalization response. Interestingly, a fifth QTL on LG10 detected in von Maydell et al. (2026b), might represent a re-detection of the previously reported QTL (further discussed below) under ESC in this study. This provides further support for the hypothesis that this QTL might represent a more general flowering-time regulator, which was assumed because this QTL had an effect within annual F<sub>2</sub> plants before vernalization and within biennial F<sub>2</sub> plants after vernalization (von Maydell et al. 2026b).

We do not want to claim that the significant detection of four QTLs by von Maydell et al. (2026b) under LSC, together with the evidence for a weak cumulative effect of these four QTLs under ESC constitutes a validation of these QTLs. Nor can we currently completely explain vernalization requirement in this wild biennial background. For example, it remains an interesting question, whether the dosage of biennial alleles affects the duration of low-temperature exposure required for vernalization, so that some F<sub>2</sub> plants with a low number of biennial alleles might be vernalized in the early outdoor period by low night temperatures. This would be similar to annual carrot, which still needs a cold stimulus in early spring, while this would not be sufficient in biennial carrot (Wohlfeiler et al. 2022). Intensive analyses will be necessary to validate these QTLs and to mechanistically understand vernalization, including its interaction with other pathways of flowering time. So far, the data of this study in combination with von Maydell et al. (2026b) support a polygenic four-loci (+x) model for vernalization requirement in our wild-type biennial accession. On a quantitative scale, these loci may exert cumulative additive effects, while their phenotypic expression depends on environmental conditions, e.g. resulting from sowing date.

### **3.3 Comparison of QTL detection with cultivated biennial background (P1)**

Already the phenotypic data for P1 and the biennial parent ‘Rekord’ suggested some differences in the genetic architecture of flowering regulation. 1) The high number of annual flowering individuals within ‘Rekord’ (25%) towards the end of the year indicated a lower penetrance of vernalization requirement. Thus, other flowering pathways associated with temperature, day length or general growth status likely surpassed vernalization requirement. However, the unusual bushy morphology and weak inflorescences indicated that the fine-tuning of flowering regulation was rather impaired in this cultivar under ESC. 2) The higher proportion of non-flowering F<sub>2</sub> plants despite the incomplete penetrance of biennial flowering in the parent indicated that a lower number of loci might explain vernalization

requirement in P1/‘Rekord’. 3) The later average flowering of P1 compared to P2 including a second peak of late flowering F<sub>2</sub> plants indicated a higher effect size of biennial alleles in the F<sub>2</sub>. Indeed, the QTL mapping identified two major QTLs on LG02 and LG06. The significant association of non-flowering F<sub>2</sub> plants with the group of double-homozygous biennial F<sub>2</sub> plants suggests an association of these loci with vernalization requirement. However, since non-flowering plants were not included in QTL mapping, the detection of these loci likely reflects their strong effects on flowering time as a quantitative trait within annual flowering F<sub>2</sub> plants. Even considering the incomplete penetrance of biennial flowering in the parent, a pure two-locus model cannot yet explain the observed segregation so that one or a few additional minor loci may contribute to vernalization response. The detected minor QTL on LG05 and LG10 do not explicitly fulfill this role since the biennial allele seems not to be strongly associated with non-flowering individuals, but given the low number of non-flowering individuals this can be a random finding.

Notably, the QTL interval of P1 on LG02 (*Cc-FT02.2*) slightly overlaps with the QTL interval of P2 detected under LSC (*Cc-FT02.1*). It cannot be excluded but it seems unlikely that both QTLs refer to the same locus, also considering that this locus was not detected in P2 under ESC. Interestingly, preliminary analyses also hint at a strong flowering-time-independent locus associated with thousand-grain weight in close proximity to *Cc-FT02.1*, indicating that this region might be very interesting for breeding of developmental and agronomic traits.

### 3.4 Diversity of flowering time regulation

While our analysis of a cultivated biennial background clearly demonstrated a different genetic architecture compared to P2/‘Carum 63’, it still does not support the one-major-locus interpretation proposed by Nemeth and Pluhar (1996), who suggested one major locus associated with flowering type

and additional minor loci associated with flowering time on a quantitative scale. At first glance this might be explained by the use of different cultivars ('Maud' in Nemeth and Pluhar (1996) vs. 'Rekord' in our study), but our data for an F<sub>2</sub> based on 'Maud' showed 97.6% (Table S3) annual flowering under ESC. Therefore, we hypothesize that the apparent simple one-locus pattern reported by Nemeth and Pluhar (1996) may have resulted from environmental effects associated with specific environmental conditions. Since we no longer have access to the cultivar 'Maud' we cannot finally check the genetic regulation in this cultivar. In addition, the study from Nemeth and Pluhar (1996) marked as 'preliminary analyses' does not provide any detailed data or method descriptions.

Although the unpublished data from 1999 and 2000 are rather old, we considered these data very valuable for the current study, also taking into account that such a number of crosses is scarcely manageable based on current resources for an orphan crop like caraway. These data add additional years and a considerable degree of diversity to this study with a focus on cultivated material. With the inclusion of four gene bank accessions (CARUM 10, 13, 16 and 63, the last in the QTL mapping population P2), which can be classified as land race or wild-type, the investigated material spans a broad range of the phylogenetic diversity of biennial caraway germplasm as presented in von Maydell et al. (2021). The usage of the same annual parent allows differences between F<sub>2</sub> populations to be attributed to differences between biennial parents.

Across all F<sub>2</sub> populations and years we can state that the proportion of annual flowering plants is always very high (>95%) under ESC. Thus, we may hypothesize that vernalization requirement in caraway is unlikely to be explained by simple one- or two-locus Mendelian models alone. We further observed that F<sub>2</sub> populations on average showed a significantly later beginning of flowering than the annual parents. This indicates that loci from the biennial parent generally delay flowering time on a quantitative scale instead of operating merely on a qualitative scale (flowering vs. non-flowering).

657  
658

### 659 **3.5 Re-detection of QTLs for flower color (in P2)**

660 Based on observations in the greenhouse von Maydell et al. (2026b) reported bleaching of flower color  
661 under high radiation. We can confirm that the intensity of rose coloration is strongly environmentally  
662 dependent as it was strongly bleached during flowering in summer. Despite the difficult phenotyping  
663 under ESC we re-detected both loci on LG01 and LG10 from von Maydell et al. (2026b). However,  
664 possibly due to phenotyping errors the explained phenotypic variance was lower and the QTL intervals  
665 were not improved, although sample size was considerably increased. Fine-mapping remains  
666 necessary. Notably, the peak of the QTL on LG10 in this study was even closer ( $<0.2$  Mb compared to  
667  $>0.6$  Mb in von Maydell et al. (2026b)) to the annotated cluster of MYB transcription factors on  
668 contig\_1950, which may contain potential candidate genes. In carrot these genes are involved in the  
669 regulation of purple pigmentation (Bannoud et al. 2019; Xu et al. 2019). Interestingly, under ESC, the  
670 QTL on LG01 had a stronger effect than the QTL on LG10, which was the dominant locus under LSC  
671 (von Maydell et al. 2026b). Based on the hypothesis that the annual allele of the QTL on LG01 has a  
672 stabilizing effect on coloration, it might be that this stabilization was more important for detecting rose  
673 coloration under bleaching conditions.

674

### 675 **3.6 Diagnostic markers for fine mapping**

676 Due to the high explained phenotypic variance and small confidence intervals of the QTLs on LG02  
677 and LG06 detected in P1, we consider that validation and fine-mapping of these QTLs as well as their  
678 final application in breeding programs might be achieved relatively quickly. Considering the high  
679 number of genes (457 and 258 for LG02 and LG06, respectively) within the 1-LOD intervals fine-  
680 mapping remains absolutely necessary. Nevertheless, Zinc finger RING/FYVE/PHD-type coding  
681 genes in close proximity to peak markers of both QTLs (Table S1I) may be interesting candidates,

considering that such genes were already detected at the peak marker of the strongest QTL potentially associated with vernalization response in P2 under LSC (von Maydell et al. 2026b). However, these genes, which can be associated with vernalization genes such as *VRN2* or *VIN3* (Gendall et al. 2001; Labandera et al. 2021; Sung and Amasino 2004), may not be exclusively targeted prior to additional evidence.

QTLs from P2 could also be very interesting to fine tune flowering time regulation, but likely this can only be achieved in a long-term process due to their polygenic and environmentally dependent background. Admittedly, we cannot be sure that QTLs from P1 are stable in different environments, but at least this can be efficiently checked by diagnostic markers in F<sub>3</sub> generation.

Based on the developed and validated PACE markers it will be possible to select F<sub>2</sub> plants homozygous for defined QTL intervals, but otherwise variable parental background. After propagation by self-fertilization the resulting F<sub>3</sub> families could be useful to analyze the effect of the QTLs in different combinations under varying conditions (e.g. sowing dates). This could be more effective than additional F<sub>2</sub> screenings, which are difficult to effort in an orphan crop. Moreover, these markers can be used for selecting recombinant F<sub>2</sub> plants for fine-mapping of QTLs from P1 in their segregating F<sub>3</sub> families. For this purpose, it is very advantageous that one F<sub>1</sub> plant can produce >2000 F<sub>2</sub> seeds.

Despite using non-normalized DNA concentrations for PACE assays, we achieved a high conversion rate between GBS and PACE results. Therefore, these markers can be used in quick and effective screening protocols in laboratories not equipped with automated sample preparation as usual for research groups or breeding companies working on orphan crops like caraway. We do not claim that these markers can be used as diagnostic markers in other crosses of biennial and annual caraway or as diagnostic markers for the flowering type in general. Preliminary analyses using diverse annual and biennial caraway germplasm from von Maydell et al. (2021) indicate that all PACE markers are

working in other genetic backgrounds, but that some markers show rather random segregation at least within the biennial gene pool. Thus, we recommend carefully testing each single plant used for crosses.

### **3.7 Limitations of the study, future prospects & conclusions**

By choosing two extremely far-related biennial parents, we focused on and succeeded in revealing potential genetic diversity in flowering time regulation in caraway. As a disadvantage, both populations scarcely cross-validate the detected QTLs so that our findings still need further validation. In particular in cases of overlapping QTL intervals, it needs to be determined whether these refer to identical or different loci. Moreover, the current knowledge is restricted to two biennial genotypes and in case of the cultivar ‘Rekord’ to ESC. Whether flowering time regulation in other biennial cultivars resembles the genetic architecture of the investigated ‘Rekord’ and whether the detected QTLs operate consistently under medium to late sowing conditions needs to be tested. Although the QTL intervals for P1 were relatively small, the number of genes potentially associated with flowering time (see Table S1 I) was still too high to narrow down to a small number of candidate genes. We consider that our study prepares the stage for several fine-mapping and more mechanistic studies including transcriptomics, which will be necessary to finally understand the genetic regulation of flowering in caraway. The developed PACE markers provide a starting tool for these future studies.

While we observe an increasing number of genome assemblies in Apiaceae as basic genetic resources (Iorizzo et al. 2016; Liu et al. 2025; Liu et al. 2024; Palumbo et al. 2018; Song et al. 2021; Song et al. 2019), QTL studies are still rare in this family despite their importance for linking genetic variation to traits relevant for classical breeding. So far, mapping studies on flowering time have focused on carrot, revealing monogenic to digenic control of vernalization (Alessandro et al. 2013; Wohlfeiler et al. 2019), with two duplicated SOC1 genes recently identified within one major QTL (Wohlfeiler et al. 2026). Therefore, this comprehensive study on flowering time in caraway sheds new light on the potential diversity and complexity of flowering time regulation including vernalization requirement in

731     Apiaceae, demonstrating a strong environmental impact associated with sowing date and a contrasting  
732     genetic architecture between a wild and a cultivated biennial background.

733

## 734 **Declarations**

### 735 **Ethics approval and consent to participate**

736 Not applicable

737

### 738 **Consent for publication**

739 Not applicable

740

### 741 **Availability of data and materials**

742

743 Phenotypic data, SNP matrices, genetic maps, detailed results from QTL mapping and PACE marker  
744 designs are added to the supplements. Information on non-used PACE marker designs will be shared  
745 on request. The GBS data (FASTQ files) generated for this study have been deposited in the European  
746 Nucleotide Archive (ENA) under the project accession numbers PRJEB108912 (P1, Parents) and  
747 PRJEB108913 (P2, Parents).

748

### 749 **Competing interests**

750

751 The authors declare that they have no competing interests.

### 752 **Funding**

753

754 Basic works of this study were funded by the "Fachagentur Nachwachsende Rohstoffe" (FNR, Funding  
755 ID: 2220NR103) on behalf of the Federal Ministry of Agriculture, Food and Regional Identity  
756 (BMLEH). A fund for open access publication was granted by the Julius Kühn Institute (JKI).

757

### 758 **Author contributions**

759 DvM wrote the original draft, designed the trial, conducted statistical analysis and genetic mapping,  
760 participated in funding acquisition and project administration.  
761 BK organized and conducted DNA extraction and several PACE assays.  
762 AP organized, conducted and validated several PACE assays  
763 FSL supported basic bioinformatic works and trouble-shooting.  
764 JZ contributed to field-data acquisition.  
765 MJ contributed to funding acquisition and project administration.  
766 FM supervised the project and contributed to funding acquisition and project administration.  
767 All authors read and approved the final version of the manuscript

## 768 **Acknowledgments**

769  
770 We thank Brian Boyle and Sharen Roland and their teams in Canada for excellent library preparation  
771 and sequencing. We thank Saba Gharaati, Thomas Heck and Lisa Lepczynski for contributing to high  
772 quality DNA extraction for GBS. We thank the team of Dr. Friedrich Pank for thorough investigation  
773 and documentation of F<sub>2</sub> data in 1999 and 2000. We thank Jenny Knibbiche for conducting the crosses  
774 for QTL mapping. Not at least we thank the team VQ, Bettina Albrecht, Katrin Wolf and Kay  
775 Fischbach contributing to the diverse field works.

776

777

778

779

780

## References

- Alessandro MS, Galmarini CR, Iorizzo M, Simon PW (2013) 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>
- Bannoud F, Ellison S, Paolinelli M, Horejsi T, Senalik D, Fanzone M, Iorizzo M, Simon PW, Cavagnaro PF (2019) 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>
- Bastien M, Légaré G, Pomerleau S, St-Cyr J, Boyle B, Belzile F (2017) Genotyping-by-Sequencing in Potato. 283–296 [https://doi.org/10.1007/978-3-319-66135-3\\_15](https://doi.org/10.1007/978-3-319-66135-3_15)
- Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J (2012) 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>
- Chen H, Zeng Y, Yang Y, Huang L, Tang B, Zhang H, Hao F, Liu W, Li Y, Liu Y, Zhang X, Zhang R, Zhang Y, Li Y, Wang K, He H, Wang Z, Fan G, Yang H, Bao A, Shang Z, Chen J, Wang W, Qiu Q (2020) Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nature Communications* 11:2494 <https://doi.org/10.1038/s41467-020-16338-x>
- Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML (2013) Special features of RAD Sequencing data: implications for genotyping. *Mol Ecol* 22:3151–3164 <https://doi.org/10.1111/mec.12084>
- Dupuis J, Siegmund D (1999) Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* 151:373–386 <https://doi.org/10.1093/genetics/151.1.373>
- Gendall AR, Levy YY, Wilson A, Dean C (2001) The VERNALIZATION 2 Gene Mediates the Epigenetic Regulation of Vernalization in Arabidopsis. *Cell* 107:525–535 [https://doi.org/https://doi.org/10.1016/S0092-8674\(01\)00573-6](https://doi.org/https://doi.org/10.1016/S0092-8674(01)00573-6)
- Huang M, Liu XL, Zhou Y, Summers RM, Zhang ZW (2019) BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* 8 <https://doi.org/10.1093/gigascience/giy154>
- Iorizzo M, Ellison S, Senalik D, Zeng P, Satapoomin P, Huang J, Bowman M, Iovene M, Sanseverino W, Cavagnaro P, Yildiz M, Macko-Podgórní A, Moranska E, Grzebelus E, Grzebelus D, Ashrafi H, Zheng Z, Cheng S, Spooner D, Van Deynze A, Simon P (2016) 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>
- Krueger F (2021) Trim Galore. In GitHub repository. GitHub. <https://github.com/FelixKrueger/TrimGalore.com/fenderglass/Flye>
- Labandera A-M, Tedds HM, Bailey M, Sprigg C, Etherington RD, Akintewe O, Kalleechurn G, Holdsworth MJ, Gibbs DJ (2021) The PRT6 N-degron pathway restricts VERNALIZATION

820 2 to endogenous hypoxic niches to modulate plant development. *New Phytol* 229:126–139  
821 <https://doi.org/https://doi.org/10.1111/nph.16477>

822 Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*  
823 e-prints:arXiv:1303.3997

824 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,  
825 Genome Project Data Processing S (2009) The Sequence Alignment/Map format and  
826 SAMtools. *Bioinformatics* (Oxford, England) 25:2078–2079  
827 <https://doi.org/10.1093/bioinformatics/btp352>

828 Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012)  
829 GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28:2397–2399  
830 <https://doi.org/10.1093/bioinformatics/bts444>

831 Liu H, Zhang J-Q, Chen C, Wang Y-H, Xu Z-S, Zhao Q-Z, Zhang J, Xue J-Y, Xiong A-S (2025) The  
832 parsley genome assembly and DNA methylome shed light on apigenin biosynthesis in the  
833 *Apiaceae*. *Plant Physiol* 197:kiaf077 <https://doi.org/10.1093/plphys/kiaf077>

834 Liu H, Zhang J-Q, Zhang R-R, Zhao Q-Z, Su L-Y, Xu Z-S, Cheng Z-M, Tan G-F, Xiong A-S (2024)  
835 The high-quality genome of *Cryptotaenia japonica* and comparative genomics analysis  
836 reveals anthocyanin biosynthesis in *Apiaceae*. *The Plant Journal* 118:717–730  
837 <https://doi.org/https://doi.org/10.1111/tpj.16628>

838 Mahboubi M (2019) Caraway as Important Medicinal Plants in Management of Diseases. *Natural*  
839 *Products and Bioprospecting* 9:1–11 <https://doi.org/10.1007/s13659-018-0190-x>

840 McKinney GJ, Waples RK, Seeb LW, Seeb JE (2017) Paralogs are revealed by proportion of  
841 heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural  
842 populations. *Mol Ecol Resour* 17:656–669 <https://doi.org/10.1111/1755-0998.12613>

843 Misginna SG, Roszlerova P, Gaoua O, Kamulegeya P, Pichova M, Jozova E, Curn V (2026)  
844 Molecular markers and genomic resources in caraway (*Carum carvi* L.): Current status,  
845 research gaps, and strategic directions for breeding. *Czech Journal of Genetics and Plant*  
846 *Breeding* 62:121–134 <https://doi.org/10.17221/16/2026-CJGPB>

847 Németh É (1998) Questions of the generative development. In: Németh E (ed) *Caraway The Genus*  
848 *Carum*, vol 7. *Medicinal and Aromatic Plants - Industrial Profiles*. harwood academic  
849 publishers, Amsterdam, pp 69–81

850 Németh E, Pluhar Z (1996) Preliminary observations on the inheritance of caraway flowering.  
851 *Beiträge zur Züchtungsforschung* 2. Jahrgang:116–118

852 Palumbo F, Galla G, Vitulo N, Barcaccia G (2018) First draft genome sequencing of fennel  
853 (*Foeniculum vulgare* Mill.): identification of simple sequence repeats and their application in  
854 marker-assisted breeding. *Mol Breed* 38:17 <https://doi.org/10.1007/s11032-018-0884-0>

855 Pank F (2012) Kümmel (*Carum carvi* L.). In: Hoppe B (ed) *Handbuch des Arznei- und*  
856 *Gewürzpflanzenbaus*, vol 4. Verein für Arznei- und Gewürzpflanzen SALUPLANTA e.V.,  
857 Bernburg, pp 728–762

858 Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-  
859 Villeda H, Sorrells M, Jannink J-L (2012) Genomic Selection in Wheat Breeding using  
860 Genotyping-by-Sequencing. *The Plant Genome* 5  
861 <https://doi.org/https://doi.org/10.3835/plantgenome2012.06.0006>

862 R Core Team (2025) R: A language and environment for statistical computing. R Foundation for  
863 Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

864 Song X et al. (2021) The celery genome sequence reveals sequential paleo-polyploidizations,  
865 karyotype evolution and resistance gene reduction in apiaceae. *Plant Biotechnol J* 19:731–744  
866 <https://doi.org/10.1111/pbi.13499>

867 Song X et al. (2019) Deciphering the high-quality genome sequence of coriander that causes  
868 controversial feelings. *Plant Biotechnol J* n/a <https://doi.org/10.1111/pbi.13310>

869 Sung S, Amasino RM (2004) Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger  
870 protein VIN3. *Nature* 427:159–164 <https://doi.org/10.1038/nature02195>

871 van Ooijen JW (2004) MapQTL 5, Software for the Mapping of Quantitative Trait Loci in  
872 Experimental Populations. Plant Research International

873 van Ooijen JW (2018) JoinMap® 5, Software for the calculation of genetic linkage maps in  
874 experimental populations.

875 von Maydell D, Beleites C, Stache A-M, Riewe D, Krähmer A, Marthe F (2024) Genetic variation of  
876 annual and biennial caraway (*Carum carvi*) germplasm offers diverse opportunities for  
877 breeding. *Industrial Crops and Products* 208:117798  
878 <https://doi.org/10.1016/j.indcrop.2023.117798>

879 von Maydell D, Brandes J, Lehnert H, Junghanns W, Marthe F (2020) Breeding synthetic varieties in  
880 annual caraway: observations on the outcrossing rate in a polycross using a high-throughput  
881 genotyping system. *Euphytica* 217:1 <https://doi.org/10.1007/s10681-020-02732-5>

882 von Maydell D, Hammer M, Junghanns W, Marthe F (2026a) The potential of winter annual  
883 cultivation of caraway (*Carum carvi* L.) under changing climate conditions: Exploring genetic  
884 resources and evaluating cultivation system and sowing density. *Journal of Applied Research*  
885 *on Medicinal and Aromatic Plants* 53:100728  
886 <https://doi.org/10.1016/j.jarmap.2026.100728>

887 von Maydell D, Lehnert H, Berner T, Klocke E, Junghanns W, Keilwagen J, Marthe F (2021) On  
888 genetic diversity in caraway: Genotyping of a large germplasm collection. *PLoS One*  
889 15:e0244666 <https://doi.org/10.1371/journal.pone.0244666>

890 von Maydell D, Lim F-S, Junghanns M, Poeschl Y, Budahn H, Marthe F, Keilwagen J, Schmutzer T  
891 (2026b) Dissecting flowering time and flower color in *Carum carvi* utilizing a long-read draft  
892 genome and a GBS-based QTL mapping. *Scientific Reports* 16:22067  
893 <https://doi.org/10.1038/s41598-026-61767-1>

894 Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York

895 Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Golemund G, Hayes A,  
896 Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D,  
897 Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to  
898 the Tidyverse. *Journal of Open Source Software* 4 <https://doi.org/10.21105/joss.01686>

899 Wohlfeiler J, Alessandro MS, Cavagnaro PF, Galmarini CR (2019) Multiallelic digenic control of  
900 vernalization requirement in carrot (*Daucus carota* L.). *Euphytica* 215:37  
901 <https://doi.org/10.1007/s10681-019-2360-2>

902 Wohlfeiler J, Alessandro MS, Morales A, Cavagnaro PF, Galmarini CR (2022) Vernalization  
903 Requirement, but Not Post-Vernalization Day Length, Conditions Flowering in Carrot  
904 (*Daucus carota* L.). *Plants* 11:1075. <https://doi.org/10.3390/plants11081075>

905 Wohlfeiler J, Morales A, Galmarini CR, Alessandro MS, Carvajal S, Senalik D, Ellison S, Macko-  
906 Podgórní A, Iorizzo M, Grzebelus D, Simon PW, Cavagnaro PF (2026) A pair of duplicated  
907 SOC1 genes, DcSOC1-1 and DcSOC1-2, control vernalization requirement in carrot. BMC  
908 Plant Biol <https://doi.org/10.1186/s12870-026-09529-z>

909 Xu ZS, Yang QQ, Feng K, Xiong AS (2019) Changing Carrot Color: Insertions in DcMYB7 Alter  
910 the Regulation of Anthocyanin Biosynthesis and Modification. Plant Physiol 181:195–207  
911 <https://doi.org/10.1104/pp.19.00523>

912 Zhou Q, Tang D, Huang W, Yang Z, Zhang Y, Hamilton JP, Visser RGF, Bachem CWB, Robin  
913 Buell C, Zhang Z, Zhang C, Huang S (2020) Haplotype-resolved genome analyses of a  
914 heterozygous diploid potato. Nat Genet 52:1018–1023 [https://doi.org/10.1038/s41588-020-](https://doi.org/10.1038/s41588-020-0699-x)  
915 0699-x

916 Zhu C, Wang C, Zhang YM (2007) Modeling segregation distortion for viability selection. I.  
917 Reconstruction of linkage maps with distorted markers. Theor Appl Genet 114:295–305  
918 <https://doi.org/10.1007/s00122-006-0432-x>

919 Zuo J-F, Niu Y, Cheng P, Feng J-Y, Han S-F, Zhang Y-H, Shu G, Wang Y, Zhang Y-M (2019)  
920 Effect of marker segregation distortion on high density linkage map construction and QTL  
921 mapping in Soybean (*Glycine max* L.). Heredity 123:579–592  
922 <https://doi.org/10.1038/s41437-019-0238-7>

923

924

## 925 **Glossary**

926 BOF: Begin of flowering

927 COL: Flower color

928 DAS: Days after sowing

929 ESC: Early sowing conditions

930 LSC: Late sowing conditions

931 MAS: Marker assisted selection

932 P1: F<sub>2</sub> Population P1 originating from the cross of the biennial cultivar 'Rekord' x annual inbred line  
933 'L32'

934 P2: F<sub>2</sub> Population P2 originating from the cross of the biennial wild-type accession 'CARUM 63' x  
935 annual inbred line 'L32'

## 936 **Figure Captions**

937 Figure 1. Field trial at JKI in Quedlinburg. A) At left side biennial parent 'Carum 63' with few  
938 flowering plants; at right side corresponding F<sub>2</sub> P2 with many plants close to end of flowering. B) At  
939 left side biennial parent 'Rekord' with few flowering plants; at right side corresponding F<sub>2</sub> P1 with  
940 many plants within full flowering or close to end of flowering. C) At left side F<sub>2</sub> P1; at right side annual  
941 parent L32. D/E) Typical morphology in the first year of 'Carum 63' and 'Rekord', respectively. C  
942 from June 30, 2025; others from Juli 28, 2025.

943 Figure 2 Density distribution of SNP markers across the 10 linkage groups (LG) of the thinned genetic  
944 maps for P1 and P2.

945 Figure 3: Comparison of genetic and physical marker positions (Marey maps) per LG. The plots  
946 visualize the relationship between genetic distance (y-axis) and estimated physical position (x-axis).  
947 Input data consist of three genetic linkage maps: population P1 (represented by purple dots), population  
948 P2 2025 (represented by gold triangles) and population P2 2023 (represented by grey stars). The  
949 physical coordinates on the x-axis represent cumulative pseudo-chromosomal lengths calculated by  
950 concatenating the unanchored contigs of the draft genome assembly according to their genetically  
951 estimated order.

952 Figure 4: Begin of flowering of A: F<sub>2</sub> population P1 (purple) and parents and B: F<sub>2</sub> population P2  
953 (yellow) and parents. Quantitative data of plants flowering in the second year (merely applicable for  
954 ‘CARUM 63’) are not presented. Legend: N = Number of presented annual flowering plants out of  
955 total number of plants, which had a healthy status until end of the first vegetation period (i.e. including  
956 biennials); percentage = proportion of annual flowering plants.

957 Figure 5 Genome-wide LOD score profiles for the traits BOF (light green) and COL (pink) obtained  
958 from rMQM. Panels are structured by linkage groups (columns) and populations/years (rows: P1 2025,  
959 P2 2025, and P2 2023). The dashed red line indicates the genome-wide significance threshold,  
960 displaying exclusively the more stringent (higher) significance level calculated for each respective  
961 population and trait.

962 Figure 6 Effect of two-locus (LG02/LG06) homozygous genotype combinations (*aa*, *ab*, *ba*, *bb*) on  
963 beginning of flowering. The lower panel illustrates normal flowering response (80–160 days) with  
964 boxplots (medians), red dashed bars (class means), and individual genotypes (jittered dots). The upper  
965 panel indicates extreme phenotypes classified as Very Late Flowering (VLF, > 200 days) or Non-  
966 Flowering (NF).

967 Figure 7 Comparison of some QTLs mapped on the consensus physical map across three population-  
968 year combinations (P2 2023, P2 2025, and P1 2025). The pseudochromosome backbones are  
969 represented by grey vertical bars. Colored bars indicate the 1-LOD (thick) and 2-LOD (thin) confidence  
970 intervals, while white diamonds denote the peak marker positions.

971 Figure 8 Relationship between the cumulative number of biennial 'b' alleles across four QTL peak loci  
972 from von Maydell et al. (2026b) and beginning of flowering. Red dashed bars show arithmetical means,  
973 and individual jittered dots represent genotypes. The solid blue line shows the linear regression model.

974 Figure 9 Effect of two-locus (LG01/LG10) homozygous genotype combinations (aa, ab, ba, bb) on  
975 flower color with 'a' representing the annual and 'b' representing the biennial allele. Red dashed bars  
976 = class means and jittered dots = individual genotypes.

977 Figure 10 Begin of flowering of 15 different F2 populations and the annual parent in the years 1999  
978 and 2000. N per genotype and year = 46-48. N of annual parent 1999 = 96. Red dashed line indicates  
979 arithmetical mean. Different letters indicate significant differences among groups according to Tukey's  
980 HSD test ( $\alpha = 0.05$ ).

981 Figure 11 Exemplary allelic discrimination by PACE marker M183 with 88 F2 samples, 6 parental  
982 samples and two negative controls (black diamonds). Orange dots represent homozygous annual (allele  
983 1/FAM) genotypes, blue squares homozygous biennial (allele 2/HEX) genotypes and green triangles  
984 heterozygous genotypes.

## 985 **Supplementary**

986 **Table S1:** Raw data and summaries from biparental mapping. Detailed explanations/definition within  
987 tables. A/B) Phenotypic data; C/D) SNP Matrices; E) Mapping Results; F) Kruskal-Wallis, G) BLINK,  
988 H) Full QTL Table, I) List of genes in 1-LOD intervals of Cc-FT02.2 and Cc-FT06.1

989     **Table S2:** A-D) Thinned and full SNP matrices of P1 and P2 and E) consensus map statistics.

990     **Table S3:** Detailed data of 15 additional F<sub>2</sub> populations: A) Raw data. B) Segregation ratio in 1999.

991     C) Mean values. D) Tukey-HSD test.

992     **Table S4:** Detailed information of PACE markers including primer sequences.

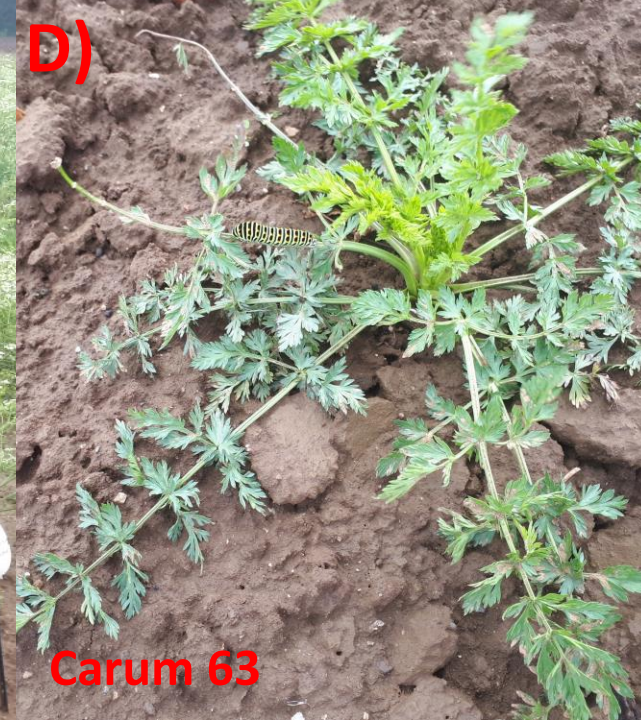
993

994     **Figure S1:** Consensus physical map showing the estimated order and orientation of contigs.

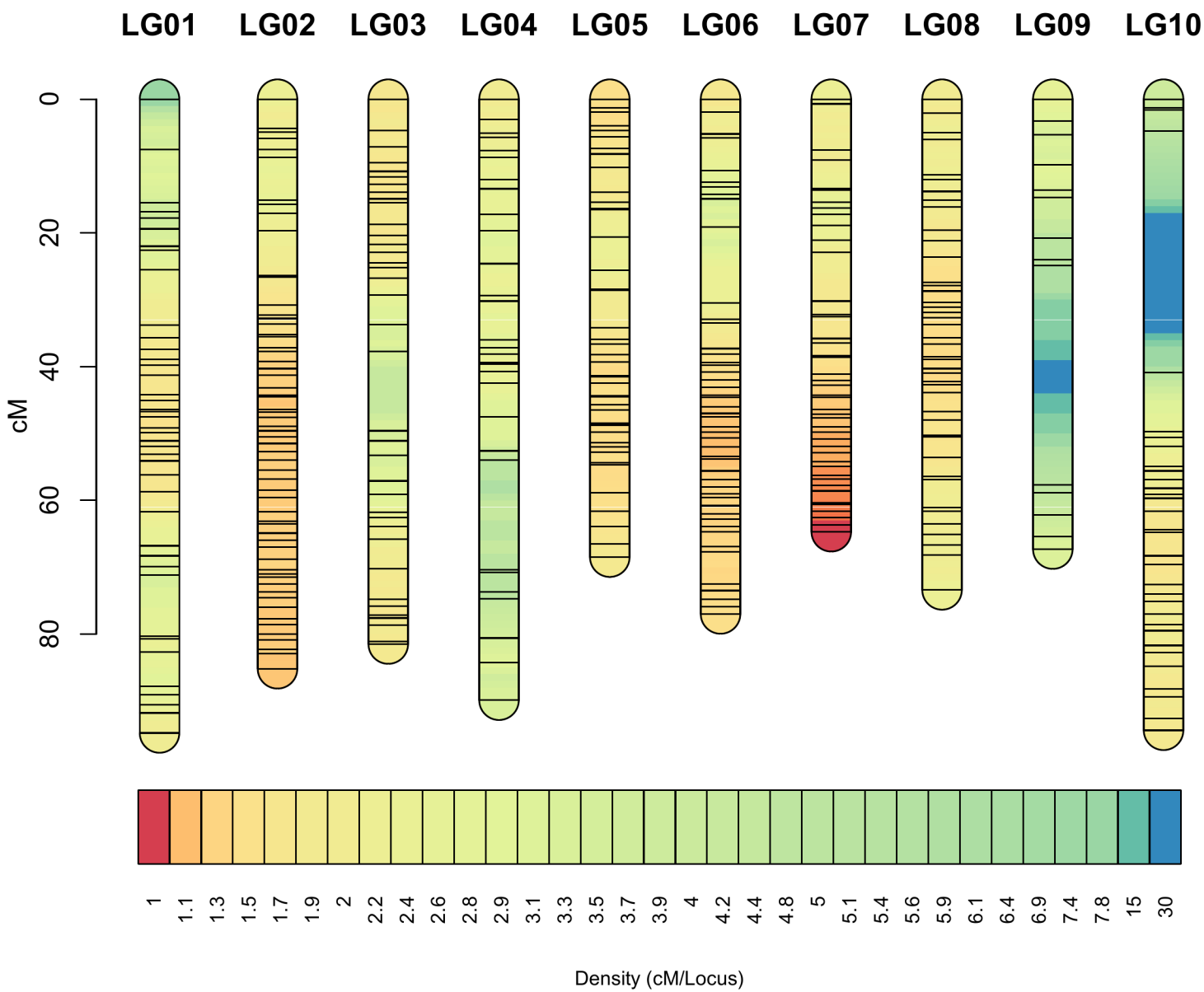
995     **Figure S2:** Effect of detected QTL on BOF or COL per genotype combinations.

996

997

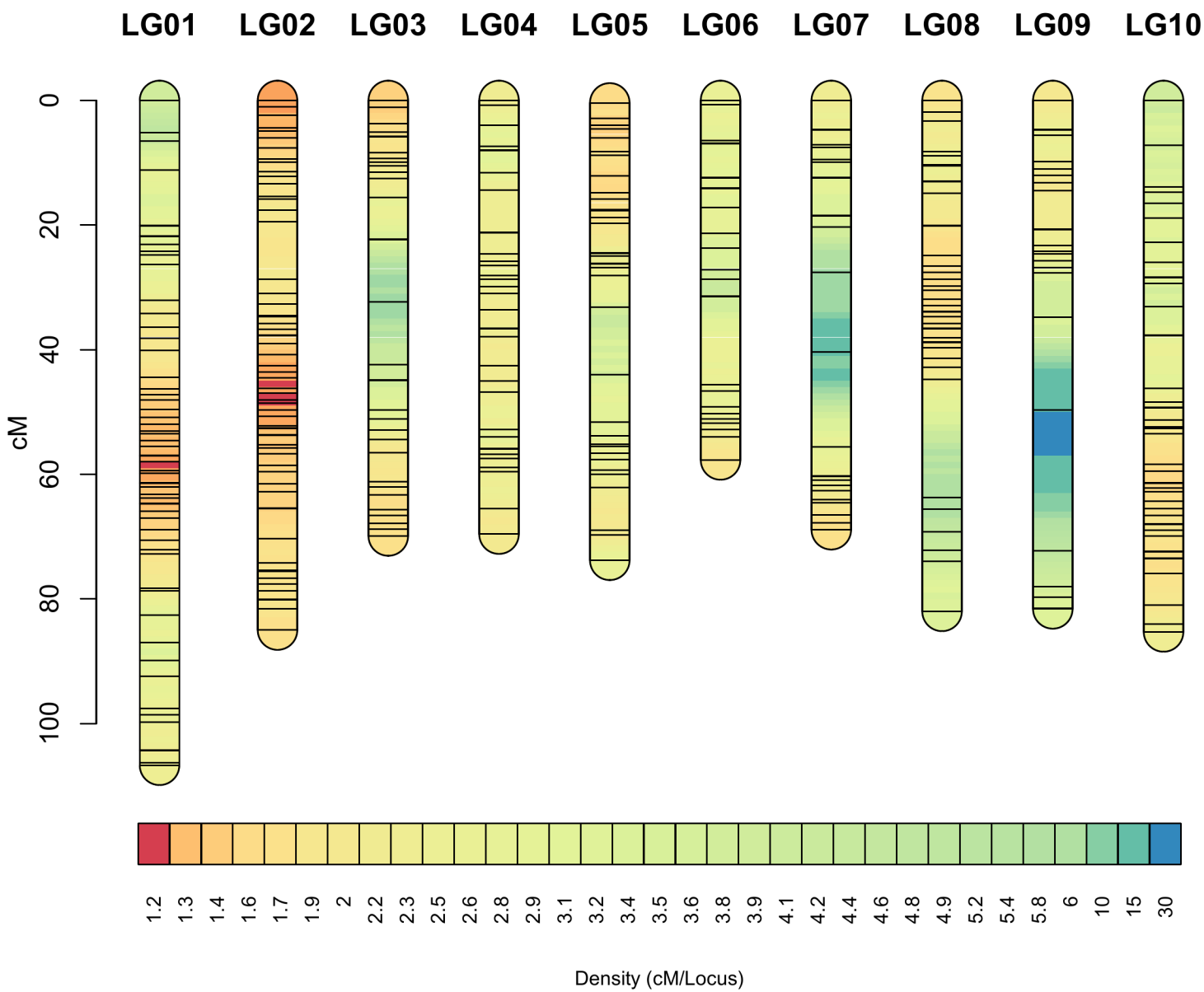


A: Map P1

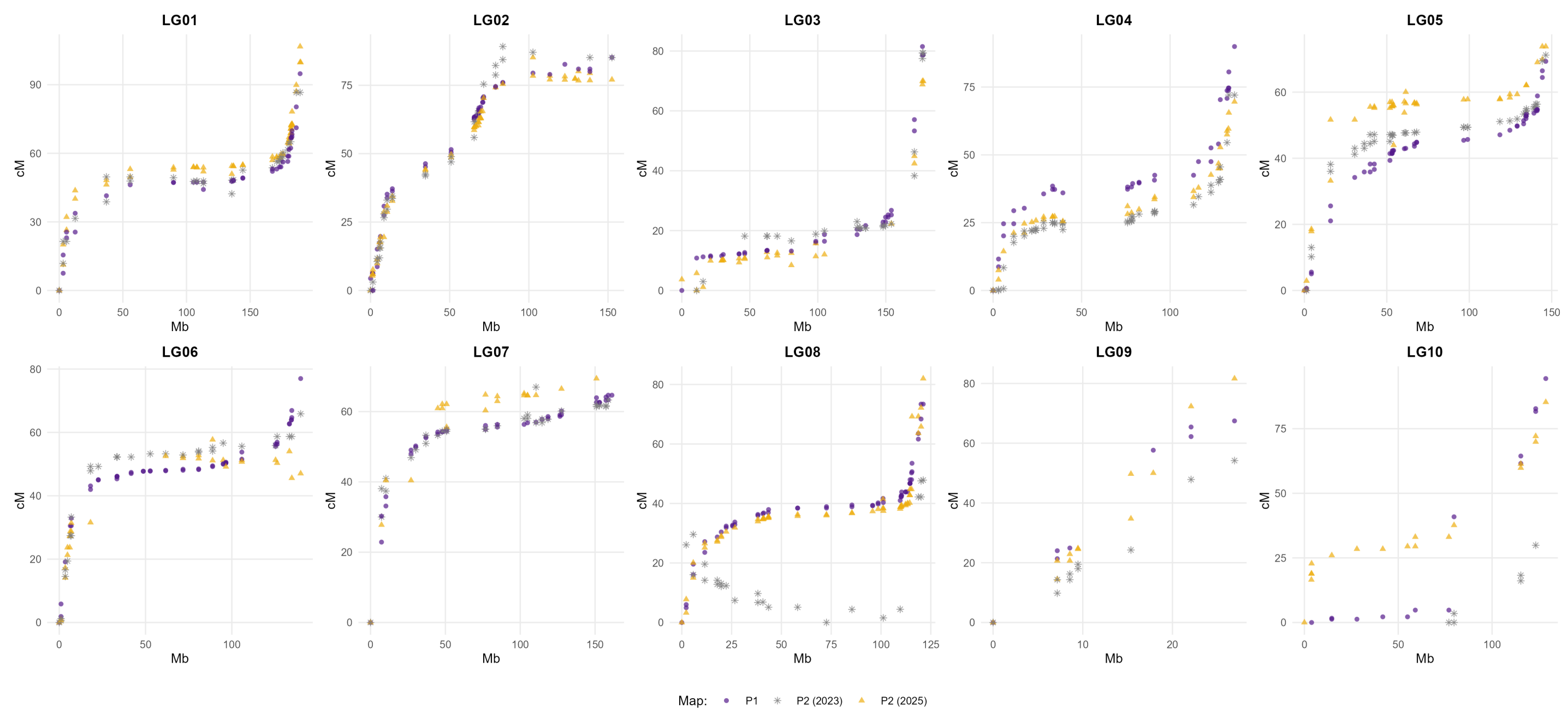


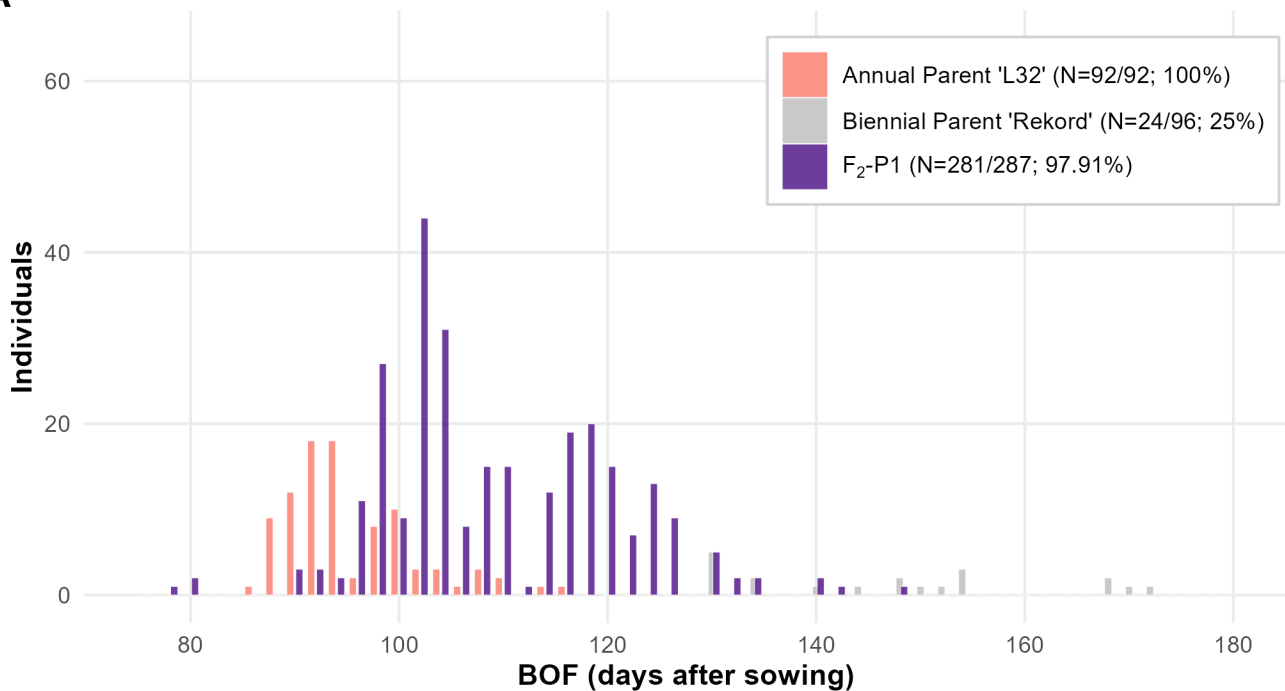
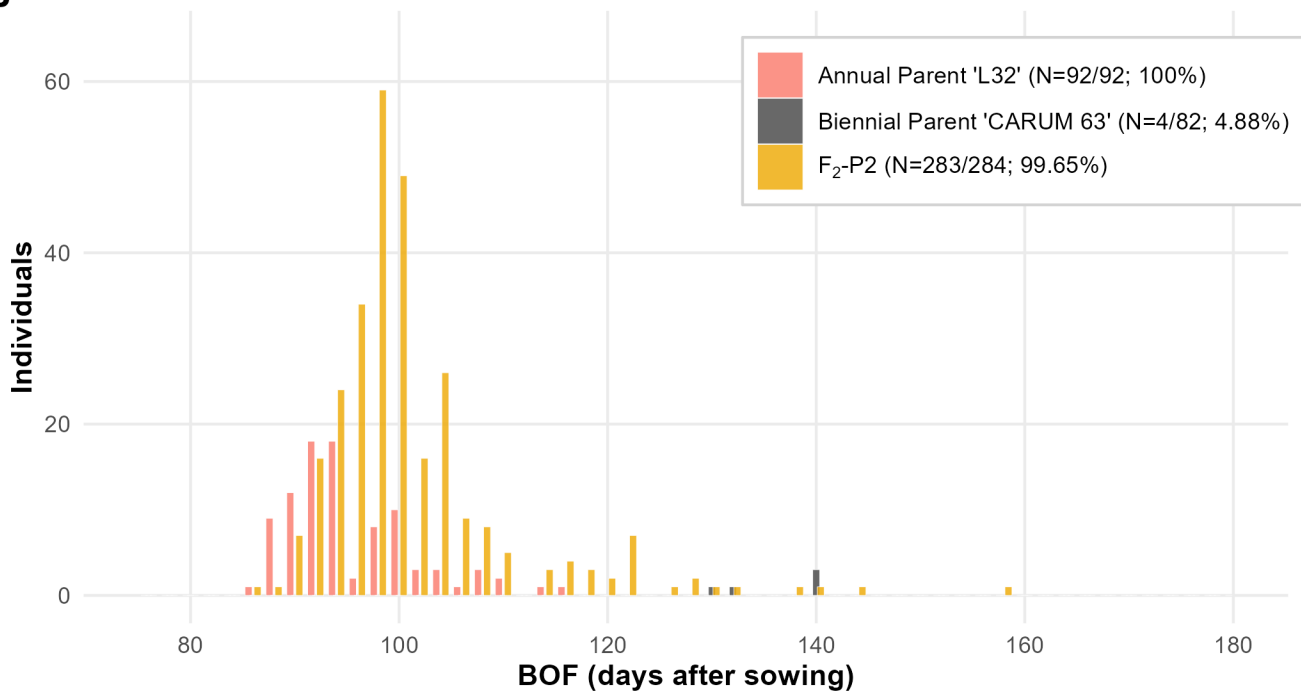
Rendered by LinkageMapView

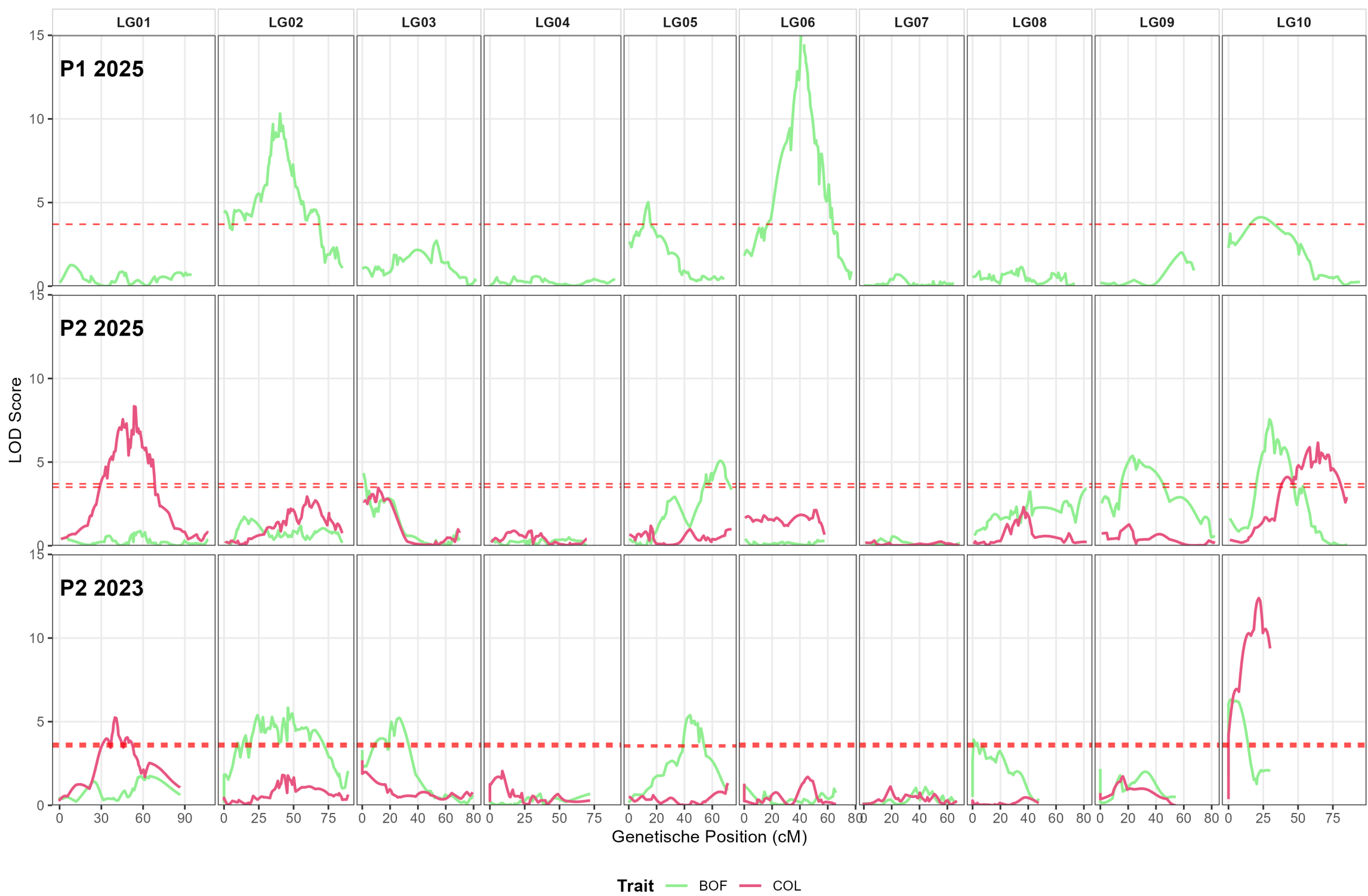
B: Map P2

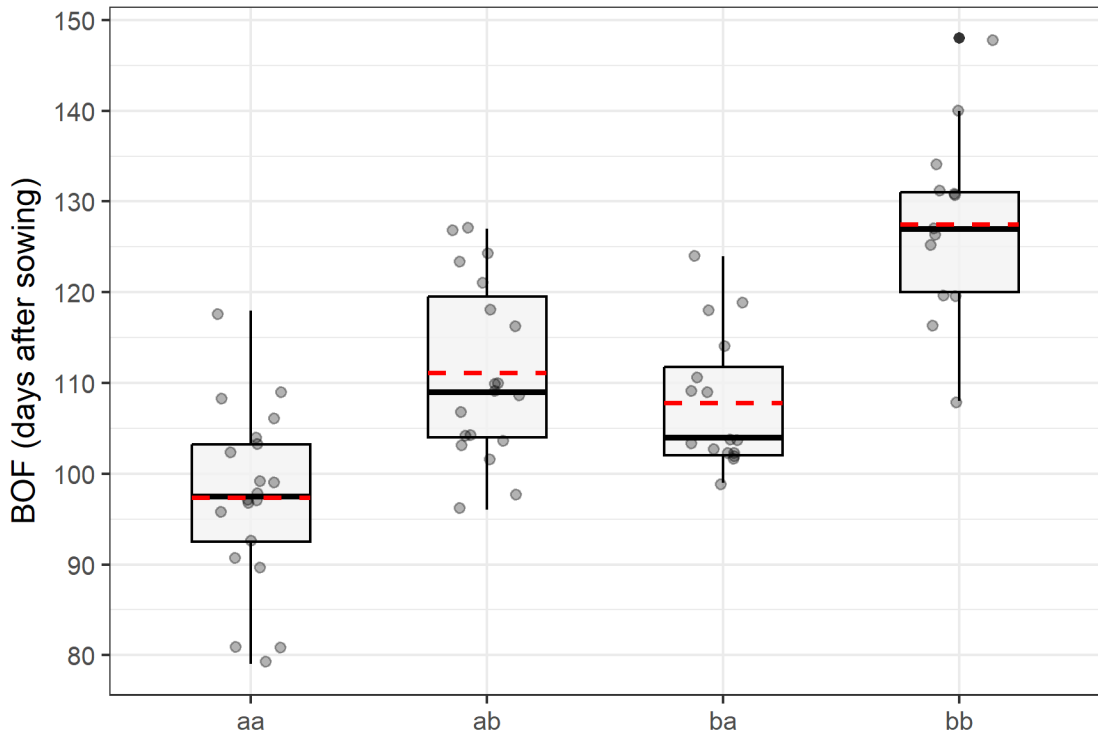


Rendered by LinkageMapView

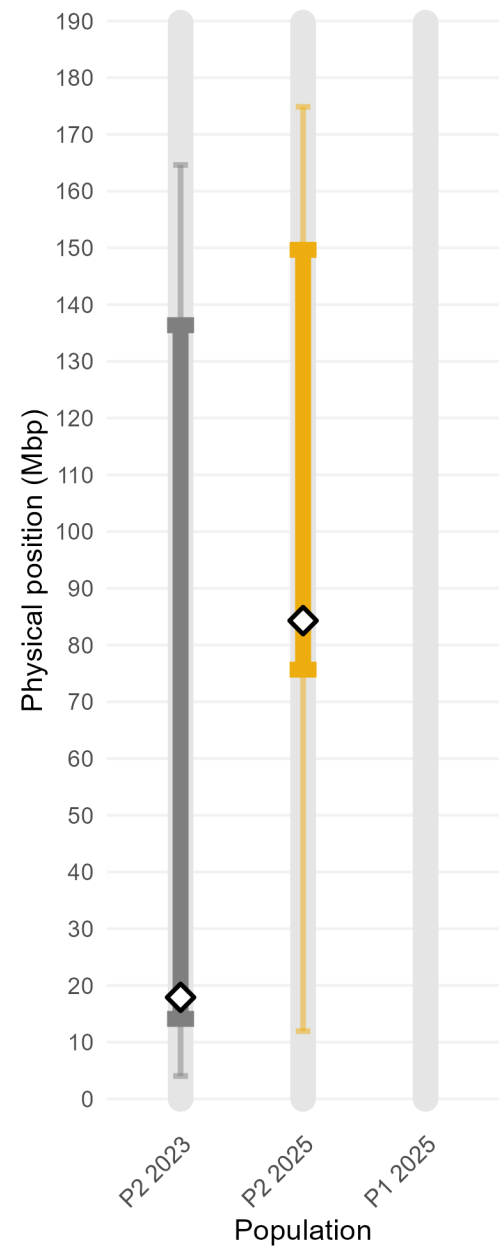


**A****B**

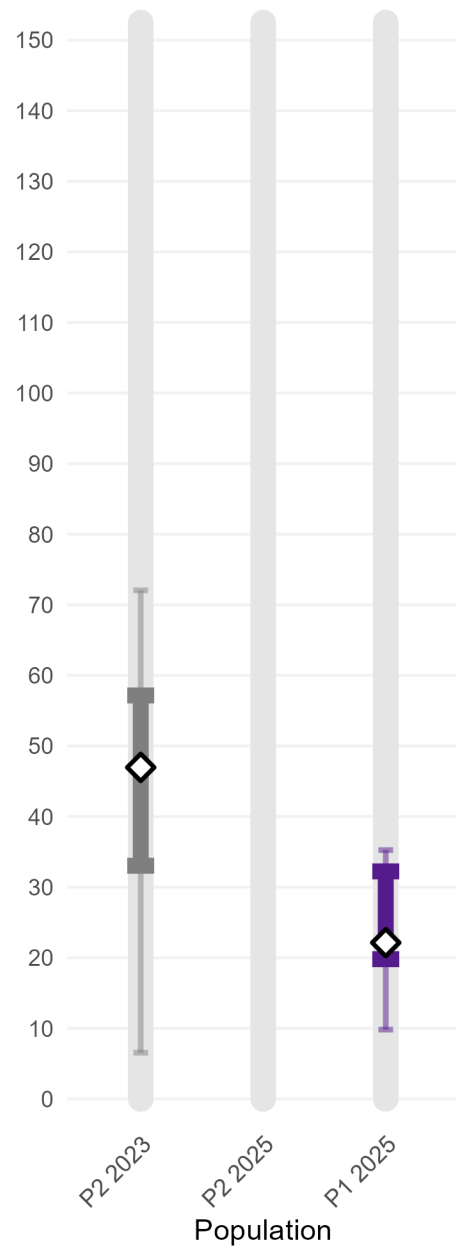




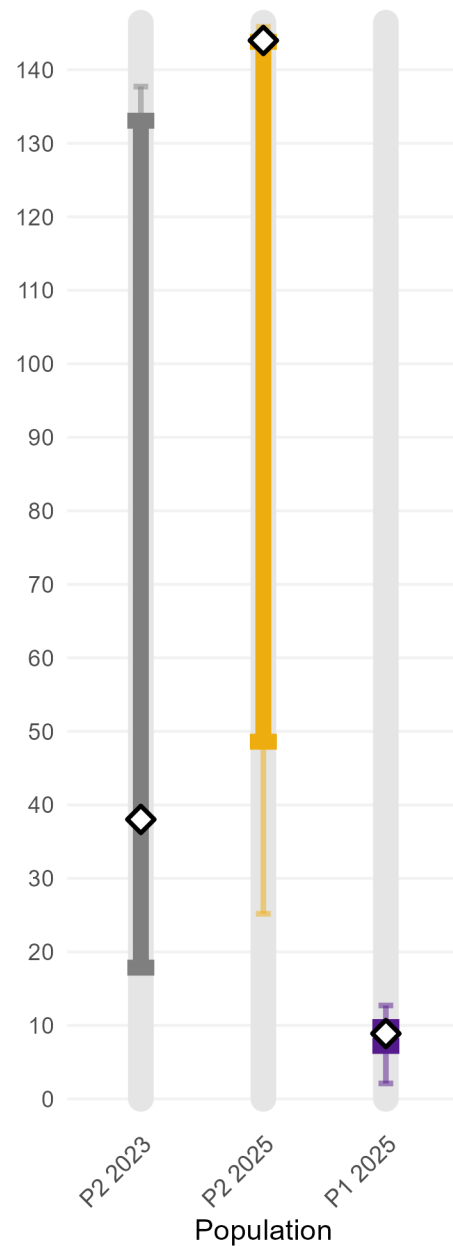
LG01 - COL



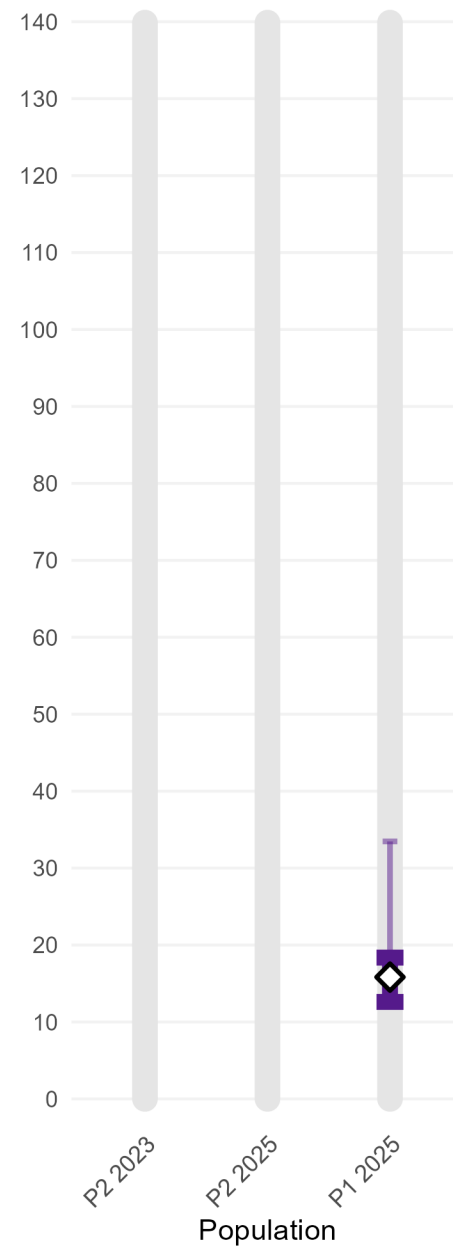
LG02 - BOF



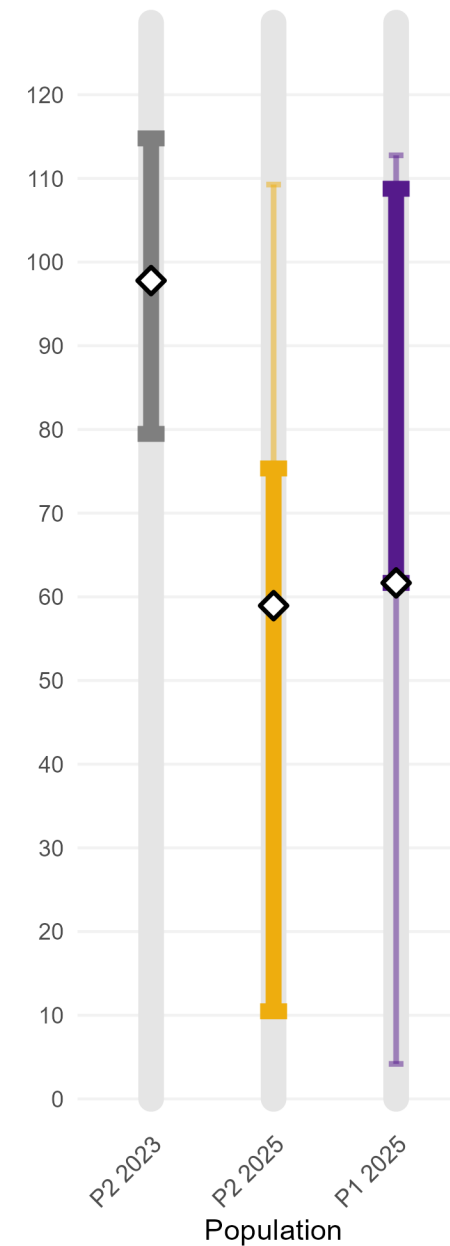
LG05 - BOF



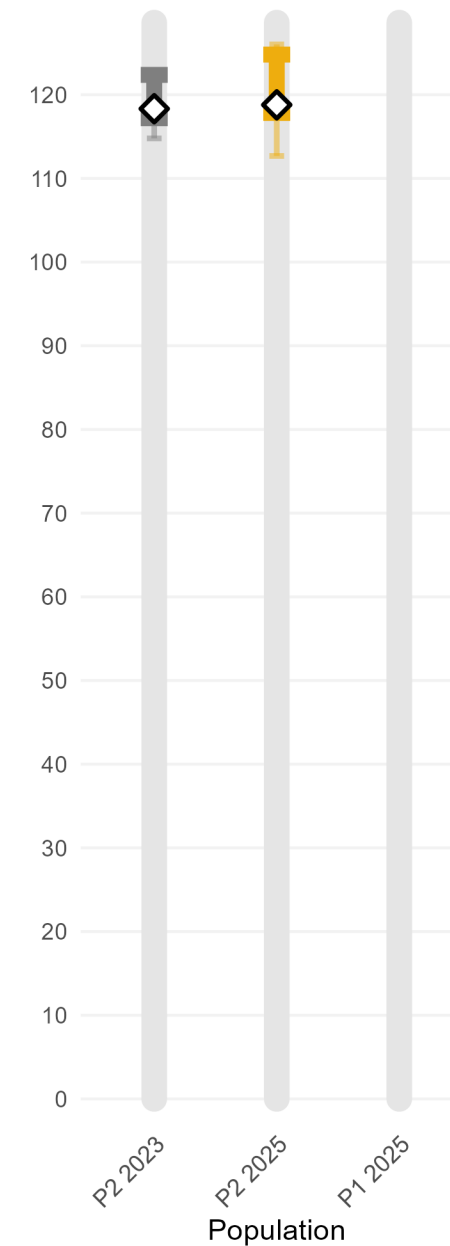
LG06 - BOF

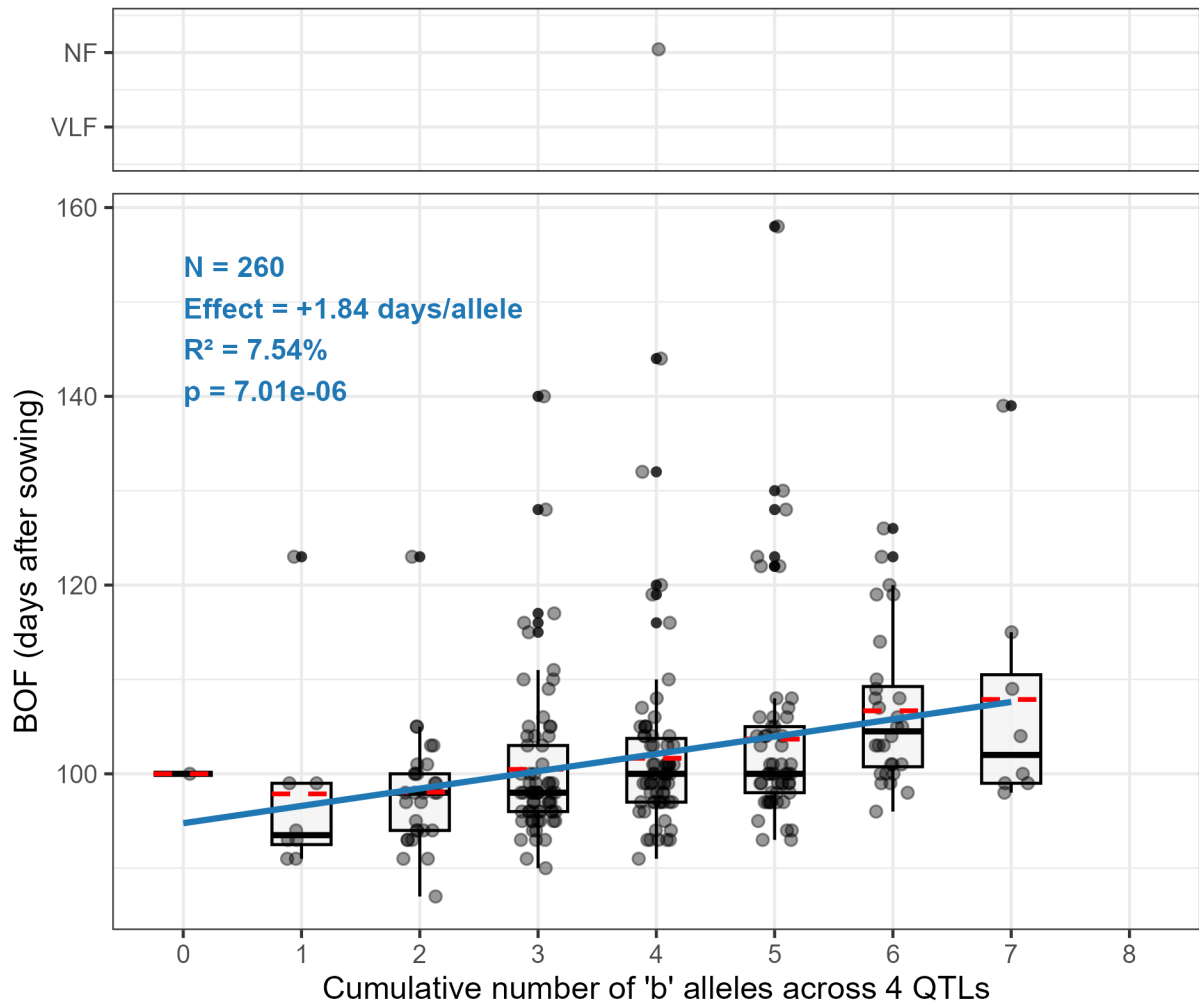


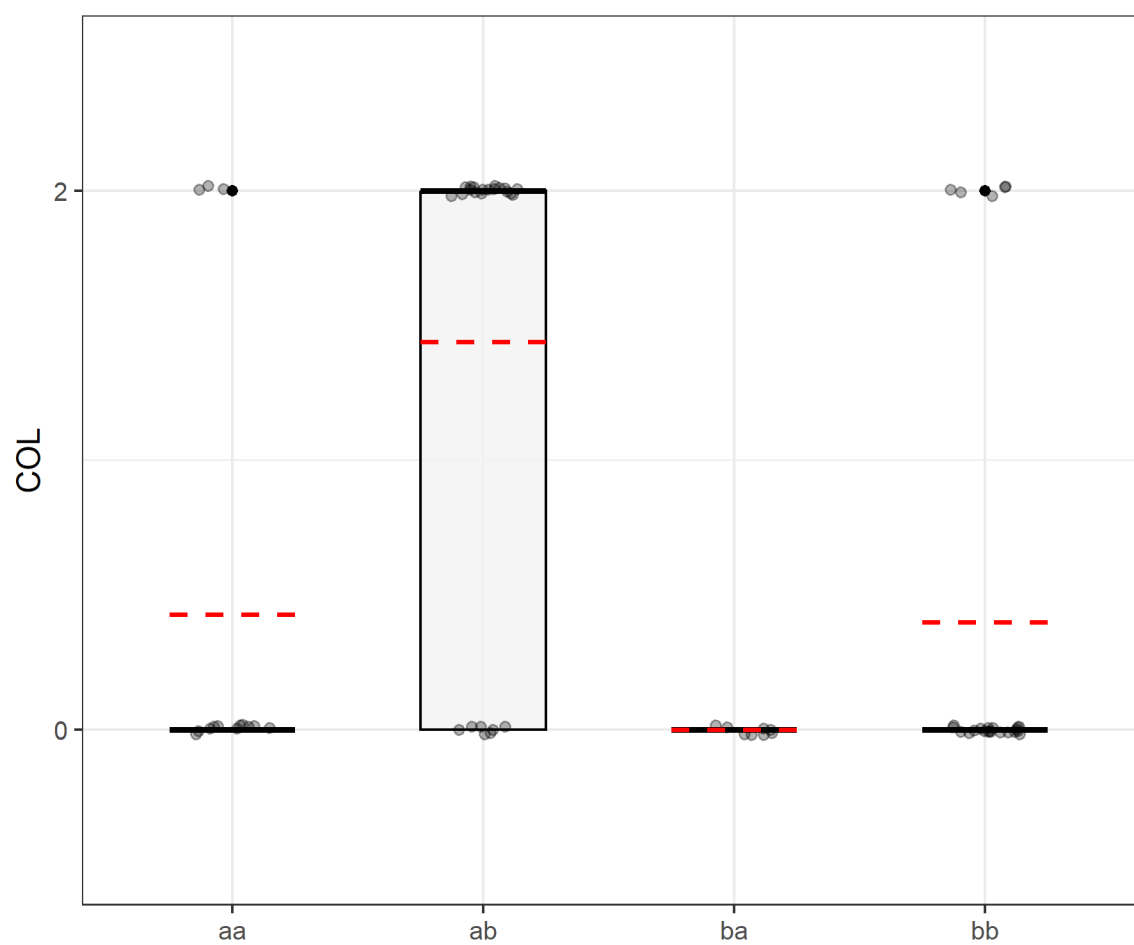
LG10 - BOF

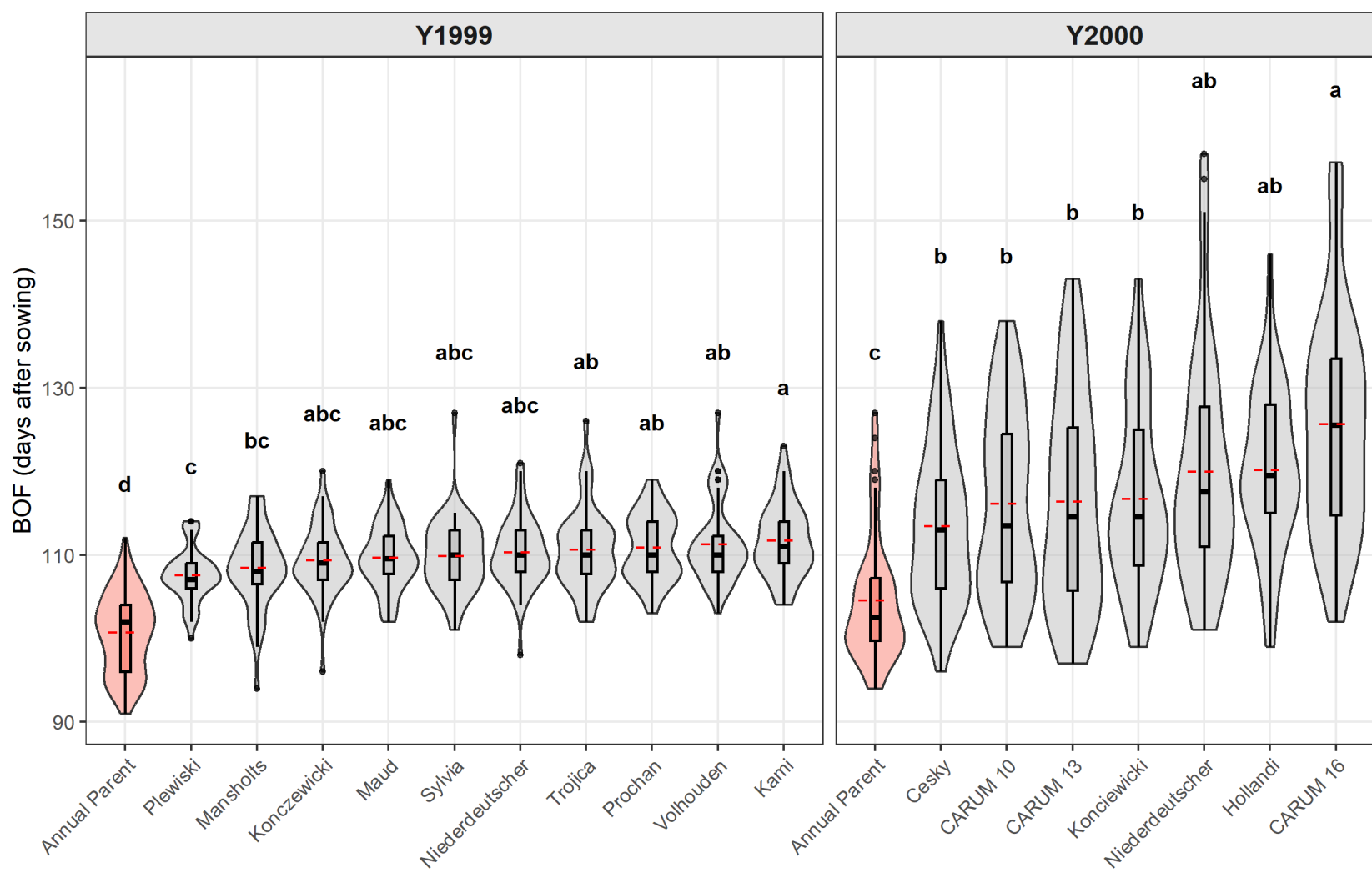


LG10 - COL

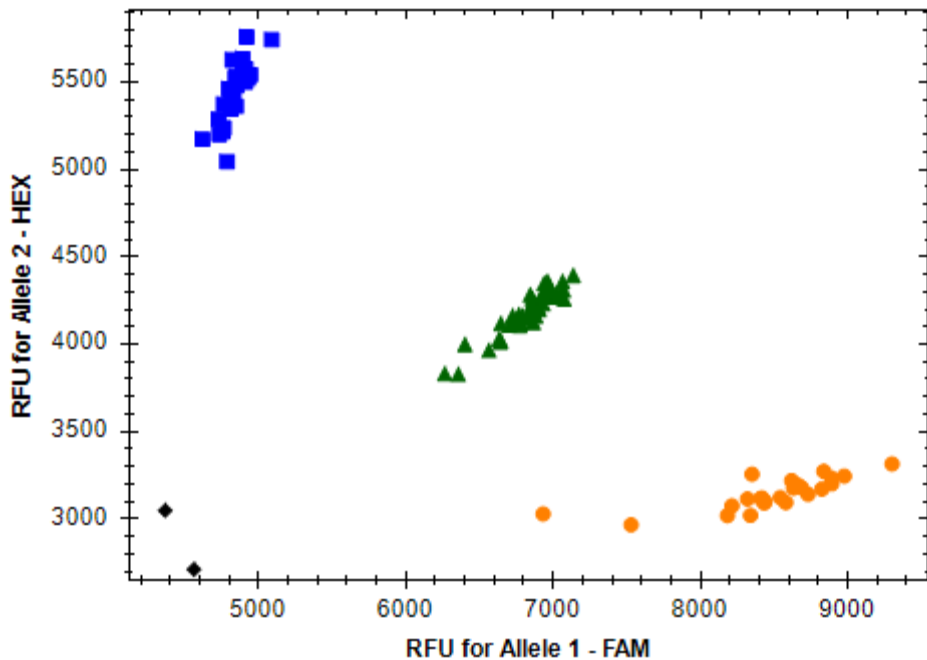








Allelic Discrimination



## Supplementary Files

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

- [FigureS1ConsensusMap.png](#)
- [FigureS2QTLBoxplots.pdf](#)
- [TableS1BiparentalMapping20260812.xlsx](#)
- [TableS2Maps.xlsx](#)
- [TableS315F2Populations20260610.xlsx](#)
- [TableS4PACEMarker.xlsx](#)