QTL mapping utilizing $F_{2:3}$ linkage mapping populations reveals regions of chromosomes 2 and 6 are significantly associated with root width in carrot

A. Vega-Alfaro, S.H. Brainard and I.L. Goldman

Department of Horticulture, University of Wisconsin-Madison, Madison Wisconsin, USA.

Abstract

This study aimed to map quantitative trait loci (QTL) associated with carrot root width using linkage mapping. Two bi-parental populations derived from the cross of carrot inbred lines L1408 and W133, followed by two generations of selfing, were combined resulting in a total of 123 F_3 individuals grown in a randomized complete block design in the years 2020 and 2021 in Randolph, Wisconsin, USA. Phenotyping of carrot root traits was performed using a digital imaging pipeline, and genotyping was done using genotyping-by-sequencing (GBS). A total of 3,953 high quality single nucleotide polymorphisms (SNPs) markers were retained. A genetic map was constructed, and composite interval mapping was performed. The results showed that three QTLs were significantly associated with carrot root width on chromosomes 2 and 6 and explain between 10 and 18% of the phenotypic variance. The results here expand the current understanding of the genetic control of carrot root shape traits.

Keywords: linkage mapping, QTL mapping, *Daucus carota*, genotyping-by-sequencing

INTRODUCTION

Carrot (*Daucus carota* subsp. *sativus*) root shape consists of a group of traits that affect the visual appearance, culinary quality, and market value of this crop. Root width is a component of carrot market class, with varying width requirements for the market segments. For example, baby-cut carrots have a narrow, long, and cylindrical shape, while processing cultivars tend to have a larger, more robust shape with wide shoulders and root tips.

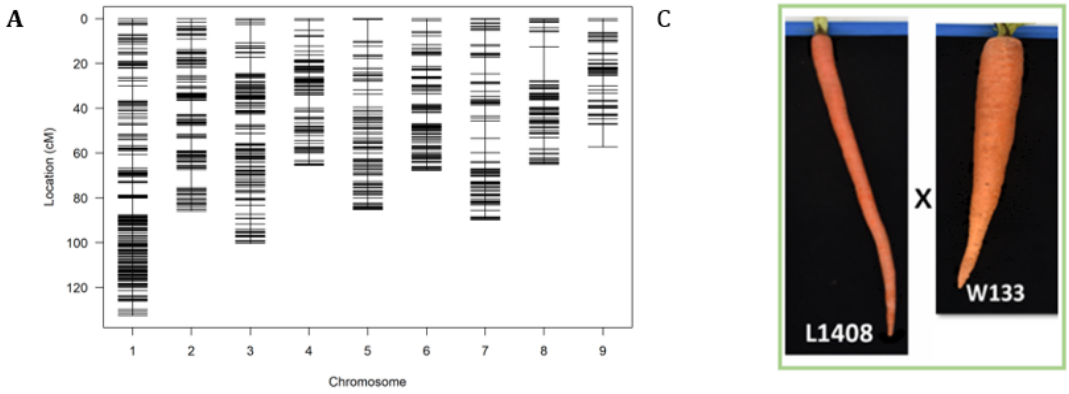
Since at least early on the 17th century, traits related to root shape have been used to classify carrots and were also likely subject to selection (Banga, 1957). However, the genetic basis of carrot root shape, and in particular root width, is not yet well understood. It is hypothesized that root shape characteristics are determined by several genes having small effects with moderate influence of environment. Recent advances in genomic tools and molecular markers have enabled researchers to map quantitative trait loci (QTL) associated with this trait. QTL mapping involves the identification of specific genomic regions that are associated with variation in a quantitative trait. Understanding the genetic basis of carrot root shape and its association with market class is crucial for carrot breeding programs aimed at improving the quality and expanding the market classes of this crop. Identifying QTLs associated with this trait can facilitate the development of marker-assisted selection strategies, which accelerate the breeding process by enabling selection for desired traits at early stages of plant development. In this study, the objective was to map QTL associated with carrot root width using a combination of genetic mapping in a combined bi-parental population of a cross between a processing carrot type and a long and cylindrical inbred line developed by the US Department of Agriculture.

MATERIALS AND METHODS

Plant material

Two bi-parental populations, derived from the cross of carrot inbreds L1408 and W133 were used for linkage mapping. L1408 is a very long 'Imperator' type developed by the US Department of Agriculture, Vegetable Crops Research Unit (USDA VCRU) and inbred W133 is

a medium length wedge-shaped inbred developed by the University of Wisconsin - Madison (Goldman, 1996) (Figure 1C). One F_1 individual from each population was self-pollinated to obtain F_2 individuals and each F_2 individual was self-pollinated to obtain F_3 families. The population was named "L1408×W133" and consisted of 123 $F_{2:3}$ progeny. The population was grown in a randomized complete block design in the years 2020 and 2021 at Jack's Pride Farms in Randolph, Wisconsin, USA. One replication per block was planted in each of two blocks per year. Ten plants per 1-m row were harvested and stored at 5°C until phenotyping.



B Genetic map statistics

Chromosome	Marker number	Average spacing (cM)	Maximum spacing (cM)	Length (cM)
1	495	0.3	7.4	132
2	231	0.4	8.3	85
3	254	0.4	8.1	100
4	130	0.5	5.9	65
5	196	0.4	9.8	85
6	288	0.2	5.0	67
7	133	0.7	7.8	90
8	209	0.3	15.2	65
9	92	0.6	10.1	57
Overall	2028	0.4	15.2	750
Individuals=123, type of markers= SNPs.				

Figure 1. (A) Position of each molecular marker in the genetic map of carrot population L1408×W133. (B) Map statistics for the genetic map shown in A. (C) Carrot inbred founders of the L1408×W133 population.

Phenotyping

123 F_3 individuals (progeny) were phenotyped using a digital imaging pipeline as described in Brainard et al. (2021). Briefly, the digital imaging pipeline acquired a high-resolution picture using a digital camera. The image was processed into a black and white image called a binary mask. The binary mask is later straightened by removing excess curving resulting in a straight mask that is the input for software that computes several carrot root shape traits. In this paper we focused on shoulder width (width at 10% of root length), mid width (width at 50% of length), and tip width (width at 90% of length, all measured starting at the crown of the root).

Genotyping and genetic map construction

~10 seeds per F_3 family were planted in conical tubes at Walnut St. Greenhouses (University of Wisconsin-Madison), and leaf tissue was sampled from ~10 plants per F_3 family and bulked. The bulked leaf tissue was stored, lyophilized, macerated, and submitted for DNA

extraction and sequencing. The genomic DNA was extracted using the QIAGEN DNeasy mericon 96 HT kit (Qiagen, germantown, MD), and DNA quantification was done using the Quant-iT PicoGreen dsDNA kit (Life Technologies, Grand Island, NY). Restriction enzyme *ApeKI* was used to digest DNA followed by annealing of sample-specific barcodes and Illumina adapters. The genotyping by sequencing (GBS) libraries were sequenced using an Illumina NovaSeq 6000 sequencer. On average, 4 million 150-bp paired-end reads were obtained per sample, and variant discovery was performed using Tassel GBS Version 2 (Glaubitz et al., 2014) and aligned to version 3 of the carrot reference genome (Iorizzo et al., 2016, 2020). After using bcftools (Li et al., 2009) and custom R scripts to filter single nucleotide polymorphisms (SNPs), a set of filtering criteria were applied. These criteria included filtering for a minor allele frequency greater than or equal to 0.05, a 95th percentile of read depths greater or equal to 20. Only common markers to both populations were retained. Further filtering was done by excluding markers with genotype frequencies outside the range of 0.10 to 0.91 and those with more than 15% missing data or non-biallelic markers. As a result, 3,953 high-quality markers were retained. Because carrot shows severe segregation distortion (Iorizzo et al., 2016), we also avoided filtering markers based on goodness of fit tests (e.g., χ^2) which could potentially filter out informative markers leading to reduced representation in chromosome sections with known segregation distortion, for example chromosome 3 (Bannoud et al., 2019).

Markers were recoded based on the known phase of the F_1 generation, designating the L1408 allele as the "A" or reference allele, and the "B" allele as coming from W133. Only A×B and B×A type markers were retained. Genetic map construction was performed using custom R scripts and the R package MapRtools (v. 0.30; Endelman, 2023).

Markers were binned using the function LDbin from MapRtools at a threshold of 0.99. This function creates marker bins based on linkage disequilibrium (LD). The binning step resulted in 2,137 unique marker bins. After forming 9 linkage groups corresponding to the 9 carrot chromosomes each linkage group was trimmed using the LG and plot_genofreq functions from MapRtools. Both functions help remove (trim) markers from a linkage group based on genotype frequencies that deviate from a fitted spline. Markers were ordered locally on chromosomes 3, 4, 7 by using the function order_markers in MapRtools which uses a matrix of recombination frequencies among markers to calculate an optimized order solving the traveling salesperson problem. The function was run using 4-6 iterations and the solution with the lowest sum of adjacent recombination frequency (SARF) was chosen. The input recombination frequency matrix between markers for the order_markers function was obtained using the MLEL function in MapRtools which estimates linkage using maximum likelihood. Markers on chromosomes 1, 2, 5, 6, 8 and 9 were ordered according to the reference genome, because the estimated optimized order and the reference genome order showed discrepancies and the reference genome order usually reduced the map length compared to a locally estimated marker order based on recombination frequency. Map distances were estimated using the Kosambi mapping function (Kosambi, 1943) and 20-point multiple regression.

Phenotypic analysis and QTL mapping

Ten roots per F_3 family were phenotyped. Phenotypic data were analyzed in R (R Core Team, 2022) based on the following model:

$$y_{ijk} = g_i + e_j + b_{k(j)} + ge_{ij} + \varepsilon_{ijk}$$

All the terms in the model were fitted as random using the lmer function from the lme4 R package (Bates et al., 2015). In the model, y_{ijk} is the best linear unbiased predictor for the response variable of interest and $BLUP[g_{ijk}]=y_{ijk}$, g_i is the genotypic value for ith genotype, e_j is the environment (defined here as a combination of location and year), $b_{k(j)}$ is the block nested within environment, ge_{ij} is the genotype by environment interaction and ε_{ijk} is the model residual term $iid\ N{\sim}(0, \sigma_e^2)$. Phenotypic correlations were estimated using Microsoft Excel using the function correl.

Haley-Knott regression was used to perform composite interval mapping using the cim function in R/qtl (Broman et al., 2003) and to generate logarithm of odds (LOD) traces. Significant thresholds were estimated based on 1000 permutations at an alpha level of 0.05 using the cim function with the following parameters: number of marker covariates equal to 1, a window size of 5 cM and Haley-Knott regression method. Effect plots were constructed using the effectplot function from R/qtl.

RESULTS AND DISCUSSION

Genetic linkage map

The number of markers per chromosome ranged from 495 in chromosome 1 to 92 in chromosome 9 (Figure 1B). The average resolution was 0.4 centimorgans (cM) between markers and the maximum spacing between two markers was 15.2 cM, on chromosome 8 (Figure 1A). Overall, the linkage map included 2,028 markers and total map length was 750 cM which compares in size with other reported maps in carrot (Bannoud et al., 2019; Parsons et al., 2015). In addition, except for chromosome 1, the total length of the chromosome was equal to or less than 100 cM (Figure 1B). It may be advantageous to obtain genetic maps with smaller linkage groups because it can simplify the interpretation of the map and reduce the likelihood of errors due to falsely linked markers (Mackay, 2004). This result contrasts with the map reported by Ellison et al. (2017) where the average linkage map size was ~130 cM, with chromosome 6 at 164 cM. The map we present here also has the advantage that it was constructed based only using SNPs derived from GBS. Compared to maps built using markers including random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLPs), a SNP-based map is more reproducible and standardized, as SNPs are genotyped using high-throughput sequencing or microarray technologies. Our map showed similar resolution (measured in markers cM^{-1}) compared with high-quality genetic maps reported in Parsons et al. (2015) and Bannoud et al. (2019).

Correlations

The measured traits were positively correlated. The correlations ranged between 0.54 and 0.87. The correlation coefficient between shoulder width and mid width was 0.87. The correlation coefficient between shoulder width and tip width was 0.54 and the correlation coefficient between mid width and tip width was 0.68.

QTL discovery

Significant QTLs were detected for shoulder width (Figure 2A), mid width (Figure 2B) and tip width (Figure 2C). Two regions, one on chromosome 2 and one on chromosome 6 were found to be associated with shoulder width (Figure 2A). The QTL on chromosome 2 for shoulder width explained 11% percent of the phenotypic variation (Table 1). The closest marker to the QTL peak for shoulder width was SDCARV3_CHR2_34447749 which coincides with the QTL peak at chromosome 2 for tip width (Figure 2C). A single QTL on chromosome 6 was detected for mid width which explained around 18.7% of the phenotypic variation. Even though shoulder width and mid width are highly correlated phenotypically, the QTL detected for shoulder width (Figure 2A) is different from the QTL identified in the same chromosome for mid width (Table 1; Figure 2B). Overall, 3 unique QTL involved in root width control were identified. Two unique QTL on chromosome 6 and one QTL on chromosome 2 controls width at proximal and distal parts of the carrot root while a single QTL con chromosome 6 controls width at the middle of the root.

Average effect of allele substitution

Overall, progeny carrying two copies of the “A” allele derived from founder L1408 tended to have narrower widths (Figures 1C and 3). The “B” allele derived from founder W133 tended to increase the width of the root at all QTL (Figure 3A-D). One copy of the “B” allele at QTL on chromosome 2 increased shoulder and tip width by ~0.40 standard deviations. One copy of the “B” allele for the QTL on chromosome 6 increased mid width by approximately 0.6

standard deviations and explained about 18% of the phenotypic variation. Taken together, the data suggest that the allele from the W133 founder increased the width of the of the root profile measured at the shoulder, mid point and tip of the storage root. In addition, one identical QTL was found to control width at the tip and the shoulder the root; however, this QTL has a stronger additive effect for the trait shoulder width compared to tip width (Figure 3A, D).

Table 1. QTL parametrization of quantitative trait loci (QTL) for widths at 10, 50 and 90% of root length measured from the root crown in the biparental caroot population "L1408×W133". Additive effect was half the difference between the two homozygous genotypes. Dominance effect was the value of the heterozygote genotype minus additive effect.

Trait	CHR[a]	Marker at QTL	LOD score	PVE[b]	Additive effect	Dominance effect	Dominance degree
Shoulder width	2	SDCARV3_CHR2_34447749[c]	4.9	11.0	0.70	1.0	1.4
Shoulder width	6	SDCARV3_CHR6_4920456	4.7	13.5	1.1	-0.5	0.5
Mid width	6	SDCARV3_CHR6_7546679	5.5	18.7	1.1	0.27	0.24
Tip width	2	SDCARV3_CHR2_34447749[c]	4.9	11.8	0.8	0.94	1.1

[a]CHR = chromosome; [b]PVE = percent variance explained by the QTL; [c]Same QTL identified for two traits.

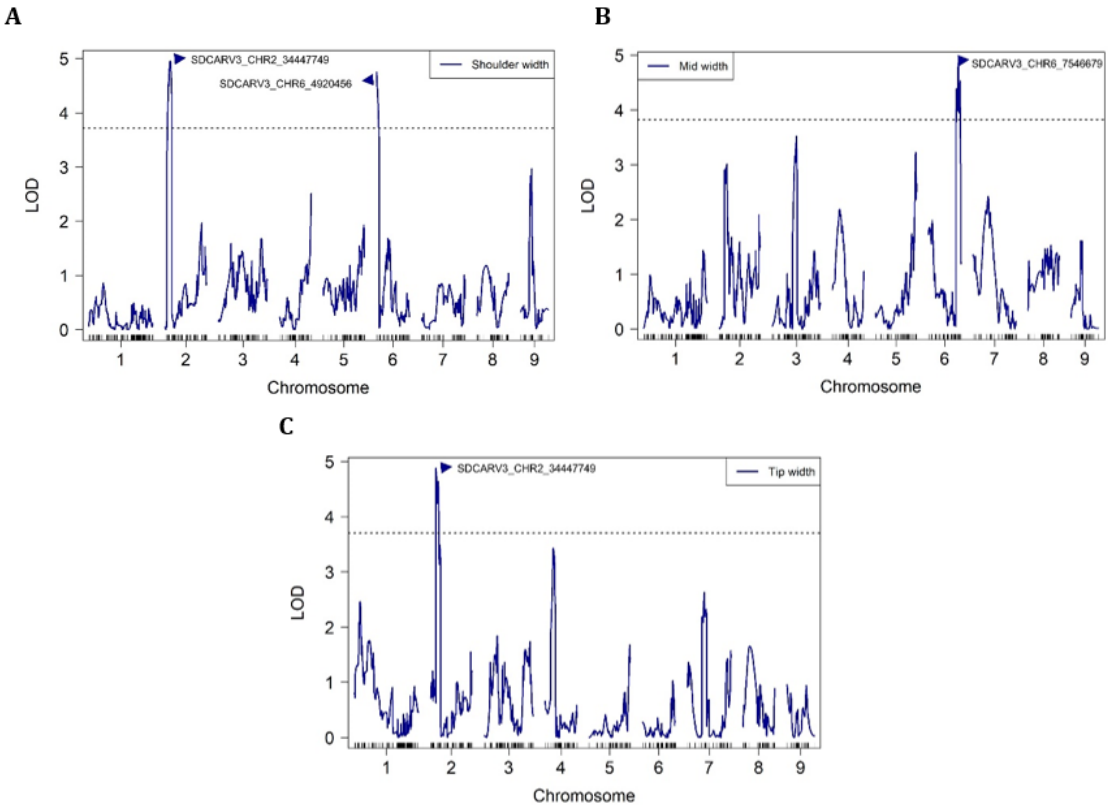


Figure 2. Logarithm of odds (LOD) trace for (A) width at 10% of the length, (B) width at 50% of the length and (C) width at 90% of the length for carrot population "L1408×W133". Parents of the population are inbreds L1408 and W133. LOD thresholds of 3.71, 3.82 and 3.70 for A, B and C respectively were obtained using 1000 permutations and are drawn as a dashed horizontal line.

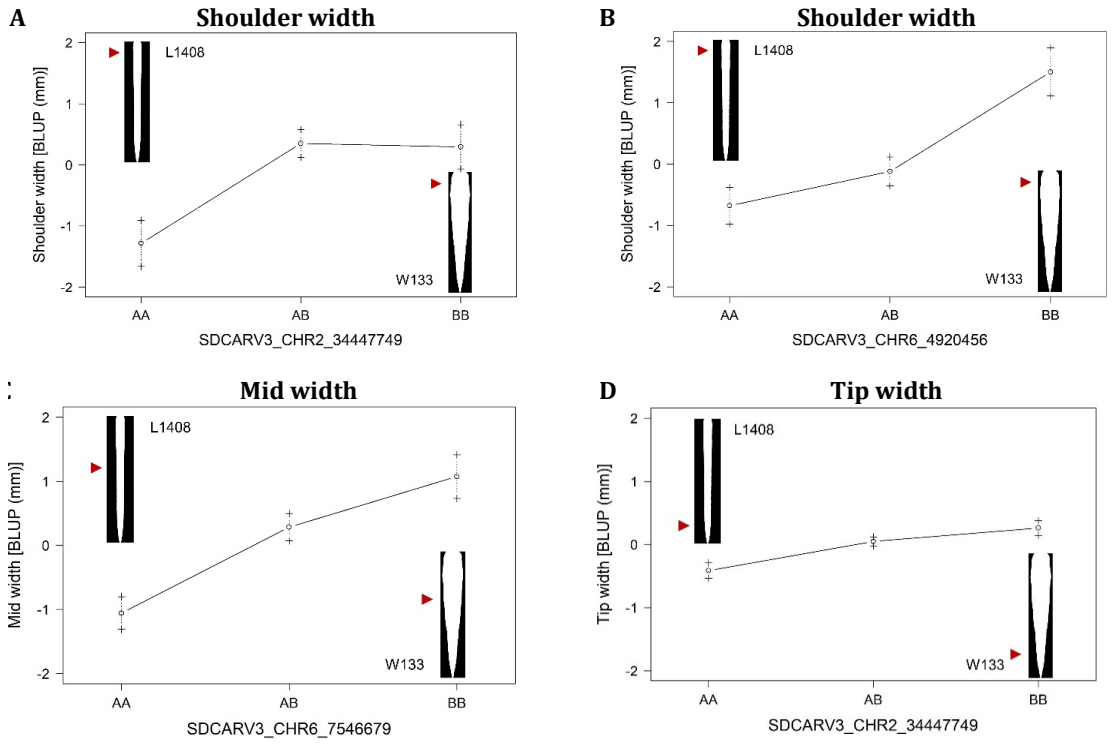


Figure 3. The effect of allele substitution of discovered quantitative trait loci (QTL) for (A-B) shoulder width (C) mid width and (D) tip width. For all panes, the standard error of the mean at each genotype is represented with "+" signs. The "AA" genotype represents the founder L1408, and "BB" represents the founder W133. A straightened mask of both parents is shown to illustrate the phenotype. The red arrow indicates where the width was measured in each response variable.

Dominance degree

Relatively high dominance values were detected for shoulder width and tip width. For shoulder width, the QTL at chromosome 2 exhibits complete dominance (Table 1; Figure 3A). The QTL at chromosome 2 for tip width (identical QTL) also exhibited complete dominance (Table 1). The rest of the QTL show an additive behavior where, as the number of copies of the "B" allele increases, the width of the root profile increases linearly (Table 1; Figure 3B-D).

CONCLUSIONS

The results provide evidence that the variation in root width in carrot is under genetic control with relatively simple inheritance. The width traits studied here are controlled by QTL on regions of chromosomes 2 and 6. In addition, the effect of allele substitution suggests that the width of the root increases with the number of copies of the "B" allele from founder W133. Taken together these results improve the current understanding of carrot root shape traits and market class. Hopefully, work like this will lead to marker development for use in carrot breeding programs to facilitate breeding for different shapes, and accessing germplasm available across market classes which is typically excluded due to the added effort of breeding for the shape of interest within the desired market class.

Literature cited

Banga, O. (1957). Horticultural Plant Breeding in the Netherlands. Mededeling No. 108 (Instituut voor de Veredeling van Tuinbouwgewassen, IVT), https://edepot.wur.nl/392199.

Bannoud, F., Ellison, S., Paolinelli, M., Horejsi, T., Senalik, D., Fanzone, M., Iorizzo, M., Simon, P.W., and Cavagnaro, P.F. (2019). Dissecting the genetic control of root and leaf tissue-specific anthocyanin pigmentation in carrot (*Daucus carota* L.). Theor Appl Genet *132* (*9*), 2485–2507 https://doi.org/10.1007/s00122-019-03366-5. PubMed

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Softw. *67* (*1*), 1–48 https://doi.org/10.18637/jss.v067.i01.

Brainard, S.H., Bustamante, J.A., Dawson, J.C., Spalding, E.P., and Goldman, I.L. (2021). A digital image-based phenotyping platform for analyzing root shape attributes in carrot. Front Plant Sci *12*, 690031 https://doi.org/10.3389/fpls.2021.690031. PubMed

Broman, K.W., Wu, H., Sen, S., and Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics *19* (*7*), 889–890 https://doi.org/10.1093/bioinformatics/btg112. PubMed

Ellison, S., Senalik, D., Bostan, H., Iorizzo, M., and Simon, P. (2017). Fine mapping, transcriptome analysis, and marker development for Y_2, the gene that conditions β-carotene accumulation in carrot (*Daucus carota* L.). G3 (Bethesda) *7* (*8*), 2665–2675 https://doi.org/10.1534/g3.117.043067. PubMed

Endelman, J. (2023). MapRtools: Tools for Genetic Mapping Teaching and Research (R package version 0.30), https://github.com/jendelman/MapRtools/.

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., and Buckler, E.S. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One *9* (*2*), e90346 https://doi.org/10.1371/journal.pone.0090346. PubMed

Goldman, I.L. (1996). A list of germplasm releases from the University of Wisconsin Carrot Breeding Program, 1964-1992. HortScience *31* (*5*), 882–883 https://doi.org/10.21273/HORTSCI.31.5.882.

Iorizzo, M., Ellison, S., Senalik, D., Zeng, P., Satapoomin, P., Huang, J., Bowman, M., Iovene, M., Sanseverino, W., Cavagnaro, P., et al. (2016). A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. Nat Genet *48* (*6*), 657–666 https://doi.org/10.1038/ng.3565. PubMed

Iorizzo, M., Bostan, H., Ellison, S., et al. (2020). Improved hybrid de novo genome assembly, gene prediction and annotation of carrot (*Daucus carota*). Paper presented at: Plant and Animal Genome Proceedings: *Apiaceae* Workshop (San Diego, CA).

Kosambi, D.D. (1943). The estimation of map distances from recombination values. Ann. Eugen. *12* (*1*), 172–175 https://doi.org/10.1111/j.1469-1809.1943.tb02321.x.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25* (*16*), 2078–2079 https://doi.org/10.1093/bioinformatics/btp352. PubMed

Mackay, T.F. (2004). The genetic architecture of quantitative traits: lessons from *Drosophila.* Curr Opin Genet Dev *14* (*3*), 253–257 https://doi.org/10.1016/j.gde.2004.04.003. PubMed

Parsons, J., Matthews, W., Iorizzo, M., Roberts, P., and Simon, P. (2015). *Meloidogyne incognita* nematode resistance QTL in carrot. Mol. Breed. *35* (*5*), 114 https://doi.org/10.1007/s11032-015-0309-2.

R Core Team. (2022). R-4.2.2 (Vienna, Austria: R Foundation for Statistical Computing), https://cran.r-project.org/bin/windows/base/.