CarrotOmics: a genetics and comparative genomics database for carrot (*Daucus carota*)

William R. Rolling [1,2], **Douglas Senalik** [1,2], **Massimo Iorizzo** [3], **Shelby Ellison** [2], **Allen Van Deynze** [4] and **Philipp W. Simon** [1,2,*]

[1]Vegetable Crop Research Unit, USDA-ARS, Moore Hall, 1575 Linden Drive, Madison, WI 53706-1514, USA
[2]Department of Horticulture, University of Wisconsin-Madison, Moore Hall, 1575 Linden Drive, Madison, WI 53706-1514, USA
[3]Department of Horticultural Science and Plants for Human Health Institute, North Carolina State University, NC Research Campus, 600 Laureate Way, Kannapolis, NC 28081, USA
[4]College of Agricultural & Environmental Sciences, Seed Biotechnology Center, University of California-Davis, 150 Mrak Hall, One Shields Avenue, Davis, CA 95616, USA

*Corresponding author: Tel: +608-262-1248; Fax: +608-262-4743; Email: philipp.simon@usda.gov

Abstract

CarrotOmics (https://carrotomics.org/) is a comprehensive database for carrot (*Daucus carota* L.) breeding and research. CarrotOmics was developed using resources available at the MainLab Bioinformatics core (https://www.bioinfo.wsu.edu/) and is implemented using Tripal with Drupal modules. The database delivers access to download or visualize the carrot reference genome with gene predictions, gene annotations and sequence assembly. Other genomic resources include information for 11 224 genetic markers from 73 linkage maps or genotyping-by-sequencing and descriptions of 371 mapped loci. There are records for 1601 Apiales species (or subspecies) and descriptions of 9408 accessions from 11 germplasm collections representing more than 600 of these species. Additionally, 204 Apiales species have phenotypic information, totaling 28 517 observations from 10 041 biological samples. Resources on CarrotOmics are freely available, search functions are provided to find data of interest and video tutorials are available to describe the search functions and genomic tools. CarrotOmics is a timely resource for the Apiaceae research community and for carrot geneticists developing improved cultivars with novel traits addressing challenges including an expanding acreage in tropical climates, an evolving consumer interested in sustainably grown vegetables and a dynamic environment due to climate change. Data from CarrotOmics can be applied in genomic-assisted selection and genetic research to improve basic research and carrot breeding efficiency.

Database URL: https://carrotomics.org/

Introduction

Well-curated 'omic' data in an openly accessible database are likely the best way for plant scientists to disseminate information for research and breeding programs (1). This is important because genomic data are not biologically meaningful without analyses and interpretation, and the raw data are not easily provided through traditional publications. Plant genomic databases have been instrumental for improved gene annotation and understanding of genome structure. This is because they make community engagement straightforward, and a broad group of researchers can access the data to manually curate computer predictions using experimental evidence for gene function (2–4). Arabidopsis is the gold standard for plant genomics, and with community engagement, over 94% of genes have a predicted function and 48% of genes have some experimental evidence for predicted function (5).

Plant genomes harbor tremendous diversity in size, structure and evolutionary history (5). As such, having genetically related species with high-quality genome assemblies can be helpful for gene prediction when assembling a new genome. Currently, many economically important families have high-quality reference genomes and databases available,

including maize (6) and rice (7) for Poaceae and Soybean (8) and Medicago (9) for Fabaceae. Rice is a great example of how a high-quality reference genome can be leveraged in the assembly of additional related genomes, and more than 100 Poaceae genomes have been sequenced following the rice genome (5).

Databases are useful for more than just reference genome analysis. For example, databases have been used to combine results of genetic mapping studies and produce more accurate estimations of the genetic positions of quantitative trait loci (QTL) using Meta-QTL analysis. This type of analysis is more practical when mining results from a database as highlighted in examples from cotton (10), corn (11) and soybean (12). Databases are also useful to generate hypotheses about which genes underlie important traits. For example, the Genome Database for Rosaceae (13) has been used as a source of data to study important gene families such as *MADS-box* in plum (14) and *NF-Y* genes in peach (15). Similarly, the CottonGen genome database (16) has been used to identify positional candidate genes following QTL mapping (17) or provide primer sequences for transcriptomic analyses of positional candidate genes (18).

In this manuscript, we describe the CarrotOmics (https://carrotomics.org/) database, which was built to disseminate carrot genomic information and encourage community involvement in carrot genetic research, breeding and germplasm study. A whole-genome sequence for carrot (*Daucus carota* L.) was produced in 2014 (19), and these sequences were initially available for download via website, but no broad functioning, publicly available genome database has been maintained to organize carrot genomics data. CarrotOmics was developed to provide unrestricted access to carrot genomic resources such as the high-quality reference genome sequence, gene prediction, annotation and expression data from an orange-colored, Nantes-type, doubled-haploid carrot (20). CarrotOmics also stores data from genetic mapping, phenomic, functional genomics and taxonomic studies.

Carrot is one of many economically important plant Apiaceae species, and the availability of the carrot genome at CarrotOmics will make future Apiaceae reference genome assembly and comparative genomic analyses more straightforward. For example, celery is closely related to carrot, and the celery reference genome was assembled using identity and coverage thresholds determined based on methods from the carrot genome assembly. The celery reference genome also used the carrot genome to consider synteny, genome size and gene content (21). A genome sequence is also available for the Apiales crops coriander (22) and ginseng (23), and the combination of data from these genomes can provide a basis for future comparative genomic studies.

CarrotOmics is also intended as a resource for carrot breeders. Carrot germplasm improvement is of interest because carrots are highly nutritious (24) and will help contribute to nutritional security in the 21st century. In addition to the fresh vegetable market, carrots are also the raw material for valuable commodities including juices, natural food coloring and baby food. The nutritional and monetary value has encouraged additional carrot cultivation, and production has increased 6-fold from 1970 to 2018, placing carrots among the 10 most produced vegetables, globally (25). This increasing production trend is likely to continue as East Asian countries are growing more carrots (24), and new production regions are being established in subtropical regions (25), both of which will require novel cultivars adapted for those agroecosystems.

Many of the new production regions have challenging climates and diverse agronomic practices for carrot production. New cultivars adapted for these agroecosystems are required. However, with each generation of seed production and selection requiring a full year, the current carrot breeding cycles take approximately 7 years to develop new lines (26). Carrot breeding programs can increase breeding efficiency or genetic gain in a breeding cycle through the application of genomic-assisted selection. Data curated on CarrotOmics can be helpful to identify variants for marker-assisted selection or produce genomic selection models.

With these goals in mind, carrot researchers have been collecting large genotypic and phenotypic datasets for many diverse accessions. One application has been to study the consequences of domestication and the population structure in carrot landraces and cultivars. These studies have found gene flow from wild populations to cultivars (27); genetic clustering of more related groups including wild, East Asian, Central Asian, Western (Europe and the Americas) and modern Western germplasm (28) and even genetic differences between market class of carrots (29).

In addition to being genotyped, many accessions have been phenotyped to identify genetic markers linked to economically important traits via genome-wide association studies (GWAS). For example, a novel carotenoid accumulation locus was identified by scoring for the presence or absence of orange pigmentation in more than 600 accessions (27). GWAS results complement mapping completed in more than 90 biparental populations, identifying 354 QTL (Spring 2022) for traits such as carotenoid accumulation (30), anthocyanin accumulation (31) and nematode resistance (32). Recent mapping studies have been improved because markers can be anchored to physical positions and positional candidate genes identified with the availability of the high-quality reference genome (20, 30–32).

The examples outlined here and many more studies have provided a large base for genomic-assisted breeding. CarrotOmics was developed to make that data and the reference genome available to breeders and the Apiaceae research community.

Database summary

(i) Genome sequences available for download
   (a) Carrot reference genome—https://carrotomics.org/bio_data/290811
   (b) Other Apiaceae and organellar genomes—https://carrotomics.org/node/175
   (c) Transcriptomic data—https://carrotomics.org/node/94
(ii) A comprehensive list of Apiaceae germplasm accessions stored by germplasm centers
   (d) Germplasm search—https://carrotomics.org/search/germplasm
(iii) Descriptions, images and phenotypic evaluations of carrot germplasm
   (e) Phenotype search—https://carrotomics.org/search/qualitative_traits
(iv) Genomic & data search tools
   (f) JBrowse—https://carrotomics.org/jbrowses
   (g) BLAST—https://carrotomics.org/blast
   (h) MegaSearch—https://carrotomics.org/tripal_megasearch
(v) Published carrot linkage maps and QTL.
   (i) Genetic maps—https://carrotomics.org/search/featuremap/list
   (j) QTLs—https://carrotomics.org/search/qtl
(vi) User help guides
   (k) User manual—https://carrotomics.org/userManual
   (l) News—https://carrotomics.org/news/community
   (m) Contact information—https://carrotomics.org/contact

CarrotOmics framework

CarrotOmics is implemented using Tripal (33), a suite of Drupal modules for creating biological websites. Internal storage

uses the GMOD/Chado v1.3 schema (34). CarrotOmics supports Tripal v3 RESTful web services (https://carrotomics.org/web-services/). This interface allows access to all publicly available data programmatically and allows communication with other Tripal sites. CarrotOmics also includes a BrAPI Tripal module providing a standardized application programming interface. Breeding API (BrAPI) (https://brapi.org/) helps different plant breeding databases communicate inputs, outputs, structure and how data are passed from the database website (35).

The CarrotOmics database was developed using Tripal Core modules and custom modules developed by the MainLab Bioinformatics Crop Genomics core (https://www.bioinfo.wsu.edu/). Custom modules from the MainLab include the Tripal MegaSearch (36), Tripal Map (37), Chado display, Chado search and MCL (38) and an Expression Module (39). These custom modules may be downloaded from https://gitlab.com/mainlabwsu. Custom features implemented for CarrotOmics (https://github.com/carrotomics/) are housed in a custom module or in modifications to existing modules (i.e. chado_search, mcl).

Other crop databases that were developed using MainLab resources include the Rosaceae (13), Cotton (16), Citrus (40), Pulse (41) and Vaccinium (https://www.vaccinium.org/). The homepages of these databases serve as hubs to access all data and tools.

The CarrotOmics homepage (Figure 1) includes community-specific news (Figure 1A), CarrotOmics updates and a *Quick Start* with direct links to data and tools (Figure 1B). Pull-down menus at the top of the homepage provide access to all database tools and data (Figure 1C). Additional news, database background information and upcoming or recent updates can be found in the 'General' pull-down menu (Figure 1D). Questions about the database or tools can be found in 'User Manual' or 'Contact' pages and in the 'Help' pull-down menu (Figure 1D).

Genome & transcript data

A Nantes-type doubled-haploid carrot (Table 1) was sequenced to produce a high-quality reference genome, with the assembly accounting for >90% of the estimated genome size and 32 113 predicted genes (20). The full genome sequence, mRNA, exon, coding sequence (CDS), protein sequences and predicted gene function are available for download (Table 1) from either the 'Genome Assemblies' or 'Data Download' links in the 'Data' pull-down menu. Additional genomic features are available from the 'Data Download' (Table 1) including the predicted pathogen receptor genes (42) and a catalog of single nucleotide polymorphisms (SNPs), mobile elements, tandem repeats and gaps in the assembly.

Gene expression data are also available via the 'Transcripts and Expression Data' link (Table 1) in the 'Data' pull-down menu. Carrot researchers can provide results of transcriptomic studies, and there are currently (Spring 2022) data available from four studies (20, 43–45). The transcriptomic
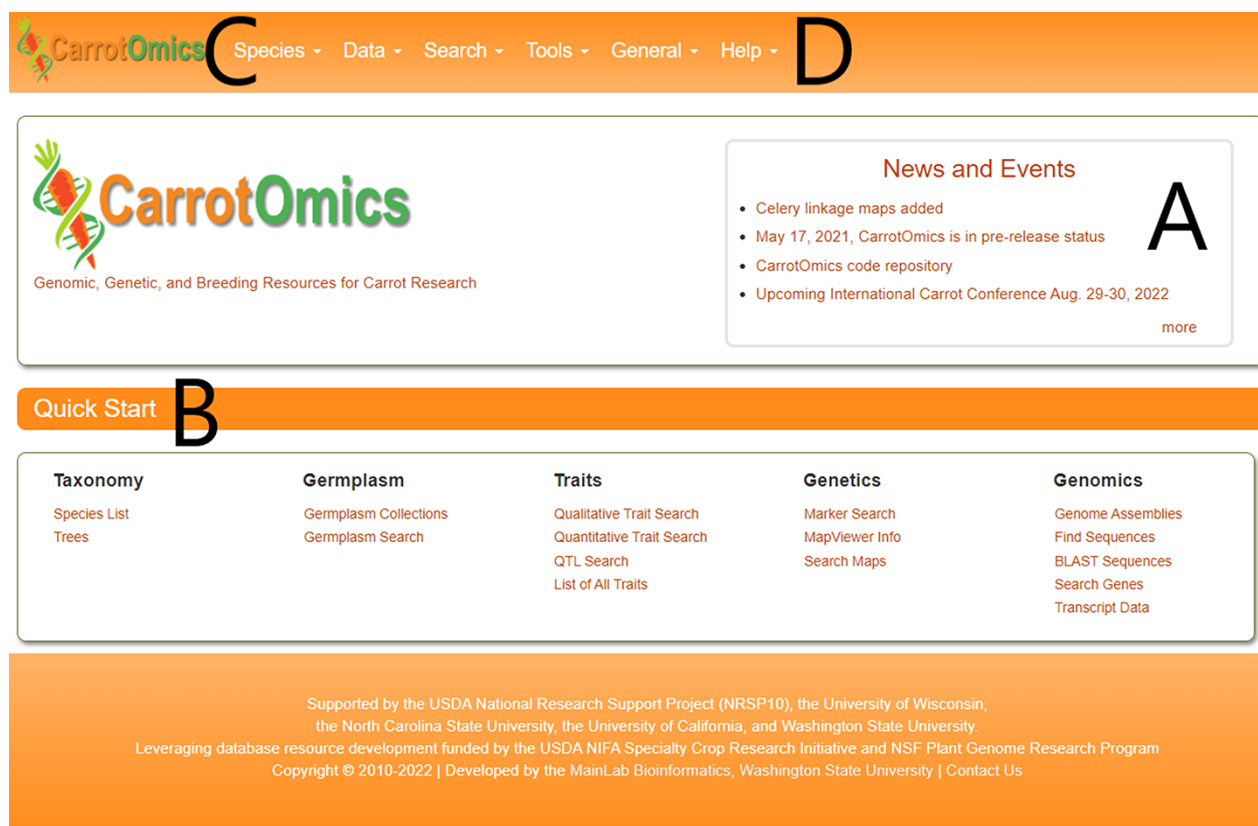
Figure 1. The CarrotOmics homepage includes (A) a news and events banner. (B) a Quick Start menu to directly access tools or data. (C) Pull-down menus to access all parts of the database. (D) A 'Help' pull-down menu to access user manuals and tutorials.

Table 1. Summary of data related to the reference genome

Description	Search location
Reference genome files	
Genome sequence[a]	https://carrotomics.org/bio_data/290811
Gene annotations	https://carrotomics.org/file/188230
Coding sequence	https://carrotomics.org/file/188228
Polypeptide sequence	https://carrotomics.org/file/188229
Results of reference genome analyses	
Tandem repeats	https://carrotomics.org/Analysis/177694
Genetic variants	https://carrotomics.org/Analysis/177695
Gene predictions	https://carrotomics.org/Analysis/79607
Pathogen receptor genes	https://carrotomics.org/Analysis/177878
Gaps & gap fill	https://carrotomics.org/Analysis/177692
	https://carrotomics.org/Analysis/177698
Transcriptomes	
Iorizzo *et al.* 2011[b]	https://carrotomics.org/bio_data/169050
DCAR V1.0 gene prediction[c]	https://carrotomics.org/bio_data/79607
NCBI	https://carrotomics.org/bio_data/70510

[a]fasta or gff3 files containing data from the reference genome.
[b]Name of study on *transcriptomics page* (https://carrotomics.org/node/94).
[c]Transcriptomes assembled to the reference genome currently stored at CarrotOmics.

Table 2. Summary of 'omic' data stored at CarrotOmics

Description	Records	Search tool location
Markers	11 224	https://carrotomics.org/search/markers
Genetic maps	73	https://carrotomics.org/search/featuremap/list
QTL	354	https://carrotomics.org/search/qtl
Publication	32 662	https://carrotomics.org/find/publications
Germplasm records	9408	https://carrotomics.org/search/germplasm
Phenotypes	28 517	https://carrotomics.org/search/qualitative_traits
Phylogenetic trees	60	https://carrotomics.org/node/73

data can be searched by gene or transcript ID using the 'Gene and Transcript Search' tool found in the 'Search' pull-down menu.

Germplasm & phenotypic data

Germplasm, genomic and phenotypic data are also deposited on CarrotOmics. A summary is in the 'Data' pull-down menu (Figure 1C), 'Data Overview' link. Currently (Spring 2022), there are 9408 germplasm accession records from 1601 species, 73 genetic maps, 11 224 genetic markers, 371 mapped loci, 28 517 phenotypic measures and 32 662 publications (Table 2).

Germplasm records include data for 881 accessions of cultivated carrot (*D. carota* subsp. *sativus* or var. *sativus*). Although carrot is the focal species, CarrotOmics also serves as a data resource for other plant species from the Apiales order. Currently, seven families are represented: (i) Apiaceae, (ii) Araliaceae, (iii) Griseliniaceae, (iv) Myodocarpaceae, (v) Pennantiaceae, (vi) Pittosporaceae and (vii) Torricelliaceae. All families can be searched in the 'Species' pull-down menu, 'Organism Summary' link (Table 2).

Species information also includes a National Center for Biotechnology Information (NCBI) Taxonomic species tree found in the 'CarrotOmics Taxonomy Tree' link and phylogenetic relationships within the *Daucus* genus found in the 'Phylogenetic Trees' link (Table 2). The combination of wild carrot, carrot landraces and crop wild relatives provides germplasm accessions to screen for novel traits to improve abiotic stress tolerance (46) and disease resistance traits or mine data from previously screened accessions [e.g. see Alternaria resistance scores (47)].

Those accessions that have been phenotyped can be queried in the 'Quantitative' or 'Qualitative Trait Search' found in the 'Search' pull-down menu (Table 2). A total of 53 qualitative measurements and 70 quantitative measurements from the United States Department of Agriculture (USDA)-National Plant Germplasm System are stored at CarrotOmics. Three additional quantitative traits including Alternaria resistance (47), canopy height (48) and flavor scores from in-house germplasm screening projects are also stored on CarrotOmics.

Phenotypic data will continue to be added and updated by database curators, and individual researchers may contribute peer-reviewed or biologically replicated preprint data via the 'Data Submission' link found in the 'Data' pull-down menu.

Genetic mapping & marker data

There are 73 genetic maps currently stored on CarrotOmics, and users will be able to submit the results of their (future) mapping studies via the 'Data Submission' page. Existing genetic maps are searchable at the 'Map Search' link found in the 'Search' pull-down menu (Table 2). Results of map searches are organized into a table with a link to the map and descriptions of the mapping population including parental accessions, mapping population size and generation. A related function in the 'Search' pull-down menu is 'Marker Search' (Table 2). Markers can be searched by (sub)species, marker type, mapping population, linkage group and, if available, the trait that the marker is associated with from the mapping study or association analysis. Published marker-trait associations can also be found with the 'QTL and MTL Search' (Table 2).

Tools & references

There are many search tools for specific types of data on CarrotOmics. However, there is also the 'MegaSearch' (36) tool in the 'Search' pull-down menu. 'MegaSearch' is a broad functioning search module that can be used to query the entirety of the database. Users can select data types of interest, filter the data, customize fields and download data from a single interface.

There are four analysis tools currently available via 'Tools' pull-down menu including (i) 'BLAST+', (ii) 'JBrowse', (iii) 'MapViewer' and (iv) 'Expression Heatmap'. The 'BLAST+' tool (49) links to blastn, blastx, tblastn and blastp analyses with options to select a database and specify advanced BLAST options.

A useful tool to visualize genetic variants and associated linkage maps is the 'MapViewer' tool, which includes a video tutorial (Figure 2). The tool can be used to visualize and compare genetic maps from different populations. For example, the 'Y2' locus related to carotenoid accumulation (45) was mapped to a region on linkage group 7 (corresponds
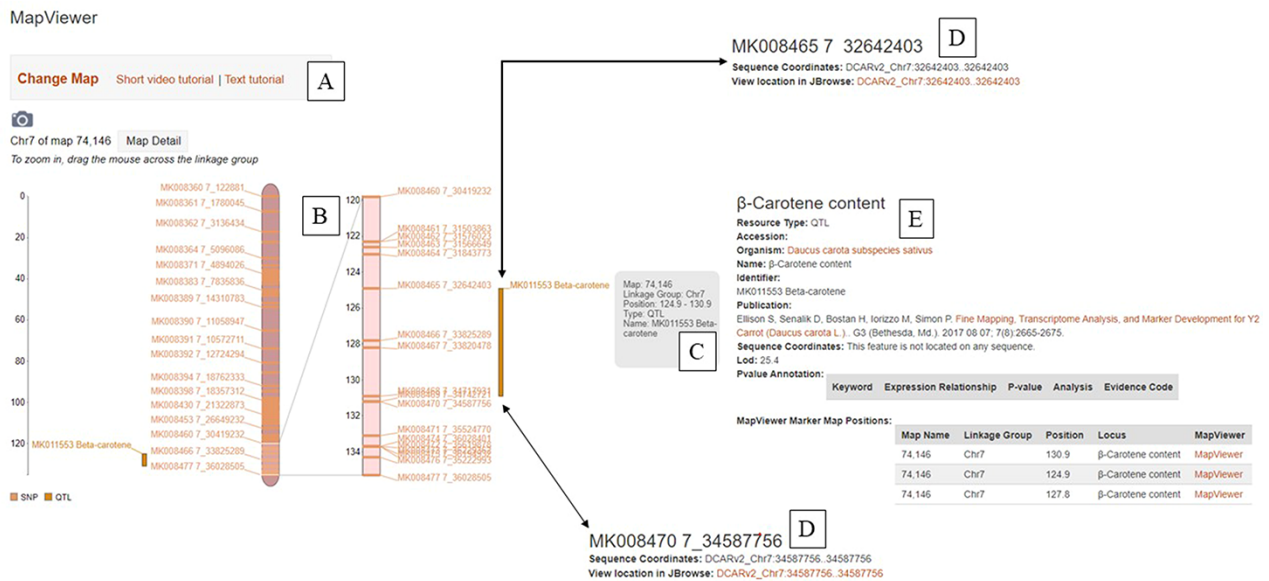
Figure 2. The 'MapViewer' tool available at CarrotOmics. (A) Links to video and text tutorials on use of 'MapViewer'. (B) Linkage map display from Y2 QTL mapping study. (C) Information about QTL displayed on interactive 'MapViewer' plot. (D) Physical positions for markers flanking Y2 locus. (E) Information about the study used to create the linkage map.
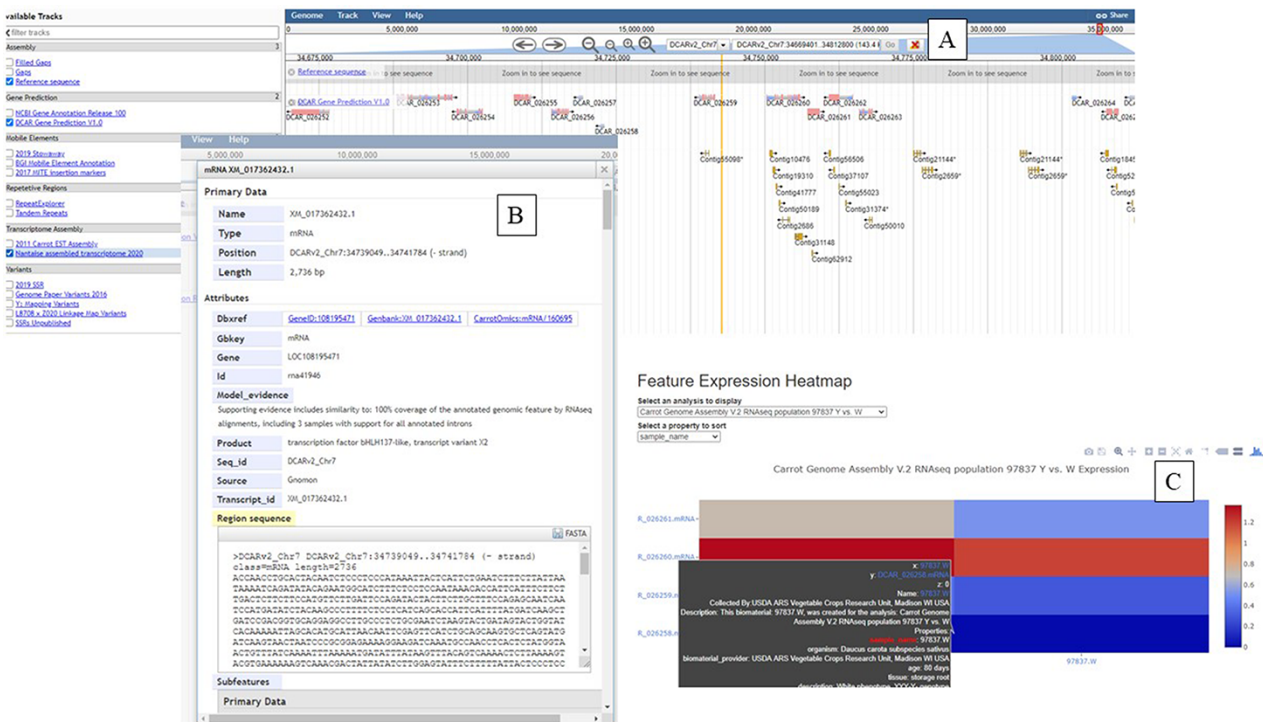

Figure 3. The 'JBrowse' tool available at CarrotOmics. (A) 'JBrowse' console with gene annotation tracks displayed. (B) Information about selected gene. (C) Heatmap display from transcriptome studies available from 'Expression Heatmap' tool.

to Chr. 07) (Figure 2B). Information about this QTL can be accessed directly in MapViewer (Figure 2C) or through a link to a page about the QTL (Figure 2D). The physical positions of markers flanking the QTL can also be accessed through links on MapViewer (Figure 2E).

JBrowse can be used to browse genome assemblies (50) (Figure 3). The carrot genome 'JBrowse' currently contains 17 tracks to visualize sequences, gaps and gene annotations from NCBI and the reference genome publication (20). There are tracks for mRNA data, including those used in the genome annotation (20) and from a more recent transcriptome study of carotenoid synthesis (44). Five tracks are available for notable genome features including transposons and repetitive elements. Another set of five tracks represents genetic variants used in genetic mapping or identified in the reference genome study (20).

The final tool is an 'Expression Heatmap' to visualize transcript expression data. Specific carrot transcripts of interest can be searched based on identifier or annotation (Figure 3C). Once transcripts of interest are selected, a heatmap displaying relative expression levels is displayed and available for export. These plots can be customized by selecting a specific transcriptome analysis or type of tissue and biomaterial used.

Conclusions and future directions

CarrotOmics was developed to serve as a hub for carrot breeding and Apiaceae genomics research. The initial release of CarrotOmics includes a high-quality reference genome, a plethora of genetic maps, germplasm descriptions and gene expression data. The database will be updated as new or improved genome assemblies are produced, more genotypic and phenotypic data become available and additional transcriptomic studies are completed.

We anticipate this database will be maintained in a similar manner as those other databases hosted by the MainLab, including CottonGen (16), Genome Database for Rosaceae (13), Citrus Genome Database (40), Pulse Crop Database (41) and Genome Database for Vaccinium (https://www.vaccinium.org/). Specifically, the USDA carrot genomic and breeding lab has been supported for over 50 years, and carrots continue to be an important vegetable crop with USDA-Vegetable Crop Research Unit support, external USDA-National Institute of Food and Agriculture (NIFA) grants and cooperative agreements with industry partners. We anticipate CarrotOmics will become an integral part of the USDA-Carrot research crop program, specifically with the addition of the breeding management system (51), which is BrAPI complaint. This function will be particularly useful as the USDA moves to digital agricultural breeding platforms (i.e. Breeding Insights).

CarrotOmics provides novel data and functionality not currently available in other Apiaceae databases including descriptions and phenotypic data from the USDA-Germplasm Resources Information Network (GRIN) germplasm collections for Apiales species, Apiales publication search tools and phylogenetic analyses. CarrotOmics also describes the results of QTL mapping studies including information about the physical positions of QTL, the mapping populations and nearby markers. Researchers from other Apiaceae species are invited to upload data to CarrotOmics and use tools not yet available on other databases including MapViewer, Expression Heatmaps and MegaSearch. We have added google analytics to track page visits and download counts to identify which areas of the database are most used and put database resources toward developing those tools.

In time, we also hope to provide links to data analysis pipelines and tutorials on how to use the deposited data for genomic studies such as genomic prediction, genome-wide association analyses, Mendelian and biparental mapping and calculation of genetic values of breeding lines. CarrotOmics will also house nontechnical resources for outreach events and educational activities, cultivars released, seed availability and video demonstrations for the culinary uses of carrot, as they become available. We hope these resources as well as the genomic tools will support an active global community of breeders, molecular biologists, growers and horticulturists.

Author contributions

D.S. formatted and uploaded data to the CarrotOmics database and developed the CarrotOmics Tripal module. W.R.R. assisted with database formatting and wrote the original version of the manuscript. M.I., S.E., A.V.D., D.S. and P.W.S. conceived the project and provided technical suggestions for the database. All authors read, revised and approved the final version of this manuscript.

Conflict of interest

The authors declare no competing interest.

Data availability

All data described in this manuscript is available at https://www.carrotomics.org/.

References

1. Southwood,D. and Ranganathan,S. (2019) Genome databases and browsers. In: Ranganathan S, Nakai K, Schonbach C (ed.). *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier, Amsterdam, pp. 251–256. 10.1016/B978-0-12-809633-8.20754-1.

2. Veeckman,E., Ruttink,T. and Vandepoele,K. (2016) Are we there yet? Reliably estimating the completeness of plant genome sequences. *Plant Cell*, **28**, 1759–1768. 10.1105/tpc.16.00349.

3. Vaattovaara,A., Leppälä,J., Salojärvi,J. *et al.* (2019) High-throughput sequencing data and the impact of plant gene annotation quality. *J. Exp. Bot.*, **70**, 1069–1076. 10.1093/jxb/ery434.

4. Kress,W.J., Soltis,D.E., Kersey,P.J. *et al.* (2022) Green plant genomes: what we know in an era of rapidly expanding opportunities. *PNAS*, **199**, e2115640118. 10.1073/pnas.2115640118.

5. Kersey,P.J. (2019) Plant genome sequences: past, present, future. *Curr. Opin. Plant Biol.*, **48**, 1–8. 10.1016/j.pbi.2018.11.001.

6. Woodhouse,M.R., Cannon,E.K., Portwood,J.L. *et al.* (2021) A pan-genomic approach to genome databases using maize as a model system. *BMC Plant Biol.*, **21**, 385. 10.1186/s12870-021-03173-5.

7. Sakai,H., Lee,S.S., Tanaka,T. *et al.* (2013) Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol.*, **54**, e6. 10.1093/pcp/pcs183.

8. Grant,D., Nelson,R.T., Cannon,S.B. *et al.* (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.*, **38**, D843–D846. 10.1093/nar/gkp798.

9. Krishnakumar,V., Kim,M., Rosen,B.D. *et al.* (2015) MTGD: the Medicago Truncatula Genome Database. *Plant Cell Physiol.*, **56**, e1. 10.1093/pcp/pcu179.

10. Said,J.I., Lin,Z., Zhang,X. *et al.* (2013) A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. *BMC Genom.*, **14**, 776. 10.1186/1471-2164-14-776.

11. Liu,R., Zhang,H., Zhao,P. *et al.* (2012) Mining of candidate maize genes for nitrogen use efficiency by integrating gene expression and QTL data. *Plant Mol. Biol. Rep.*, **30**, 297–308. 10.1007/s11105-011-0346-x.

12. Rolling,W., Lake,R., Dorrance,A.E. *et al.* (2020) Genome-wide association analyses of quantitative disease resistance in diverse sets of soybean [*Glycine max* (L.) Merr.] plant introductions. *PLoS One*, **15**, e0227710. 10.1371/journal.pone.0227710.

13. Jung,S., Lee,T., Cheng,C.H. *et al.* (2019) 15 years of GDR: new data and functionality in the Genome Database for Rosaceae. *Nucleic Acids Res.*, **47**, D1137–D1145. 10.1093/nar/gky1000.

14. Quesada-Traver,C., Guerrero,B.I., Badenes,M.L. *et al.* (2020) Structure and expression of bud dormancy-associated MADS-box genes (DAM) in European plum. *Front. Plant Sci.*, **11**, 1288. 10.3389/fpls.2020.01288.

15. Li,M., Li,G., Liu,W. *et al.* (2019) Genome-wide analysis of the NF-Y gene family in peach (*Prunus persica* L.). *BMC Genom.*, **20**, 612. 10.1186/s12864-019-5968-7.

16. Yu,J., Jung,S., Cheng,C.H. *et al.* (2021) CottonGen: the community database for cotton genomics, genetics, and breeding research. *Plants*, **10**, 2805. 10.3390/plants10122805.

17. Yuan,Y., Zhang,H., Wang,L. *et al.* (2019) Candidate quantitative trait loci and genes for fiber quality in *Gossypium hirsutum* L. detected using single- and multi-locus association mapping. *Ind. Crops Prod.*, **134**, 356–369. 10.1016/j.indcrop.2019.04.010.

18. Liu,W., Song,C., Ren,Z. *et al.* (2020) Genome-wide association study reveals the genetic basis of fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *BMC Plant Biol.*, **20**, 1–13. 10.1186/s12870-020-02611-0.

19. Xu,Z.S., Tan,H.W., Wang,F. *et al.* (2014) CarrotDB: a genomic and transcriptomic database for carrot. *Database*, **2014**, bau096. 10.1093/database/bau096.

20. Iorizzo,M., Ellison,S., Senalik,D. *et al.* (2016) A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat. Genet.*, **48**, 657–666. 10.1038/ng.3565.

21. Li,M.Y., Feng,K., Hou,X.L. *et al.* (2020) The genome sequence of celery (*Apium graveolens* L.), an important leaf vegetable crop rich in apigenin in the Apiaceae family. *Hort. Res.*, **7**, 9. 10.1038/s41438-019-0235-2.

22. Song,X., Nie,F., Chen,W. *et al.* (2020) Coriander Genomics Database: a genomic, transcriptomic, and metabolic database for coriander. *Hort. Res.*, **7**, 55. 10.1038/s41438-020-0261-0.

23. Jayakodi,M., Choi,B.S., Lee,S.C. *et al.* (2018) Ginseng Genome Database: an open-access platform for genomics of Panax ginseng. *BMC Plant Biol.*, **18**, 62. 10.1186/s12870-018-1282-9.

24. Simon,P.W. (2019) Economic and academic importance. In: Simon P, Iorizzo M, Grzebelus D *et al.* (eds). *The Carrot Genome*. Springer, Berlin, pp. 1–8. 10.1007/978-3-030-03389-7_1.

25. Simon,P.W. (2000) Domestication, historical development, and modern breeding of carrot. In: Janick J (ed). *Plant Breeding Reviews*, Vol. 19. Wiley, Hoboken, pp. 157–190.

26. Peterson,C.E. and Simon,P.W. (1986) Carrot breeding. In: Basset MJ (ed). *Breeding Vegetable Crops*. AVI Publishing, Westport, 321–356.

27. Ellison,S.L., Luby,C.H., Corak,K.E. *et al.* (2018) Carotenoid presence is associated with the or gene in domesticated carrot. *Genetics*, **210**, 1497–1508. 10.1093/genetics/210.4.NP.

28. Coe,K.M., Ellison,S.L., Rolling,W.R. *et al.* (2022) New insights into the genetic mechanisms controlling carotenoid accumulation in carrot. *Plant & Animal Genome Conference XXIX*. San Diego, CA.

29. Stelmach,K., Macko-Podgórni,A., Allender,C. *et al.* (2021) Genetic diversity structure of western-type carrots. *BMC Plant Biol.*, **21**, 200. 10.1186/s12870-021-02980-0.

30. Just,B.J., Santos,C.A.F., Yandell,B.S. *et al.* (2009) Major QTL for carrot color are positionally associated with carotenoid biosynthetic genes and interact epistatically in a domesticated × wild carrot cross. *Theor. Appl. Genet.*, **119**, 1155–1169. 10.1007/s00122-009-1117-z.

31. Cavagnaro,P.F., Iorizzo,M., Yildiz,M. *et al.* (2014) A gene-derived SNP-based high-resolution linkage map of carrot including the location of QTL conditioning root and leaf anthocyanin pigmentation. *BMC Genom.*, **15**, 1118. 10.1186/1471-2164-15-1118.

32. Parsons,J., Matthews,W., Iorizzo,M. *et al.* (2015) *Meloidogyne incognita* nematode resistance QTL in carrot. *Mol. Breed.*, **35**, 114. 10.1007/s11032-015-0309-2.

33. Spoor,S., Cheng,C.H., Sanderson,L.A. *et al.* (2019) Tripal v3: an ontology-based toolkit for construction of FAIR biological community databases. *Database*, **2019**, baz077. 10.1093/database/baz077.

34. Mungall,C.J. and Emmert,D.B. (2007) The FlyBase consortium, a Chado case study: an ontology-based modular schema for representing genome-associated biological information. *Bioinformatics*, **23**, i337–i346. 10.1093/bioinformatics/btm189.

35. Selby,P., Abbeloos,R., Backlund,J.R. *et al.* (2019) BrAPI – an application programming interface for plant breeding applications. *Bioinformatics*, **35**, 4147–4155. 10.1093/bioinformatics/btz190.

36. Jung,S., Cheng,C.H., Buble,K. *et al.* (2021) Tripal MegaSearch: a tool for interactive and customizable query and download of big data. *Database*, **2021**, baab023. 10.1093/database/baab023.

37. Buble,K., Jung,S., Humann,J.L. *et al.* (2019) Tripal MapViewer: a tool for interactive visualization and comparison of genetic maps. *Database*, **2019**, baz100. 10.1093/database/baz100.

38. Jung,S., Lee,T., Cheng,C.H. *et al.* (2017) Extension modules for storage, visualization and querying of genomics, genetic and breeding data in Tripal databases. *Database*, **2017**, bax092. 10.1093/database/bax092.

39. Chen,M., Henry,H., Almsaeed,A. *et al.* (2017) New extension software modules to enhance searching and display of transcriptome data in Tripal databases. *Database*, **2017**, bax052. 10.1093/database/bax052.

40. Gmitter,F.G., Chen,C., Machado,M.A. *et al.* (2012) Citrus genomics. *Tree Genet. Genomes*, **8**, 611–626. 10.1007/s11295-012-0499-2.

41. Humann,J., Jung,S., Cheng,C.H. *et al.* (2016) Cool Season Food Legume Genome Database: a resource for pea, lentil, faba bean and chickpea genetics, genomics and breeding. In Cool Season Food Legume Genome Database: a resource for pea, lentil, faba bean and chickpea genetics, genomics and breeding. In: *Plant & Animal Genome Conference XXIV*. San Diego, CA.

42. Osuna-Cruz,C.M., Paytuvi-Gallart,A., Di Donato,A. *et al.* (2018) PRGdb 3.0: a comprehensive platform for prediction and analysis of plant disease resistance genes. *Nucl. Acids Res.*, **46**, D1197–D1201. 10.1093/nar/gkx1119.

43. Iorizzo,M., Senalik,D.A., Grzebelus,D. *et al.* (2011) De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. *BMC Genom.*, **12**, 389. 10.1186/1471-2164-12-389.

44. Arias,D., Maldonado,J., Silva,H. *et al.* (2020) de novo transcriptome analysis revealed that photomorphogenic genes are required for carotenoid synthesis in the dark-grown carrot taproot. *Mol. Genet. Genomics*, **295**, 1379–1392. 10.1007/s00438-020-01707-4.

45. Ellison,S., Senalik,D., Bostan,H. *et al.* (2017) Fine mapping, transcriptome analysis, and marker development for Y2, the gene that

conditions β-carotene accumulation in carrot (*Daucus carota* L.). *G3 (Bethesda)*, **7**, 2665–2675. 10.1534/g3.117.043067.

46. Simon,P.W., Rolling,W.R., Senalik,D. *et al.* (2021) Wild carrot diversity for new sources of abiotic stress tolerance to strengthen vegetable breeding in Bangladesh and Pakistan. *Crop Sci.*, **61**, 163–176. 10.1002/csc2.20333.

47. Tas,P.M. (2016) Evaluating resistance to *Alternaria Duaci* and related traits among diverse germplasm of *Daucus Carota*. Dissertation UW-Madison.

48. Corak,K.E., Ellison,S.L., Simon,P.W. *et al.* (2019) Comparison of representative and custom methods of generating core subsets of a carrot germplasm collection. *Crop Sci.*, **59**, 1107–1121. 10.2135/cropsci2018.09.0602.

49. Altschul,S.F., Gish,W., Miller,W. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410. 10.1016/S0022-2836(05)80360–80362.

50. Buels,R., Yao,E., Diesh,C.M. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66. 10.1186/s13059-016-0924-1.

51. Jung,S., Lee,T., Gasic,K. *et al.* (2021) The Breeding Information Management System (BIMS): an online resource for crop breeding. *Database*, **2021**, baab054. 10.1093/database/baab054/.