CALIFORNIA LEAFY GREENS RESEARCH PROGRAM

April 1, 2017, to March 31, 2018

GENETIC VARIATION IN LETTUCE

Richard W. Michelmore
The Genome Center and
The Department of Plant Sciences
University of California, Davis
rwmichelmore@ucdavis.edu

SUMMARY:

We continue to develop and apply novel methods for detecting, analyzing, and manipulating genetic variation in lettuce. This project aims to ensure that lettuce benefits from the application of genomic and biotechnological techniques and has four components: (i) Transgene expression and genome editing in lettuce. (ii) Cloning and characterization of disease resistance genes. (iii) Genetic mapping with the goal of locating and developing markers for most of the disease resistance genes known in lettuce. (iv) Comparative genomics to identify candidate genes controlling horticultural traits. We completed the characterization of genome editing events of a gene that inhibits germination at high temperatures and knock-outs of several other genes have been made and are currently being characterized. We continue to make extensive use of high-throughput sequencing and marker technologies. Ten new genes for resistance to downy mildew have been mapped to the lettuce genome. Two major and multiple minor quantitative trait loci (QTLs) have been identified and mapped to the lettuce genome. The v7 version of the reference genome assembly of lettuce cv. Salinas has been published in 2017 and is publically accessible. The v9 version with nine chromosomal scaffolds is being prepared for release. We have sequenced ~90 additional lettuce genotypes to assess allelic variation for horticulturally important genes and provided molecular markers for selection in breeding programs. Genome sequences have been and are being mined for candidate genes for traits such as disease resistance, development, and horticulturally important traits such as tipburn resistance. We are curating several databases that include genetic, molecular marker, cultivar, phenotypic and sequence data for lettuce.

CALIFORNIA LEAFY GREENS RESEARCH PROGRAM

April 1, 2017 to March 31, 2018

PROJECT TITLE: **GENETIC VARIATION IN LETTUCE**

PRINCIPAL INVESTIGATOR: **Richard W. Michelmore**
The Genome Center and
The Department of Plant Sciences
University of California, Davis
rwmichelmore@ucdavis.edu

COOPERATING PERSONNEL: **Keri Cavanaugh**
María José Truco
Dean Lavelle
Lorena Parra
Pauline Sanders
Alex Kozik
Huaqin Xu
Sebastian Reyes Chin Wo
Lien Bertier
Brett Pike
Byron Froman
UC Davis Genome Center and
The Department of Plant Sciences
University of California, Davis
Ivan Simko
Beiquan Mou
Steve Klosterman
UDSA-ARS, Salinas
Richard Smith
Michael Cahn
UC Cooperative Extension, Monterey County
Krishna Subbarao
The Department of Plant Pathology
University of California, Davis

OBJECTIVES:

To develop and apply new methods for detecting, analyzing and manipulating variation in lettuce. We continue to pursue these objectives in four sub-projects:

1) Analysis of transgenes in lettuce and genome editing.

2) Molecular cloning of genes for disease resistance and other horticultural traits.

3) Development of a detailed genome assembly and identification of reliable, readily assayed markers linked to disease resistance genes.

4) Utilization of comparative genomics to identify candidate genes controlling horticultural traits and development of robust molecular markers for them.

In the first three projects we are emphasizing either novel forms of disease resistance or increasing the efficiency of selection for disease resistant genotypes. The fourth objective includes a wide range of horticultural traits. Some of these studies have been funded from Federal grants

and support from seed companies. All projects were initiated with CLGRP funds and application of the results to lettuce improvement is supported by CLGRP funds. All projects impact improvement of both crisphead and leafy types.

PROCEDURES AND RESULTS:

Transgene Expression and Genome Editing in Lettuce

Our studies on transgene expression remain a low priority, in part because there is not a major breeding objective that can only be addressed by transgenic lettuce and therefore commercial deployment of transgenes in lettuce is not a near-term need. Introduction of genes into lettuce using *A. tumefaciens* is routine; however, transgenes that expressed well in some other plant species are expressed poorly after multiple generations in lettuce. We now use the Arabidopsis ubiquitin promoter in preference to the CaMV 35S promoter that is prone to silencing in lettuce. We continue to generate transgenics as components of other projects (see below); these provide additional data on transgene expression and stability.

Since 2012, a technology for precise genome editing based on the CRISPR/Cas9 system has been developed for use in plants and animals. This technology can be used to create gene knockouts, deletions and replacements as well as for introduction of new genes and sequences. Gene knock-outs are currently much easier than sequence replacements or additions but the enabling technology is advancing rapidly. We continue to develop and apply genome editing technology for lettuce. We have used four generations of constructs for generating gene knockouts (Fig. 1). Knock-outs of NCED4 resulting in high temperature germination as described in previous CLGRB reports has been published (Bertier *et al.* 2018. High-resolution analysis of the efficiency, heritability and editing outcomes or CRISPR-Cas9-induced modifications of NCED4 in lettuce (*Lactuca sativa*). *G3: Genes, Genomes, Genetics* **8**:1513-1521; https://doi.org/10.1534/g3.117.300396). We are currently using version 4, which uses a polycistronic gRNA containing multiple gRNAs in a single construct. This enables us to target multiple genes at once and also increases the efficiency of generating knock-outs in T_1 plants because multiple gRNAs targeting the same gene can be used. These constructs also contain a dsRED gene as a fluorescent reporter for the T-DNA, enabling the easy selection of CAS9-free T_2 plants.

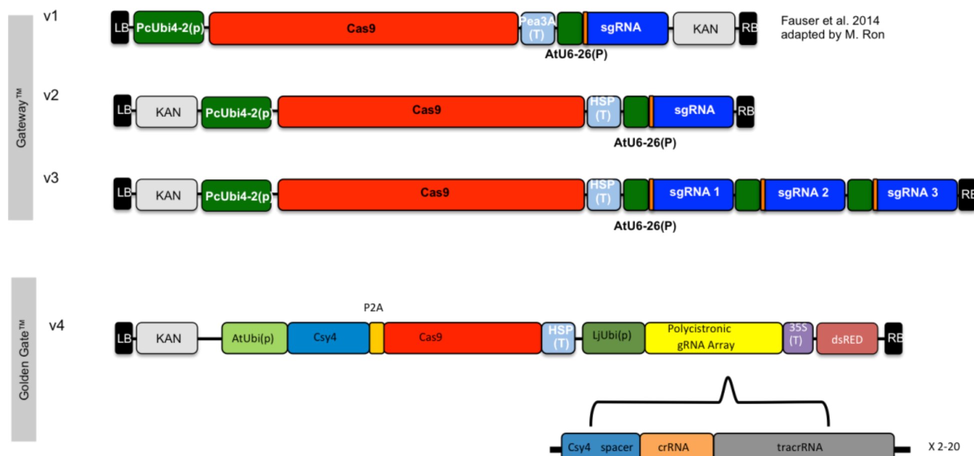


Figure 1. Maps of the four different constructs used for generating knock-outs. Versions 1 and 2 contain only one gRNA. Version 3 has 3 gRNAs, each driven by a separate promoter. Version 4 can contain up to 20 gRNAs that are transcribd from the same promoter, after which they are cleaved into single gRNAs.

Table 1. Genes in lettuce currently being studied using knock-outs.

Candidate Gene	Phenotype	Construct used	Collaborators
NCED4	Germination thermosensitivity	v1	Kent Bradford
ERF1	Germination thermosensitivity	v2	Kent Bradford
Vert1 (3 genes)	*Verticillium* wilt resistance (race 1)	v2, v4	Steve Klosterman
Phytoene desaturase	Chlorophyl biosynthesis	v2	
XTH (6 genes)	Cell wall biosynthesis	v3	Gail Taylor
Cycloidea (3 genes)	Transcription factor involved in flower shape	v4	
Cor (4 genes)	Bacterial corky root resistance	v4	
Novel miRNA	Unknown	v4	Blake Meyers
Tipburn resistance (3 genes)	Physiological tissue breakdown	v4	

Technologies for targeted gene insertion and replacement in plants are still not available. Efficiencies are often far below 1%. High-throughput, quantitative investigation of multiple procedural variables in plants is lacking. Possible experimental variables include the design and amounts of gRNAs and donor template as well as the importance of the number (1 or 2), placement, and spacing of the double strand break(s). Also, the size and position of the repair template for homology directed repair relative to the targeted SNP(s), its complementarity to the non-target or target strand, and targeting the transcriptionally active strand are potentially important variables. We are optimizing these variables using a fluorescent protein conversion assay in protoplasts.

In the longer term, we will use genome editing of lettuce to create stacks of resistance genes containing several R genes at single chromosomal positions so that they will be inherited as single Mendelian loci in breeding programs. Such stacks could contain multiple resistance genes effective against all known pathotypes of downy mildew and multiple viral, bacterial, fungal pathogens as well as insect pests and nematodes. This will enhance the durability of resistance by increasing the evolutionary hurdle that pathogens will have to overcome. Gene stacks could be expanded as more resistance genes become available and genes replaced when overcome by changes in the pathogens. A herbicide resistant gene (e.g. *ALS*) could be used as selectable marker for the gene stack. Gene stacking will greatly simplify breeding for disease resistance so that breeders can focus on more complex traits such as water and nitrogen use efficiencies and nutritional quality. However, in order for this to happen genes for resistance to each disease need to be identified at the molecular level and technology for inserting genes, preferably without tissue culture, need to be developed as well as intellectual property and regulatory issues need to be clarified.

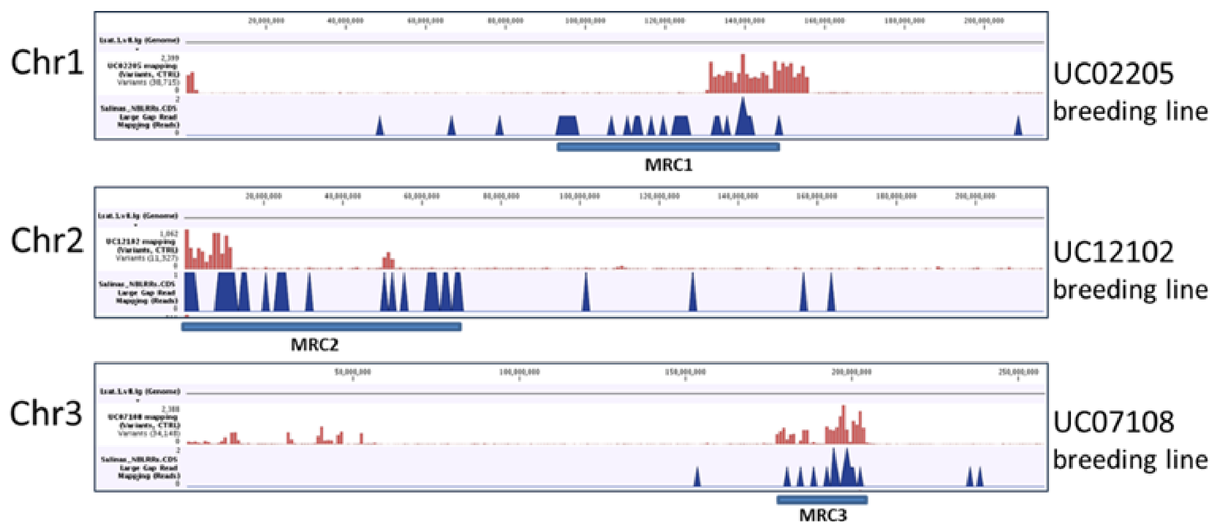
Resistance Gene Identification

We continue to map loci for resistance to downy mildew (DM), corky root, *Fusarium* and *Verticillium* wilts, onto the consensus genetic map and place them on the genome sequence. We are continuing to develop molecular markers to assist the selection of resistance genes. Of the over 50 phenotypic resistance genes mapped in lettuce, most co-localize to one of five major resistance clusters on chromosomes 1, 2, 3, 4, and 8 (MRCs 1, 2. 3, 4, and 8 respectively). The majority of these resistance phenotypes are linked to NB-LRR-encoding (NLR) genes as described in previous CLGRB reports that provide markers for these resistances.

Resistance to Downy Mildew

We used whole genome sequencing to identify chromosome segments introgressed from wild species during breeding for DM resistance. Numerous genes for resistance to DM have been introduced into cultivated lettuce from wild species (*L. serriola, L. saligna* and *L. virosa*) by repeated backcrossing (often six or more backcrosses) by public and commercial breeders. This has resulted in near-isogenic lines (NILs) that only differ for small chromosome regions that are potentially associated with resistance. Twelve of these advanced breeding lines show resistance to many isolates of *B. lactucae* in California, including recent highly virulent isolates. Mapping sequencing reads from these 12 advanced breeding lines against the *L. sativa* reference genome (cv. Salinas) allowed identification of introgressed regions as indicated by a high density of single nucleotide polymorphisms (SNPs) (Fig. 2).

Figure 2. Examples of polymorphic regions in advanced breeding lines with resistance introgressed from wild species. Introgressed segments have a high density of SNPs (red) relative to the reference genome and are co-located with MRCs. NLR density in the reference genome sequence is represented in dark blue.



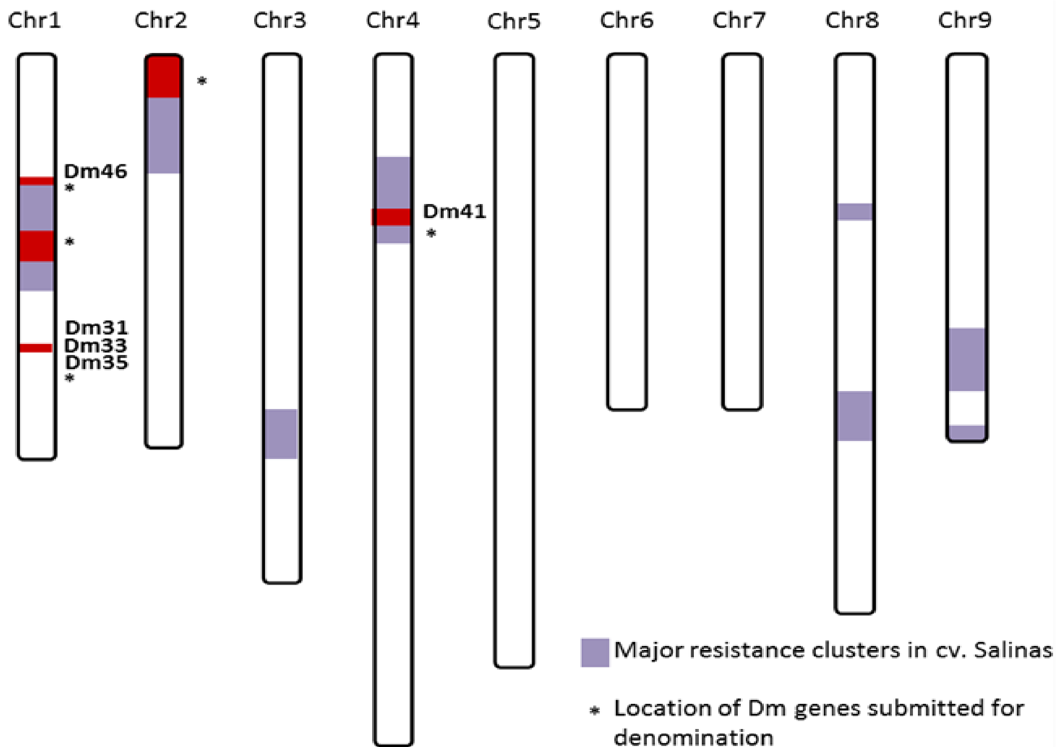
We have generated F_2 populations from crosses between these NILs and cv. Salinas in order to confirm which of these polymorphic regions are required for resistance. Most populations segregate 3:1 for resistant and susceptible individuals, consistent with one dominant resistant gene determining resistance to downy mildew. Segregation ratios from three populations are consistent with the presence of two dominant genes required for resistance, which could be indicating the presence of one resistant gene and one gene that is monitored ("guarded") by the resistance gene.

Ten F_2 progeny from these crosses were genotyped by sequencing and the resistance genes were mapped into the reference genome (Table 2, Fig. 3). New resistance genes from lines UC02202, UC02204, UC07105, UC07107, UC07108 were named as numbered *Dm* genes consistent with their prior resistance factor denominations described in Parra *et al*. (2016, *Euphytica* **210**:309. https://doi.org/10.1007/s10681-016-1687-1). New resistance genes from the lines UC12100, UC12101, UC12102 and UC12103 will be submitted to the International Bremia Evaluation Board (IBEB) for denomination.

Table 2. Advanced breeding lines with resistant donor accessions, *Dm* gene denominations and chromosomal locations. S.f.d. = to be submitted for denomination.

ID	Donor	*Dm* gene	Location
UC02202	*L. saligna* LJ85314	*Dm33*	Chr1
UC02204	*L. virosa* LJ85289	*Dm35*	Chr1
UC02206	*L serriola* W66331A	*Dm31*	Chr1
UC07105	*L. sativa* PI491226	*Dm41*	Chr4
UC07107	*L. saligna* PI491206	*Dm46*	Chr1
UC07108	*L. saligna* PI491208	*Dm47*	Chr2, Chr4
UC012100	*L. saligna* CGN9311	s.f.d.	Chr1
UC012101	*L. saligna* CGN5318	s.f.d.	Chr1, Chr2
UC012102	*L. saligna* CGN5282	s.f.d.	Chr4
UC012103	*L. saligna* CGN5147	s.f.d.	Chr1

Figure 3: The genomic locations of new resistance genes on lettuce chromosomes.



In order to identify the resistance genes present in the advanced breeding lines and distinguish these genes from resistance genes previously described in other cultivars, we have used long fragment sequencing (PacBio) for the characterization of the resistance gene repertoires present in each breeding line and the different lettuce cultivars, carrying known *Dm* genes. Resistance gene enrichment sequencing (RenSeq) allowed the identification of numerous NBS-LRR sequences from ~ 50 genotypes, including the UC advanced breeding lines. This data in combination with phenotypic evaluations will allow the identification of candidate genes for known and new resistances against *Bremia lactucae*. Resistance genes will be validated using CRISPR-mediated knock-outs of the candidate gene.

Genetic Analysis

We have been constructing detailed genetic maps using genotyping by sequencing on populations segregating for a variety of traits in collaboration with other researchers (See Table 2, 2016-2017 CLGRB report). This year, we genotyped and analyzed a single population developed by German Sandoya at University of Florida that segregates for resistance to bacterial leaf spot.

Genetic Analysis of Tipburn Resistance

A QTL for resistance to tip burn had been previously identified in linkage group 5 in a RIL population from a cross between cvs. Emperor and El Dorado (Jenni *et al*., 2013. *Theor. Appl. Genet*. 126:3065-3079). Subsequently, we fine mapped this QTL by phenotyping lines selected as being recombinant using codominant molecular markers flanking and within the QTL region and phenotyping those recombinant lines in field experiments in different environments (see previous CLGRB reports for more information). These recombinant lines were sequenced and analysis of chromosomal recombination points reduced the genomic interval associated with tipburn resistance to a region containing 21 candidate genes. RNAseq studies of expression level differences in the recombinant lines, further reduced the number of candidate genes. We are currently developing CRISPR-mediated knock-out lines for three candidate genes to test whether they affect tipburn.

In collaboration with I. Simko, R. Hayes USDA and S. Jenni, Canada, seven recombinant inbred line (RIL) populations had been assessed over the past 15 years for tipburn in multiple environments and years (Table 3). Several other morphological traits were also assessed including core height, head firmness, head closure, leaf crinkliness, plant fresh weight, and leaf savoy. These populations were genotyped by sequencing and analyzed to elucidate the genetic architecture of resistance to tipburn and to identify QTL for tipburn resistance.

Table 3. RIL populations used to study tipburn and identified QTLs for resistance.

RIL population	# Trials	QTLs for tipburn
Iceberg x Saladin	2	-
F_1 (Valmaine x Salinas 88) x Salinas	5	LG2, LG7
Salinas 88 x La Brillante	4	**LG1**, LG4, **LG5**, LG9
Emperor x El Dorado	1	**LG5**, LG7
Salinas x Calicel	2	**LG1**, LG2, LG3, **LG5**, LG7, LG8
Grand Rapids x Iceberg	2	LG2, LG3, LG4, **LG5**, LG7, LG9
Salinas x *L.serriola* UC96US23	3	LG3, **LG5**, LG7

Analysis of these seven populations revealed that the genetic architecture for tipburn in lettuce consists of major, intermediate, and minor QTL scattered throughout the lettuce genome (Fig. 4). The significance of these QTL varied from population to population, possibly because either causal genes were not segregating in some populations, or that they sometimes had small undetected effects, or because they were influenced by the environment.

The RIL populations that segregated for head type revealed more QTL (mostly minor to intermediate QTL distributed throughout the genome) than RIL populations that did not segregate for head type (one major QTL with some unlinked minor QTL). Two major QTL for tipburn incidence were identified in linkage groups (LG) 1 and 5 that determined up to 45 and 66% of the phenotypic variance. The major QTL in LG 1 collocated with head firmness QTL. The major QTL in LG 5 collocated with QTLs for core height, leaf crinkliness, and head firmness (Fig. 4).

Figure 4. QTLs identified for tipburn and colocation with other traits. The QTL are color coded by population: SxB = Black, VxS = Brown, SxC = Red, GxI = Blue, IxS = Dark yellow, SxL = Purple, ExD = Dark Green. The QTL names represent the "Trait_Chromosome_Location Year and Treatment". The trait represented by letter code: Firmness = FIRM, Core Height = CH, Head Closure = CLOS, Leaf Crinkliness = LC , Tipburn Incidence = TBIN, Tipburn Severity = TBSV, Plant Fresh Weight = FW. The location Salinas Spence Farm is represented by SPE, Yuma, Arizona = AZ, Salinas Carr Lake, California = CL, Quebec, Canada = QC. The GxI population was planted in three water and nitrogen treatments, when planted under high water/high nitrogen (HWHN), low water/ high nitrogen (LWHN), and high water/low nitrogen (HWLN).

Water Use Efficiency (WUE) and Nitrogen Use Efficiency (NUE)

Starting in 2011, we have conducted a series of large field trials at Spence, USDA Salinas, to investigate NUE and WUE as part of USDA SCRI-funded projects (with matching support from the CLGRB) as described in prior CLGRB reports. QTLs for several traits were identified (See 2013-2014 report). The last trial conducted in summer 2016 at Spence Farm focused on NUE to test if the QTL identified previously were reproducible across years. Major clusters of QTLs for NUE were identified in LG3, LG7 and LG9, some of which collocated with QTLs from previous studies.

Genome and Transcriptome Sequencing

The genome of lettuce was initially sequenced in collaboration with the BGI, Shenzen, China, funded by an international consortium of ten companies as well as the BGI. This resulted in 2.4 Gb of genomic sequence assigned to genetic bins ordered along the nine lettuce chromosomes representing 97% of the total length of assembled scaffolds. Numerous phenotypes have been placed relative to genetically validated sequence scaffolds. The genome was annotated to provide *ca*. 41,000 high-confidence gene models. The v7.0 version of the lettuce genome sequence has been published (Fig. 8; Reyes-Chin-Wo *et al*. 2017. *Nat. Comm*. **8**:14953) and is publically available at https://lgr.genomecenter.ucdavis.edu/ as well as several public databases such as GenBank, CoGe, and Phytozome.

The second phase of the Lettuce Genome Project funded the International Lettuce Genome Consortium comprised of 17 large and small seed companies concluded in March 2018. This focused on refining the draft genome sequence using several new approaches in order to provide greater contiguity and better resolution of each genetic bin of scaffolds as well as extensive allelic sequence variation. For this we collaborated with Dovetail Genomics (Santa Cruz) to generate chromosome-scale genomic scaffolds. The v9.0 version of the lettuce genome has nine chromosomal scaffolds and is being prepared for release. In collaboration, with Ivan Simko and others we sequenced ~90 lettuce cultivars and wild accessions and identified SNPs relative to the reference genome that are being made available through our website.

In collaboration with academic groups working on different aspects of lettuce biology, we are conducting RNAseq profiling experiments to provide an atlas of genes expressed in lettuce at different developmental stages, under different abiotic stresses, and during resistance and susceptibility to diverse diseases. We now have over 500 tracks of RNAseq data from multiple diverse experiments. These data are being made available through our website.

Databases

We continue to curate several publicly accessible databases for lettuce. The Compositdb contains several searchable databases for lettuce (accessible throughhttp://compositdb.ucdavis.edu/) and is the main portal for distributing information generated by the Compositae Genome Project (CGP). Chiplett provides access to data from the Affymetrix lettuce Genechip project as well as the ultra-dense genetic map (http://chiplett.ucdavis.edu/). The G2G site (http://scri.ucdavis.edu/) provides access to information generated as part of the Next-Generation Lettuce Breeding: Genes to Growers (G2G) and CLGRP-funded projects. Our GBrowse genome viewer (http://gviewer.gc.ucdavis.edu/cgi-bin/gbrowse/lettucePublic/) provides access to the ultra-dense map as genetic chromosomal pseudomolecules. Lettcv (http://compositdb.ucdavis.edu/database/lettcv2/display/) archives extensive genetic, passport and performance data on lettuce cultivars. The CGP database (http://compgenomics.ucdavis.edu/compositae_index.php) contains extensive sequence and related information as well as links to lettuce genetic maps (lettuce genetic map viewer) and marker

information. Morphodb (http://compgenomics.ucdavis.edu/morphodb_index.php) is an archive of and provides access to phenotypic information on *Lactuca* species; this database utilizes standard ontologies to facilitate searches across databases. The lettuce v7 genome assembly is publically available at https://lgr.genomecenter.ucdavis.edu/. These databases are being modified to facilitate access to marker information for breeding purposes from disease-centric, breeder-oriented perspectives. The Bremia Database displays virulence phenotypes, mating type and fungicide sensitivity for Californian isolates of *B. lactucae* characterized from 2001 to the present (http://bremia.ucdavis.edu/bremia_database.php).