CALIFORNIA LEAFY GREENS RESEARCH PROGRAM
April 1, 2020 to March 31, 2021

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, John Emerson, Rongkui Han, Alex Kozik, Huaqin Xu, Sagayamary Sagayaradj, Juliana Gil, Kyle Fletcher**
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
Department of Plant Pathology, UC Davis

ABSTRACT:

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) genetic mapping, cloning, and characterization of disease resistance genes; (iii) developing genomic resources; and (iv) comparative genomics to identify candidate genes controlling horticultural traits. We continue to make extensive use of high-throughput sequencing, marker technologies, and genome editing for gene identification. Genome sequences are being mined for candidate genes for traits, such as disease resistance, development, and horticulturally important traits such as tipburn resistance. We have initiated a program to study salinity tolerance in lettuce. The v10 version of the reference genome assembly of cv. Salinas with nine chromosomal scaffolds that has much improved contiguity and sequence accuracy has been distributed to the International Lettuce Genomics Consortium. We continue to curate several databases that include genetic, molecular marker, cultivar, phenotypic and sequence data for lettuce.

OBJECTIVES:

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

1) Analysis of transgenes in lettuce and genome editing.
2) Molecular identification of genes for disease resistance and other horticultural traits.
3) Development of genomic resources.
4) Utilization of comparative genomics to identify candidate genes controlling horticultural traits and development of molecular markers for them.

In the first three projects we are mostly 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. Several 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. We continue to generate transgenics as components of other projects (see below); these provide additional data on transgene expression and stability.

Since 2012, technology for precise genome editing based on the CRISPR/Cas9 system has been developed for use in plants and animals. 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 as part of our gene identification strategy.

In addition to validating potential candidate genes for disease resistance, we are targeting a subset of the many genes that have been previously functionally characterized in other plant species, particularly Arabidopsis. We are using CRISPR-mediated gene knockouts to deduce whether the homologous genes in lettuce have a similar function. We are initially investigating genes potentially controlling traits such as nutrient content and leaf, flower, and root development.

In the longer term, once the technology has been adequately developed, we will use genome editing of lettuce to create stacks of resistance genes containing several resistance and other genes at single chromosomal positions so that they will be inherited as single Mendelian loci in breeding programs. 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, technology for inserting genes, preferably without tissue culture, needs to be developed and genes for resistance to each disease need to be identified at the molecular level.

Molecular identification of genes for disease resistance and other horticultural traits

We previously mapped multiple genes for resistance to downy mildew (DM), corky root, and *Fusarium* and *Verticillium* wilts, onto the consensus genetic map and placed them on the genome sequence. This provides 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 (MRCs) on Chromosomes 1, 2, 3, 4, and 8 (Fig. 1). The majority of these resistance phenotypes are linked to NB-LRR-encoding (NLR) genes as described in previous CLGRB reports that provide markers for selecting for these resistances. No new genes were mapped in the past year.

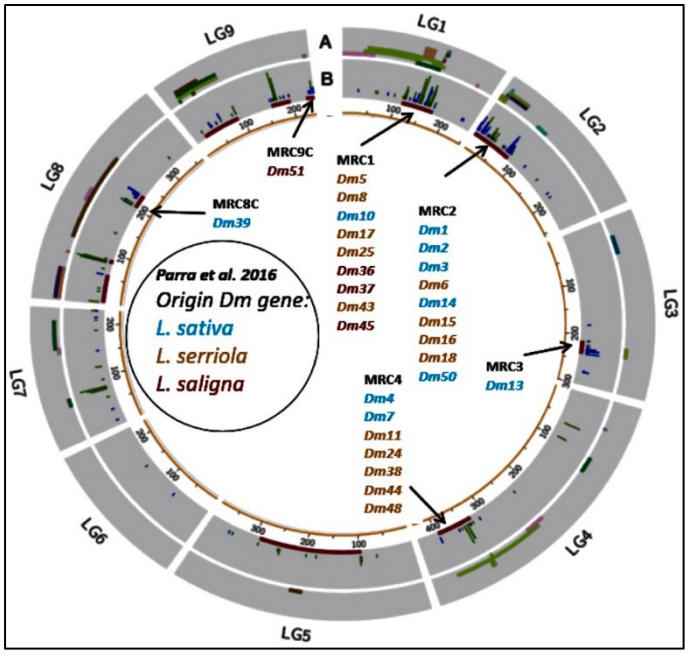


Figure 1. Chromosomal distribution of QTL (A) and *Dm* genes (B) for resistance to *Bremia lactucae* in lettuce. The colors indicate the donor species (Parra et al., 2016. *Euphytica* **210**:309-326), illustrating the importance of wild species as a source of resistance genes.

To assist in genetic mapping experiments, we developed the Assembly Free Linkage Analysis Pipeline (AFLAP) that enables the construction of ultra-high-density genetic maps without calling single nucleotide polymorphisms (SNPs) against a reference genome assembly (Fletcher *et a*l. 2021. Genome Biology 22:115). Maps are generated using unique single copy k-mers in each parent as markers rather than SNPs. AFLAP works with multiple types of populations, including F_1 (pseudotest-cross), F_2, and RIL populations. This approach is fast and widely applicable to diverse populations and species. We tested the pipeline on published genetic data of Arabidopsis. We used AFLAP on a lettuce F_6 population of 235 RILs that had been genotyped by GBS; the resultant genetic map with 8,191 markers was collinear with the lettuce genome assembly. We also validated AFLAP on two crosses of *B. lactucae*. This genetic map of *B. lactucae* contained over 90,000 markers ordered in 18 linkage groups; this was used to improve the reference genome assembly by making it genetically consistent. The resultant genome assembly of *B. lactucae* is highly syntenic with the telomere-to-telomere assembly that we generated for *Peronospora effusa* (spinach downy mildew) in collaboration with Allen VanDeyzne and others (Fig. 2).

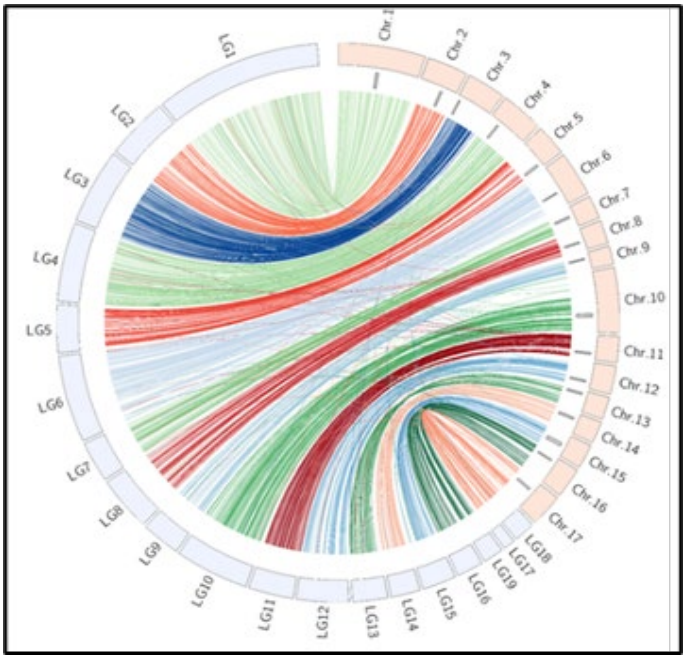


Figure 2. High collinearity between chromosome-scale genome assemblies of *Bremia lactucae* and *Peronospora effusa.* The genetically-validated chromosomal sequences of *B. lactucae* are shown in blue and the telomere-to-telomere chromosomes of *P. effusa* in red. The lines join genes with sequence similarity.

We were successful in cloning the *Dm11* gene for resistance to *B. lactucae*. The genome of *B. lactucae* was mined for genes that might encode effectors, proteins important for virulence, that also may be detected by resistance genes. Transient *Agrobacterium*-mediated expression in a large panel of wild and cultivated lettuce lines detected several effectors that elicited

a resistance response. One effector in particular, BSW04, elicited a necrotic response characteristic of resistance in lines expressing *Dm11*. Genetic fine mapping, RNAi silencing, and k-mer analysis all pointed to the same NBS-LRR encoding gene. Therefore, this gene was cloned from cv. Ninja and when it was transiently co-expressed with BSW04, there was a necrotic response (Fig. 3). This is the first time that co-expression of a resistance gene with its cognate effector gene has been demonstrated to result in a resistance reaction in lettuce. This opens up the possibility of efficient cloning of more resistance genes from lettuce.

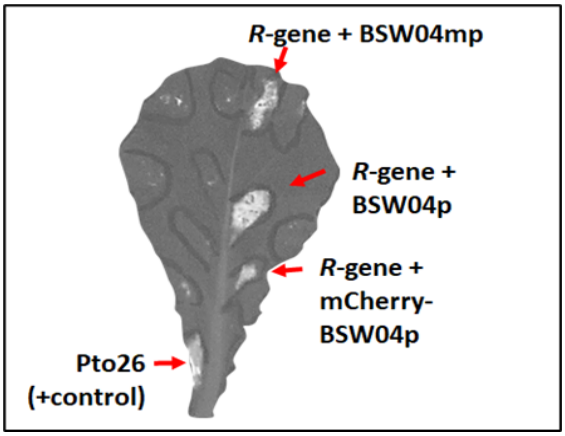


Figure 3. The *Dm11* gene from cv. Ninja elicits necrosis in cv. Cobham Green when co-expressed with BSW04 paralogs recognized by Ninja. The leaf has been cleared and viewed under UV light to visualize the necrotic areas. Only areas co-infiltrated with *Agrobacterium* carrying constructs to express both the resistance gene and the recognized effector showed necrosis. Other areas were infiltrated with only one construct or other combinations and controls that were not expected to result in necrosis.

Resistance to corky root

Many years ago, we showed that resistance to corky root was determined by the recessive *cor* gene on Chromosome 3. Fine mapping and genomic analysis of resistant and susceptible lines identified four candidate genes for resistance to corky root, one or more of which could encode a susceptibility factor. Constructs for CRISPR-Cas9-mediated gene knock-outs with guide RNAs targeting each of these candidate genes in the *cor* region were transformed into cv. Salinas. We recovered T_2 seed from 32 transgenic lines that are ready to be evaluated for resistance to *Rhizorhapis suberifaciens* strain CA1.

We needed a pathogenic stock of *R. suberifaciens* CA1 in order to screen the transgenic knock-out lines for resistance. *R. suberifaciens* is difficult to maintain in culture and unfortunately our stock of CA1 no longer elicited the characteristic reactions on susceptible and resistant lines of lettuce. Fortunately, we were able to obtain another sample of CA1 from Isolde Francis, California State University, Bakersfield that was derived from our original CA1 strain and had been transferred to the Belgian Coordinated Collections of Microorganisms (BCCM), conveyed lyophilized to Florida, and then to CSU Bakersfield, where it was revived. This sample of CA1 elicits the appropriate reactions on susceptible and resistant lines and therefore will be used for screening the knock-out lines.

One of the challenges when trying to determine whether we had a pathogenic strain was the lack of reliable diagnostics. Therefore, we sequenced CA1 and multiple other strains of *R. suberifaciens* and related species in collaboration with Isolde Francis. We have generated a complete, high-quality assembly of the *R. suberifaciens* CA1 genome using PacBio HiFi reads sequenced on a Sequel II by the DNA Technologies Core at UC Davis. A total of 44K HiFi reads were produced and the genome was assembled *de novo*, comprising five circular contigs totaling 3.405 Mb, averaging 80-fold coverage. CA1 has a large circular genome of ~3.1 Mb and four circular, low-copy plasmids ranging in size from 15 to 69 Kb. This aligns well with the fragmented assembly of CA1 that had been previously generated from short reads and is available from NCBI

GenBank (GCA_014200045.1). We are currently annotating the genome and will search for potential pathogenicity-associated genes by comparing pathogenic and non-pathogenic strains.

We are also assembling the genomes of 10 pathogenic and non-pathogenic strains of *Rhizorhapis* species and related genera obtained from Isolde Francis using PacBio HiFi reads. In addition, we are sequencing two strains provided by companies that are being used for screening for resistance to corky root. Companies interested in submitting additional strains for sequencing should contact Maria Ferrer (mferrerruiz@ucdavis.edu). Preliminary comparative analysis has revealed considerable genomic variation, particularly regarding plasmid composition. Using these data, we will develop diagnostic assays for pathogenic strains of *R. suberifaciens* that cause corky root in lettuce, so that we can reliably screen for resistance. These assays will also be useful for diagnosis in the field as well as epidemiological studies.

Genetics of salt tolerance in wild *Lactuca* species

We have initiated a genetic analysis of salt tolerance in order to breed lettuce cultivars with increased salt tolerance. We previously screened wild *Lactuca* germplasm for salinity tolerance. Phenotypic variation in fresh biomass reduction was observed among *Lactuca* germplasm when grown with moderate salinity (100 mM NaCl). Eight accessions with contrasting sensitivity have been selected as parents for RIL populations based on fresh biomass to identify QTLs and candidate genes for salinity tolerance.

In the past year, we conducted replicated evaluations of salinity tolerance in the greenhouse using the eight selected parental accessions to characterize the salinity stress response in *Lactuca* species as well as to determine key traits to evaluate on the segregating populations. PI274564 had the smallest reduction in fresh shoot weight (-14.1%) compared to the other accessions, which ranged from (-31.3%–49.5%; Table 1). Phenotyping of ionic accumulation and biochemical responses are underway. PI274564 has been crossed with W622607 and W637132 to generate mapping populations.

Table 1. Average % reduction of fresh shoot weight in select accessions of *Lactuca* germplasm irrigated with 75 mM NaCl for 3 weeks (two replications, n = 12).

Accession	% Reduction of Fresh Shoot Weight in 75 mM NaCl
PI 274564	-14.1%
PI 667823	-31.3%
W6 37132	-34.2%
PI 667815	-36.9%
Salinas	-39.2%
W6 22607	-40.1%
UC96US23	-43.2%
L. aculeata	-49.5%

Analysis of nitrogen and water use efficiencies (NUE and WUE)

Our studies on robustness, the ability to perform well across multiple environments (phenotypic plasticity) under changing nitrogen and water applications, that were described in previous reports to the CLGRB have been published (Macias-González *et al*. 2021. Crop Science 61: 1582-1619). We analyzed 50 lettuce lines and a RIL population derived from cvs. Grand Rapids x Iceberg in multiple experiments spread over several years. Efficiency, sensitivity, and plant N uptake and utilization were important components of robustness under changing N application. The genetics of efficiency, sensitivity, nutrient concentration, nutrient content, dry-to-fresh biomass percentage, bolting, and shoot biomass production were studied by QTL analysis of the RIL population. Major clusters of QTLs were identified for these traits. These data indicate that breeding for robustness under fluctuating N or water availability can improve NUE and WUE.

Genetic analysis of sensitivity to tipburn

Two papers describing our studies on tipburn sensitivity have been published. The first integrates data from seven RIL populations that had been analyzed in multiple environments and years to identify QTLs for tipburn sensitivity (Macias-González, M., *et al.* 2019. *Theor. Appl. Genet*. 132:2209-2222). Twenty-three major, intermediate, and minor unique QTLs for tipburn were identified in one or more populations scattered throughout the genome. Two major QTLs for tipburn incidence were identified in linkage groups (LGs) 1 and 5, which determined up to 45 and 66% of the phenotypic variance. The major QTL in LG 1 collocated with a QTL for head firmness. The major QTL in LG 5 collocated with QTLs for core height, leaf crinkliness, and head firmness. The beneficial alleles at the QTLs in LGs 1 and 5 are present in cv. Salinas, the genotype sequenced for the reference genome assembly.

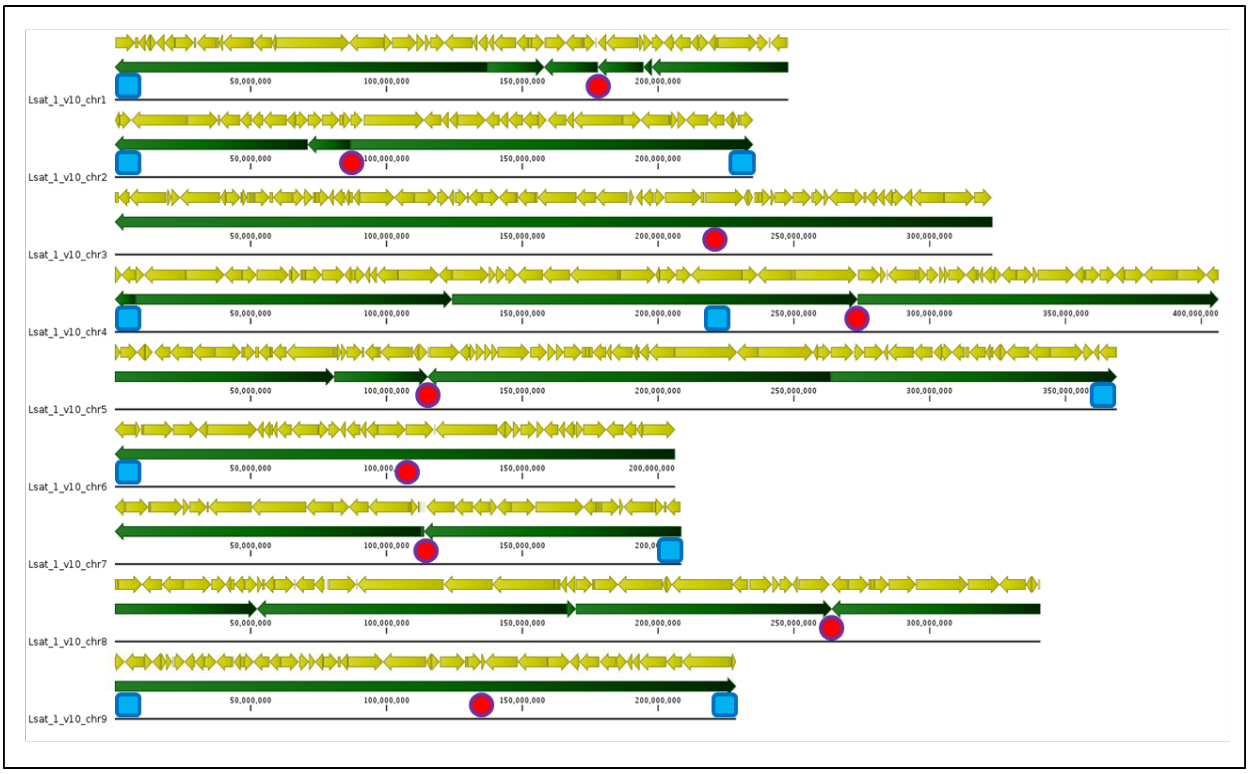
The second paper describes the genetic dissection of the major locus for tipburn (in)sensitivity on LG 5 of cv. Salinas (Macias-González, M. *et al*. 2021. *G3. Genes|Genomes|Genetics*, jkab097, https://doi.org/10.1093/g3journal/jkab097). This QTL was dissected genetically to identify candidate genes for tipburn by creating lines with recombination events within the QTL and assessing their resistance to tipburn. By comparing lines with contrasting haplotypes, the genetic region was narrowed down to ~877 Kb that was associated with a reduction of tipburn by ~60%. Analysis of the lettuce reference genome sequence revealed 12 genes in this region, one of which is a calcium transporter with a single nucleotide polymorphism in an exon between haplotypes with contrasting phenotypes. RNA-seq analysis of recombinants revealed two genes that were differentially expressed between contrasting haplotypes consistent with the tipburn phenotype. One encodes a Teosinte branched1/Cycloidea/Proliferating Cell factor transcription factor; however, differential expression of the calcium transporter was not detected. The phenotypic data indicated that there is a second region outside of the ~877 Kb region but within the quantitative trait locus, at which a haplotype from the susceptible parent decreased tipburn by 10 to 20%. A recombinant line was identified with beneficial haplotypes in each region from both parents that showed greater tipburn resistance than the resistant parent; this line could be used as the foundation for breeding cultivars with more resistance than is currently available.

Development of genomic resources

We have generated a new, high-quality reference genome assembly for lettuce that provides the foundation for understanding the genetic architecture of traits in *Lactuca* spp. Genome assembly and annotation is a reiterative process that provides increasing completeness and accuracy. Our publicly available v8 assembly was generated from short Illumina reads and

was composed of fragmented contigs genetically oriented into chromosome-scale ordered scaffolds. We have used a combination of Oxford Nanopore (~92x coverage), and PacBio HiFi (~14x coverage) long-read sequences as well as Illumina sequences to build a highly contiguous v10 reference assembly. Bionano optical mapping and Hi-C proximity ligation were used to resolve repeat regions and to scaffold this complex, highly repetitive genome. The final v10 reference genome assembly spans 2.575 Gb across 1,791 contigs with a contig N_{50} of 8.1 Mb. This genome assembly includes more than 98.4% of the 1,375 BUSCO embryophyte genes. Analysis of repeat sequences and Hi-C data revealed the positions of candidate centromeres and telomere-like sequences at the ends of most but not all chromosomes (Fig. 4).

Figure 4. Schematic of the v10 reference genome assembly of cv. Salinas. The light green boxes are 576 contigs totaling 2.57 Gb, the dark green boxes are 29 scaffolds after Bionano analysis and genetic orientation into the nine near-complete chromosomes. The red circles are repeated regions representing likely centromeres. The blue squares represent telomere-like sequences. A total of 1,151 small unplaced contigs totaling 9.3 Mb are not shown.



To capture genetic diversity within *Lactuca* spp., nine diverse wild and cultivated lettuce lines are currently being assembled using long-read Oxford Nanopore and short read Illumina sequences. These high-quality reference assemblies of *L. sativa* cvs. Salinas, Ninja, La Brillante, PI251246, VIAE, Flashy Troutback, and Merlot as well as *L. serriola* accessions Armenian and US96UC23 will be compared to reveal differences in gene content, particularly with regard to their resistance gene repertoires.

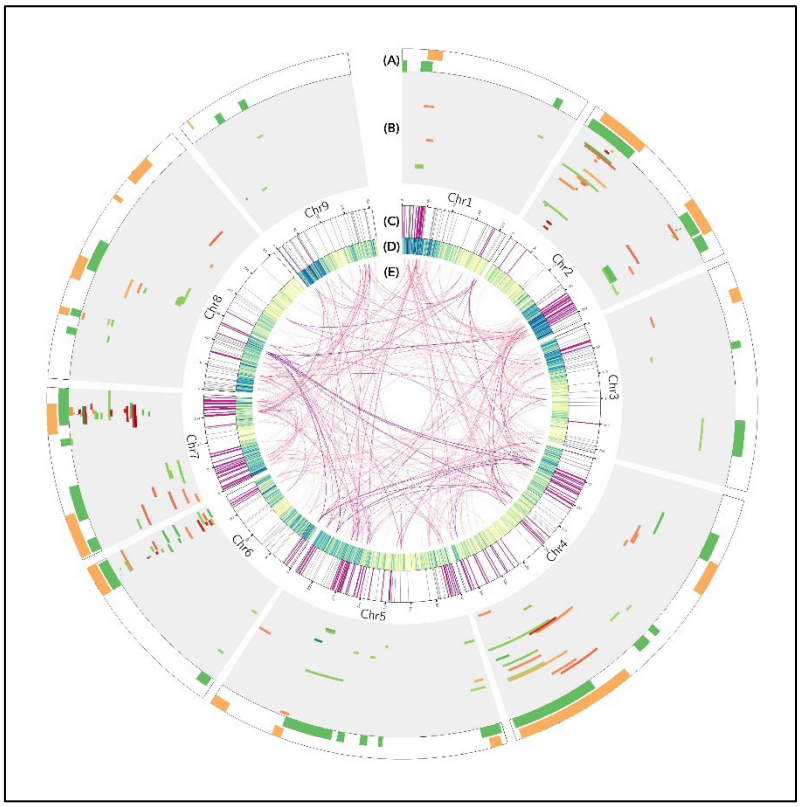
We continue to characterize gene expression using high-quality PacBio Iso-Seq data. The full-length transcript sequences provided by Iso-Seq increases the accuracy of genome annotation. Iso-Seq data had been generated for cvs. Salinas, Ninja, Diana, Cobham Green, LS102, and PI251246 as well as *L. serriola* accessions Armenian, US96UC23, and other lines. The Iso-seq data was critical to the cloning of the full length *Dm11* gene. In collaboration with other academic groups, we currently have multiple Illumina-based short-read RNA-seq experimental tracks of *L. sativa* cv. Salinas (243), *L. serriola* US96UC23 (31), *L. saligna* (4), and *L. virosa* (4). These transcriptomic analyses are providing an atlas of genes expressed in lettuce at different developmental stages, under different abiotic stresses, and during resistance and susceptibility to pathogens.

Utilization of comparative genomics to identify candidate genes controlling horticultural traits

We completed a QTL analysis of the genetic basis of daylength sensitivity that has recently been published (Han, R. *et al.* 2021. *Theor. Appl. Genet.* https://doi.org/10.1007/s00122-021-03908-w). Lettuce is a facultative long-day plant which changes in flowering time in response to photoperiod. Variations exist in both flowering time and the degree of photoperiod sensitivity among accessions of wild (*L. serriola*) and cultivated (*L. sativa*) lettuce. An F_6 population of 236 RILs derived from a cross between a photoperiod-sensitive *L. serriola* accession from Armenia and a photoperiod-insensitive *L. sativa* accession, PI251246, was planted under long-day (LD) and short-day (SD) conditions in a total of four field and screenhouse trials; the developmental phenotype was scored weekly in each trial. QTL mapping revealed five flowering time QTLs that together explained more than 20% of the variation in flowering time under LD conditions. Using two independent statistical models to extract the photoperiod sensitivity phenotype from the LD and SD flowering time data, an additional five QTLs were identified that together explained more than 30% of the variation in photoperiod sensitivity in the population. Orthology and sequence analysis of genes within the ten QTLs revealed potential functional equivalents in the lettuce genome to the key regulators of flowering time and photoperiodism, *FD* and *CONSTANS*, respectively, in Arabidopsis.

We reviewed flowering time regulation in lettuce to consolidate current knowledge of the genetic and molecular control of flowering time in lettuce (Han *et al.* 2021. Frontiers in Plant Science 12: 632708). We integrated information from 52 QTL mapping experiments conducted by five different research groups for bolting and flowering time with populations derived from cultivars and wild accessions. A total of 67 consolidated QTLs were identified for bolting and flowering time distributed over all nine chromosomes of lettuce, with Chromosomes 2 and 7 having the highest numbers of QTLs. The majority of the reported QTLs were environmentally sensitive, such that their phenotypic effects were only detected in subsets of experiments. Nine of the identified QTLs had a pleiotropic effect on both bolting and flowering time. In the model plant species, *Arabidopsis thaliana*, flowering time is controlled by a network of genes in pathways that respond to diverse environmental and developmental signals including temperature, day length, gibberellic acid, carbohydrates, and age. Key floral integrator genes regulate these pathways and their expression initiates the vegetative-to-floral shift of the apical meristem. We identified 405 genes in the lettuce genome that were orthologous to 306 genes controlling flowering in Arabidopsis. These were distributed throughout the lettuce genome; 115 are co-located with QTLs for bolting and flowering (Fig. 5; see Han *et al*., 2021 for details).

Figure 5. Chromosomal distribution of QTL and candidate genes controlling flowering.



(A) Physical location of 35 bolting time (green) and 29 flowering time (yellow) consensus QTLs in lettuce.
(B) Physical location of 142 QTLs reported in each of 49 field and greenhouse experiments. Each track represents one experiment. Bolting time QTLs are indicated by green color blocks and flowering time QTLs by orange blocks.
(C) Location of lettuce orthologs of genes with flowering time function in Arabidopsis. Flowering time orthologs within known QTLs are highlighted in fuchsia.
(D) Gene density in the lettuce genome.
(E) Flowering time orthologs within the same orthogroup are connected. Darker shade connections indicate larger orthogroups.

Databases

We continue to curate several publicly accessible databases for lettuce accessible through http://michelmorelab.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. These databases continue to be revised 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).