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 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 have demonstrated that the ubiquitin promoter is less prone to gene silencing than the CaMV 35S promoter. We have characterized genome editing events by knocking out a gene that inhibits germination at high temperatures as well as several other genes. We continue to make extensive use of high-throughput sequencing and marker technologies. The genome of lettuce has been sequenced and assembled and ~41,000 genes annotated. Genotyping by sequencing of the core mapping population has assigned over 97% of the assembled genome to genetic bins ordered along chromosomal linkage groups. The v7 version of the reference genome assembly of cv. Salinas has been published and is publically accessible. We have sequenced additional lettuce genotypes to assess allelic variation for horticulturally important genes and provide 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 have developed and are curating several databases for lettuce as part of the Genes for Growers Project that include genetic, molecular marker, cultivar, phenotypic and sequence data for lettuce.
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: María José Truco
Keri Cavanaugh
Miguel Macías González
Dean Lavelle
Lorena Parra
Germán Sandoya
Pauline Sanders
Alex Kozik
Huaqin Xu
Sebastian Reyes Chin Wo
Lien Bertier
Paola Perez
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 genetic map 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 were expressed poorly after multiple generations in lettuce. Factors influencing the stability of transgene expression over multiple generations in lettuce are not well understood and warrant further investigation. Multiple studies by us and others have utilized the CaMV 35S promoter to express transgenes. Other labs have recently shown that this is prone to transcriptional silencing. In collaboration with David Tricoli (Parsons Plant Transformation Facility, UC Davis) we generated plants expressing the DsRED fluorescent protein from the Arabidopsis ubiquitin promoter. This provided a readily scorable reporter to study gene silencing (Fig. 1). Gene expression remained stable over multiple generations of selfing and outcrossing indicating that this promoter is not as prone to silencing as the CaMV 35S promoter. Therefore the ubiquitin promoter will be used in the future. We continue to generate transgenics as components of other projects (see below); these provide additional data on transgene expression and stability.

Figure 1. Detection of expression of the DsRed transgene that encodes the red fluorescent protein.
A and B: Lettuce seedling and leaf discs viewed under normal light.
C and D: The same materials viewed under fluorescent conditions.
Leaf discs in B and D were taken from individual plants of a T2 family that segregated for the transgene and therefore the fluorescence phenotype.
Since 2012, a new technology for precise genome editing based on the CRISPR/Cas9 system of bacterial immunity has been developed and adapted for use in plants and animals. This technology can be used to create gene knockouts, deletions and replacements and 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 are developing and applying genome editing technology for lettuce. To demonstrate the efficiency of generating monoallelic and biallelic gene knockouts in lettuce with CRISPR/Cas9, we targeted a gene involved in seed germination thermotolerance (NCED4); this gene had previously been cloned by Kent Bradford’s group and provides a whole-plant selectable phenotype. A total of 59 transgenics were generated of two lettuce genotypes (cvs. Cobham Green and Salinas) transformed with three different CAS9 constructs each containing one guide RNA (gRNA) targeting NCED4. The NCED4 gene was sequenced in the T1 primary transformants. Of the 47 T1 plants containing the CAS9 gene, 26 plants (53%) showed editing of at least one NCED4 allele. Editing efficiency was similar in both genotypes, while the different gRNAs varied in efficiency. Homozygous knockouts of NCED4 resulted in large increases in their maximum temperature for germination, with both cultivars germinating over 70% at 37°C (Fig. 2). Moreover, because LsNCED4 can provide a simple, whole-plant selectable marker for use in a CRISPR co-editing strategy, simultaneous targeting of LsNCED4 and an unrelated gene of interest would enable the selection of plants enriched for editing events in the gene of interest simply by germinating seeds at high temperature.

CRISPR-mediated knockouts (KOs) have been made for other genes and similar efficiencies have been observed. The T2 progeny are currently being phenotyped. KO of Phytoene desaturase results in white sectors in leaves and shoots and therefore provides a visual marker for editing. KOs of LsVert1 and LsVert3 will confirm which gene determines resistance to Verticillium race 1. LsERF1 is another gene cloned by the Bradford group that confers the lack of thermoinhibition of seed germination.

In the longer term, we will develop 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 (e.g. Fig. 3). 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.
Figure 2. Editing of NCED4 results in germination of lettuce cv. Salinas at high temperature. A. Distribution of CRISPR mutation types in the T₁ generation for Salinas (26 T₁ plants) and Cobham Green (21 T₁ plants). B, C: Average germination of 6 homozygous edited T₃ families at different temperatures for Cobham Green (B) and Salinas (C) compared to wildtype. D. Germination at 35°C of T₂ seed from three independent lines plus an unedited control (top left).

Figure 3. Example of a conceptual stack of resistance genes conferring resistance to multiple pathogens plus a herbicide resistance gene (ALS) as a selectable marker.

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 the ultra-dense map as well as align them with 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 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. Comprehensive descriptions of all resistance genes in lettuce have been published (Christopoulou et al., 2015. Dissection of two complex clusters of resistance genes in lettuce (Lactuca sativa). Molec. Pl.-Micobe Interact. 28: 751-765; Christopoulou et al., 2015. Genome-wide architecture of disease resistance genes in lettuce. G3, Genes, Genomes, Genetics 5:2655-269.). These papers provide extensive enabling data for marker development.
Resistance to Downy Mildew

We have reviewed the available information on sources of resistance to downy mildew and rationalized it into 51 resistance genes and factors plus 10 new uncharacterized sources and 15 quantitative trait loci along with supporting documentation as well as genetic and molecular information. This involved multiple rounds of consultation with many of the original authors. This has been published and provides a foundation for future nomination of new Dm genes (Parra et al., 2016. Rationalization of genes for resistance to Bremia lactucae in lettuce. *Euphytica* 210:309-326. DOI 10.1007/s10681-016-1687-1).

We are using a whole genome sequencing (WGS) approach 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. Low-pass WGS of NILs provides the opportunity to efficiently identify these regions in order to map the resistance gene(s) and provide markers for marker-assisted selection (MAS). Over the past 30 years we have generated advanced breeding lines that constitute NILs of new Dm genes from the wild species, *L. serriola* and *L. saligna*, in the *L. sativa* cv. Salinas background. Twelve of these advanced breeding lines show resistance to many isolates representative of *B. lactucae* in California, including recent highly virulent isolates; one of our advanced lines was resistant to all of the isolates assayed (Table 1).

**Table 1. Reactions of advanced breeding lines to old and recent isolates of *B. lactucae* from California.**

Green = resistant reaction.

Red = susceptible reaction.
We have obtained low-coverage whole genome sequence of these 12 advanced breeding lines. Mapping the sequencing reads from these 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). In some cases, NILs seem to have multiple polymorphic regions (Fig. 4).

**Figure 4. Polymorphic regions resulting from introgression of resistance into an advanced breeding line from *L. serriola*.** Two regions coincide with Major Resistance Cluster 1 and 2 (MRC1 and MRC2). Additional regions that do not coincide with any canonical resistance gene sequences and may or may not be involved in resistance.

<table>
<thead>
<tr>
<th>Chr1</th>
<th>Chr2</th>
<th>Chr3</th>
<th>Chr4</th>
<th>Chr5</th>
<th>Chr6</th>
<th>Chr7</th>
<th>Chr8</th>
<th>Chr9</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
</tbody>
</table>

We have generated F2 populations from crosses between these NILs and cv. Salinas in order to determine which of these polymorphic regions are required for resistance. Most of the populations show a segregation ratio between resistant and susceptible individuals of 3:1, consistent with one dominant resistant gene involved in the resistance to downy mildew. However, 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. F2 progeny from these crosses are being low-pass sequenced to determine which regions confer resistance and to provide markers for selecting for resistance genes in order to generate resistance gene pyramids.
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 identification of the repertoire of resistance genes 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* (Fig. 5). Resistance genes will be validated using CRISPR-mediated knock-outs of the candidate gene.

**Figure 5: Data inputs for the identification of candidate genes for resistance to *B. lactucae*.**

![Genetic Analysis](chart)

**Genetic Analysis**

We have been constructing detailed genetic maps using Genotyping by Sequencing (GBS) on populations segregating for a variety of traits in collaboration with other researchers (Table 2). GBS uses methylation-sensitive enzymes to avoid repetitive regions of the genome and reduce the amount of sequencing required. We implemented a bioinformatics pipeline using Tassel and custom scripts for fast and accurate SNP haplotyping and the development of maps for each population.

**Table 2: Populations analyzed with GBS for mapping traits**

<table>
<thead>
<tr>
<th>Population</th>
<th>Trait</th>
<th>PI/Collaborators</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI251246 x Arm. <em>L. serriola</em></td>
<td>RIL Verticillium race 2 resistance</td>
<td>UCD</td>
</tr>
<tr>
<td>Lolla Rosa x Salinas</td>
<td><em>F</em>₂,₃ Fusarium resistance</td>
<td>UCD</td>
</tr>
<tr>
<td>Iceberg x PI491224</td>
<td><em>F</em>₂,₃ LDM field resistance</td>
<td>UCD/USDA</td>
</tr>
<tr>
<td>Grand Rapids x Salinas</td>
<td><em>F</em>₂,₃ LDM field resistance</td>
<td>UCD/USDA</td>
</tr>
<tr>
<td>Salinas x Calicel</td>
<td>RIL Tipburn</td>
<td>UCD</td>
</tr>
<tr>
<td>PI171674 x PI204707</td>
<td><em>F</em>₂,₃ Verticillium race 2 resistance</td>
<td>USDA</td>
</tr>
<tr>
<td>Reine des Glaces x Delsay</td>
<td>RIL Xanthomonas resistance</td>
<td>INRA, France</td>
</tr>
<tr>
<td>Salinas x Green Towers</td>
<td><em>F</em>₂,₃ Fusarium resistance</td>
<td>UCD</td>
</tr>
<tr>
<td>Red Tide x Lolla Rosa</td>
<td><em>F</em>₂,₃ Fusarium resistance</td>
<td>UCD</td>
</tr>
</tbody>
</table>
Vanda x Gisele F2:3 Fusarium resistance Brazil
Grand Rapids x Iceberg F2:3 LDM field resistance/NUE,WUE UCD/USDA
P171674 x Salinas F2:3 Verticillium race 2 Resistance USDA
Reine des Glaces x Eruption F2:3 Xanthomonas Resistance USDA
RH08-0111 x PI491108-1 F2:3 USDA
Saladin x Iceberg RIL U. of Warwick, UK
Salinas x L. serriola RIL Core population UCD
King Louie x Autumn Gold F2:3 Fusarium resistance USDA/UCD
Ninja x Valmaine F2:3 Effector recognition variation UCD

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). A new trial in summer 2016 focused on NUE was conducted at Spence Farm. to test if the QTL identified previously are reproducible across years. Data are currently being analyzed using a new genetic map generated with GBS data.

In collaboration with Blue River Technology, a drone was used to collect remote data on the plants growing in the 2016 trial. The objective was to test the correlation of the remote data with data taken on the ground. Blue River Technology provided estimates for biomass (Kg) and a proxy for leaf area. Average plant weight, percent solids, percent nitrogen, percent phosphorous and percent potassium were measured on the ground. Only one of the flights, conducted two days before harvest, was proofed with ground data. Data analysis continues. We will use the drone proxy for leaf area in QTL analysis for NUE. Leaf area is an important trait in NUE since it affects the amount of light a plant can capture.

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). We fine mapped this QTL by phenotyping lines selected as being recombinant using codominant molecular markers flanking and within the QTL region. Informative recombinants homozygous for five markers within the QTL interval were phenotyped for tipburn in Yuma, AZ during the winter of 2015-2016 and in Salinas, CA during the spring of 2016. Data were collected for tipburn incidence and severity. Haplotypes differed in tipburn in both field trials (Fig. 7). The middle marker in the haplotype was highly correlated with tipburn. Haplotypes carrying the resistance allele (D allele from El Dorado) at this marker showed a decrease on tip burn incidence up to six fold compared to those carrying the susceptible allele (E allele from Emperor).

These recombinant lines were sequenced and analysis of chromosomal recombination points has further reduced the interval associated with tipburn resistance. The genomic interval currently contains 21 candidate genes. We are analyzing these genes for expression differences that could be associated with the differences in resistance to tip burn. Candidate genes will be validated using CRISPR-mediated KOs and transgenic complementation.
Figure 7: Tip burn incidence per recombinant haplotype in Yuma and Salinas trials. Parental lines El Dorado and Emperor are haplotypes DDDDD and EEEEEE respectively.

We also developed a RIL population of 382 families from a cross between the crisphead cultivars Salinas (tipburn resistant) and Calicel (tipburn susceptible) specifically to analyze resistance to tipburn in cultivars adapted for California production. This population was evaluated in Salinas, CA in 2015 for tipburn resistance and genotyped using GBS. Five phenotypes were measured: plant fresh weight, tipburn incidence and severity, leaf serration and head firmness. QTLs were identified for all of these traits. A major QTL tipburn on LG5 is in the same region as the QTL identified in the Emperor x El Dorado population. Therefore, the major QTL that we are analyzing in the Emperor x El Dorado population also seems to be relevant in Californian cultivars.

Genome and Transcriptome Sequencing

The genome of lettuce was initially sequenced in collaboration with the BGI, Shenzhen, China, funded by an international consortium of ten companies as well as the BGI. We now have 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. We have placed genetically validated scaffolds relative to numerous phenotypes. The genome has been annotated to provide ca. 41,000 high-confidence gene models. The 7.0 version of the 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.

**Figure 8. Overview of the v7 reference genome assembly of lettuce cv. Salinas.** Track A: Number of scaffolds in 1 Mb intervals. Blue, SOAPdenovo scaffolds; Red, HiRise superscaffolds. B: Chromosomal pseudomolecules. Dark areas indicate 63% of genome positioned and oriented accurately. C: Gene density. D: Repeat density. E: Density of SNPs used for genetic map construction. F: Size of tandem gene arrays. Black blocks show MRC regions. Colored lines in center link syntenic blocks of five plus genes due to whole genome triplication.
The second phase of the Lettuce Genome Project is underway that is 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. This three-year endeavor is funded by the International Lettuce Genome Consortium that comprises of 17 large and small breeding companies. For this we are working in collaboration with Dovetail Genomics (Santa Cruz) to provide chromosome-scale genomic scaffolds. In collaboration, with Ivan Simko and others we have sequenced ~90 lettuce cultivars and wild accessions. The SNPs relative to the reference genome are being made available through our website.

In collaboration with 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. These data are being made available through our website.

**Databases**