

[My Desktop](#)
[Prepare & Submit Proposals](#)
[Proposal Status](#)
[Proposal Functions](#)
[Awards & Reporting](#)
[Notifications & Requests](#)
[Project Reports](#)
[Submit Images/Videos](#)
[Award Functions](#)
[Manage Financials](#)
[Program Income Reporting](#)
[Grantee Cash Management Section Contacts](#)
[Administration](#)
[Lookup NSF ID](#)

Preview of Award 0820451 - Final Project Report

[Cover](#) |
[Accomplishments](#) |
[Products](#) |
[Participants/Organizations](#) |
[Impacts](#) |
[Changes/Problems](#)

Cover

Federal Agency and Organization Element to Which Report is Submitted:	4900
Federal Grant or Other Identifying Number Assigned by Agency:	0820451
Project Title:	Comparative Genomics of Phenotypic Variation in the Compositae
PD/PI Name:	Loren H Rieseberg, Principal Investigator Kent J Bradford, Co-Principal Investigator Richard V Kesseli, Co-Principal Investigator Richard W Michelmore, Co-Principal Investigator
Recipient Organization:	Indiana University
Project/Grant Period:	03/15/2009 - 02/28/2015
Reporting Period:	03/01/2014 - 02/28/2015
Submitting Official (if other than PD\PI):	Loren H Rieseberg Principal Investigator
Submission Date:	05/22/2015
Signature of Submitting Official (signature shall be submitted in accordance with agency specific instructions)	Loren H Rieseberg

Accomplishments

*** What are the major goals of the project?**

Long Term Goals of Compositae Genome Project:

- To establish tools and resources for important crops and weeds in Compositae that will be the basis of genomic investigations throughout the family.
- To develop comprehensive gene catalogs and ultimately fully sequenced genomes for economically and evolutionarily

significant members of the Compositae.

- To develop detailed physical and genetic maps that integrate phenotypic and genotypic data.
- To characterize and exploit agriculturally useful alleles from wild species.
- To understand genome evolution and phenotypic diversification in the Compositae.

Specific Goals of This Proposal:

- To extend the Compositae sequence database emphasizing the gene space of key species and allelic variation across the family.
- To determine the relative frequencies of nucleotide versus copy number variation.
- To quantify the effects of paleopolyploidy on diversification rates.
- To exploit parallel phenotypic transitions in crops and weeds to test evolutionary hypotheses.
- To determine syntenic relationships within the Compositae and with other flowering plants.
- To identify and characterize candidate genes for domestication and weed traits.
- To train students at all levels from diverse backgrounds in evolutionary genomics.

* What was accomplished under these goals (you must provide information for at least one of the 4 categories below)?

Major Activities:

Gene Space - We adapted a duplex-specific nuclease protocol for generating genomic libraries depleted for repeated elements to sequence the gene space of lettuce, sunflower, safflower and gerbera to a depth of ~20X as proposed. We used this information to design a set of sequence capture probes that target conserved orthologous sequences in the Compositae, with the goal of developing an approach for efficient phylogenomic analyses. We also developed a bioinformatic and phylogenetic workflow for processing and analyzing the resulting data. We applied this approach to 15 species from across the Compositae, resulting in the production of phylogenetically informative sequence data from 763 loci.

Genome Sequencing - We have generated draft whole genome assemblies for ten Compositae species: cultivated lettuce, cultivated sunflower, wild lettuce, horseweed, fleabane, common ragweed, chicory, endive, globe artichoke, and safflower. The lettuce, sunflower, and fleabane genome assemblies were primarily funded from other sources. Most of these assemblies relied on deep Illumina sequencing (circa 100x depth) and long mate-pair libraries (to aid with scaffolding). For several species, a Hi-C based library preparation (ragweed) or longer sequence reads from the 454 (horseweed and sunflower) and/or PacBio (horseweed) platforms were used to increase contig and scaffold lengths. For most of the species (lettuce, sunflower, fleabane, ragweed, chicory, endive, artichoke, and safflower), scaffolds have been placed and ordered on linkage groups through the generation and use of high-density genetic maps (see below). For sunflower, a sequence-based physical map permitted additional fine-scale ordering and orientation of scaffolds.

Expressed Sequence Tags - We developed protocols for normalized cDNA library development for both the Illumina and 454 sequencing platforms. Using these protocols, we obtained ESTs for 76 genotypes, representing 43 different taxa (Table 2). These include ESTs for 11 food crops, 1 medicinal crop, 2 horticultural crops, 1 industrial crop, 10 weeds, 13 crop or weed progenitor species, a representative of a previously unsampled subfamily, and four outgroup species that represent the most related families to the Compositae. These outgroup species were sampled to determine the placement of ancient genome duplications that we previously had shown to be close to the base of the family.

We also used a PacBio RSII to sequence large cDNA libraries derived from several genotypes of cultivated lettuce. The cDNAs were size-selected prior to sequencing to enrich for transcripts longer than 2 kb. More than 90% of the error-corrected reads were

full-length. The long read lengths are being used to assist in gene annotation and to identify paralogs and splice variants that were poorly assembled in the previous transcriptome assemblies derived from small insert Illumina libraries. At least 85% of these error-corrected transcripts are longer than the predicted gene models due to additional information on UTRs and alternative splicing events.

Copy Number Variation – The genomes of cultivated and wild lettuce were compared and CNVs with genes missing or increased in either cultivated lettuce or wild lettuce were identified. We have characterized the cultivated sunflower germplasm collection and identified a hierarchical core set of lines (i.e., diversity panel) that captures as much allelic diversity as possible for a given sample size. We have re-sequenced 288 of these lines to 5-20x depth and have analyzed the resulting data for CNVs.

Paleopolyploidy and genetic diversity - We developed a highly automated pipeline to infer gene family phylogenies from EST or genomic DNA sequences and to summarize the age of duplication events (http://evopipes.net/dup_pipe.html). We used this pipeline, in combination with synteny information, to show that the Compositae are ancient hexaploids. By analyzing the age distribution of duplicate genes in the outgroup families, we were further able to show that the Compositae and Calyceraceae share a common paleotetraploidy event. Significant improvements in marker densities on genetic maps for sunflower, safflower, and lettuce (below) have confirmed and extended earlier findings related to ancient polyploidization.

Additional evolutionary analyses were conducted by the CGP team that exploited the large amounts of sequence data generated by the project. These include the detection of genes under selection during domestication and speciation in the Compositae, identification of the factors controlling the formation and size of islands of differentiation during speciation, intron gains and losses across the family, and determination of the wild parents of Jerusalem artichoke.

RILs, genetic maps, and synteny – We employed genotyping by sequencing approaches to develop ultra-dense genetic maps for 11 Compositae taxa. For cultivated lettuce, cultivated sunflower, wild lettuce, chicory, endive, globe artichoke, and safflower we employed low coverage whole genome shotgun (WGS) sequencing data for genotyping, whereas for fleabane, noug, ragweed, and starthistle, genome complexity was reduced with restriction enzymes prior to sequencing. The genetic maps were employed to resolve paleopolyploidy events, estimate syntenic relationships across the family, and map QTLs for phenotypic traits (see below). For several species (lettuce, sunflower, chicory, endive, and safflower), mapping populations were advanced to generate permanent resources (i.e., recombinant inbred lines).

Genetic analyses of similar phenotypic transitions - A major effort of the CGP has been the phenotypic and genetic characterization of the lettuce and sunflower diversity panels, as well as RIL populations of both species. For lettuce, these have been genotyped and distributed to international collaborators to facilitate the genetic dissection of numerous domestication, developmental, and physiological traits. These include abiotic stress resistance, flowering time, heat stress, secondary metabolism, water use efficiency, seedling traits, root structure, seed dispersal, leaf shape, spininess, high temperature seed germination, and light requirement for germination. We are fine-mapping several domestication genes using WGS data from 96 RILs for *in silico* bulked segregant analysis and by saturating the regions with recombinants to refine the genetic intervals. In addition, a distinct major QTL (*Htg9.1*) for high temperature seed germination has been identified in a second RIL population and associated with a candidate gene. This QTL is adjacent to a second one for variation in seed germination thermotolerance due to differing maternal environments during seed

maturation.

As reported previously, sunflower RIL and testcross populations have been phenotyped extensively and used for QTL analyses of agronomic and biochemical traits. The resulting data have provided starting points for map-based cloning of genes involved in several key traits including ray floret identity, flowering time variation, heterosis, chromosome pairing, and circadian regulation. We have developed efficient methods for transformation and RNAi in sunflower, which are being employed for functional characterization of these genes. We also have developed a permanent, publicly available association mapping resource for sunflower with funding from other sources. The population has been genotyped by WGS sequencing (see above), which resulted in the discovery of 4.1 million high quality SNPs. Analyses of the population using the 4.1 million SNPs revealed numerous genetic associations for a variety of important agronomic traits, ranging from seed oil content, to flowering time, to branching. As we had hoped, in many cases associations can be narrowed down to a single gene.

Specific Objectives:

1. Characterize the gene-space in the Compositae.
2. Extend the EST database with an emphasis on allelic diversity.
3. Analyze the relative frequencies of nucleotide versus copy number variation.
4. Analyze paleopolyploidy and genetic diversity in an evolutionary framework.
5. Develop RILs and genetic maps, and conduct asyntenic analyses.
6. Identify the genetic changes underlying similar phenotypic transitions in the Compositae.
7. Communicate our results externally.
8. Educate students in the technical and analytical aspects of evolutionary genomics.

Significant Results:

Gene space - Targeted sequence capture of COS loci identified from our gene space assemblies facilitated phylogenetic analyses based on a large number of genes across the Compositae. We generated usable sequence for 763 loci and recovered a phylogeny consistent with known relationships within the family. Moreover, our workflow was successful in reconstructing relationships within the *Helianthus* demonstrating that this method is useful for both broader- and finer-scaled phylogenetic reconstruction.

Whole Genome Assemblies. In collaboration with the broader Compositae scientific community, we generated draft whole genome assemblies for ten Compositae species.

1. The **lettuce** assembly includes 21,513 chromosome-anchored scaffolds comprising 2.4 Gb of the 2.7 Gb genome with a scaffold N50 of 461 Kb. Lettuce genome annotation identified 40,000 high-confidence gene models.
2. The assembly of **wild lettuce** covers 1.99 Gb of its 2.7 Gb genome and is comprised of ~45,020 scaffolds > 1 Kb with a scaffold N50 of ~107 Kb.
3. The **safflower** assembly spanned roughly two thirds of the expected 1.3 Gb genome size, but in ~2 million contigs.
4. The draft assembly of **common ragweed** covers 1.3 Gb, which is slightly larger than the estimated genome size, with N50 = 248 Kb.
5. The **fleabane** assembly is 1.4 Gb in length, which is circa 99% of the expected genome size, with N50 = 88 Kb).
6. The latest version of the **sunflower** assembly includes 14,940 chromosome-anchored scaffolds that cover 3.1 Gb of the 3.6 Gb genome, with N50 = 476 Kb. The genome includes >98% of core Eukaryotic Genes (CEGs), of which ~90% are full length. We also assembled a complete mitochondrial genome sequence.
7. The **chicory** was sequenced and assembled into 44,027 scaffolds with an N50 of 25 Kb covering approximately half of the 1.28 Gb genome, a quarter of which were

- anchored to chromosomes. Over 98% of CEGs were present.
8. **Endive** was sequenced and assembled into 21,766 scaffolds with an N50 of 61 Kb covering approximately half of the 1.07 Gb genome, a third of which were anchored to chromosomes. Over 98% of CEGs were present.
 9. The reference genome assembly of **Globe artichoke** was generated comprising of 13,662 scaffolds that covered 0.725 of the 1.08 Gb genome and encoded 26,906 predicted genes. Circa 73% of the scaffolds were anchored to chromosomes.
 10. The small 0.35 Gb **horseweed** genome was assembled into 13,966 scaffolds with N50 = 34 Kb. The assembly covered 92.3% of the genome and comprises 44,592 protein coding genes.

Expressed Sequence Tags - We evaluated coverage of the 76 Compositae transcriptomes using BLASTX analysis of a set of ultra-conserved orthologs, and recovered most of these genes (88–98%). We found a correlation between contig length and read length, suggesting that longer reads can aid in the assembly of more complete transcripts. We compared the divergence of orthologs at synonymous sites (K_s) between Compositae crops and their wild relatives and found greater divergence when the progenitors were self-incompatible. We also found greater divergence between pairs of taxa that had some evidence of postzygotic isolation. For several more distantly related congeners, such as chicory and endive, we identified a signature of introgression in the distribution of K_s values.

Copy number variation - We identified ~400 genes that were present in cultivated lettuce but missing in wild lettuce, including CNVs in disease resistance genes. In sunflower, we found that ~14% of coding sequences varied in copy number. CNVs were over-represented for genes involved in self-defense. Such variation may be advantageous in the presence of changing environmental conditions.

Paleopolyploidy and genetic diversity - See next section for genetic map-based insights into paleopolyploidy.

Evolutionary analyses. We employed project data to address questions regarding the origin(s) of Compositae crops. Using evidence from multiple domestication genes and from neutral markers, we demonstrated a single domestication event for extant cultivated sunflower in eastern North America. This finding further established eastern North America as an independent domestication center.

The Jerusalem Artichoke was also cultivated in eastern North America before European contact. Competing hypotheses regarding its origin implicate the occurrence of polyploidization with or without hybridization and list six species as potential parents. Using a genome skimming approach, we demonstrated that this species originated recursively from perennial sunflowers of eastern North America, via hybridization between tetraploid Hairy Sunflower and diploid Sawtooth Sunflower.

We addressed fundamental evolutionary questions and showed that 1) effective population size is an important predictor of adaptive evolution; (2) the functional architecture of genomes plays a key role in shaping patterns and repeatability of genomic divergence, and (3) the repeatability of genomic divergence is greater at macro- than micro-evolutionary scale. We also showed that harmful mutations are accumulating in the genomes of Compositae crops as a consequence of domestication bottlenecks.

Lastly, we developed a computational toolkit for identifying TE families from unassembled DNA sequence reads and showed that there has been a directional increase in copy number of Gypsy retrotransposons since the origin of the family.

Develop RILs, genetic maps, and synteny - Analyses of the sequence-based genetic maps for six Compositae species revealed a 3x polyploidy event that occurred after the 3x polyploidy event shared by most dicots. Sunflower had an additional 2x polyploidy event that occurred after the two aforementioned 3x polyploidization events. Comparison of genetic maps within *Helianthus* revealed that species differ by numerous major rearrangements. Genomic scans revealed elevated genetic differentiation in rearranged portions of the genome, indicating that such rearrangements restrict gene flow. In contrast, extensive synteny was observed between artichoke, lettuce, and chicory, indicating lower rates of chromosomal evolution in other parts of the family.

Genetic analyses of similar phenotypic transitions – Genetic comparisons of the two major oilseeds in the Compositae, sunflower and safflower, revealed that QTL for oil composition and flowering time map to homologous genomic regions in both species (as well as lettuce for flowering time). The evolution of these traits appears to have been driven by parallel genetic changes across species.

In lettuce, we confirmed that the major QTL associated with high temperature seed germination in wild lettuce (*Htg6.1*) is due to altered regulation of *LsNCED4*, encoding a key regulated gene in the ABA biosynthetic pathway. We characterized expression of all four *LsNCED* family members during seed development and germination under different temperature conditions. *LsNCED2* and *LsNCED3* respond to water stress while *LsNCED4* responds to high temperature.

In sunflower, we identified a member of the *Cycloidea* gene family (*Cyc2c*) responsible for both the tubular and double-flowered mutations in sunflower. The “*tubular*” mutation (in which the ray florets exhibit radial instead of bilateral symmetry) results from a transposon insertion within the *Cyc2c* coding region that that knocks out the function of this gene. An upstream insertion in this same gene causes a loss of ray-specific expression. Instead, this allele is expressed across the composite inflorescence, resulting in disc flowers that develop into ray flowers, similar to Van Gogh’s famous sunflowers.

We also showed that four paralogs of the *FT/TFL1* family have been targets of selection during sunflower domestication and underlie flowering time QTL. A frameshift mutation in *HaFT1* also causes heterosis in sunflower and Arabidopsis. Our BAC-based analysis of the genomic region surrounding *fad2-1*, a fatty acid desaturase, revealed a selective sweep spanning ≥ 95 kb.

Key outcomes or Other achievements:

Sequenced and assembled the gene space of multiple genotypes from three crop lineages.

Sequenced, assembled, and annotated the genomes of ten genotypes from six crop lineages.

Generated high quality EST/transcriptomic data for 76 genotypes from 42 taxa.

Demonstrated extensive copy number variants (CNVs) in lettuce and sunflower.

Established the position of multiple polyploidization events during Compositae evolution.

Developed and characterized >10 mapping populations for multiple species.

Cloned and characterized domestication and agriculturally important genes.

Developed lettuce germplasm for distribution with an introgressed native allele and mutations in *LsNCED4* to improve seed germination at high temperatures.

Continued development and curation of the CGP website and database.

Trained 148 undergraduate, graduate, and postdoctoral students in genomics/bioinformatics (34 during current reporting period).

Published 102 papers (18 during current reporting period).

*** What opportunities for training and professional development has the project provided?**

This project has contributed to the training of 19 undergraduate, 11 graduate, and 4 postdoctoral students in the areas of plant genomics and bioinformatics over the past year. Additionally, it contributed to the training and experience of 3 technicians and 8 PhD-level project scientists. Common to all labs, each of the students that have worked on this project gained skills and experience in a number of field and laboratory techniques associated with genetics, genomics, genetic mapping and bioinformatics. Below we provide a brief description of each trainee's contributions to the project.

Postdocs

Savithri Nambesan worked on candidate gene and physiological analyses related to plant architecture and other domestication traits in sunflower (Burke lab).

Adam Bewick took the lead on the processing of the sunflower re-sequencing data and the associated CNV analyses (Burke lab).

Evan Staton worked on bioinformatic analyses of variation in genome composition in sunflower and related species (Burke lab).

Qian Yu analyzed the molecular basis of heterosis in sunflowers (Rieseberg lab).

Graduate Students

Jan Chang completed work on papers from his study of intron evolution across the Compositae (Kesseli lab).

Trudi Gulick is characterizing epigenomic variation and transgenerational inheritance in weedy and domesticated lettuce (Kesseli lab).

Caitlin Ishibashi is responsible for characterizing traits related to abiotic stress tolerance in wild and cultivated and sunflower (Burke lab).

Ed McAssey is responsible for sequencing and evolutionary analyses of candidate genes for domestication traits in sunflower (Burke lab).

Rishi Masalia has been involved in evolutionary analyses of sequence diversity across the Compositae (Burke lab).

Dina Tsirelson is contributing to the genetic mapping of QTL in lettuce (Kesseli lab).

Fei-Yian Yoong identified a likely candidate gene associated with a seed dormancy QTL in RILs derived from *L. sativa* cv. Salinas x accession PI251246, a primitive landrace of *L. sativa*. She also analyzed a large RNASeq dataset for gene expression correlation networks from the same genotypes. She completed her PhD thesis in March 2015. (Bradford lab).

Tomas Zavada characterized genetic diversity in the wild and domesticated accessions of chicory and endive from Europe and weedy accessions of chicory across North America (Kesseli lab).

Sebastian Reyes Chin Wo is responsible for gene annotations and assists in genome assemblies of lettuce, chicory and

endive (Michelmore lab).

Marilena Christopoulou has contributed the large cDNA sequencing component of lettuce using the PacBio RS (Michelmore lab).

Emilio Ortiz worked on assembly and comparative genomics of Gerbera species (Still lab).

Undergraduate Students

Celina Abundis graduated and has begun a PhD program. She contributed to the genetic mapping of QTL in lettuce (Kesseli lab).

Corey Aguilar will graduate in May 2015. He worked on the natural history and dispersal pathways of weedy chicory (Kesseli lab).

Lisette Bayona graduated. She worked in the spring and summer 2014 with graduate student Gulick on comparative weed-domesticated epigenetics and led efforts to develop teaching materials for the undergraduate Genetics course at UMB based on lettuce RILs developed in this project (Kesseli lab).

Caleb Botta assisted with DNA extractions, greenhouse work, and phenotyping of sunflower mapping populations (Burke lab).

Jeremiah (Seth) Bradley assisted with DNA extractions, greenhouse work, and phenotyping of sunflower mapping populations (Burke lab).

Kody Brindley assisted with DNA extractions, greenhouse work, and phenotyping of sunflower mapping populations (Burke lab).

Teresa Jardini assisted with candidate gene analyses in lettuce (Michelmore lab).

Florence Lip Ka Ying was involved in fine-mapping the *Htg9.1* QTL and extracting RNA for RNA-Seq project (Bradford lab).

Rondy Malik graduated and began a PhD program at Indiana University. He studied genetic diversity in weeds (Chicory) under the guidance of graduate student Zavada and is co-author of one manuscript that has been submitted for publication (Kesseli lab).

Bao Nguyen assisted in hormone-related experiments including germination assays, ethylene measurements, seed counting, and hormone preparation and Simon Foucher applied different bioinformatic methods to analyze RNA-seq data. Bao Nguyen presented an oral report at the UC Davis Undergraduate Research Conference on her work. (Bradford lab).

Adolfo Vargas assisted with candidate gene analyses in lettuce (Michelmore lab).

Diana Zafra worked on the introgression of *NCED4* alleles from *L. serriola* into *L. sativa* germplasm (Still lab).

Alvaro Martinez helped Maria DNA extractions of segregating populations (Michelmore lab).

Fatumata Bah, a Ronald McNair fellow, worked with graduate students Gulick and Stuart Morey on seedling and root QTL in lettuce (Kesseli lab).

Alexander Winnett has been developing a bioinformatics pipeline to help students identify candidate genes for QTL in studies being conducted in the Kesseli lab. He is also working with graduate student Gulick on documenting methylation patterns in lettuce.

William Fatherley will graduate May 2015. He conducted a study on plasticity in lettuce as part of an REU summer program at UMB (Kesseli lab).

Zahra Ammar was a student in the Bridges to Baccalaureate program at UMB. She completed a summer research project with graduate student Tsirelson on root traits and responses to changing light conditions in lettuce (Kesseli lab).

Blake Stark evaluated germination of lettuce in response to far-red light and salinity (Still lab)

Deborah Lopez evaluated germination of lettuce seeds in response to salinity, evaluated root growth under reduced water potentials (Still lab).

Fabian Perez worked on phenotyping and genotyping to assess effects of introgression of *NCED4* alleles from *L. serriola* into *L. sativa* germplasm (Still lab).

Technicians

Tyler Fletcher provided molecular biology and field/greenhouse support for sunflower and safflower research (Burke lab).

Huaqin Xu developed and managed the database for the Composite project (Michelmore lab).

Mirna Calderon worked on the introgression of *NCED4* alleles into commercial lettuce cultivars (Still lab).

PhD-level Project Scientists

Dr. John Bowers conducted bioinformatics work related to SNP and SFP mapping, sequence analysis of EST datasets, and genetic map construction (Burke lab).

Dr. Peetambar Dahal developed recombinant introgression lines of the native *LsNCED4* allele from US96UC23 into Salinas background. (Bradford lab).

Alex Kozik was the lead bioinformatician for the Composite project (Michelmore lab). Dean Lavelle characterized candidate genes (Michelmore lab).

Maria-Jose Truco is contributing to the integration of molecular maps developed on multiple lettuce populations segregating for a variety of traits including resistance genes and morphological traits. She also is contributing to QTL mapping of domestication traits of agricultural importance relative to candidate genes (Michelmore lab).

Tadeusz Wroblewski analyzed candidate genes in lettuce (Michelmore lab).

Lutz Freonike developed the algorithms for haplotyping and mapping genomic scaffolds (Michelmore lab).

Dean Lavelle coordinated the resequencing efforts and candidate gene identification (Michelmore lab).

Youngsook You assessed genetic and epigenetic mechanisms responsible for trait stability/phenotypic stability in response to environmental stress in a lettuce recombinant inbred line. Has been responsible for bioinformatics and assembly of *Gerbera* genomes (Still lab).

*** How have the results been disseminated to communities of interest?**

Information generated by the project is disseminated to the scientific community through the CGP website (http://compgenomics.ucdavis.edu/compositae_database.php), publications (18 during reporting period), and talks and posters at scientific meetings. In addition, Michelmore and Burke co-organize the Compositae Workshop at the annual Plant and Animal Genome Conference. Last year we increased the number of datasets housed within the CGP, including access to EST collections and transcriptome assemblies, as well as links to the latest genetic maps of lettuce and sunflower. We plan to add all current NGS based mapping projects to the website including the maps of chicory, endive, artichoke, safflower, and noug.

The CGP has broad impacts through undergraduate education. We regularly employ undergraduates to assist in projects, which exposes them to the latest molecular and computational approaches. All of our institutions have strong, proactive programs to encourage application, recruitment and research participation of under-represented and disadvantaged groups. U Mass Boston (UMB) and Cal Poly Pomona (CPP), in particular, have high proportions of student groups traditionally under-represented in the sciences.

The Kesseli lab is an active participant in four major programs at UMB that are aimed at recruiting and improving the success of students from groups underrepresented in the sciences. The Ronald McNair Program is based within the College of Math and Sciences and the Kesseli lab has been a regular location for the placement of these students. The NIH Initiative to Maximize Student Diversity (IMSD) is housed in the Biology department and supplies financial and academic support for students planning to pursue graduate careers in biological sciences. Kesseli is on the advisory board for this program. The NIH funded Bridges to Baccalaureate program housed in the Biology department partners UMB with two community colleges in the Boston area. The goal is to give students at the community colleges the opportunity to attend and succeed in 4-year colleges. The students from these programs have successfully completed their degrees and continued to be involved in research. Former students Monir Ejemel and Roba Hersi graduated with honors from UMB. Rondy Malik also came from first from the Bridges program and then entered the IMSD program; he graduated from UMB in June 2013 and began a graduate program in plant biology at Indiana University. He is still collaborating with former members of the CGP and is co-author on a manuscript recently submitted to the American Journal of Botany. Celina Abundis graduated in December 2013 and is now working in a plant biology research lab at University of Massachusetts, Amherst as part of a pre-doctoral training program. She worked last year with graduate student Stuart Morey on a manuscript examining stress responses QTL in lettuce. The Biology department also houses an NSF funded REU site program and the Kesseli lab has supported close to two dozen students in that program and many have worked within the CGP. These students have presented their work regularly at annual national meetings catering to under-represented minorities such as Society for the Advancement of Hispanic/Chicanos and Native Americans in Science (SACNAS) and Annual Biomedical Research Conference for Minority Students (ABRCMS).

Discussions in the literature (e.g. Redfield 2012) have highlighted the need to introduce more engaging and relevant tools and technologies into the teaching curricula. These discussions and the CGP's strong commitment to outreach and to enhancing educational opportunities for students, have prompted Kesseli's group to incorporate resources generated within the CGP into our classrooms. The recombinant inbred lines (RILs) generated from a cross between domesticated (Salinas) and wild lettuce have been used previously in summer workshops for training high school teachers. Kesseli introduced these and other tools into sophomore level genetics labs at UMB and nearly 700 students have participated in these labs over the past 6 semesters.

Scores of traits segregate in the RILs and students can choose and analyze "their" trait. With simple Excel files containing previously mapped single nucleotide polymorphisms of the RILs, students were able to perform genome wide association studies and map their trait.

The plasticity of lettuce allows for plants to be grown in flats of 2 x 2 cm pots and scored as seedlings or later as flowering adults. The system provides opportunities to map genes, characterize epistatic interactions, and conduct follow-up studies in bioinformatics and molecular genetics. The basic steps to produce this lab are outlined below.

Prior to the class, RILs and parents are planted in rows (4 individuals/RIL) with in flats.

1. The Lab manual and teaching assistant described RIL populations, single nucleotide polymorphisms (SNPs), and concepts of genome wide association studies (GWAS), quantitative traits and epistasis.
2. Teams of three students were provided with 3 flats containing 60 RILs and 2 parents.
3. Students choose their traits by identifying either differences between the parents or unique segregating traits within the RIL population (likely caused by epistatic interactions).
4. Students are provided with Excel files with the RIL and the SNP data.
5. Students copy and paste the columns of RILs with similar extreme phenotypes into adjacent columns to visualize genomic regions common to one phenotypic category and differing from the other.
6. Students identify RILs belonging to extreme phenotypic classes.

Outcomes:

Students were able to identify candidate regions for several traits not previously mapped.

Some traits chosen by students proved to be complex or difficult to score and clear candidate genomic regions were not identified. For many others, however, students identified one or more candidate regions.

Student interest in the mapping and gene interaction lab increased three fold with the addition of the RILs and familiar lettuce

system over the previous year.

Conceptual understanding seems improved based on test and in class iclicker questions though further study is needed.

Co-PI Still integrated topics from the CGP into the course Plant Breeding (PLT 404). Still organized and hosted the Agriculture Research Institute Research Showcase (November 2014), which promoted agriculture-based research of undergraduate and graduate students at CPP. He also edited articles for the College of Agriculture's magazine "Agricolumn" featuring agriculture-related research projects at CPP.

Caitlin Ishibashi, Ed McAssey, and Rishi Masalia (graduate students, Burke lab) have worked with other graduate students in the UGA Department of Plant Biology to establish and maintain an outreach program with a seventh grade agriscience class at a public middle school in Athens, Georgia. Through a series of lessons about the ecology, domestication, and taxonomy of different crop plants, this program aims to improve science literacy through teaching about the impact that humans have had on plants and vice versa.

One undergraduate at UC Davis (Bradford lab) presented her work under this project at the 26th Annual Undergraduate Research, Scholars and Creative Activities Conference in May, 2015: Bao Nguyen, Identification of a Candidate Gene for a High Temperature Germination Trait in Lettuce.

A PhD student at UC Davis, Fei-Yian Yoong (Bradford lab), presented her work under this project at the Annual Meeting of the American Society of Plant Biologists, Portland, Oregon, July 12-15.

Project scientist Youngsook You (Still Lab) presented his work from this project at the Plant & Animal Genome XXIII Conference, San Diego CA, January 10-14, 2015 and at the American Society of Plant Biologists Annual Meeting, Portland, Oregon. July 12-16.

Products

Books

Book Chapters

Conference Papers and Presentations

You Y, Ellis DR, Still DW (2015). *Breeding lettuce to maximize nutritional content across multiple environments*. Plant & Animal Genome XXIII Conference. San Diego CA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

You Y, Ellis DR, Still DW (2014). *Maternal environment during seed maturation affects seed germination under stress: genetics or epigenetics?*. American Society of Plant Biologists Annual Meeting. Portland, Oregon. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Inventions

Journals

Barb, J.G., J.E. Bowers, S. Renaut, J.I. Rey, S.J. Knapp, L.H. Rieseberg, and J.M. Burke (2014). Chromosomal evolution and patterns of introgression in *Helianthus*. *Genetics*. 197 969. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Baute G.J., N.C. Kane, C. Grassa, Z. Lai, and L.H. Rieseberg (2015). Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives. *New Phytologist*. 206 830. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Hartman, Y., Hooftman, D.A.P. , Uwimana, B., Schranz, M.E., van de Wiel, C.M., Smulders, M.J.M., Visser, R.G.F., Michelmore, R.W., van Tienderen , P.H. (2014). Abiotic stress QTLs in lettuce crop-wild hybrids: comparing greenhouse and field experiments.. *Ecology and Evolution*. 4 2395. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Hodgins, K.A., D.G. Bock, M.A. Hahn, S.M. Heredia, K.G. Turner, and L.H. Rieseberg (2015). Comparative genomics in the Asteraceae reveals little evidence for parallel evolutionary change in invasive taxa. *Molecular Ecology*. 24 2226. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Mandel, J.R., E.V. McAssey, S. Nambeesan, E. Garcia-Navarro, and J.M. Burke (2014). Molecular evolution of candidate genes for crop-related traits in sunflower (*Helianthus annuus* L.). *PLoS ONE*. 9 e99620. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Marchand, G., V.A. Huynh-Thu, N. Kane, Nolan; S. Arribat, D. Varès, D. Rengel, S. Balzergue, L.H. Rieseberg, P. Vincourt, P. Geurts, M. Vignes, and N. Langlade (2014). Bridging physiological and evolutionary time scales in a gene regulatory network. *New Phytologist*. 203 685. Status = PUBLISHED; Acknowledgment of Federal Support = No ; Peer Reviewed = Yes

Nambeesan, S.U., J.R. Mandel, J.E. Bowers, L.F. Marek, D. Ebert, J. Corbi, L.H. Rieseberg, S.J. Knapp, and J.M. Burke (2015). Association mapping in sunflower (*Helianthus annuus* L.) reveals independent control of apical vs. basal branching. *BMC Plant Biology*. 15 . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: doi:10.1186/s12870-015-0458-9

Pearl, S.A. and J.M. Burke (2014). Genetic diversity in *Carthamus tinctorius* (Asteraceae; safflower), an underutilized oilseed crop. *American Journal of Botany*. 101 1640. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Peng, Y., Z. Lai, T. Lane, M. Nageswara-Rao, M. Okada, M. Jasieniuk, H. O'Geen, R.W. Kim, R.D. Sammons, L. Rieseberg, and C.N. Stewart, Jr. (2014). De novo genome assembly of the economically-important weed *Conyza canadensis* using integrated data from multiple sequencing platforms. *Plant Physiology*. 166 1241. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Renaut, S., and L.H. Rieseberg (2015). The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other Compositae crops. *Molecular Biology and Evolution*. . Status = ACCEPTED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Renee L. Eriksen, J. L. Hierro, O. Eren, K. Andonian, K. Torok, P. I. Becerra, D. Montesinos, L. Khetsuriani, A. Diaconu, Rick Kesseli (2014). Dispersal Pathways and Genetic Differentiation among Worldwide Populations of the Invasive Weed *Centaurea solstitialis* L. (Asteraceae). *PLoS One*. 9 (12), e114786. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1371/journal.pone.0114786

Staton, S.E. and J.M. Burke (2015). Transposome: a toolkit for annotation of transposable element families from unassembled sequence reads. *Bioinformatics*. . Status = AWAITING_PUBLICATION; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: doi: 10.1093/bioinformatics/btv059

Whitney, K.D, K.W. Broman, N. Kane, S.M. Hovick, R.A. Randell, L.H. Rieseberg (2015). QTL mapping identifies candidate alleles for adaptive introgression between wild sunflower species. *Molecular Ecology*. 24 2194. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Yakimowski, S.B., and L.H. Rieseberg (2014). The role of homoploid hybridization in evolution: A century of studies synthesizing genetics and ecology. *American Journal of Botany*. 101 1247. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Licenses

Other Products

Other Publications

Wei, Z., Julkowska, M.M., Laloë, J.-O., Hartman, Y., de Boer, G.-J., Michelmore, R.W., van Tienderen, P., Testerink, C.S., Schranz, M.E. (2014). *A mixed model QTL analysis for salt tolerance in crop-wild hybrids of lettuce seedlings.* journal article. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Patents

Technologies or Techniques

Thesis/Dissertations

Fei-Yian Yoong. *Natural variation and genetic analyses of high temperature germination in lettuce: from QTL mapping to candidate gene identification*. (2015). University of California, Davis. Acknowledgement of Federal Support = Yes

Websites

Participants/Organizations

What individuals have worked on the project?

Name	Most Senior Project Role	Nearest Person Month Worked
Rieseberg, Loren	PD/PI	1
Bradford, Kent	Co PD/PI	1
Kesseli, Richard	Co PD/PI	2
Michelmore, Richard	Co PD/PI	1
Burke, John	Co-Investigator	1
Still, David	Co-Investigator	1
Bewick, Adam	Postdoctoral (scholar, fellow or other postdoctoral position)	12
Nambeesan, Savithri	Postdoctoral (scholar, fellow or other postdoctoral position)	8
Staton, Evan	Postdoctoral (scholar, fellow or other postdoctoral position)	12
You, Youngsook	Postdoctoral (scholar, fellow or other postdoctoral position)	2
Yu, Qian	Postdoctoral (scholar, fellow or other postdoctoral position)	12
Bayona, Lissette	Technician	4
Fletcher, Tyler	Technician	12
Xu, Huaqin	Technician	6
Bowers, John	Staff Scientist (doctoral level)	10

Chang, Jian	Staff Scientist (doctoral level)	1
Dahal, Peetambar	Staff Scientist (doctoral level)	1
Kozik, Alex	Staff Scientist (doctoral level)	8
Lavelle, Dean	Staff Scientist (doctoral level)	12
Truco, Maria-Jose	Staff Scientist (doctoral level)	6
Wroblewski, Tadeusz	Staff Scientist (doctoral level)	0
Gulick, Trudi	Graduate Student (research assistant)	6
Ishibashi, Caitlin	Graduate Student (research assistant)	12
Masalia, Rishi	Graduate Student (research assistant)	12
McAssey, Ed	Graduate Student (research assistant)	12
Ortiz, Emilio	Graduate Student (research assistant)	9
Tsirelson, Dina	Graduate Student (research assistant)	3
Yoong, Fei-Yian	Graduate Student (research assistant)	12
Zavada, Tomas	Graduate Student (research assistant)	4
Abundis, Celina	Undergraduate Student	1
Bah, Fatoumata	Undergraduate Student	1
Botta, Caleb	Undergraduate Student	3
Bradley, Jeremiah (Seth)	Undergraduate Student	3
Brindley, Kody	Undergraduate Student	3
Fatherley, William	Undergraduate Student	3
Foucher, Simon	Undergraduate Student	5
Jardini, Teresa	Undergraduate Student	0
Lopez, Deborah	Undergraduate Student	3

Malik, Rondy	Undergraduate Student	1
Martinez, Alvaro	Undergraduate Student	2
Nguyen, Bao	Undergraduate Student	1
Perez, Fabian	Undergraduate Student	3
Stark, Blake	Undergraduate Student	3
Vargas, Adolfo	Undergraduate Student	2
Winnett, Alexander	Undergraduate Student	1
Ying, Florence	Undergraduate Student	1
Bruch, Ryan	High School Student	1

Full details of individuals who have worked on the project:

Loren H Rieseberg

Email: lriesebe@indiana.edu

Most Senior Project Role: PD/PI

Nearest Person Month Worked: 1

Contribution to the Project: Directed Overall Project

Funding Support: This project

International Collaboration: Yes, Canada

International Travel: Yes, Canada - 0 years, 9 months, 1 days

Kent J Bradford

Email: kjbradford@ucdavis.edu

Most Senior Project Role: Co PD/PI

Nearest Person Month Worked: 1

Contribution to the Project: Supervised research and dissertation preparation of PhD student who completed degree in this reporting period. Papers from this work are in preparation.

Funding Support: This person was not personally supported by this project. His participation was provided supported by the University of California, Davis.

International Collaboration: No

International Travel: No

Richard V Kesseli

Email: rick.kesseli@umb.edu

Most Senior Project Role: Co PD/PI

Nearest Person Month Worked: 2

Contribution to the Project: Worked on weed genomics and evolution. Supported outreach activities and developed teaching modules for genetics courses using materials developed from this project.

Funding Support: My support was from University of Massachusetts, Boston.

International Collaboration: No

International Travel: No

Richard W Michelmore

Email: rwmichelmore@ucdavis.edu

Most Senior Project Role: Co PD/PI

Nearest Person Month Worked: 1

Contribution to the Project: coPI

Funding Support: CA State funded salary

International Collaboration: Yes, Netherlands, Singapore, United Kingdom

International Travel: No

John Burke

Email: jmburke@uga.edu

Most Senior Project Role: Co-Investigator

Nearest Person Month Worked: 1

Contribution to the Project: Directed research on safflower and sunflower genetics and genomics

Funding Support: This project

International Collaboration: Yes, Canada

International Travel: No

David Still

Email: dwstill@csupomona.edu

Most Senior Project Role: Co-Investigator

Nearest Person Month Worked: 1

Contribution to the Project: Contributed to studies of lettuce physiology. Directed outreach to Cal Poly Pomona

Funding Support: This award and CPP.

International Collaboration: No

International Travel: No

Adam Bewick

Email: bewickaj@uga.edu

Most Senior Project Role: Postdoctoral (scholar, fellow or other postdoctoral position)

Nearest Person Month Worked: 12

Contribution to the Project: Molecular evolutionary analyses of candidate genes; bioinformatic analyses of gene copy number variation in the sunflower genome.

Funding Support: NSF - this project

International Collaboration: No

International Travel: No

Savithri Nambeesan

Email: sunamb@uga.edu

Most Senior Project Role: Postdoctoral (scholar, fellow or other postdoctoral position)

Nearest Person Month Worked: 8

Contribution to the Project: Sunflower gene space sequencing; SFP mapping in sunflower using the Affymetrix array; characterization of the Cycloidea gene family and functional analysis of the roles of these genes in flower development; candidate gene and physiological analyses related to plant architecture and other domestication traits in sunflower. Has also led the genotypic/phenotypic characterization of sunflower testcross RILs and subsequent data analyses.

Funding Support: NSF - this award

International Collaboration: No

International Travel: No

Evan Staton

Email: statonse@uga.edu

Most Senior Project Role: Postdoctoral (scholar, fellow or other postdoctoral position)

Nearest Person Month Worked: 12

Contribution to the Project: Bioinformatic analyses of sequence composition of sunflower, lettuce, and Taraxacum shotgun and BAC sequences. He is also involved in bioinformatic analyses of EST and gene space sequences.

Funding Support: NSF - this award University of Georgia

International Collaboration: No

International Travel: No

Youngsook You

Email: yyou@csupomona.edu

Most Senior Project Role: Postdoctoral (scholar, fellow or other postdoctoral position)

Nearest Person Month Worked: 2

Contribution to the Project: Assessed genetic and epigenetic mechanisms responsible for trait stability/phenotypic stability in response to environmental stress in a lettuce recombinant inbred line

Funding Support: NSF - this award

International Collaboration: No

International Travel: No

Qian Yu

Email: qianyu@umail.iu.edu

Most Senior Project Role: Postdoctoral (scholar, fellow or other postdoctoral position)

Nearest Person Month Worked: 12

Contribution to the Project: Analyses of heterosis in sunflower

Funding Support: This project

International Collaboration: No

International Travel: No

Lissette Bayona

Email: lissmar08@gmail.com

Most Senior Project Role: Technician

Nearest Person Month Worked: 4

Contribution to the Project: Helped with field projects and developing resources for teaching modules.

Funding Support: NSF - this grant University of Massachusetts

International Collaboration: No

International Travel: No

Tyler Fletcher

Email: fletcht@uga.edu

Most Senior Project Role: Technician

Nearest Person Month Worked: 12

Contribution to the Project: Genotypic and phenotypic characterization of sunflower mapping populations; functional characterization of members of the Cycloidea gene family.

Funding Support: NSF - this award

International Collaboration: No

International Travel: No

Huaqin Xu

Email: huaxu@ucdavis.edu

Most Senior Project Role: Technician

Nearest Person Month Worked: 6

Contribution to the Project: Developed and managed databases

Funding Support: This project.

International Collaboration: No

International Travel: No

John Bowers

Email: jebowers@uga.edu

Most Senior Project Role: Staff Scientist (doctoral level)

Nearest Person Month Worked: 10

Contribution to the Project: Conducted bioinformatics work related to SNP and SFP mapping, sequence analysis of EST datasets, and genetic map construction.

Funding Support: NSF - this award

International Collaboration: No
International Travel: No

Jian Chang

Email: chang.jan@gmail.com

Most Senior Project Role: Staff Scientist (doctoral level)

Nearest Person Month Worked: 1

Contribution to the Project: Characterized intron evolution across the Compositae (Kesseli lab).

Funding Support: University of Massachusetts International Funding

International Collaboration: Yes, Taiwan

International Travel: No

Peetambar Dahal

Email: pdahal@ucdavis.edu

Most Senior Project Role: Staff Scientist (doctoral level)

Nearest Person Month Worked: 1

Contribution to the Project: Dr. Peetambar Dahal developed recombinant introgression lines of the native LsNCED4 allele from UC96US23 into Salinas background. (Bradford lab).

Funding Support: USAID-Horticulture Innovation Laboratory Agricultural Research Institute, Cal Poly University Pomona

International Collaboration: No

International Travel: No

Alex Kozik

Email: akozik@gmain.com

Most Senior Project Role: Staff Scientist (doctoral level)

Nearest Person Month Worked: 8

Contribution to the Project: Alex Kozik was the lead bioinformatician for the Compositae project (Michelmore lab).

Funding Support: CA state funding

International Collaboration: No

International Travel: No

Dean Lavelle

Email: addressdolavelle@ucdavis.edu

Most Senior Project Role: Staff Scientist (doctoral level)

Nearest Person Month Worked: 12

Contribution to the Project: Characterized candidate genes (Michelmore lab).

Funding Support: This project.

International Collaboration: No

International Travel: No

Maria-Jose Truco**Email:** mjtruco@ucdavis.edu**Most Senior Project Role:** Staff Scientist (doctoral level)**Nearest Person Month Worked:** 6

Contribution to the Project: Contributed to the integration of molecular maps developed on multiple lettuce populations segregating for a variety of traits including resistance genes and morphological traits. She also is contributing to QTL mapping of domestication traits of agricultural importance relative to candidate genes (Michelmore lab).

Funding Support: CA State funding**International Collaboration:** No**International Travel:** No**Tadeusz Wroblewski****Email:** wroblewski@ucdavis.edu**Most Senior Project Role:** Staff Scientist (doctoral level)**Nearest Person Month Worked:** 0

Contribution to the Project: Analyzed candidate genes in lettuce (Michelmore lab).

Funding Support: This project**International Collaboration:** No**International Travel:** No**Trudi Gulick****Email:** tagulick@gmail.com**Most Senior Project Role:** Graduate Student (research assistant)**Nearest Person Month Worked:** 6

Contribution to the Project: Characterized epigenomic variation and transgenerational inheritance in weedy and domesticated lettuce (Kesseli lab).

Funding Support: NSF - this grant University of Massachusetts**International Collaboration:** No**International Travel:** No**Caitlin Ishibashi****Email:** bashi@uga.edu**Most Senior Project Role:** Graduate Student (research assistant)**Nearest Person Month Worked:** 12

Contribution to the Project: Characterization of traits related to abiotic stress tolerance in wild and cultivated and sunflower.

Funding Support: NSF GRFP University of Georgia**International Collaboration:** No**International Travel:** No

Rishi Masalia**Email:** masalia@uga.edu**Most Senior Project Role:** Graduate Student (research assistant)**Nearest Person Month Worked:** 12**Contribution to the Project:** Involved in evolutionary analyses of sequence diversity across the Compositae (Burke lab).**Funding Support:** NSF and teaching assistantship**International Collaboration:** No**International Travel:** No**Ed McAssey****Email:** emcassey@plantbio.uga.edu**Most Senior Project Role:** Graduate Student (research assistant)**Nearest Person Month Worked:** 12**Contribution to the Project:** Sequencing and evolutionary analyses of candidate genes for domestication traits in sunflower.**Funding Support:** NSF - this award University of Georgia**International Collaboration:** No**International Travel:** No**Emilio Ortiz****Email:** eaalvarado@cpp.edu**Most Senior Project Role:** Graduate Student (research assistant)**Nearest Person Month Worked:** 9**Contribution to the Project:** Worked on assembly and comparative genomics of Gerbera species (Still lab).**Funding Support:** This grant**International Collaboration:** No**International Travel:** No**Dina Tsirelson****Email:** dinushka@gmail.com**Most Senior Project Role:** Graduate Student (research assistant)**Nearest Person Month Worked:** 3**Contribution to the Project:** QTL analysis of domestication and weedy traits in lettuce**Funding Support:** NSF - this grant University of Massachusetts**International Collaboration:** No**International Travel:** No**Fei-Yian Yoong****Email:** fyoong@ucdavis.edu**Most Senior Project Role:** Graduate Student (research assistant)

Nearest Person Month Worked: 12

Contribution to the Project: Mapped genes associated with seed dormancy and quality traits in RILs derived from *L. sativa* cv. Salinas x accession PI251246, a primitive landrace of *L. sativa*. She also developed and analyzed a large RNASeq dataset to compare gene expression patterns during imbibition/germination at different temperatures. (Bradford lab). Completed PhD dissertation and degree in March 2015.

Funding Support: NSF - this award University of California, Davis

International Collaboration: No

International Travel: No

Tomas Zavada

Email: Tomas.Zavada@umb.edu

Most Senior Project Role: Graduate Student (research assistant)

Nearest Person Month Worked: 4

Contribution to the Project: Characterized genetic diversity in the wild and domesticated accessions of chicory and endive (Kesseli lab).

Funding Support: NSF - this award University of Massachusetts

International Collaboration: No

International Travel: No

Celina Abundis

Email: celina.abundis@gmail.com

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 1

Contribution to the Project: Contributed to the genetic mapping of QTL in lettuce (Kesseli lab).

Funding Support: NSF - this grant NIH - IMSD program University of Massachusetts

International Collaboration: No

International Travel: No

Fatoumata Bah

Email: bfatoumata42@gmail.com

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 1

Contribution to the Project: Working with graduate students Trudi Gulick and Dina Tsirelson on quantitative genetics in lettuce

Funding Support: Ronald McNair Fellow and this NSF grant

International Collaboration: No

International Travel: No

Caleb Botta

Email: cbotta@uga.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 3

Contribution to the Project: Assisted with laboratory and greenhouse work.

Funding Support: NSF - this award

International Collaboration: No

International Travel: No

Jeremiah (Seth) Bradley

Email: sbradley@uga.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 3

Contribution to the Project: Assisted with laboratory and greenhouse work.

Funding Support: NSF - this award

International Collaboration: No

International Travel: No

Kody Brindley

Email: brindley@uga.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 3

Contribution to the Project: Assisted with laboratory and greenhouse work.

Funding Support: NSF - this award

International Collaboration: No

International Travel: No

William Fatherley

Email: William.Fatherley001@umb.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 3

Contribution to the Project: Worked with graduate student Trudi Gulick on epigenetic and stress in lettuce

Funding Support: This NSF grant plus funding from University of Massachusetts, Boston

International Collaboration: No

International Travel: No

Simon Foucher

Email: sofoucher@ucdavis.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 5

Contribution to the Project: Assisted graduate student in utilizing different methods and software to analyze RNA-Seq

data.

Funding Support: Unpaid intern.

International Collaboration: No

International Travel: No

Teresa Jardini

Email: tjardnini@ucdavis.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 0

Contribution to the Project: Assisted in candidate gene analysis in lettuce

Funding Support: This project

International Collaboration: No

International Travel: No

Deborah Lopez

Email: lopez@cpp.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 3

Contribution to the Project: Evaluated germination of lettuce seeds in response to salinity, evaluated root growth under reduced water potentials (Still lab).

Funding Support: This grant.

International Collaboration: No

International Travel: No

Rondy Malik

Email: Rondy.Malik001@umb.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 1

Contribution to the Project: Studied genetic diversity in weeds (Chicory) under the guidance of graduate student Zavada (Kesseli lab)

Funding Support: NSF - this grant University of Massachusetts NIH IMSD program

International Collaboration: No

International Travel: No

Alvaro Martinez

Email: amartinez@ucdavis.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 2

Contribution to the Project: Helped with DNA extractions

Funding Support: this project

International Collaboration: No

International Travel: No

Bao Nguyen

Email: btknguyen@ucdavis.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 1

Contribution to the Project: She assisted graduate student in hormonal-related experiments including germination assays, ethylene measurement, seed counting, and hormone preparation.

Funding Support: Unpaid intern.

International Collaboration: No

International Travel: No

Fabian Perez

Email: fabianperez@cpp.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 3

Contribution to the Project: Worked on phenotyping and genotyping to assess effects of introgression of NCED4 alleles from *L. serriola* into *L. sativa* germplasm (Still lab).

Funding Support: This grant.

International Collaboration: No

International Travel: No

Blake Stark

Email: bwstark@cpp.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 3

Contribution to the Project: Evaluated germination of lettuce in response to far-red light and salinity (Still lab)

Funding Support: This grant

International Collaboration: No

International Travel: No

Adolfo Vargas

Email: avargas@ucdavis.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 2

Contribution to the Project: Assisted in candidate gene analysis in lettuce

Funding Support: This project

International Collaboration: No

International Travel: No

Alexander Winnett

Email: Alexander.Winnett001@umb.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 1

Contribution to the Project: Working with graduate student Trudi Gulick in the Kesseli lab on bioinformatics and epigenetics in the lettuce genome.

Funding Support: This NSF grant and University of Massachusetts, Boston internal grants.

International Collaboration: No

International Travel: No

Florence Lip Ka Ying

Email: klip@ucdavis.edu

Most Senior Project Role: Undergraduate Student

Nearest Person Month Worked: 1

Contribution to the Project: Involved in fine-mapping the Htg9.1 QTL and extracting RNA for RNA-Seq project.

Funding Support: Unpaid

International Collaboration: No

International Travel: No

Ryan Bruch

Email: rbm32@comcast.net

Most Senior Project Role: High School Student

Nearest Person Month Worked: 1

Contribution to the Project: helped analyze a candidate gene

Funding Support: none

International Collaboration: No

International Travel: No

What other organizations have been involved as partners?

Name	Type of Partner Organization	Location
Arcadia Biosciences, Inc.	Industrial or Commercial Firms	California
BGI	Industrial or Commercial Firms	China
CIDA	Other Organizations (foreign or domestic)	Canada
CSU-Agriculture Research Institute	Academic Institution	California

Genome Canada	Other Organizations (foreign or domestic)	Canada
Global Crop Diversity Trust	Other Nonprofits	Bonn
USDA-ARI	State or Local Government	USA
USDA-NIFA	State or Local Government	USA
Western Regional Seed Physiology Research Group	Other Nonprofits	California

Full details of organizations that have been involved as partners:

Arcadia Biosciences, Inc.

Organization Type: Industrial or Commercial Firms

Organization Location: California

Partner's Contribution to the Project:

Collaborative Research

More Detail on Partner and Contribution: Arcadia Biosciences, Inc. screened a TILLING population in lettuce to identify putative mutants in LsNCED4, which we further characterized phenotypically.

BGI

Organization Type: Industrial or Commercial Firms

Organization Location: China

Partner's Contribution to the Project:

Collaborative Research

More Detail on Partner and Contribution: The whole lettuce genome was sequenced by the BGI funded by a consortium of 10 companies independent of the CGP. We coordinated the efforts and received the assemblies, raw reads, and annotation from their efforts. In turn, we are contributing our reads and validating their assemblies using genetic information.

CIDA

Organization Type: Other Organizations (foreign or domestic)

Organization Location: Canada

Partner's Contribution to the Project:

Financial support

Collaborative Research

Personnel Exchanges

More Detail on Partner and Contribution: We have collaborated with the Canadian International Development Agency on the conservation, characterization, and improvement of noug, a Compositae oilseed crop indigenous to Ethiopia. The crop is grown by greater than a million farmers in Ethiopia, and CIDA has funded the collection and genetic and phenotypic characterization of noug germplasm, whereas the current project has developed genomic tools and resources for most promising noug germplasm that was discovered and trained Ethiopian scientists in genomics and bioinformatics.

Together, the NSF and CIDA sponsored work has jump-started a modern breeding program for noug improvement, highlighted by the development of self-compatible cultivars and genomic selection for yield.

CSU-Agriculture Research Institute

Organization Type: Academic Institution

Organization Location: California

Partner's Contribution to the Project:

Collaborative Research

More Detail on Partner and Contribution:

Genome Canada

Organization Type: Other Organizations (foreign or domestic)

Organization Location: Canada

Partner's Contribution to the Project:

Financial support

Collaborative Research

More Detail on Partner and Contribution: We are partnering with Genome Canada to sequence the sunflower genome. Genome Canada has provided approximately \$7.5 million for sequencing and bioinformatics, but the sequencing project will build on genetic mapping and EST resources created by the Compositae Genome Project to assemble and annotate the genome.

Global Crop Diversity Trust

Organization Type: Other Nonprofits

Organization Location: Bonn

Partner's Contribution to the Project:

Financial support

Collaborative Research

More Detail on Partner and Contribution: The Global Crop Diversity Trust is leading a 10 year project to collect, conserve, and evaluate crop wild relatives of 21 major crops, as well as conduct pre-breeding for a subset of these, targeting traits of interest. Due to the germplasm and genomic resources developed by the present project, sunflower has been chosen as the focus of pilot project to determine the optimal methodology and workflow for pre-breeding and evaluation. Thus, we are working closely with the Trust and will manage the genotyping and pre-breeding of sunflower for the project (funded by the Trust), whereas evaluations will be conducted in Africa.

USDA-ARI

Organization Type: State or Local Government

Organization Location: USA

Partner's Contribution to the Project:

Financial support

More Detail on Partner and Contribution: We have received USDA/ARI funding to screen Composite species for their

medicinal potential.

USDA-NIFA

Organization Type: State or Local Government

Organization Location: USA

Partner's Contribution to the Project:

Financial support

More Detail on Partner and Contribution: Characterization of the LsNCED4 gene has also been supported by a USDA-NIFA project (2008-02509) to K.J. Bradford, which ended in August 2011.

Western Regional Seed Physiology Research Group

Organization Type: Other Nonprofits

Organization Location: California

Partner's Contribution to the Project:

Financial support

More Detail on Partner and Contribution: The Western Regional Seed Physiology Research Group, a voluntary organization of 13 seed companies, supports research in the Bradford lab. This group is supporting additional seed physiology studies using the mapping populations developed by the CGP. This project will attempt to map loci that determine the sensitivity of seed dormancy to the maternal environment during seed development.

What other collaborators or contacts have been involved?

Nothing to report

Impacts

What is the impact on the development of the principal discipline(s) of the project?

This project is the primary contributor to the development genomic sequence databases representing the Compositae that enables a wide range of studies in evolution, genomics, physiology, pathology, and crop improvement. Genetic populations and genotyping resources have been and are being developed that enable mapping of numerous domestication-related traits and identification of the underlying genes and genetic variation. Specific genes associated with flowering time, flower morphology, seed dormancy and seed oil accumulation have been identified and their regulation and evolution are being further characterized. In addition to resources for the Compositae, we have made technological advances, especially in the development of normalization strategies for transcriptome and gene space sequencing, that are now widely employed.

Likewise, we have developed popular bioinformatic pipelines for the processing next generation sequencing data, as well as for assembly and subsequent evolutionary analyses.

What is the impact on other disciplines?

This project has made important contributions to the discipline of evolutionary biology by developing a robust method for detecting both recent and ancient hybridization from analyses of a single genome and for distinguishing between auto- and allopolyploidization events. The project has also contributed to the discipline of taxonomy by providing useful molecular markers for phylogenetic studies of the Compositae, the largest and most successful family of flowering plants.

The Still laboratory partnered with a computer science professor at Cal Poly Pomona (Dr. Amar Raheja) to develop

algorithms to differentiate between germinated and not-germinated seeds in digital images. This laboratory also partnered with an electrical engineering professor at Cal Poly Pomona to develop and manufacture programmable controllers that control the light cycles, intensity and ratios of LED panels consisting of red, far-red and blue LEDs. This project was the culmination of two engineering students' senior projects. In addition, DW Still began a new project with a faculty member (Tom Thoen) of the Engineering Technology Department. The objective of this project is to develop a pick-and-place robot capable of placing individual seeds within Petri dishes, each placed equidistant to facilitate automated image acquisition. This project will serve as a senior project for a team of three Cal Poly Pomona seniors. If these two projects are successful it will greatly facilitate phenotyping of large populations.

What is the impact on the development of human resources?

As documented in the Project Participants section of this report, this project has contributed to the training of numerous postdocs, graduate and undergraduate students, and technicians. All of these participants are learning skills that will enable them to be productive scientists in the range of disciplines encompassed by the project. The Project's emphasis on targeting minority-serving institutions is recruiting more of these under-represented students into graduate schools and scientific careers.

What is the impact on physical resources that form infrastructure?

Nothing to report.

What is the impact on institutional resources that form infrastructure?

Nothing to report.

What is the impact on information resources that form infrastructure?

The sequence databases developed by this project constitute the vast majority of genomic information available in the Compositae, making it available publicly for a wide array of applications in evolution, taxonomy, breeding, physiology and molecular biology. A number of projects are extending information from model systems, such as developmental regulatory mechanisms, to specific examples in Compositae species. Improved methods for sample preparation have enhanced the efficiency and expanded the capacity of state of the art sequencing technologies

What is the impact on technology transfer?

Germplasm developed through this project, such as RILs and NILs, are being distributed and utilized by other labs and by commercial seed companies to study and improve lettuce and sunflower. Similarly, molecular markers associated with domestication and product quality traits are being utilized in commercial breeding programs to enhance the efficiency and speed of variety improvement.

What is the impact on society beyond science and technology?

The information developed by our project, especially the whole genome sequences for five food crop lineages, is likely to be heavily used by commercial seed companies to facilitate breeding and improvement.

Changes/Problems

Changes in approach and reason for change

Nothing to report.

Actual or Anticipated problems or delays and actions or plans to resolve them

Nothing to report.

Changes that have a significant impact on expenditures

Nothing to report.

Significant changes in use or care of human subjects

Nothing to report.

Significant changes in use or care of vertebrate animals

Nothing to report.

Significant changes in use or care of biohazards

Nothing to report.