California Leafy Greens Research Program
Final Report
April 1, 2020 to March 31, 2021

Project Title: IMPROVING THE POST-HARVEST QUALITY AND SHELF LIFE OF LETTUCE

Principle Investigator: Gail Taylor, University of California, Davis, CA
gtaylor@ucdavis.edu
Cooperating Personnel: Annabelle Damerum, University of California, Davis, CA **adamerum@ucdavis.edu**; Professor Richard Michelmore, University of California, Davis, CA; Dr Ivan Simko, USDA, Salinas
Technical Staff: Suzanne Milner, University of California, Davis, CA **smilne@ucdavis.edu**

ABSTRACT

Post-harvest quality, including shelf life and nutritional quality, are complex, quantitative traits controlled by multiple genes, making it difficult to determine the underlying genetic regulation and markers for plant breeding. Such traits impact on the profitability and success of the crop, but are currently overlooked in breeding programs. By characterizing the variation in post-harvest quality traits, including shelf life and nutritional quality, we use tools such as quantitative trait loci (QTL) mapping and genome wide association studies (GWAS) to link phenotype to genotype, to elucidate the genetic architecture of these traits which will inform marker-assisted breeding programs. During the third year of this project, we have completed two field trials in Salinas (Fall '20) and El Centro (Winter '21) utilizing the GWAS population in collaboration with Ivan Simko (USDA), to characterize this population for whole head lettuce post-harvest traits, adding to the data we have previously collected for baby leaf lettuce on this same population. We have refined key regions of the lettuce genome associated with shelf life in multiple different environments and uncovered novel candidate genes underpinning shelf life, moving closer to developing molecular markers which can be utilized in future breeding programs.

OBJECTIVES

1. Perform genome-wide association studies to identify significant marker-trait associations

Deliverable: Robust QTL (genomic regions) for post-harvest quality traits.
Progress: completed

2. Refine our major QTL for shelf life, evaluating extensive phenotype datasets using meta-QTL analysis and other functional approaches

Deliverable: *Refined candidate gene targets in particular for QTL on chromosome 6, determining shelf life in baby lettuce.*
Progress: completed

3. Develop markers linked to QTL of interest for future breeding lines for enhanced shelf life.
   Deliverable: fine mapped QTL for shelf life in whole baby lettuce

Progress: Currently being initiated and to be completed in current project

PROCEDURES

Objective 1

We continued to utilize a diverse panel of 493 lettuce accessions developed by the lab of Ivan Simko[1] (USDA) to conduct genome-wide association studies (GWAS), performing two field trials at the USDA research site in Salinas (August-November, 2020) and the UC Desert Research Center in El Centro (February-April, 2021):

1) For the Salinas trial, seedlings were sown in a greenhouse at UC Davis on August 9th-10th and were watered by mist irrigation three times daily. On September 1st, seedlings were transported to Salinas and transplanted into the field and were harvested 9 weeks later on the 2-3rd of November.
2) For the El Centro trial, seedlings were sown from February 1st-3rd in a greenhouse at the UC Desert Research Center and were misted for 1 minute every hour. On February 16th, seedlings were transplanted into the field and were harvested 7 weeks later from the 5th-7th April.

For both trials, lettuce lines were planted in a fully randomized, three block trial design. Each 40-inch bed comprised two planting rows, with each row spaced 12 inches apart. The two outer beds acted as guard rows, along with four plants at the beginning and end of each row. For the Salinas trial, two plants constituted one replicate (planted consecutively in a row), with six beds of experimental plants each 274 feet long and two beds of guard plants (Figure 1A). For the El Centro trial, three plants constituted one replicate (planted consecutively in a row), with nine beds of experimental plants 322 feet long with two guard rows.
In addition to this population, we incorporated 23 current cultivars utilized in the industry, provided to us by the Calgreens board (Table 1). For both trials, we utilized the PaperPot Co. transplanting system (**https://paperpot.co/**). In this system, seedlings are sown into collapsible paper-chain pots into solid-bottom germination trays at 12 inch spacing. Seedlings in the biodegradable paper-chain pots can then be rapidly transplanted directly into the field using the mechanical transplanter.

For the Salinas trial, at head harvest maturity, three leaves of varying ages (outer, middle, inner) were flash-frozen in a sterile manner into liquid nitrogen and stored on dry ice. Samples were transported back to the Taylor laboratory at UC Davis on dry ice, where they were stored at -80ºC for nutritional quality and microbiome analyses. The rest of the head was harvested to measure phenotypic traits including shelf life, cell wall strength, elasticity and plasticity, leaf area, leaf surface traits (stomata density, epidermal cell number and area) and nutritional quality using a spectrophotometric pigment meter (see Appendix A). Shelf life was measured by scoring on days 1, 7, 14, 21 and 28 post-processing from 1 (no damage) to 3 (end of shelf life).

For the El Centro trial, at head harvest maturity, one whole head was frozen on dry ice and transported to the Taylor laboratory at UC Davis where they were stored at -20ºC until processing for microbiome analyses. Another whole head was harvested to measure phenotypic traits previously described (Appendix A). Shelf life was scored daily from day 7 post-harvest for 1 (no damage) to 3 (end of shelf life).

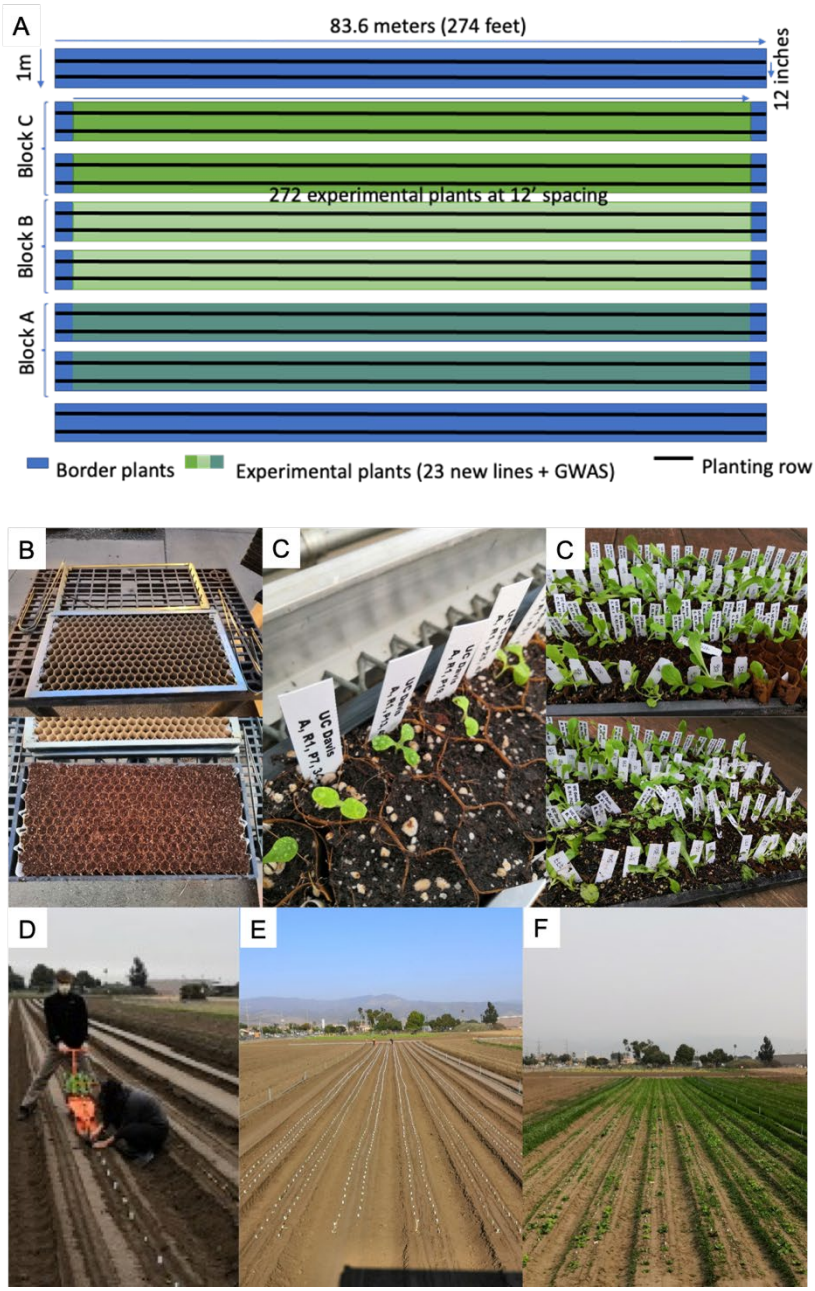


Figure 1 Example trial layout for the Salinas 2020 trial (A). Paper pot cells are spread and filled with soil (B) and a single seed sewn per cell at 12 inch spacing (C). Seedlings ready to transplant (C) using the paper pot transplanter (D). Trial 1-week (E) and 4-weeks (F) post-transplant.

For data processing, all available phenotype data was tested for normality using the Shapiro-Wilk test and data transformation was applied if necessary, using R statistics [2]. Genotype data for the diversity panel were collected using tunable Genotyping by Sequencing technology, as described in Kandel et al. [1], yielding 50,325 polymorphic sites. These sites were filtered to remove any with >10% missing data, >10% heterozygosity and a minor allele frequency of <0.05. The resultant 4,966 SNPs were utilized to determine the population structure using STRUCTURE[3] using an admixture model, with a burn-in period of 100,000 and 100,000 Markov chain Monte Carlo repetitions. Optimal sub-population structure (K) was determined by the Evanno method, implemented in STRUCTURE HARVESTER[4]. A general linear

model (GLM) was run using the software TASSEL (v5)[5], correcting for population structure by incorporation of sub-population coefficients (Q) calculated in STRUCTURE. Significant marker-trait associations (SMTAs) were identified after Bonferonni P value correction and Manhattan plots were drawn in TASSEL.

Objective 2

In order to investigate the genetic basis of shelf life in lettuce, we have previously completed several field trails with two lettuce populations and performed QTL analyses and GWAS (Trials summarized in Table 1).
The first population is the core recombinant inbred line mapping population, developed by the Michelmore laboratory [6]. This is a population of recombinant inbred lines (RILs) developed from a cross between the Salinas cultivar and the expected wild progenitor of lettuce, *Lactuca serriola*, hereby referred to as the SxS RIL population. In the second year of this project, we developed a high-density genetic linkage map and conducted QTL mapping analyses with data collected from all previous trials, to create an extensive dataset of QTL. Since 2019, we have utilized the diverse panel developed by the Simko lab (as discussed in Objective 1) for GWAS.

In this second objective, we analyzed both QTL and GWAS data collected for shelf life thus far to identify genomic regions identified in multiple studies, thus providing robust regions for further analyses. A region was considered of interest if QTL or a significant marker-trait association were identified within 10 Mb. From these selected regions, gene coding sequences (CDS) were retrieved from the *Lactuca sativa* genome sequence (V8) via COGE (https://genomevolution.org/) and used in a BLASTx (BLAST version 2.9.0) search against the *Helianthus annuus* genome sequence (HanXRQr1.0) with an E value cut-off of $< 1 \times 10^{-5}$ to assign gene annotation. Gene ontology (GO) terms were assigned to each gene and any significantly enriched terms were identified using ShinyGO v0.61 (ref), with an FDR-corrected P-value cut-off of 0.05.

Objective 3
We completed a literature review of all existing data on breeding for post-harvest quality in lettuce, identifying the key post-harvest targets for future breeding and highlighting the key resources available for lettuce breeding and research [7].

To further refine candidate gene regions, alongside the work completed in objective 2 we have also identified a selection of lettuce lines from the GWAS diversity panel with extreme high or low values for shelf life, consistent across trials. Taking together the shelf life data collected in four trials on the USDA GWAS population (Davis 2019, Salinas 2019, Salinas 2020 and El Centro 2021) we identified accessions consistently demonstrating high or low shelf life in these environments. For each trial, accessions were ranked for shelf life and any with a shelf life below 9 days from the day of harvest were included in the “low shelf life” group and those with a shelf life greater than 14 days were included in the “high shelf life” group. Groups were compared amongst each trial and accessions appearing in the high or low shelf life groups in 2 or more trials were retained for further analyses.

These lines will be utilized in an RNA sequencing experiment to identify any genes which are differentially expressed and could be candidate genes for the high or low shelf life phenotype.

Table 1 Summary of all trials conducted and phenotype data collected to date

Location	Year	Population	Harvest age	Traits measured
Alresford, UK	2003	*Salinas* x *L. serriola* RILs	Baby leaf	Shelf life, cell wall strength, leaf physiology, leaf morphology
Azenha do Mar, Portugal	2003	*Salinas* x *L. serriola* RILs	Baby leaf	Shelf life, cell wall strength, leaf physiology, leaf morphology
Greenhouse, UK	2004	*Salinas* x *L. serriola* RILs	Baby leaf	Cell wall strength, pigment analyses
Greenhouse, UK	2006	*Salinas* x *L. serriola* RILs	Baby leaf	Cell wall strength, nutritional quality, pigment analyses
Azenha do Mar, Portugal	2009	*Salinas* x *L. serriola* RILs	Baby leaf	Nutritional quality, pigment analyses
Greenhouse, UK	2012	*Salinas* x *L. serriola* RILs	Baby leaf	Leaf physiology
Davis, California	2018	*Salinas* x *L. serriola* RILs	Baby leaf	Shelf life, cell wall strength, leaf physiology, leaf morphology, pigment analyses, nutritional quality

Davis, California	2019	GWAS	Baby leaf	Shelf life, cell wall strength, leaf physiology, leaf morphology, pigment analyses, nutritional quality
Salinas, California	2019	GWAS	Baby leaf	Shelf life, cell wall strength, leaf physiology, leaf morphology, pigment analyses, nutritional quality
Salinas, California	2020	GWAS	Whole head	Microbiome analyses, shelf life, stomata and leaf morphology traits, pigment analyses, nutritional quality
El Centro, California	2021	GWAS	Whole head	Microbiome analyses, shelf life, stomata and leaf physiology traits, pigment analyses

* details of trait measurement can be found in the Appendix A.

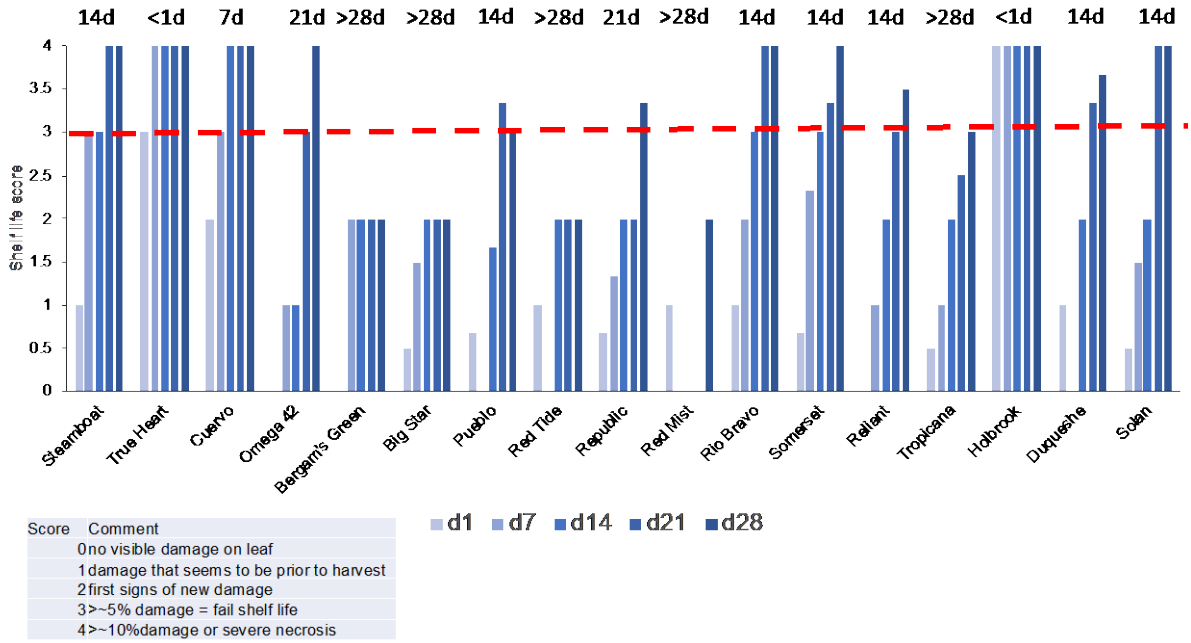
RESULTS AND DISCUSSION

Objective 1

During our Salinas trial we observed high disease pressure, largely resulting from INSV and often in conjunction with Pythium wilt (**Figure 2**), meaning we were able to collect shelf life data for 237/492 accessions of the GWAS population and 17/23 of the commercial lines. We have, however, collected samples for sequencing the total bacteria and fungi inhabiting the leaves, which will give us an interesting insight into the microbiome and may help us to understand the high disease pressure observed. Average shelf life was 12.3 days and ranged from 1->28 days.

Figure 2 High INSV and pythium disease observed in the field in Salinas, October-November 2020.

For the commercial varieties investigated, Bergam's Green, Big Star, Red Tide, Red Mist and Tropicana were determined to have the greatest shelf life of >28 days (did not reach end of shelf life by 28 days; Figure 3). True Heart, Holbrook and Cuervo had the poorest shelf life of 1-7 days post-harvest.



Score	Comment
0	no visible damage on leaf
1	damage that seems to be prior to harvest
2	first signs of new damage
3	>~5% damage = fail shelf life
4	>~10%damage or severe necrosis

Figure 3 Shelf life measured for 17 commercial lettuce varieties. Bars indicate shelf life score ranging from 1-4, measured on 1, 7, 14, 21 and 28 days post-harvest. A score >3 was determined as the threshold for end of shelf life, indicated by the red line. Shelf life score is given at the top of each bar.

Significant marker-trait associations (SMTAs) were identified for the nutritional quality traits anthocyanin, carotenoids and chlorophyll (Figure 4 and Table 3) on chromosomes 1, 2, 4, 5, 7, 8 and 9, measured using a handheld spectrometer (Appendix A). Two markers on chromosome 9 at 140.34 and 150.96 Mb were significantly associated with both anthocyanin and chlorophyll content and will be important for future investigation.

Unlike the Salinas 2020 trial, we observed low disease pressure in our trial in El Centro, 2021. Shelf life of the GWAS population ranged from **7-23.5 days**, averaging at 12 days. For the commercial varieties, shelf life ranged from 9.4-16.3 days, averaging at 12.4 days (Figure 5). Interestingly, varieties Holbrook and Cuervo, previously identified to rank poorly in shelf life had the 2nd and 4th highest shelf life, respectively, likely to reflect disease susceptibility.

GWAS identified three SMTAs, on chromosome 4 at position 204.65 Mb and 339.22 Mb and chromosome 5 at 271.08 Mb (Figure 6). Continuing work aim to identify SMTAs for nutritional quality traits, along with sequencing the bacterial and fungal microbiomes.

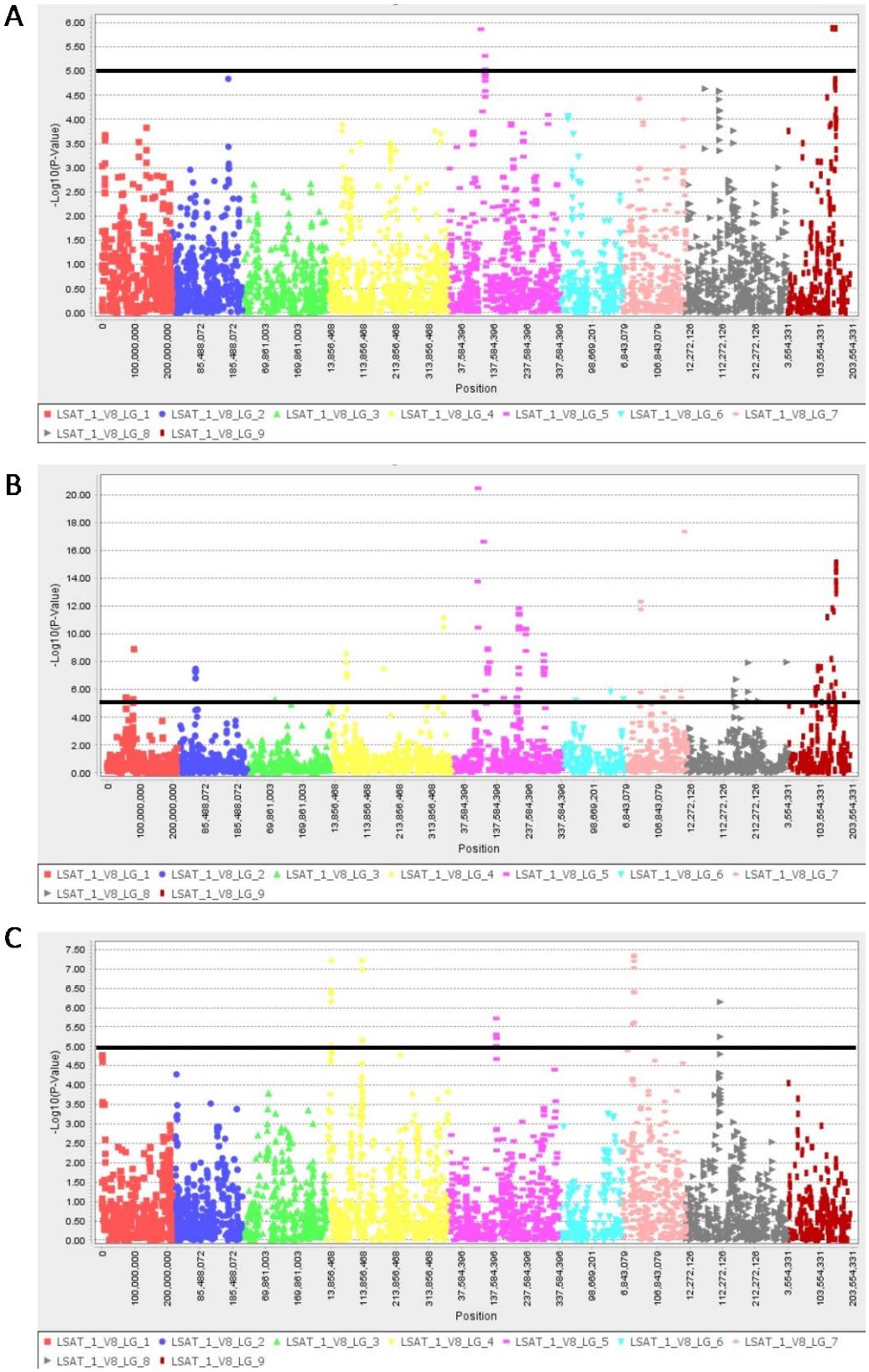


Figure 4 Manhattan plots for A) chlorophyll, B) anthocyanin and C) carotenoid reflectance indexes, measured in the Salinas 2020 trial. Each point represents a molecular marker (SNP), colored according to chromosome number. The solid black horizontal line represents the threshold for marker significance.

Table 3 Significant marker-trait associations identified for nutritional quality traits measured in the Salinas 2020 trial.

Trait	Chr	Position (V8, bp)	P value (corrected)	R^2
Anthocyanin reflectance index	1	81520681	0.001	0.1103
Anthocyanin reflectance index	2	56958622	0.005	0.09012
Carotenoid reflectance index	4	15456613	0.024	0.08594
Carotenoid reflectance index	4	15708533	0.01	0.0914
Anthocyanin reflectance index	4	49412612	0.002	0.10253
Anthocyanin reflectance index	4	52952965	0.011	0.09301
Carotenoid reflectance index	4	109726822	0.01	0.09931
Anthocyanin reflectance index	4	165525482	0.005	0.09734
Anthocyanin reflectance index	4	350681680	0.001	0.14286
Anthocyanin reflectance index	5	79676849	0.001	0.1687
Anthocyanin reflectance index	5	80705220	0.001	0.22707
Anthocyanin reflectance index	5	98499421	0.001	0.19185
Anthocyanin reflectance index	5	103000198	0.047	0.07319
Anthocyanin reflectance index	5	111090973	0.001	0.10954
Anthocyanin reflectance index	5	207165885	0.001	0.13947
Anthocyanin reflectance index	5	227977132	0.001	0.13225
Anthocyanin reflectance index	5	284384339	0.003	0.10102
Carotenoid reflectance index	7	36177690	0.008	0.09348
Anthocyanin reflectance index	7	52373894	0.001	0.15601
Anthocyanin reflectance index	7	129680620	0.047	0.07451
Anthocyanin reflectance index	7	176330032	0.047	0.07426
Anthocyanin reflectance index	7	187707340	0.001	0.19487
Carotenoid reflectance index	8	105639831	0.033	0.0872
Anthocyanin reflectance index	8	146636413	0.047	0.07484
Anthocyanin reflectance index	8	152027458	0.018	0.08862
Anthocyanin reflectance index	8	189447032	0.003	0.09919
Anthocyanin reflectance index	9	88818540	0.039	0.07497
Anthocyanin reflectance index	9	95311149	0.005	0.09433
Anthocyanin reflectance index	9	124429664	0.001	0.13163
Anthocyanin reflectance index	9	136934549	0.003	0.1017
Anthocyanin reflectance index	9	140341145	0.001	0.14708
Chlorophyll reflectance index	9	140341145	0.049	0.07976
Anthocyanin reflectance index	9	144734499	0.001	0.13976
Anthocyanin reflectance index	9	150955364	0.023	0.08535
Chlorophyll reflectance index	9	150955364	0.049	0.08168
Anthocyanin reflectance index	9	151782423	0.001	0.17696

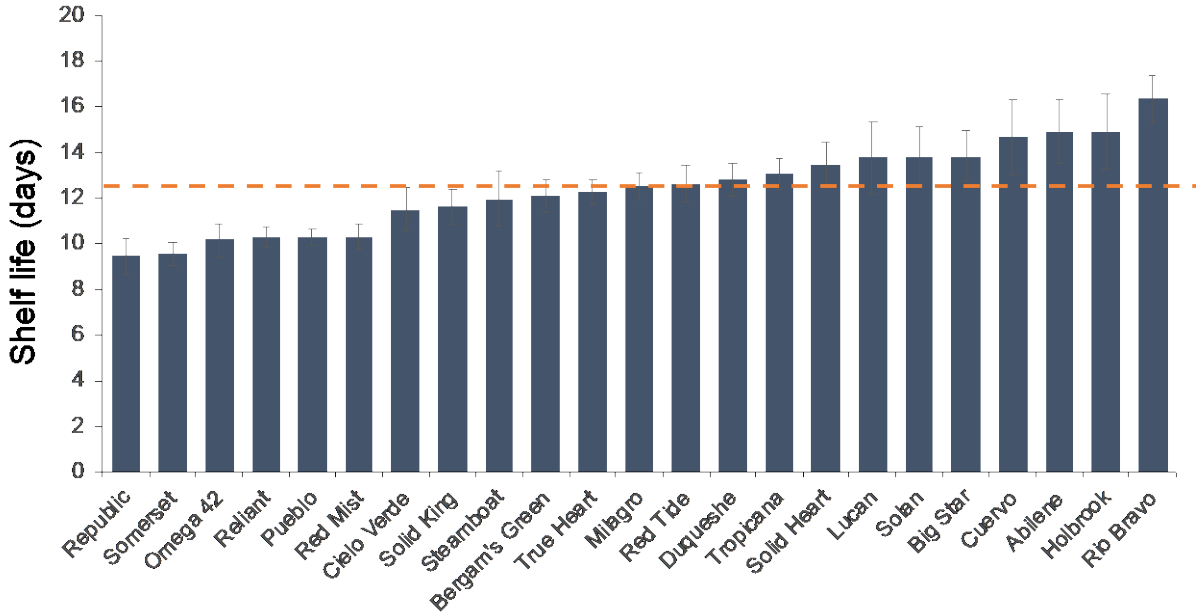
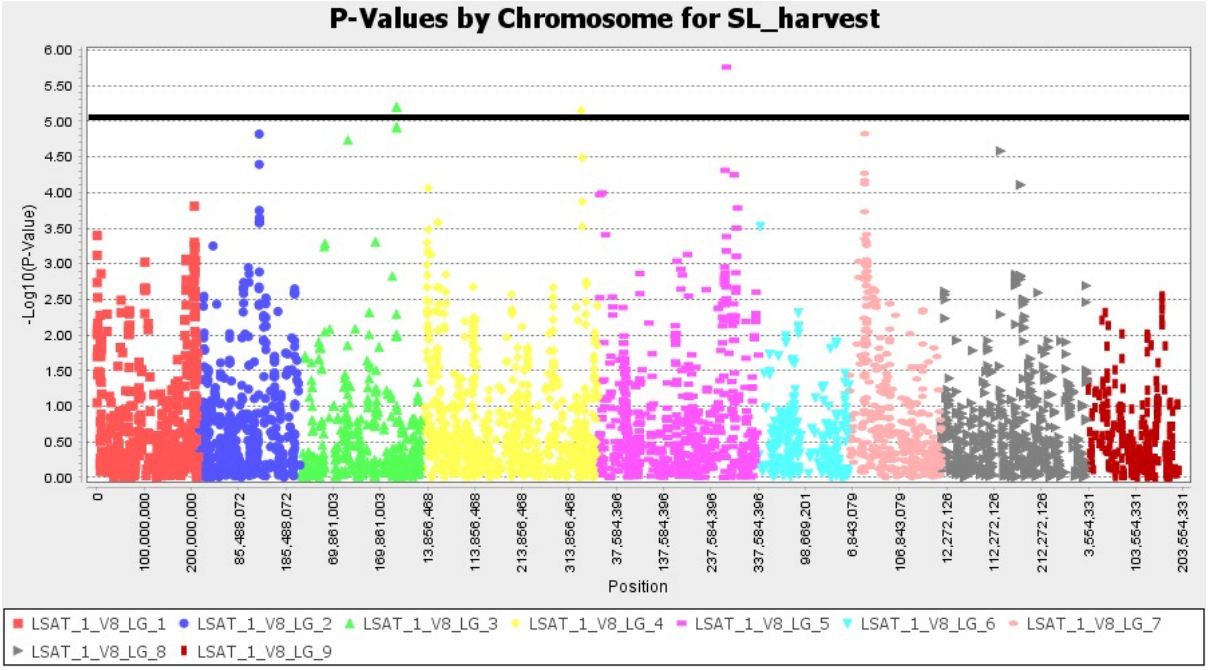


Figure 5 Shelf life of 23 commercial lettuce varieties assessed in the El Centro 2021 trial. Bars represent mean shelf life ± standard error, with dashed orange line demonstrating average shelf life of 12.4 days.



Trait	Chromosome	Position (V8, bp)	P value	R^2
Shelf life	3	204654410	6.38E-06	0.05225
Shelf life	4	339224387	7.07E-06	0.0483
Shelf life	5	271081931	1.74E-06	0.05929

Figure 6 Manhattan plot for shelf life measured in the El Centro 2021 trial. Each point represents a molecular marker (SNP), colored according to chromosome number. The solid black horizontal line represents the threshold for marker significance. Table underneath provides details of the SMTAs.

Objective 2

Taking together both our studies on the RIL core mapping population (Michelmore lab, UCD) and the lettuce diversity panel (Simko lab, USDA) **we identified six key regions in the lettuce genome harboring QTL or significant marker-trait associations** (SMTAs, Table 4). Three SMTAs were identified on chromosome 4 within a 1 Mb for total shelf life and number of days until leaf damage was observed for whole baby leaves from both the Salinas and Davis 2019 trials (candidate region ID 4.1). Another SMTA was identified for total shelf life measured in the Salinas 2019 trial on chromosome 4 at 280.7 Mb, close to the SMTAs identified by the Simko group for fresh-cut lettuce[1] (4.2). A region on chromosome 5 from 67.6-72.1 Mb contained SMTAs for shelf life of fresh-cut lettuce[1] and the number of days until leaf damage was observed for whole baby leaves in the Davis 2019 trial (5.1). Two other important regions were identified on chromosome 5; one containing a shelf life QTL detected using the RIL population and a SMTA, both measured from whole baby leaves (5.2) and another containing a newly detected SMTA for the shelf life of head lettuce from the El Centro 2021 trial within 10 Mb of a SMTA for whole baby leaf shelf life detected in the Davis 2019 trial (5.3). The final region from 140.1-151.1 Mb on chromosome 6 contained two QTL for whole baby leaf shelf life detected from the RIL population in two different trials and a SMTA for number of days until damage was observed in whole baby leaves from the Davis 2019 trial (6.1).

All protein coding sequences within candidate gene regions were extracted and genes were annotated (see Appendix B for complete list). Interesting candidate genes included a stress-induced protein found to negatively regulate programmed cell death (HVA22-like protein) within region 4.1, a wound response family protein within candidate region 4.2 and a defense response-related protein (jacalin-like lectin domain-containing protein) within region 5.1. Region 5.3 contained 4 xyloglucan endotransglucosylase/hydrolases which are involved in cutting and rejoining the hemicellulose chains in the plant cell wall, impacting on cell wall assembly and growth. We previously identified that silencing of these genes could increase the shelf life of commercial lettuce by 1.5-2 days[8] and more recently we have utilized CRISPR/Cas9 genome editing to knockout four XTH genes in lettuce, which are currently being bulked for field trialing. Other cell-wall influencing genes included three pectinesterase inhibitor domain-containing proteins within region 6.1 which influence cell wall loosening and strengthening and can impact on growth, fruit ripening and response to abiotic stress.

Gene ontology (GO) terms are functional annotations given to genes, divided into three categories: biological processes, molecular functions and cellular components. GO enrichment analyses identifies any GO terms in a gene list (i.e. a candidate gene region) which are significantly overrepresented compared to the number of times that GO term is identified in the whole genome. We identified a number of significantly enriched GO terms within the candidate regions investigated (Appendix C), including those for carbohydrate metabolism, transcription regulation, oxidation-reduction processes and proteolysis.

Table 4 Genomic regions explored in candidate gene mining

Population	Trial	Trait	Chr	Position (Mb)	Candidate region ID
USDA GWAS	Salinas Fall 2019	Shelf life	4	186.4	4.1
USDA GWAS	Davis Spring 2019	Shelf life	4	187.3	
USDA GWAS	Davis Spring 2019	Days to damaged	4	187.3	
USDA GWAS	Salinas Fall 2019	Shelf life	4	280.7	4.2
USDA GWAS	Kandel (2020) Shelf life (Spe16)	Shelf life	4	283.5	
USDA GWAS	Kandel (2020) Shelf life (Mar17)	Shelf life	4	283.5	
USDA GWAS	Kandel (2020) Shelf life (Gon19)	Shelf life	4	283.5	
USDA GWAS	Kandel (2020) Shelf life (Spe16)	Shelf life	5	67.6	5.1
USDA GWAS	Davis Spring 2019	Days to damaged	5	71.4-72.1	
SxS RILs	Portugal 2003	Shelf life	5	249.9	5.2
USDA GWAS	Davis Spring 2019	Shelf life	5	250.5-250.6	
USDA GWAS	El Centro 2021	Shelf life	5	271.1	5.3
USDA GWAS	Davis Spring 2019	Shelf life	5	280.9	
SxS RILs	UK 2003	Shelf life	6	140.1	6.1
SxS RILs	Davis 2018	Shelf life	6	145.4	
USDA GWAS	Davis Spring 2019	Days to damaged	6	151.1	

Objective 3

Taking together the shelf life data collected in four trials on the USDA GWAS population (Davis 2019, Salinas 2019, Salinas 2020 and El Centro 2021) we identified accessions consistently demonstrating high or low shelf life in these environments.

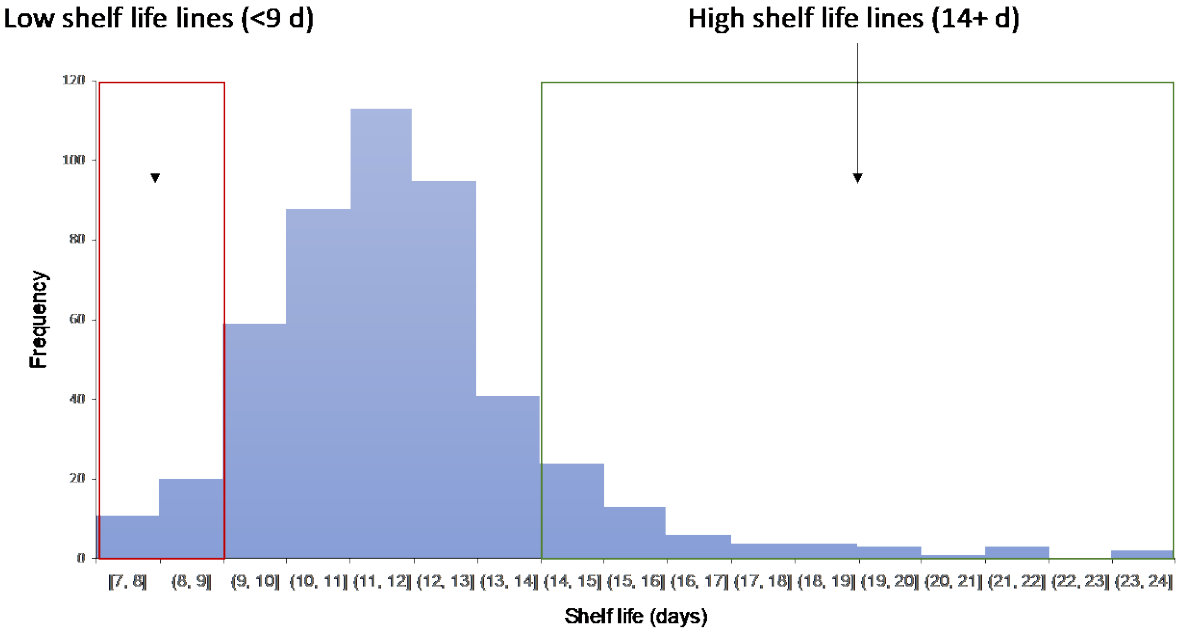


Figure 7 Selecting high and low shelf life accessions. Example frequency distribution of shelf life in days, with accessions with shelf below 9 days from date of harvest identified as "low shelf life" (red box) and above 14 days as "high shelf life" (green box).

We identified 4 "high shelf life" accessions which demonstrated a shelf life of >14 days in 3/4 of the trials investigated and 45 accessions which demonstrated a shelf life of >14 days in 2/4 trials (49 accessions out of a total of 493). For the "low shelf life" group, we identified one accession which demonstrated a shelf life of <9 days for all trials investigated, 6 accessions with a shelf life of <9 days in 3/4 trials and 27 accessions in 2/4 trials (34 accessions total). We will next work to perform RNA sequencing of these extreme lines in order to identify genes which are differentially expressed between these two groups and therefore are potential candidates for controlling this trait.

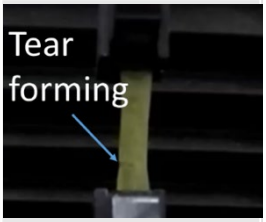
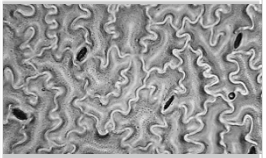
Conclusions

In this challenging third year of this project, with a pandemic and difficulties with travel and lab access, we were able to conduct two successful field trials in the Salinas and Imperial Valleys. We collected data to assess the post-harvest of ~3000 lettuce plants, including a diverse lettuce panel of 493 accessions and 23 commercial varieties, which we utilized to identify SMTAs for shelf life and nutritional quality traits. We made significant steps towards refining candidate genomic regions which are important for shelf life and identified underlying candidate genes. Further work in the next year of this project will involve refining the areas of the genome further with respect to shelf life and developing markers for key shelf traits for hand-over to industry and partners.

Literature

1. Sthapit Kandel, J., Peng, H., Hayes, R. J., Mou, B. & Simko, I. Genome-wide association mapping reveals loci for shelf life and developmental rate of lettuce. *Theor. Appl. Genet.* (2020) doi:10.1007/s00122-020-03568-2.
2. R Core Team. R: A Language and Environment for Statistical Computing. vol. 7 275–286 (2017).
3. Pritchard, J. K., Stephens, M., Rosenberg, N. a & Donnelly, P. Association mapping in structured populations. *Am. J. Hum. Genet.* **67**, 170–81 (2000).
4. Earl, D. A. & VonHoldt, B. M. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361 (2012).
5. Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
6. Truco, M. J. *et al.* An Ultra High-Density, Transcript-Based, Genetic Map of Lettuce. *G3 Genes | Genomes | Genet.* **3**, 617–631 (2013).
7. Damerum, A., Chapman, M. A. & Taylor, G. Innovative breeding technologies in lettuce for improved post-harvest quality. *Postharvest Biol. Technol.* **168**, 111266 (2020).
8. Wagstaff, C. *et al.* Modification of cell wall properties in lettuce improves shelf life. *J. Exp. Bot.* **61**, 1239–1248 (2010).

Appendix A
Details of traits analyzed.

Trait	Analyses	
Shelf life	Evaluation of the number of days after harvest until the first signs of visible damage. Damage found included yellowing, bruising, necrosis and pinking/browning	
Cell wall strength, elasticity and plasticity	Tensiometer analysis. The tensiometer pulls a leaf section three times which informs: 1) elasticity and plasticity 2) elasticity and 3) leaf strength where elasticity is the reversible stretching, plasticity is the initial irreversible stretching and the leaf strength is an indication of the cell wall strength and subsequent toughness of the leaf	
Leaf physio-logy	Imprints of the cell patterning on the leaf surface revealing the size and ratio of epidermal cells to stomata. Some of the leaves had fewer big cells and some of the leaves had more smaller cells, both with a range of stomata numbers	
Leaf morph-ology and yield	Leaf area images for differences in leaf sizes, partly determined by leaf shape	
Nutrition al quality	FRAP- Total antioxidant capacity analysis using a color change assay	
Pigment analysis	In the 2018 trial a CCM200 (opti-sciences) was used which measures the chlorophyll content of a leaf. In the 2019 trial a polypen (photo systems instruments) was used which measures additional pigment data such as greenness, and anthocyanin (stress pigment) content	

Appendix B

Genes identified within each candidate region investigated and annotation according to BLASTx against the *Helianthus annuus* genome.

Candidate region ID	*L. sativa* feature ID	*H. annus* gene ID	Sequence identity (%)	E value	Gene description
4.1	108900.1	OTF90738	90.667	2.76E-99	HVA22-like protein
4.1	108921.1	OTG16193	72.786	0	Putative cytochrome P450 82A4
4.1	107981.1	OTF90741	79.121	1.31E-46	Putative ZF-HD homeobox protein, Cys/His-rich dimerization domain-containing protein
4.1	107960.1	OTG16237	87.31	0	Aspartokinase (EC 1.1.1.3) (EC 2.7.2.4) (Homoserine dehydrogenase)
4.1	107581.1	OTG35787	76	1.11E-37	Putative GPCR kinase (Putative cGMP-dependent kinase)
4.2	142660.1	OTF98392	97.287	0	Putative 14-3-3-like protein 16R
4.2	142681.1	OTG06775	84.371	0	Putative ARM repeat superfamily protein
4.2	142700.1	OTG14167	86.617	0	Rhamnogalacturonan endolyase (EC 4.2.2.23)
4.2	142741.1	OTF85747	84.12	2.35E-117	Uncharacterized protein
4.2	142200.1	OTG15408	68.821	0	Putative tetratricopeptide repeat (TPR)-like superfamily protein
4.2	142020.1	OTG15272	88.92	0	Trehalose 6-phosphate phosphatase (EC 3.1.3.12)
4.2	144000.1	OTG15129	78.056	1.54E-173	Putative wound-responsive family protein
4.2	143840.1	OTG15130	71.145	0	Putative ankyrin repeat protein
4.2	143781.1	OTG30057	71.992	0	Putative tetratricopeptide-like helical domain-containing protein
4.2	143760.1	OTG15131	86.941	0	Phospholipase D (EC 3.1.4.4)
4.2	143721.1	OTG36236	27.301	1.93E-34	Putative nucleotide-binding alpha-beta plait domain-containing protein
4.2	143700.1	OTG15191	53.315	3.18E-109	Putative F-box domain-containing protein
4.2	143680.1	OTG28088	95.093	0	Serine/threonine-protein kinase TOR (EC 2.7.11.1)
4.2	143641.1	OTG15138	88.014	8.64E-177	Putative 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
4.2	144340.1	OTG20023	33.951	9.61E-19	Putative RNA-directed DNA polymerase, eukaryota
4.2	144300.1	OTF99677	66.087	7.88E-46	Putative DEA(D/H)-box RNA helicase family protein
4.2	144281.1	OTG02452	50.909	7.89E-45	Putative DNA mismatch repair protein, putative
4.2	144260.1	OTG26816	63.768	3.93E-09	Putative histone acetyltransferase of the TAFII250 family 2
4.2	144240.1	OTG22349	75.362	1.35E-49	Putative HAC13 protein (HAC13)
4.2	144221.1	OTG15142	90.566	9.09E-143	Putative RING/U-box superfamily protein
4.2	144561.1	OTG29289	54.701	1.01E-22	Uncharacterized protein
4.2	144541.1	OTG29289	82.759	3.31E-08	Uncharacterized protein
4.2	144520.1	OTG15148	83.778	0	Putative eukaryotic aspartyl protease family protein
4.2	144501.1	OTG22138	92.617	0	DNA replication licensing factor MCM7 (EC 3.6.4.12)
4.2	144441.1	OTF95841	83.409	0	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit (Oligosaccharyl transferase 48 kDa subunit)
4.2	144421.1	OTG15147	88.083	0	Protein-serine/threonine phosphatase (EC 3.1.3.16)
4.2	144401.1	OTG15146	62.443	0	Uncharacterized protein

5.1	31741.1	OTG01493	65.934	2.46E-60	Putative zinc finger, RanBP2-type
5.1	31640.1	OTG34890	67.677	3.05E-96	Putative alpha/Beta hydrolase fold protein
5.1	31600.1	OTG19515	53.857	0	Putative CRC domain, Lin-54 family
5.1	31580.1	OTF90294	71.93	1.22E-22	Putative WD repeat-containing protein 48
5.1	31560.1	OTG29957	89.485	0	Ammonium transporter
5.1	31541.1	OTG19994	52.423	1.15E-60	Putative plant Tudor-like RNA-binding protein
5.1	31501.1	OTG29595	83.929	1.28E-26	Putative ARM repeat superfamily protein
5.1	31481.1	OTG30227	80.328	8.61E-56	Cytochrome b-c1 complex subunit 7
5.1	31460.1	OTG09511	50.185	6.47E-54	Putative DNA-binding domain-containing protein
5.1	31800.1	OTF99361	68.323	7.57E-65	Putative limit dextrinase
5.1	32401.1	OTG22434	63.816	1.06E-51	Putative ATPase, AAA-type, core, P-loop containing nucleoside triphosphate hydrolase
5.1	32380.1	OTG23340	66	1.79E-86	Choline monooxygenase, chloroplastic (EC 1.14.15.7)
5.1	32361.1	OTG14738	50	6.95E-13	Putative DNA-binding domain-containing protein
5.1	32281.1	OTG38567	99.095	1.34E-155	GTP-binding nuclear protein
5.1	32261.1	OTF97007	54	1.90E-37	Putative SGNH hydrolase-type esterase domain-containing protein
5.1	32241.1	OTG13084	58.43	0	Putative potassium channel, voltage-dependent, EAG/ELK/ERG, Transient receptor potential channel
5.1	32220.1	OTG19997	86.544	0	Putative bicarbonate transporter, eukaryotic
5.1	32121.1	OTG12049	91.398	3.21E-59	Putative rlpA-like double-psi beta-barrel domain-containing protein
5.1	32061.1	OTG13488	57.406	0	Uncharacterized protein
5.1	32041.1	OTG38467	49.02	2.35E-101	Putative IQ motif, EF-hand binding site, Protein kinase-like domain protein
5.1	32021.1	OTG13488	60.896	0	Uncharacterized protein
5.1	32001.1	OTF99574	54.808	1.15E-113	Putative jacalin-like lectin domain-containing protein
5.1	31860.1	OTG18478	36.905	4.71E-47	Putative ulp1 protease family, C-terminal catalytic domain-containing protein
5.1	32481.1	OTG01438	82.799	0	Putative ABC transporter G family member 39
5.1	32461.1	OTG01438	82.654	0	Putative ABC transporter G family member 39
5.1	32440.1	OTG19636	85.484	0	Protein-serine/threonine phosphatase (EC 3.1.3.16)
5.1	32800.1	OTG19471	88.587	0	Putative LIS1 homology motif, WD40/YVTN repeat-like-containing domain protein
5.1	32821.1	OTG37890	30.827	2.71E-07	Putative nucleic acid-binding, OB-fold protein
5.1	32880.1	OTG38513	41.453	1.27E-42	Putative DNA-binding pseudobarrel domain-containing protein
5.1	32681.1	OTG31971	66.776	7.56E-153	Putative concanavalin A-like lectin/glucanase domain-containing protein
5.1	32641.1	OTG32487	67.308	9.10E-154	Putative C2 domain, GPCR kinase
5.1	32561.1	OTG19470	91.123	0	Malic enzyme
5.1	32541.1	OTG19469	89.769	0	Protein kinase domain-containing protein
5.1	32500.1	OTG01447	81.883	0	Putative homeobox protein
5.1	33821.1	OTF99593	81.042	0	Non-specific serine/threonine protein kinase (EC 2.7.11.1)
5.1	33740.1	OTG22022	94.737	2.24E-06	Putative glycosyl transferase, family 1
5.1	33721.1	OTG28216	59.786	1.52E-136	GATA transcription factor

5.1	33680.1	OTF91314	79.894	1.00E-107	Aminoacyl-tRNA hydrolase (EC 3.1.1.29)
5.1	33661.1	OTG19468	75.862	1.41E-92	SWIM-type domain-containing protein
5.1	33641.1	OTG19467	63.289	0	Putative regulator of chromosome condensation 1/beta-lactamase-inhibitor protein II
5.1	33620.1	OTG19466	52.155	1.33E-53	Uncharacterized protein
5.1	33600.1	OTG19635	77.822	0	Putative tocopherol cyclase, chloroplast / vitamin E deficient 1 (VTE1) / sucrose export defective 1 (Tocopherol cyclase)
5.1	32960.1	OTF94217	93.548	8.72E-13	Putative cullin-associated and neddylation dissociated protein
5.1	33001.1	OTG12080	77.143	3.51E-31	Putative ribosomal protein L6 family
5.1	33021.1	OTG09698	83.036	0	L-gulonolactone oxidase (EC 1.1.3.8)
5.1	33061.1	OTG06408	66.443	3.85E-52	Putative RNA helicase family protein
5.1	33420.1	OTG27964	68.432	0	Putative GAT domain, ENTH/VHS
5.1	33400.1	OTG27963	75.19	0	Putative aminotransferase class IV
5.1	33360.1	OTG27395	32.743	3.26E-07	Putative phospholipase-like protein
5.1	33341.1	OTG19634	73.927	7.74E-104	Uncharacterized protein
5.1	33301.1	OTG19633	95.56	0	Eukaryotic translation initiation factor 3 subunit L (eIF3l)
5.1	33281.1	OTG19619	90.783	0	Aspartate aminotransferase (EC 2.6.1.1)
5.1	33260.1	OTG19629	47.155	0	Putative homeodomain-like protein
5.1	33240.1	OTG08521	80.522	0	Aspartyl-tRNA synthetase (EC 6.1.1.12)
5.1	33201.1	OTG21260	63.545	0	Putative pentatricopeptide repeat protein
5.1	33160.1	OTG20169	77.604	0	Putative la protein 1
5.1	33141.1	OTG12216	72.727	1.07E-62	Putative ras-related protein
5.1	34021.1	OTG19998	95.413	5.18E-128	NFACT-R_1 domain-containing protein
5.1	34121.1	OTG23685	45.223	3.61E-23	Putative RAD3/XPD family
5.1	34141.1	OTG09328	83.948	0	Putative sulfite exporter TauE/SafE family protein
5.1	34221.1	OTG09328	76.271	1.87E-132	Putative sulfite exporter TauE/SafE family protein
5.2	124140.1	OTF91081	56.512	2.67E-131	Putative WRKY family transcription factor
5.2	124161.1	OTG05440	77.354	1.42E-173	Putative SLAC1-like protein
5.2	124281.1	OTG12240	57.456	4.70E-86	Endoplasmic reticulum transmembrane protein
5.2	124321.1	OTG12239	89.189	3.30E-13	Uncharacterized protein
5.2	124061.1	OTF91050	52.542	3.78E-34	Putative homeodomain-like protein
5.2	124021.1	OTF91050	39.815	6.82E-24	Putative homeodomain-like protein
5.2	124001.1	OTF91050	61.062	9.35E-38	Putative homeodomain-like protein
5.2	125321.1	OTG12081	63.91	9.85E-51	Uncharacterized protein
5.2	125300.1	OTG12080	93.396	1.98E-148	Putative ribosomal protein L6 family
5.2	125260.1	OTG12078	81.403	0	Glutamate receptor
5.2	125241.1	OTG12079	78.378	2.08E-126	Putative lojap-related protein
5.2	125200.1	OTG12077	69.672	7.96E-102	Putative B-box-type zinc finger
5.2	125081.1	OTG05442	74.695	1.77E-174	Uncharacterized protein
5.2	125061.1	OTG03701	73.684	4.40E-23	Thioredoxin-dependent peroxiredoxin (EC 1.11.1.24)
5.2	125041.1	OTG05439	67.935	8.24E-118	Putative interactor of constitutive active ROPs

5.2	125020.1	OTG31611	90.037	4.59E-178	Tyrosine--tRNA ligase (EC 6.1.1.1)
5.2	125361.1	OTG12074	81.25	0	Putative isopenicillin N synthase
5.2	125381.1	OTG22400	92.593	2.51E-49	Putative ribosomal protein L12/ ATP-dependent Clp protease adaptor protein ClpS family protein
5.3	141721.1	OTG03686	91.429	6.72E-63	Putative ELF4-like 3
5.3	141780.1	OTG13358	81.818	3.22E-42	Replication protein A subunit
5.3	141840.1	OTG13424	71.014	4.60E-53	Putative zinc finger, RING/FYVE/PHD-type
5.3	141861.1	OTG03687	95.111	1.47E-123	Putative charged multivesicular body protein 2a
5.3	141901.1	OTG13356	59.211	2.55E-79	Uncharacterized protein
5.3	141920.1	OTG03704	71.191	0	Mechanosensitive ion channel protein
5.3	141941.1	OTG29791	68.132	8.70E-66	Putative BREVIS RADIX-like 4
5.3	141980.1	OTG03703	88.61	0	Putative lecithin:cholesterol/phospholipid:diacylglycerol acyltransferase
5.3	142000.1	OTG13354	89.922	8.56E-146	Putative zinc finger, RING/FYVE/PHD-type, Alfin
5.3	142021.1	OTG04124	29.545	4.88E-14	Putative ankyrin repeat-containing domain, PGG domain protein
5.3	142140.1	OTG13353	93.491	0	Glycerophosphocholine acyltransferase 1
5.3	142161.1	OTG03702	53.344	1.12E-165	Putative tyrosine-protein kinase, non-receptor Jak2
5.3	141661.1	OTG03726	91.209	0	Threonine synthase (EC 4.2.3.1)
5.3	141540.1	OTF91360	85.714	2.99E-164	Putative root cap
5.3	141561.1	OTG03670	57.658	4.21E-95	Uncharacterized protein
5.3	143521.1	OTG03727	74.913	6.67E-129	Putative zinc ion binding protein
5.3	143500.1	OTG13351	91.473	1.62E-76	Uncharacterized protein
5.3	143480.1	OTF86047	77.419	6.68E-11	Putative cytochrome P450 85A1
5.3	143381.1	OTG13350	90.52	0	Putative serine/threonine/dual specificity protein kinase, catalytic domain-containing protein
5.3	143341.1	OTG13348	83.095	0	Putative FAD-binding Berberine family protein
5.3	143260.1	OTG14342	63.772	0	DDE Tnp4 domain-containing protein
5.3	143240.1	OTG30942	85.987	1.27E-66	Putative amino acid permease family protein
5.3	143221.1	OTG03721	79.57	5.82E-105	Putative lactoylglutathione lyase / glyoxalase I family protein
5.3	143200.1	OTG03720	76.462	0	E3 ubiquitin ligase (RING-type E3 ubiquitin transferase) (Serine/threonine-protein kinase)
5.3	143181.1	OTG13347	85.091	8.37E-140	MYB transcription factor
5.3	143161.1	OTG13346	78.342	0	Putative DWD (DDB1-binding WD40 protein) hypersensitive to ABA 1
5.3	143140.1	OTG24687	64.432	0	Putative cytochrome P450
5.3	143121.1	OTG13345	99.746	0	Putative RNA helicase HrpA
5.3	143101.1	OTG13344	97.97	0	Putative eukaryotic initiation factor 4A-8
5.3	143061.1	OTG13343	84.4	0	O-acyltransferase
5.3	143041.1	OTG13428	66.935	1.72E-43	Uncharacterized protein
5.3	143020.1	OTF86624	76.617	1.12E-98	Putative asparagine synthase family protein
5.3	143000.1	OTG03738	78.794	0	Uncharacterized protein
5.3	142981.1	OTG23092	47.917	1.00E-16	Uncharacterized protein

5.3	142961.1	OTG03736	82.818	0	Putative regulatory protein (NPR1)
5.3	142940.1	OTG13342	73.098	1.22E-179	Putative winged helix-turn-helix DNA-binding domain, Heat shock transcription factor family
5.3	142921.1	OTG13341	92	1.06E-80	Putative ribosomal protein S16 family protein
5.3	142901.1	OTG03730	84.109	3.57E-167	Putative alpha/beta-Hydrolases superfamily protein
5.3	142881.1	OTG13338	68.539	3.66E-157	Putative alpha/Beta hydrolase fold protein
5.3	142820.1	OTG30068	67.17	0	Cytokinin dehydrogenase (EC 1.5.99.12)
5.3	142720.1	OTG36643	52.381	1.02E-22	Uncharacterized protein
5.3	142620.1	OTG13432	88.542	4.78E-115	Translation machinery-associated protein 22
5.3	142561.1	OTG13337	82.916	0	Putative serine-threonine/tyrosine-protein kinase catalytic domain-containing protein
5.3	142541.1	OTG29807	81.467	0	Putative transducin/WD40 repeat-like superfamily protein
5.3	142520.1	OTG13333	89.627	0	Putative von Willebrand factor, type A
5.3	142501.1	OTG13332	68.387	9.40E-55	Uncharacterized protein
5.3	142361.1	OTG03758	60.287	0	Uncharacterized protein
5.3	142340.1	OTG13330	89.601	0	Beta-galactosidase (EC 3.2.1.23)
5.3	142241.1	OTG13329	47.692	3.65E-78	Uncharacterized protein
5.3	145340.1	OTG03753	63.842	3.61E-83	Putative zinc finger, NHR/GATA-type
5.3	145360.1	OTG16128	57.317	1.25E-44	Uroporphyrinogen-III synthase (EC 4.2.1.75)
5.3	145381.1	OTG21679	76.068	3.96E-58	Putative peptidoglycan-binding LysM domain-containing protein
5.3	145420.1	OTG03752	86.082	9.05E-121	Putative thioredoxin
5.3	145440.1	OTG13327	92.784	2.03E-63	Putative embryo defective protein
5.3	145481.1	OTG13326	80.066	2.33E-168	Cytidine deaminase (EC 3.5.4.5)
5.3	145561.1	OTG03750	85.502	1.89E-148	Putative tetraspanin2
5.3	145621.1	OTG03748	74.403	5.05E-139	Putative zinc finger, CCHC-type, Myb domain, plant, Homeodomain-like protein
5.3	145641.1	OTG03747	77.713	9.19E-169	Putative bacterial transferase hexapeptide repeat-containing protein
5.3	145661.1	OTG03746	78.849	0	Putative vascular endothelial growth factor receptor 2 (VEGFR2)
5.3	145680.1	OTG03745	65.217	0	Putative effector of transcription2
5.3	145700.1	OTG03744	88.621	2.46E-167	Serine O-acetyltransferase (EC 2.3.1.30)
5.3	145720.1	OTG03743	70.694	0	Putative DYW domain-containing protein
5.3	145741.1	OTG03742	85.404	0	Putative graves disease carrier protein
5.3	145761.1	OTF87658	88.67	5.79E-115	3-oxoacyl-[acyl-carrier-protein] synthase
5.3	145800.1	OTG03769	85.393	1.78E-109	Germin-like protein
5.3	145821.1	OTG03806	95.21	0	Putative NDRG, Alpha/Beta hydrolase fold protein
5.3	145840.1	OTG13314	83.548	4.34E-171	Saposin B-type domain-containing protein
5.3	145881.1	OTG13312	81.693	0	Putative metallopeptidase M24 family protein
5.3	145921.1	OTG04656	68.421	5.21E-19	Putative DNA-binding pseudobarrel domain-containing protein
5.3	145980.1	OTG13311	93.483	0	Putative pyridoxal phosphate-dependent transferase
5.3	146000.1	OTG03803	85.714	1.92E-85	Putative ATPase, F0 complex, subunit G
5.3	146040.1	OTG03802	73.306	0	Histone acetyltransferase (EC 2.3.1.48)

5.3	146061.1	OTG03801	81.046	2.02E-166	Uncharacterized protein
5.3	146080.1	OTG30126	95	5.63E-67	Putative ribosomal protein L31e family protein
5.3	146100.1	OTG12021	77.083	1.44E-20	Putative P-loop containing nucleoside triphosphate hydrolase
5.3	146120.1	OTG24097	85.271	2.58E-74	Putative 18 kDa seed maturation protein
5.3	146141.1	OTG00390	62.329	3.06E-114	Putative zinc finger, CCCH-type, K-like domain-containing protein
5.3	146160.1	OTG13308	100	4.40E-39	40S ribosomal protein S30
5.3	146261.1	OTG13306	91.324	4.11E-127	Meiotic nuclear division protein 1 homolog
5.3	146280.1	OTG03797	84.211	0	Uncharacterized protein
5.3	146321.1	OTG37964	81.481	5.64E-09	Putative phosphatidic acid phosphatase type 2/haloperoxidase
5.3	146340.1	OTG03796	83.122	0	Putative gamma-glutamyl transpeptidase 4
5.3	146360.1	OTG03793	89.888	0	Putative mitochondrial-processing peptidase subunit alpha
5.3	146440.1	OTG03790	83.94	0	Putative NAC domain containing protein 75
5.3	146461.1	OTF90994	76.19	9.52E-19	Coatomer subunit gamma
5.3	146481.1	OTF98436	90	2.04E-09	Putative homeodomain-like, Helicase/SANT-associated domain protein
5.3	146540.1	OTG13299	82.767	0	Putative leucine-rich repeat (LRR) family protein
5.3	144781.1	OTG03789	71.901	2.52E-126	Putative HAD superfamily, subfamily IIIB acid phosphatase
5.3	144820.1	OTG13501	83.75	1.64E-139	Derlin
5.3	144840.1	OTG13289	85.714	5.81E-86	Profilin
5.3	145241.1	OTG13290	93.077	5.58E-78	Profilin
5.3	145221.1	OTG13291	92.09	0	Putative MAP kinase kinase 6
5.3	145201.1	OTG13292	53.289	7.98E-32	Uncharacterized protein
5.3	145181.1	OTF96421	100	4.68E-19	60S ribosomal protein L36
5.3	145141.1	OTG13294	78.947	0	Pectinesterase (EC 3.1.1.11)
5.3	145160.1	OTF84746	55.882	2.39E-25	DNA-directed RNA polymerase subunit beta'' (EC 2.7.7.6) (PEP) (Plastid-encoded RNA polymerase subunit beta'') (RNA polymerase subunit beta'')
5.3	145120.1	OTG03786	58.952	6.28E-66	Putative IQ motif, EF-hand binding site
5.3	145100.1	OTG02678	91.837	9.57E-27	Putative peptidase S59, nucleoporin
5.3	145081.1	OTF87628	80.369	0	Putative calcium-dependent channel, 7TM region, putative phosphate
5.3	145061.1	OTG03784	90.717	0	Phosphotransferase (EC 2.7.1.-)
5.3	145040.1	OTG13295	90.847	0	Putative 6-phosphogluconate dehydrogenase family protein
5.3	144981.1	OTF95199	96.078	0	Chlorophyll a-b binding protein, chloroplastic
5.3	144940.1	OTF95199	96.078	0	Chlorophyll a-b binding protein, chloroplastic
5.3	144920.1	OTG13509	65.028	0	Ceramide glucosyltransferase (EC 2.4.1.80)
5.3	144900.1	OTG13298	83.039	1.89E-139	Putative RNA methyltransferase TrmH family
5.3	144881.1	OTF91326	81.279	1.52E-105	Uncharacterized protein
5.3	144600.1	OTF90819	69.754	0	RNA-dependent RNA polymerase (EC 2.7.7.48)
5.3	144641.1	OTG02093	77.114	4.71E-111	Putative amino acid transporter, transmembrane domain-containing protein

5.3	144661.1	OTG13239	69.328	5.72E-105	Uncharacterized protein
5.3	144680.1	OTF95732	52.632	2.87E-50	Post-GPI attachment to proteins factor 3
5.3	144720.1	OTF84369	90.164	1.77E-76	50S ribosomal protein L14, chloroplastic
5.3	144761.1	OTG23155	78.378	3.03E-53	Calcium-transporting ATPase (EC 7.2.2.10)
5.3	143861.1	OTG13242	87.037	0	Putative leucine-rich repeat (LRR) family protein
5.3	143841.1	OTG13243	93.885	0	Dolichyl-diphosphooligosaccharide--protein glycotransferase (EC 2.4.99.18)
5.3	143581.1	OTG13244	70.13	1.58E-17	Putative integrase-type DNA-binding superfamily protein
5.3	143721.1	OTG13247	75.753	0	Putative exostosin family protein
5.3	143741.1	OTG03778	80.886	0	Putative exostosin-like protein
5.3	144140.1	OTG26775	78.495	5.61E-45	Putative pentatricopeptide repeat (PPR) superfamily protein
5.3	144181.1	OTG13251	90.234	0	Putative sugar transporter 14
5.3	144220.1	OTF99427	64.583	3.00E-75	Putative AGAMOUS-like 61
5.3	144340.1	OTG13252	67.849	2.06E-167	Putative RING/U-box superfamily protein
5.3	144380.1	OTG13511	80.573	0	Lactoylglutathione lyase (EC 4.4.1.5) (Glyoxalase I)
5.3	144400.1	OTG13259	85.223	0	Xyloglucan endotransglucosylase/hydrolase (EC 2.4.1.207)
5.3	144420.1	OTG13259	85.567	0	Xyloglucan endotransglucosylase/hydrolase (EC 2.4.1.207)
5.3	144460.1	OTG13259	85.223	0	Xyloglucan endotransglucosylase/hydrolase (EC 2.4.1.207)
5.3	147380.1	OTG31606	39.45	3.07E-53	Putative transposase, Ptta/En/Spm, plant, Transposase, Tnp1/En/Spm-like protein
5.3	147400.1	OTG13257	85.926	3.70E-59	Uncharacterized protein
5.3	147421.1	OTG13255	86.538	1.01E-24	Xyloglucan endotransglucosylase/hydrolase (EC 2.4.1.207)
5.3	147441.1	OTG34910	56.797	0	Putative galactose oxidase/kelch, beta-propeller
5.3	147480.1	OTG09780	91.304	1.69E-39	Putative transducin family protein / WD-40 repeat family protein
5.3	147500.1	OTG09780	67.153	2.09E-58	Putative transducin family protein / WD-40 repeat family protein
5.3	147581.1	OTG21290	79.032	4.40E-23	Putative transmembrane amino acid transporter family protein
5.3	147641.1	OTG04060	50.785	1.65E-21	Uncharacterized protein
5.3	147680.1	OTG15650	67.368	2.05E-33	Putative HEAT repeat-containing protein
5.3	147721.1	OTG15104	72.368	9.46E-66	Putative senescence/dehydration-associated protein-related protein
5.3	147760.1	OTG13278	71.078	0	Elongation factor Ts, mitochondrial (EF-Ts) (EF-TsMt)
5.3	147860.1	OTG13277	80.229	0	Putative nucleoside phosphorylase
5.3	147900.1	OTG13276	74.686	0	RING-type E3 ubiquitin transferase (EC 2.3.2.27)
5.3	147260.1	OTG03775	90.741	6.48E-109	Glutaredoxin-dependent peroxiredoxin (EC 1.11.1.25)
5.3	147281.1	OTG11499	93.502	0	Putative lipoamide dehydrogenase 1
5.3	147300.1	OTF91680	95.987	0	Serine/threonine-protein phosphatase (EC 3.1.3.16)
5.3	147321.1	OTF91681	72.783	2.18E-173	Peroxidase (EC 1.11.1.7)
5.3	147341.1	OTF91681	73.077	3.52E-172	Peroxidase (EC 1.11.1.7)

5.3	146660.1	OTG11492	85.475	7.74E-94	Putative nucleotide-binding alpha-beta plait domain-containing protein
5.3	146681.1	OTG11357	93.617	0	Putative 5'-AMP-activated protein kinase subunit gamma-1
5.3	146901.1	OTF99818	57.407	3.37E-08	Uncharacterized protein
5.3	146881.1	OTG22755	42.105	1.90E-28	DNA polymerase alpha subunit B
5.3	146840.1	OTG11356	87.9	0	Putative regulator of chromosome condensation (RCC1) family protein
5.3	146820.1	OTF91699	87.179	0	Sulfate adenylyltransferase (EC 2.7.7.4)
5.3	146781.1	OTG11041	78.15	0	O-fucosyltransferase family protein
5.3	146760.1	OTG11040	64.219	0	Putative acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger protein
5.3	146920.1	OTG10695	58	7.53E-15	Putative bulb-type lectin domain-containing protein
5.3	146940.1	OTG10700	49.153	8.11E-08	Receptor-like serine/threonine-protein kinase (EC 2.7.11.1)
5.3	147020.1	OTF87411	69.014	7.44E-28	Putative serine/threonine/dual specificity protein kinase, catalytic domain-containing protein
5.3	147040.1	OTF97309	48.925	1.18E-37	Putative ulp1 protease family, C-terminal catalytic domain-containing protein
5.3	147080.1	OTG11037	91.077	0	Cysteine synthase (EC 2.5.1.47)
5.3	147140.1	OTG11030	82.704	0	Putative inosine/uridine-preferring nucleoside hydrolase domain-containing protein
5.3	147161.1	OTG11030	89.825	0	Putative inosine/uridine-preferring nucleoside hydrolase domain-containing protein
6.1	87700.1	OTF94534	50	1.64E-12	Non-specific serine/threonine protein kinase (EC 2.7.11.1)
6.1	87800.1	OTG02766	85.065	0	Putative MBOAT (Membrane bound O-acyl transferase) family protein (Putative lysophosphatidylcholine acyltransferase 1)
6.1	87861.1	OTF84746	97.992	0	DNA-directed RNA polymerase subunit beta'' (EC 2.7.7.6) (PEP) (Plastid-encoded RNA polymerase subunit beta'') (RNA polymerase subunit beta'')
6.1	87881.1	OTG32366	92.359	0	Putative nucleotide-rhamnose synthase/epimerase-reductase
6.1	87901.1	OTG14569	68	4.56E-23	Putative zinc finger, MYND-type
6.1	87921.1	OTG02272	89.655	1.51E-12	Putative nudix hydrolase
6.1	87960.1	OTG22059	46.62	5.03E-102	Putative RING/U-box superfamily protein
6.1	87981.1	OTG32365	81.231	0	Putative NAC domain-containing protein
6.1	88140.1	OTG02763	75.84	0	Putative poltergeist like 1
6.1	88041.1	OTG07764	47.881	1.69E-48	Putative WW domain-containing protein
6.1	88181.1	OTG35323	38.298	3.06E-06	Putative zinc finger, GRF-type
6.1	88261.1	OTG02762	84.583	6.59E-154	Putative haloacid dehalogenase-like hydrolase (HAD) superfamily protein
6.1	88281.1	OTG02758	41.053	1.52E-10	Putative PLAC8 motif-containing protein
6.1	88321.1	OTG32356	79.061	5.94E-134	Putative tetraspanin family protein
6.1	88341.1	OTG32355	90.071	0	Methyltransferase (EC 2.1.1.-)
6.1	88360.1	OTF90608	55.914	6.57E-104	Peroxidase (EC 1.11.1.7)
6.1	88380.1	OTG32354	79.501	0	Putative S-norcoclaurine synthase 1
6.1	89461.1	OTG32456	87.544	0	Pyruvate kinase (EC 2.7.1.40)

6.1	89421.1	OTG32462	89.403	0	Putative RNA helicase, putative
6.1	89400.1	OTG32461	88.514	0	Lon protease homolog 2, peroxisomal (EC 3.4.21.-)
6.1	89361.1	OTG32458	69.231	1.08E-64	Uncharacterized protein
6.1	89321.1	OTG32458	41.83	1.25E-26	Uncharacterized protein
6.1	88940.1	OTG29435	82.857	5.29E-96	Putative membrane trafficking VPS53 family protein
6.1	88920.1	OTG09233	53.465	2.34E-25	Putative acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger domain-containing protein
6.1	89141.1	OTG01855	91.579	1.39E-54	AMP deaminase (EC 3.5.4.6)
6.1	89121.1	OTG32457	59.412	2.39E-84	Putative SBP (S-ribonuclease binding protein) family protein
6.1	88460.1	OTG02806	77.907	1.62E-76	FLZ-type domain-containing protein
6.1	88480.1	OTG32491	77.083	0	Putative endoribonuclease XendoU-like protein
6.1	88500.1	OTF96161	42.781	3.37E-34	Uncharacterized protein
6.1	88521.1	OTG32512	75.679	0	Putative peptidase S8, subtilisin-related protein
6.1	88540.1	OTG32501	76.433	3.30E-80	Putative cupredoxin
6.1	88560.1	OTG32501	77.07	5.21E-91	Putative cupredoxin
6.1	88580.1	OTG32503	72.611	1.45E-80	Putative cupredoxin
6.1	88600.1	OTG32501	69.128	7.03E-71	Putative cupredoxin
6.1	88621.1	OTG02809	73.356	8.16E-150	Putative SNARE associated Golgi protein
6.1	88640.1	OTG32498	82.845	0	Putative zinc ion binding protein
6.1	88660.1	OTG32497	62.139	1.69E-136	Putative sodium/metabolite cotransporter BASS5 protein
6.1	88680.1	OTG32497	73.976	0	Putative sodium/metabolite cotransporter BASS5 protein
6.1	88700.1	OTG32497	85.075	1.81E-76	Putative sodium/metabolite cotransporter BASS5 protein
6.1	88720.1	OTG02810	69.586	0	Putative bile acid:sodium symporter/arsenical resistance protein Acr3
6.1	88780.1	OTG04247	68.333	7.34E-23	Putative HSP20-like chaperones superfamily protein
6.1	88800.1	OTG07426	65.244	2.23E-66	Putative A20/AN1-like zinc finger family protein
6.1	88880.1	OTG34599	77.354	0	Putative ACT domain repeat 8
6.1	88901.1	OTG02822	85.841	6.47E-71	Putative cornichon
6.1	89680.1	OTG08017	54.918	2.22E-30	Putative LCCL domain, Histone deacetylation protein Rxt3
6.1	89561.1	OTG02818	80.748	0	Amine oxidase (EC 1.4.3.-)
6.1	90041.1	OTG02819	85.616	0	Amine oxidase (EC 1.4.3.-)
6.1	89861.1	OTG02815	64.126	0	Putative cytochrome P450
6.1	89841.1	OTG02815	72.124	0	Putative cytochrome P450
6.1	89820.1	OTG02815	68.212	0	Putative cytochrome P450
6.1	89800.1	OTG02815	72.532	0	Putative cytochrome P450
6.1	89780.1	OTG02815	73.218	0	Putative cytochrome P450
6.1	89760.1	OTG02815	71.017	0	Putative cytochrome P450
6.1	89741.1	OTG11703	65.909	2.99E-14	Putative DNA-binding pseudobarrel domain-containing protein
6.1	89500.1	OTG18999	72.872	4.56E-85	Rab3 GTPase-activating protein catalytic subunit

6.1	89700.1	OTG37853	68.874	5.70E-58	Leucine zipper-EF-hand-containing transmembrane protein 1 (Mitochondrial proton/calcium exchanger protein)
6.1	90461.1	OTG02797	76.66	0	Putative cytochrome P450
6.1	90481.1	OTG02800	76.449	8.31E-148	Putative myb domain protein 42
6.1	90520.1	OTG32558	76.119	7.37E-97	Putative pectinesterase inhibitor domain-containing protein
6.1	90540.1	OTG32558	74.627	1.06E-90	Putative pectinesterase inhibitor domain-containing protein
6.1	90640.1	OTG32557	77.612	3.90E-118	Putative pectin methylesterase inhibitor
6.1	90680.1	OTG02801	80.952	1.61E-84	Putative pectinesterase inhibitor domain-containing protein
6.1	90700.1	OTG32555	93.579	0	Elongation factor G, chloroplastic (cEF-G)
6.1	90720.1	OTG32554	90.678	0	Putative hsp70-Hsp90 organizing protein
6.1	90100.1	OTG02812	72.345	0	Putative peptidase C13, legumain
6.1	90121.1	OTG23746	67.857	2.90E-43	Putative mo25 family protein
6.1	90160.1	OTG02812	72.124	0	Putative peptidase C13, legumain
6.1	90200.1	OTG32551	70.886	0	Putative peptidase C13, legumain
6.1	90220.1	OTG32553	70.021	0	Putative peptidase C13, legumain
6.1	90240.1	OTG32547	80.226	0	Putative NAC domain-containing protein
6.1	90261.1	OTG25382	43.094	2.15E-90	Uncharacterized protein
6.1	90281.1	OTG32546	89.882	0	Putative cupredoxin
6.1	90301.1	OTG32544	80.804	0	Putative glycosyl hydrolases family 32 protein
6.1	90321.1	OTG32543	89.182	0	Transaldolase (EC 2.2.1.2)
6.1	90381.1	OTG34782	76.533	0	Trehalose 6-phosphate phosphatase (EC 3.1.3.12)
6.1	91661.1	OTG00786	53.731	1.95E-34	Putative plastid lipid-associated protein
6.1	91561.1	OTG32617	84.651	8.18E-132	Putative ybaK/aminoacyl-tRNA synthetase-associated domain-containing protein
6.1	91520.1	OTG32616	62.121	3.79E-49	Uncharacterized protein
6.1	91481.1	OTG06192	84.483	2.07E-30	Putative homeodomain-like, carrier protein
6.1	91441.1	OTG32614	65.263	5.12E-38	Putative heavy metal-associated domain, HMA

Appendix C

Significantly enriched gene ontology terms identified within investigated candidate gene regions.

Candidate region ID	**Category**	**Enrichment FDR**	**Genes in list**	**Total genes**	**Functional Category**
4.1	Not enough genes for enrichment analyses (5)				
4.2	Molecular	6.10E-04	2	24	RNA polymerase II general transcription initiation factor activity
4.2	Molecular	6.10E-04	2	21	TBP-class protein binding
4.2	Molecular	6.10E-04	2	23	General transcription initiation factor binding
4.2	Molecular	7.20E-04	2	30	General transcription initiation factor activity
4.2	Molecular	1.60E-02	2	173	Transcription factor binding

4.2	Molecular	1.60E-02	3	587	Phosphoric ester hydrolase activity
4.2	Molecular	2.80E-02	4	1552	Hydrolase activity, acting on ester bonds
4.2	Molecular	2.90E-02	2	274	Helicase activity
4.2	biological processes	1.70E-03	2	31	RNA polymerase II preinitiation complex assembly
4.2	biological processes	1.70E-03	3	167	Protein-DNA complex assembly
4.2	biological processes	1.70E-03	3	187	Protein-DNA complex subunit organization
4.2	biological processes	1.90E-03	2	38	Transcription preinitiation complex assembly
4.2	biological processes	3.90E-03	2	61	Transcription initiation from RNA polymerase II promoter
4.2	biological processes	4.50E-03	2	72	Pollen development
4.2	biological processes	5.80E-03	2	89	Gametophyte development
4.2	biological processes	6.10E-03	2	97	DNA-templated transcription, initiation
4.2	biological processes	3.40E-02	3	813	Cellular protein-containing complex assembly
4.2	biological processes	3.40E-02	2	260	DNA recombination
4.2	biological processes	3.60E-02	3	872	Anatomical structure development
4.2	biological processes	3.60E-02	3	888	Protein-containing complex assembly
4.2	biological processes	4.40E-02	3	987	Developmental process
4.2	biological processes	4.40E-02	3	1015	Protein-containing complex subunit organization
4.2	Cellular component	1.50E-03	2	25	Transcription factor TFIID complex
4.2	Cellular component	2.60E-03	5	1372	Membrane-enclosed lumen
4.2	Cellular component	2.60E-03	5	1372	Organelle lumen
4.2	Cellular component	2.60E-03	5	1372	Intracellular organelle lumen
4.2	Cellular component	2.60E-03	4	797	Transferase complex
4.2	Cellular component	3.60E-03	2	97	RNA polymerase II, holoenzyme
4.2	Cellular component	3.60E-03	2	102	RNA polymerase II transcription regulator complex
4.2	Cellular component	3.80E-03	8	4655	Protein-containing complex
4.2	Cellular component	4.10E-03	3	505	Nucleoplasm
4.2	Cellular component	4.10E-03	4	1121	Nuclear lumen
4.2	Cellular component	4.10E-03	2	137	Nuclear DNA-directed RNA polymerase complex
4.2	Cellular component	4.30E-03	2	147	DNA-directed RNA polymerase complex

4.2	Cellular component	4.30E-03	2	154	Transcription regulator complex
4.2	Cellular component	4.40E-03	2	162	RNA polymerase complex
4.2	Cellular component	7.40E-03	4	1440	Catalytic complex
4.2	Cellular component	1.50E-02	2	326	Transferase complex, transferring phosphorus-containing groups
5.1	Cellular	6.70E-04	6	728	Cell-cell junction
5.1	Cellular	6.70E-04	6	728	Plasmodesma
5.1	Cellular	6.70E-04	6	729	Cell junction
5.1	Cellular	6.70E-04	6	728	Symplast
5.1	Cellular	6.70E-04	6	728	Anchoring junction
5.1	Cellular	1.70E-02	3	321	Plastid stroma
5.1	Cellular	1.70E-02	3	317	Chloroplast stroma
5.1	Molecular	3.80E-02	14	4870	Purine ribonucleoside triphosphate binding
5.1	Molecular	4.30E-02	12	4449	ATP binding
5.1	Molecular	4.30E-02	2	86	Transaminase activity
5.1	Molecular	4.30E-02	6	1564	Pyrophosphatase activity
5.1	Molecular	4.30E-02	2	86	Transferase activity, transferring nitrogenous groups
5.1	Molecular	4.30E-02	6	1578	Hydrolase activity, acting on acid anhydrides
5.1	Molecular	4.30E-02	6	1570	Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides
5.1	Molecular	4.30E-02	6	1488	Nucleoside-triphosphatase activity
5.1	Molecular	4.30E-02	12	4850	Adenyl nucleotide binding
5.1	Molecular	4.30E-02	12	4824	Adenyl ribonucleotide binding
5.1	Molecular	4.70E-02	6	1648	Protein serine/threonine kinase activity
5.2	molecular function	1.60E-02	2	176	Gated channel activity
5.2	molecular function	2.30E-02	2	330	Ion channel activity
5.2	molecular function	2.30E-02	2	431	Channel activity
5.2	molecular function	2.30E-02	2	431	Passive transmembrane transporter activity
5.3	Biological process	2.10E-03	22	2573	Organonitrogen compound biosynthetic process
5.3	Biological process	2.90E-02	5	202	Cytoplasmic translation
5.3	Biological process	2.90E-02	2	8	Sequestering of actin monomers
5.3	Biological process	2.90E-02	11	1040	Carbohydrate derivative metabolic process
5.3	Biological process	3.60E-02	2	17	Entrainment of circadian clock
5.3	Biological process	3.60E-02	2	20	Cysteine biosynthetic process
5.3	Biological process	3.60E-02	3	60	Cytoplasmic translational initiation

5.3	Biological process	3.60E-02	2	23	Negative regulation of actin filament polymerization
5.3	Biological process	3.60E-02	2	23	Negative regulation of protein-containing complex assembly
5.3	Biological process	3.60E-02	2	23	Negative regulation of protein polymerization
5.3	Biological process	3.60E-02	2	15	Nucleobase-containing small molecule catabolic process
5.3	Biological process	3.60E-02	2	23	Regulation of circadian rhythm
5.3	Biological process	3.60E-02	12	1497	Amide biosynthetic process
5.3	Biological process	3.60E-02	2	23	Negative regulation of cytoskeleton organization
5.3	Biological process	3.60E-02	2	23	Negative regulation of supramolecular fiber organization
5.3	Biological process	3.60E-02	3	74	Liposaccharide metabolic process
5.3	Biological process	3.60E-02	2	15	Cysteine biosynthetic process from serine
5.3	Biological process	3.60E-02	3	74	Glycolipid metabolic process
5.3	Biological process	3.60E-02	2	15	Nucleoside catabolic process
5.3	Biological process	3.60E-02	3	72	Glycolipid biosynthetic process
5.3	Biological process	4.00E-02	2	25	Glycosyl compound catabolic process
5.3	Biological process	4.60E-02	9	1010	Regulation of biological quality
5.3	Biological process	4.60E-02	2	28	Cysteine metabolic process
5.3	cellular processes	1.80E-02	18	2491	Non-membrane-bounded organelle
5.3	cellular processes	1.80E-02	18	2491	Intracellular non-membrane-bounded organelle
5.3	molecular function	2.20E-02	2	8	Actin monomer binding
5.3	molecular function	2.20E-02	9	774	Transferase activity, transferring acyl groups
5.3	molecular function	3.80E-02	4	158	O-acyltransferase activity
6.1	biolofical processes	1.40E-04	3	17	Vacuolar protein processing
6.1	biolofical processes	3.50E-04	4	93	Protein processing
6.1	biolofical processes	1.10E-03	4	139	Protein maturation
6.1	biolofical processes	1.50E-02	6	890	Protein catabolic process
6.1	biolofical processes	1.50E-02	6	870	Cellular protein catabolic process
6.1	biolofical processes	1.50E-02	7	1126	Cellular macromolecule catabolic process

6.1	biolofical processes	1.50E-02	6	867	Proteolysis involved in cellular protein catabolic process
6.1	biolofical processes	2.50E-02	12	3611	Oxidation-reduction process
6.1	biolofical processes	2.50E-02	7	1374	Macromolecule catabolic process
6.1	biolofical processes	2.60E-02	8	1866	Cellular catabolic process
6.1	biolofical processes	2.60E-02	8	1845	Proteolysis
6.1	biolofical processes	2.60E-02	9	2331	Catabolic process
6.1	biolofical processes	2.90E-02	5	815	Regulation of catalytic activity
6.1	biolofical processes	2.90E-02	5	821	Regulation of molecular function
6.1	biolofical processes	2.90E-02	6	1148	Organonitrogen compound catabolic process
6.1	biolofical processes	3.40E-02	8	2079	Organic substance catabolic process
6.1	cellular component	2.70E-02	4	415	Plastid envelope
6.1	cellular component	2.70E-02	4	356	Chloroplast envelope
6.1	cellular component	2.70E-02	3	245	Anchored component of membrane
6.1	cellular component	2.70E-02	3	240	Anchored component of plasma membrane
6.1	cellular component	3.20E-02	3	321	Plastid stroma
6.1	cellular component	3.20E-02	3	317	Chloroplast stroma
6.1	cellular component	3.20E-02	3	347	Intrinsic component of plasma membrane
6.1	cellular component	3.20E-02	5	1032	Organelle envelope
6.1	cellular component	3.20E-02	5	1032	Envelope
6.1	cellular component	4.20E-02	4	744	Vacuole
6.1	cellular component	4.60E-02	7	2104	Chloroplast
6.1	molecular function	3.40E-04	19	4937	Cation binding
6.1	molecular function	3.40E-04	19	4896	Metal ion binding
6.1	molecular function	7.90E-04	4	165	Cysteine-type endopeptidase activity
6.1	molecular function	7.90E-04	2	12	Primary amine oxidase activity
6.1	molecular function	7.90E-04	2	12	Tryptamine:oxygen oxidoreductase (deaminating) activity
6.1	molecular function	7.90E-04	2	12	Aminoacetone:oxygen oxidoreductase(deaminating) activity

6.1	molecular function	7.90E-04	2	12	Aliphatic-amine oxidase activity
6.1	molecular function	7.90E-04	2	12	Phenethylamine:oxygen oxidoreductase (deaminating) activity
6.1	molecular function	1.10E-03	2	15	Oxidoreductase activity, acting on the CH-NH2 group of donors, oxygen as acceptor
6.1	molecular function	1.60E-03	6	687	Endopeptidase activity
6.1	molecular function	2.20E-03	4	266	Cysteine-type peptidase activity
6.1	molecular function	2.30E-03	10	2169	Transition metal ion binding
6.1	molecular function	2.40E-03	7	1093	Peptidase activity
6.1	molecular function	2.40E-03	2	27	Oxidoreductase activity, acting on the CH-NH2 group of donors
6.1	molecular function	2.70E-03	3	131	Copper ion binding
6.1	molecular function	9.30E-03	5	723	Enzyme regulator activity
6.1	molecular function	3.50E-02	3	352	Enzyme inhibitor activity
6.1	molecular function	3.50E-02	10	3366	Oxidoreductase activity
6.1	molecular function	3.50E-02	4	661	Catalytic activity, acting on RNA
6.1	molecular function	3.80E-02	2	132	Quinone binding
6.1	molecular function	4.40E-02	2	148	Enzyme activator activity
6.1	molecular function	4.50E-02	5	1149	Zinc ion binding