

# Dark period transcriptomic and metabolic profiling of two diverse *Eutrema salsugineum* accessions

Jie Yin<sup>1</sup>

Email: [jyin@purdue.edu](mailto:jyin@purdue.edu)

Michael J. Gosney<sup>2</sup>

Email: [mjgosney@purdue.edu](mailto:mjgosney@purdue.edu)

Brian P. Dilkes<sup>3\*</sup>

\* Corresponding author

Email: [bdilkes@purdue.edu](mailto:bdilkes@purdue.edu)

Michael V. Mickelbart<sup>1,2\*</sup>

\* Corresponding author

Email: [mickelbart@purdue.edu](mailto:mickelbart@purdue.edu)

<sup>1</sup> Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, Indiana 47907, USA

<sup>2</sup> Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana 47907, USA

<sup>3</sup> Department of Biochemistry, Purdue University, West Lafayette, Indiana 47907, USA

## Abstract

*Eutrema salsugineum* is a model species for the study of plant adaptation to abiotic stresses. Two accessions of *E. salsugineum*, Shandong (SH) and Yukon (YK), exhibit contrasting morphology, biotic, and abiotic stress tolerance. Transcriptome and metabolic profiling from tissue samples collected during the dark period were used to investigate the molecular and metabolic bases of these contrasting phenotypes. RNA sequencing identified 17,888 expressed genes, of which 157 were not in the published reference genome and 65 were detected for the first time. Differential expression was detected for only 31 genes. The RNA sequencing data contained 14,808 single nucleotide polymorphisms (SNPs) in transcripts, 3,925 of which are newly identified. Among the differentially expressed genes, there were no obvious candidates for the physiological or morphological differences between SH and YK. Metabolic profiling indicated that YK accumulates free fatty acids and long-chain fatty acid derivatives as compared to SH; whereas sugars are more abundant in SH. Metabolite levels suggest that carbohydrate and respiratory metabolism, including starch degradation, is more active during the first half of the dark period in SH. These metabolic differences may explain the greater biomass accumulation in YK over SH. The accumulation of 56% of the identified metabolites was lower in F<sub>1</sub> hybrids than the mid-parent averages and the accumulation of 17% of the metabolites in F<sub>1</sub> plants transgressed the level in both parents. Concentrations of several metabolites in F<sub>1</sub> hybrids agree with previous studies and suggest a role for primary metabolism in heterosis. The improved annotation of the *E. salsugineum* genome and newly-identified high-quality SNPs will permit accelerated studies using the standing variation in this species to elucidate the mechanisms of its diverse adaptations to the environment.

**Keywords:** extremophile, differentially expressed genes, transcriptome, single nucleotide polymorphisms, metabolomics, fatty acid metabolism, transgressive variation

## Introduction

*Eutrema salsugineum* (formerly *Thellungiella halophila*) is a model species for the study of plant stress tolerance (Amtmann, 2009; Bressan et al., 2001; Griffith et al., 2007; Oh et al., 2010; Orsini et al., 2010; Pilarska et al., 2016; Taji et al., 2004; Volkov et al., 2003; Wong et al., 2005). The two most commonly studied accessions, Shandong (SH) and Yukon (YK), are native to the Yellow River region of China (Bressan et al., 2001; Inan et al., 2004) and the Yukon territories of Canada (Wong et al., 2005), respectively. These accessions contrast in cold tolerance (Lee et al., 2012), water stress tolerance (MacLeod et al., 2014; Xu et al., 2014), and disease resistance (Yeo et al., 2014). In response to water stress for example, YK accumulates more cuticular wax (Xu et al., 2014), exhibits delayed wilting due to higher leaf water content, and maintains a higher leaf area, as compared to SH (MacLeod et al., 2014). Some differences in adaptive mechanisms have been linked to metabolism, such as a more pronounced increase in fructose and proline content after cold acclimation in YK compared to SH (Lee et al., 2012).

RNA-seq has been used to identify differentially expressed genes (DEGs) that contribute to genotypic variation in response to physiological conditions (Bazakos et al., 2012; Huang et al., 2012a; Huber-Keener et al., 2012; Niederhuth et al., 2013; Stein and Waters, 2012; Wang et al., 2009). When comparing expression between polymorphic accessions of a species, RNA-seq allows for the simultaneous identification and quantification of single nucleotide polymorphisms (SNPs) (Nielsen et al., 2011; Wang et al., 2009). These data can later be utilized as genetic markers for linkage mapping (Chen et al., 2011; Hancock et al., 2011; Hyten et al., 2009; Lisec et al., 2008; Trick et al., 2012) and to assess allele-specific expression (Oshlack et al., 2010; Pickrell et al., 2010).

Plant metabolic profiling permits the simultaneous measurement of multiple intermediates and the end-products of biochemical pathways. Similar to RNA-seq, metabolic profiling can be used to investigate the metabolic and physiological status of biological systems (Fiehn et al., 2000) and may provide biochemical bases for differences in growth and physiology (Meyer et al., 2007). Metabolite profiling has revealed correlations between particular metabolites and growth in *Arabidopsis thaliana* (Meyer et al., 2007), but it is not clear whether these metabolites are more universally linked to heterosis.

The use of combined transcriptome and metabolome analyses have been successfully utilized to identify novel genes and enzymes in plant metabolite biosynthesis pathways (Boke et al., 2015; Sumner et al., 2015). In addition, this approach has also provided insights into pathways that are affected by gene mutation (Etalo et al., 2013; Masclaux-Daubresse et al., 2014; Page et al., 2016; Satou et al., 2014), that are related to responses to biotic stresses (Gurkok et al., 2015; Liu et al., 2016), and abiotic stresses (Bielecka et al., 2014; Hamaishi et al., 2015).

Together, transcriptome and metabolite profiling provide complementary experimental evidence to guide the construction of rational hypotheses for the biochemical basis of variation in growth. The goal of this study is to identify the metabolic and transcription bases for the growth differences between the two *E. salsugineum* accessions. We utilized contrasting genotypes to

identify genetic differences, expression divergence and metabolic compounds associated with observed phenotypic variation. We obtained gene expression and metabolite concentration data from SH and YK accessions during the dark period by RNA-seq and gas chromatography mass spectrometry (GC/MS), respectively. We utilized RT-PCR to identify DEGs implicated by our RNA-seq experiment, validating 23 of 25 candidates. By mining the RNA-seq experiment, we validated previously identified SNPs and discovered additional SNPs that differentiate these *E. salsugineum* accessions. We propose that observed differences in metabolite accumulation could contribute to differences in biomass.

## Materials and Methods

### Plant material and growth conditions

Seeds of *E. salsugineum* SH and YK accessions were obtained from individual selfed plants. To generate SH x YK F<sub>1</sub> seed, closed flower buds of YK plants were manually opened, anthers removed, and pollen from SH plants was applied to the stigma. Multiple crosses (ca. 20) were made on a single plant, which was kept in isolation until seed set. Seeds from this plant were used for F<sub>1</sub> experiments.

Seeds were stratified in the dark for 10 days at 4°C then sown in a 4:1 mix of Fafard 52 Mix (transcriptome and first metabolome experiments) or Fafard 2 Mix (second metabolome experiment) soilless media (Conrad Fafard, Inc., Agawam, MA, USA) and Turface calcined clay (Profile Products LLC., Buffalo Grove, IL, USA). Plants were grown in 50 mL conical tubes (USA Scientific, Inc., Ocala, FL, USA). A 0.5 cm hole was drilled at the bottom of each tube where seeds were sown. Tubes were then closed with a mesh cap and placed cap-side-down to allow for sub-irrigation. Tubes were placed in a mist room for 12 days, at which point seedlings had reached the 4-leaf stage. Plants were then grown in a growth chamber (E15; Conviron, Pembina, ND, USA) set at 60% relative humidity under a 12 h photoperiod (approximately 230  $\mu\text{mol m}^{-2} \text{s}^{-1}$  provided by both fluorescent and incandescent bulbs) with light and dark temperatures of 22 and 20°C, respectively. Irrigation water was a 3:1 mix of two water-soluble fertilizers (15N–2.2P–12.5K and 21N–2.2P–16.6K; The Scotts Co., Marysville, OH, USA) to supply (in  $\text{mg L}^{-1}$ ): 200 N, 26 P, 163 K, 50 Ca, 20 Mg, 1.0 Fe, 0.5 Mn and Zn, 0.24 Cu and B, and 0.1 Mo. 76% of the nitrogen was provided as nitrate. 93% sulfuric acid (Brenntag North America, Inc., Reading, PA, USA) at 0.08  $\text{mg L}^{-1}$  was mixed in irrigation water to maintain pH between 5.8 and 6.2. All plants were grown in the Purdue University Horticulture Plant Growth Facility (<https://ag.purdue.edu/hla/Hort/Greenhouse/Pages/Default.aspx>).

### Transcriptome sequencing and analysis

For RNA extraction, tissue was collected in the middle of the dark period. Four biological replicates consisting of five plants each were collected. Whole rosettes of 4-week-old plants were frozen in liquid nitrogen and stored at –80°C. Tissue was finely ground in liquid nitrogen using a mortar and pestle. Approximately 800 mg ground tissue was combined with 1 ml of TRIzol reagent (Gibco/BRL Life Technologies, Invitrogen, Carlsbad, CA, USA) and RNA was extracted according to the manufacturer's instructions. Quality of RNA was estimated by the ratio of absorbance at 260 to 280 nm and 260 to 230 nm by spectrophotometer (DU 730; Beckman

Coulter, Inc., Indianapolis, IN, USA) with both ratios between 1.8 and 2.2. RNA was reverse-transcribed into cDNA using the Poly(A) Purist protocol (Ambion, Austin, TX, USA).

The cDNA samples were fragmented into 300–500 bp molecules and sequencing libraries were constructed for the 454 GS-FLX instrument (454 Life Sciences, Roche Company, Branford, CT, USA) at the Purdue University Genomics Facility (West Lafayette, IN, USA). One sequencing run was performed (Margulies et al., 2005). Raw sequences were trimmed to remove adaptor sequences, and an initial quality trimming was performed using GS De Novo Newbler (v2.5.3; default parameters) (Part C – GS De Novo Assembler, GS Reference Mapper, SFF Tools, 2010). Trimmed reads were further trimmed using the FASTX-Toolkit (Gordon and Hannon, 2010), with a minimum quality value of 12 and minimum read length of 50 bp. Trimmed reads were aligned to the Joint Genome Institute (JGI) *E. salsugineum* genome (Phytozome v9.1: *Thellungiella halophila*: <http://www.phytozome.net/thellungiella.php>; Yang et al., 2013) using the splice-aware aligner GMAP v2012-11-27 with default parameters (Wu and Watanabe, 2005). Only uniquely mapped reads were used for transcript abundance estimates and single nucleotide polymorphism (SNP) calling.

SNPs were detected using mpileup from SAMtools v0.1.18 with mapping quality  $\geq 15$ , and depth  $\geq 3$  (Li et al., 2009). 454 sequencing has a high error rate for detecting indels (Margulies et al., 2005), so only SNPs resulting from substitutions were retained. The two accessions are substantially inbred lines and should be homozygous at each base position. Hence, only monomorphic base positions within each accession were considered for detection of differences between the two accessions. Custom Perl scripts were used to remove SNPs 1) that were heterozygous within either accession, 2) that were not biallelic between accessions, 3) that were supported by fewer than 3 sequence reads, 4) for which the alternative allele accounts for fewer than 10% of aligned reads, and 5) that were heterozygous between the SH accession and the JGI SH reference. If more than 4 SNPs were detected within a 100 bp region using the VariantFiltration module from GATK v2.4.9 (McKenna et al., 2010), they were not included in the final SNP dataset. SNPs that had mpileup quality score 999 based SAMtools were deemed “high quality” SNPs. On are Sanger sequencing data of the YK accession available from the National Center for Biotechnology Information (NCBI) (Wong et al., 2005) was also aligned to the reference genome using SSAHA2 (Ning et al., 2001). SNPs were called using SAMtools and filtered for clustered SNPs (4 SNPs within 100 bp region) using GATK as indicated. SNPs that were not biallelic or were heterozygous within YK were removed.

Genes were identified via a reference annotation-based transcript assembly method using the Cufflinks package (Roberts et al., 2011; Trapnell et al., 2010). Reads from SH and YK were assembled separately and then merged using the *cuffmerge* command (Roberts et al., 2011; Trapnell et al., 2010). The *intersect* function within BEDTools v2.17.0 was used to identify genes not annotated in the JGI *E. salsugineum* genome (newly annotated genes). Same-strandedness was not enforced when identifying newly annotated genes because of the non-strand-specific protocol for 454 library preparation. Newly annotated genes that are unique from or overlap genes annotated by Champigny et al. (2013) but are present in the JGI reference genome were also identified using the same method (Supplementary Table 1).

The number of reads uniquely aligned to each gene was determined with htseq-count within

HTSeq v0.5.4p5 (<http://www-huber.embl.de/users/anders/HTSeq/doc/count.html>) using *union* mode. The bioconductor package “DESeq” v.1.14.0 was used to identify genes likely to be differentially expressed between SH and YK without biological replicates (Anders and Huber, 2010). Gene expression was normalized, and the significance threshold for differential expression was based on a 0.2 false discovery rate (FDR) (Benjamini and Hochberg, 1995).

Genes were annotated by the best BLAST (Altschul et al., 1990) hit of *A. thaliana*. For *E. salsugineum* genes predicted by the JGI genome, annotation was taken from Phytozome v9.1 (Phytozome v9.1: *Thellungiella halophila*: <http://www.phytozome.net/thellungiella.php> Yang et al., 2013). For newly identified genes, nucleotide databases of *A. thaliana* version TAIR 10 (Lamesch et al., 2012; TAIR10: [ftp://ftp.arabidopsis.org/home/tair/Sequences/blast\\_datasets/TAIR10\\_blastsets/](ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/TAIR10_blastsets/)), *A. lyrata* (Hu et al., 2011; Phytozome v9.1: *Arabidopsis lyrata*: <http://www.phytozome.net/alyrata.php>), and *S. parvula* (Dassanayake et al., 2011; *Thellungiella parvula* genome: <http://www.thellungiella.org/data>) were built and used for similarity searching by BLASTN v2.2.28+ (Altschul et al., 1990). Genes were annotated by the best BLAST hit with the following threshold parameters:  $E \leq 1^{-30}$ ; sequence identity  $\geq 30\%$ ; sequence aligned  $\geq 30\%$  of query sequence.

## Quantitative real-time PCR

For quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR), tissue was collected as described for 454 sequencing, RNA was extracted using the RNeasy Plant Mini Kit (QIAGEN, Valencia, CA, USA), and genomic DNA was removed using the TURBO DNA-free™ Kit (Ambion, Grand Island, NY, USA). The quality and quantity of mRNA was assessed using a NanoDrop 2000 (Thermo Fisher Scientific Inc., Wilmington, DE, USA). All samples were diluted to 200 ng  $\mu\text{l}^{-1}$ , and 260/280 and 260/230 ratios were between 1.8 and 2.2. cDNA was synthesized using a High-Capacity cDNA Reverse Transcription Kit (Invitrogen, Grand Island, NY, USA). Primers were designed using Primer Express Software (v3.0.1). Primer specificity was then estimated by BLASTN using the *E. salsugineum* genome with all primer pairs. Primer efficiency was tested for all pairs of primers. cDNA was diluted 5 times by a 5-fold gradient, and then used as template for qRT-PCR, the threshold cycles ( $C_T$ ) were regressed against cDNA concentration (log). Slope of the regression line was estimated and the efficiency was calculated as  $10^{-(1/\text{slope})} - 1$ . For genes expressed in both accessions, primer efficiency was between 80 and 110% in both accessions. For genes that were only expressed in one accession based on RNA-seq data, primer efficiency was tested on both accessions, but only the accession with detected expression exhibited efficiency between 80 and 110%. Supplementary Table 2 contains all primer sequences except gene XLOC\_004723, for which acceptable qRT-PCR primers could not be designed.

All qRT-PCR reactions were conducted in StepOnePlus™ Real-Time PCR Systems (Applied Biosystems, Invitrogen, Grand Island, NY, USA). Relative gene expression of target genes was quantified by the  $\Delta C_T$  method (Livak and Schmittgen, 2001). Relative gene expression was calculated as:

$$\text{Relative expression} = (1 + E)^{-\Delta C_T}$$



$$\Delta C_T = C_{T,X} - C_{T,R}$$

where  $E$  is the primer efficiency for each pair of qRT-PCR primers.  $C_{T,X}$  and  $C_{T,R}$  is the threshold cycle of the target gene and the reference gene *Actin2* (Thhalv10020906m.g), respectively.

## Metabolite profiling and data analysis

Two metabolite profiling experiments were conducted. One experiment was performed using the same rosette tissue used for RNA-seq analysis (see above). A second metabolite profiling experiment was performed using tissue from SH, YK, and YK x SH  $F_1$  plants with 3 replicates of 5 pooled plants per replicate. In both two cases, identical extraction, derivatization, and analysis methods were used. Approximately 800 mg of frozen ground tissue was incubated in methanol at 65°C in 1.75 ml tubes and centrifuged at 13,300  $r\ min^{-1}$ . The supernatant, containing polar molecules, was decanted into a new tube. Chloroform was added to the pellet and incubated at 37°C for 15 min to solubilize nonpolar metabolites. Samples were then dried at room temperature for about 6 h (polar) and 2 h (nonpolar) in a centrifuge at 1725  $r\ min^{-1}$  and 30  $\mu M$  Hg vacuum. Samples were stored at -80°C until being sent to the Metabolomics Center at the University of Illinois (<http://www.biotech.uiuc.edu/metabolomics/>). Samples were measured by gas chromatography/mass spectrometry (GC/MS; Agilent 6890 N/5973 MSD, Palo Alto, CA, USA) after purification and trimethylsilylation with MSTFA (*N*-methyl-*N*-trimethylsilyl-trifluoroacetamide) (Gullberg et al., 2004; Singh et al., 2011). Data was analyzed by peak identification via comparison with spectra from standards and relative concentrations of metabolites were obtained by comparison with internal standard peak area (Singh et al., 2011)..

Pair-wise comparisons within each experiment were performed by two-tailed  $t$ -tests between SH and YK (experiments 1 and 2), and SH or YK and YK x SH  $F_1$  (experiment 2). In the second experiment, multiple comparisons among all three genotypes were conducted by Tukey's studentized range test (Tukey, 1949). A  $t$ -test of  $F_1$  against the mid-parental average of SH and YK was done using the variance estimated from  $F_1$  hybrids. Within each experiment, genotype was treated as the only main factor. A nested analysis of variance (ANOVA) was also conducted using data from the two experiments. In the nested analysis, experiment and genotype were the two main factors. The experiment by genotype interaction was not included in the model. All identified metabolites are presented in Supplementary Table 3.

## Pathway analysis

Pathway analysis was conducted on the annotated DEGs identified in transcriptome profiling and metabolites that differed between accessions in both metabolite experiments with the same direction. The fold differences between SH and YK was used to indicated up-regulation or down-regulation in YK compared to SH. The list of DEGs or metabolites with the fold changes were imported to MapMan software (Thimm et al., 2004). Significant pathways in which genes or metabolites were divergent from a 50/50 up/down-regulation were identified using an uncorrected Wilcoxon signed-rank (Wilcoxon, 1945).

## Results

### Novel genes and single nucleotide polymorphisms (SNPs) were identified by transcriptome sequencing

Whole rosettes of 4-week-old YK and SH plants grown in a 12h:12h light dark cycle were harvested in the middle of the dark period. Libraries of cDNA isolated from these rosettes were sequenced by pyrosequencing. These reads were aligned to the *E. salsugineum* reference genome, which was constructed using DNA from the SH accession (Yang et al., 2013). More than 1 million cDNA sequence reads, 95% of which aligned to the reference genome, were used for a reference-directed assembly of the transcriptome (Table 1), identifying 17,888 expressed genes (Table 2 and Supplementary Table 1). Of these, 65 genes were novel and not predicted in the reference genome (Yang et al., 2013) nor detected in a previous transcriptome analysis conducted by Champigny et al. (2013). Only 20 of these 65 genes have annotated orthologs in the related species *A. thaliana*, *A. lyrata*, and/or *Schrenkiella parvula* (Supplementary Table 1). Presence-absence variation (PAV), defined as zero reads aligned to one of the two parents, was observed for 18.5% of the detected expressed genes with roughly equal numbers of genes detected only in SH or YK (Table 2).

The transcript assemblies were processed to detect SNPs between SH and YK. 42% of shared genes contained a total of 14,808 SNPs, of which 4,873 were deemed “high quality” (Table 3; Supplementary Table 4). We compared our results with the SNP detection results of Champigny et al. (2013). Of the SNPs detected in our experiment, 73% (10,883 positions) and 79% (3861 positions) of the low and high stringency SNPs, respectively, were also identified by Champigny et al. (2013). We also compared our SNPs to available Sanger sequencing of cDNA clones from the YK accession (Wong et al., 2005) and identified 468 putative SNPs with reference to the SH reference genome. Of these, 441 have corresponding sequence data in the YK transcriptome we assembled and 88% (388 SNPs) had the same sequence variation in our assembly and the YK Sanger sequencing data (Supplementary Table 4).

### Transcriptome profiling and qRT-PCR identify differentially expressed genes (DEGs) between SH and YK

To assess gene expression differences between SH and YK, we determined the number of reads aligned to each gene. Less than 0.2% (thirty-one genes) of the expressed genes were identified as candidate DEGs between SH and YK (Table 4). Sixteen of these thirty-one candidates do not have homologous genes in *A. thaliana*, *A. lyrata*, or *S. parvula* and twenty have been annotated in the reference genome (Table 4). Of those with homologs in one of these species, none have been associated with previously reported trait differences in SH and YK.

To confirm DEGs based on RNA-seq, we designed quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR) primers that matched unique positions in the reference genome based on BLAST analysis (Altschul et al., 1990). Unique primers could not be designed for five genes (Thhalv10022994m.g, Thhalv10022932m.g, Thhalv10014933m.g, Thhalv10019398m.g, and Thhalv10029246m.g) due to paralogs with high sequence similarity



and no acceptable primer pair was identified for XLOC\_004723. In total, qRT-PCR data confirmed our RNA-seq data for 23 of 25 genes. For the 5 DEGs that had no close paralogs and at least 4 reads in both SH and YK, expression differences based on RNA-seq were confirmed by qRT-PCR in 4 of 5 genes (Table 4). Among the 20 genes that had fewer than 4 reads in either SH or YK, expression differences by qRT-PCR was consistent with RNA-seq data in 19 (Table 4). In addition, when the low accession had fewer than 4 reads there was no amplification in 16 of 20 cases (Table 4).

### **Metabolite profiling reveals higher accumulation of fatty acids and amino acids in YK and enhanced soluble carbohydrate accumulation in SH**

To identify potential metabolites and metabolic pathways that contribute to phenotypic and physiological differences between SH and YK, metabolite profiling was conducted in two independent experiments. In one experiment, metabolite concentrations in F<sub>1</sub> plants of a YK x SH cross were also determined.

Concentrations of free fatty acids and long-chain fatty acid derivatives were higher in YK than SH (Table 5 and Supplementary Table 3). The concentration of ferulic acid was also greater in YK than SH, indicating a potential for greater suberin and/or cutin accumulation in the YK accession.

The products of starch degradation were more abundant in SH than YK in both profiling experiments (Table 5). Maltose and glucose, primary products of starch degradation, were elevated in SH along with fructose, glycerol-3-phosphate, raffinose, and an unresolved disaccharide. This suggested more active 6-carbon metabolite catabolism via glycolysis in SH compared to YK during the night (Table 5 and Supplementary Table 3). SH accumulated higher concentrations of the disaccharides isomaltose and gentiobiose while YK accumulated a very small amount of melibiose in one of the two screens. Tricarboxylic acid (TCA) cycle intermediates, on the other hand, were not consistently different between the accessions. Citric acid and fumaric acid were accumulated at higher levels in YK while alpha-ketoglutaric acid was greater in SH (Table 5 and Supplementary Table 3). Ascorbic acid was higher in SH but the ascorbic acid degradation product tartaric acid (2,3 dihydroxybutanedioic acid, Table 5 and Supplementary Table 3) was higher in YK.

All amino acids that differed between SH and YK were higher in YK (Table 5 and Supplementary Table 3). These included alanine, glycine, serine, threonine, and valine. Of these, only threonine was differentially accumulated in both profiling experiments. The amino acids alanine, glycine, and serine were only detected in one of the two experiments. Valine was detected in both experiments, but the difference between SH and YK was only statistically significant in the first experiment. Concentrations of the other detected amino acids (isoleucine, leucine, and proline) were not different in the two accessions.

Pathway analyses were conducted on both transcriptome (Table 4) and metabolome data (Table 5) to identify differences between SH and YK during the dark period and any correlation between transcriptome and metabolome analyses of the two accessions. The metabolite pathway analysis identified two functional categories: major carbon degradation (down-regulated in YK)

and amino acid synthesis (up-regulated in YK). Glucose, maltose, and isomaltose are within the major carbon degradation pathway, while aspartic acid, alanine, N-acetylglutamic acid, threonine, serine, and glycine are within the amino acid synthesis pathway. However, of the 19 DEGs annotated in the reference genome, one unknown functional category including nine DEGs (Thhalv10015083m.g, Thhalv10002969m.g, Thhalv10023491m.g, Thhalv10021522m.g, Thhalv10018393m.g, Thhalv10009345m.g, Thhalv10015718m.g, Thhalv10022932m.g, and Thhalv10022943m.g) and one category related to cell organization of two DEGs (Thhalv10029390m.g and Thhalv10029346m.g) were identified. Hence, our analyses did not show any correlation between the *E. salsugineum* transcriptome and metabolome.

## Metabolomic phenotypes in the F<sub>1</sub> hybrid exhibit transgressive heterosis

To better understand the genetic basis for observed differences in metabolite concentrations between SH and YK, we measured metabolites in F<sub>1</sub> hybrids. Concentrations of all measured metabolites in both experiments are presented in Supplementary Table 3. Of the 144 metabolites measured in the experiment comparing the two parents and hybrids, 56% of all metabolites were lower in hybrids than the predicted mid-parent value. This included 65% of the fatty acids and 49% of carbohydrates detected (Supplementary Table 3). We performed a two-way contingency test to determine if an observed difference in the accumulation of a metabolite was predictive of heterosis for that metabolite (Supplementary Table 5). We found that metabolites with accumulation differences between the parents were neither more nor less likely to exhibit accumulation differences between the F<sub>1</sub> and the mid-parental values (Supplementary Table 5).

Hybridization can result in transgressive heterosis in which phenotypic values for the hybrids fall outside of the range of parental values. Of the 144 metabolites measured, transgressive heterosis was observed for 28. Of these, 4 metabolites were not observed in one of the two parents and 24 were detected in both parents and the hybrids (Figure 1). Of the 24 detected in all three genotypes, the transgressive heterosis was more likely to be negative (Figure 1;  $P$ -value  $\leq 0.0001$  based on Binomial exact test) and more frequently affected metabolites that did not differ in concentration between the parents (Supplementary Table 5;  $P$ -value  $\leq 0.05$  based on  $\chi^2$  test). Thus, heterosis for the metabolome was manifested by a decrease in metabolite pool sizes in hybrids and was not preferentially associated with metabolites that contributed to the variation between the two parents. This is consistent with our observation of decreased availability of primary metabolites in the faster growing YK (Table 5 and Supplementary Table 3). We propose that a metabolic consequence of enhanced growth is a reduction in pool sizes of primary metabolites and greater resource utilization for anabolic metabolism.

## Discussion

In this study, we profiled both transcript and metabolite accumulation to identify genetic and biochemical variation during the dark period in two *E. salsugineum* accessions: SH and YK. We annotated novel genes present in the reference genome and identified DEGs between SH and YK in the middle of the dark period. We found that YK accumulates more fatty acids than SH, while SH accumulates sugars at higher concentrations. Although the transcriptomic and metabolic profiling results do not offer links to each other, they do offer insight into genetic and physiological differences between these accessions. Furthermore, we identified additional SNPs,

and provide validation of previously described SNPs, varying between these accessions that can be utilized for future research.

### Validity of identified SNPs

Based on the predicted transcriptome size (Yang et al., 2013), the SNP density from our analysis is 1 SNP per 10 kb of transcribed sequence. 14,887 SNPs identified from our transcriptome sequencing data were also present in Champigny et al. (2013). However, our SNP cluster filter, which removes neighboring SNPs to account for misaligned reads at insertion-deletion polymorphisms, removed 3,604 overlapping SNPs. The removal of this set of SNPs, likely enriched for false positives, contributed to the lower SNP number detected in our study. We also acknowledge that this contributed to our false negative rate. Of the 388 SNPs that matched between our analyses of pyrosequencing data and the Sanger data, 99 were removed by our procedure. We provide a compact file containing the subset of “high-quality” (see Materials and Methods) SNPs in the supplementary data. These have a SNP density of 1 SNP per 25 kb of transcript (Supplementary Table 4). This set of validated SNPs can be used for QTL mapping and fine-mapping studies (Matsuda et al., 2015; Trick et al., 2012; Wu et al., 2016).

### Identification of genes expressed in the dark period in *E. salsugineum*

We provided expression data supporting 63% and 67% of the predicted genes in the two published reference genomes (Wu et al., 2012; Yang et al., 2013). This is likely an underestimate of the expressed genes in this species because we sampled only leaf tissue and only at night (Birnbaum et al., 2003; Czechowski et al., 2004; Doherty and Kay, 2010; Huang et al., 2012b). We identified 66% of the genes identified in another transcriptome analysis of *E. salsugineum* (Champigny et al., 2013). Despite our study relying on lower coverage (4X vs. 8X), and quantifying expression of genes only expressed in leaves at night, we annotated 65 genes not predicted in the reference genome or identified in the previous transcriptome characterization (Table 2 and Supplementary Table 1). Novel genes identified in this transcriptome study but not that of Champigny et al. (2013) are likely either only expressed at night or otherwise not expressed under the conditions of the previous study which included a 21 hr day and low dark period temperatures.

More than 97% of the genes identified in our study are homologous to genes in *A. thaliana* (Supplementary Table 1), consistent with the relatively close phylogenetic association of the two species (Al-shehbaz et al., 1999). Of the 65 newly-annotated genes in this study, 47 of them do not have homologs in *A. thaliana*. One of these novel genes has been annotated in *A. lyrata* and two have been annotated in *S. parvula* (Supplementary Table 1). Therefore, we may have identified genes that are unique to this extremophile species (Wu et al., 2012).

More than 80% of the expressed genes identified in this study were detected in both SH and YK (Table 2). This indicates a conserved transcriptome between these two accessions and is consistent with previous *E. salsugineum* transcriptome comparisons (Champigny et al., 2013; Lee et al., 2013). PAVs, as scored by read count in the transcript profiling experiment, accounted for 19 and 12% of all genes in this experiment and Champigny et al. (2013), respectively. Only 13% of these were consistently detected as PAV over the two datasets. These genes are likely

“true” PAV genes (Supplementary Table 1). Other genes detected in only one of the replicates were likely due to low mRNA abundance. PAV structural variation has been observed in *A. thaliana* (Bush et al., 2014), maize (Springer et al., 2009; Swanson-Wagner et al., 2010), and soybean (Haun et al., 2011). PAV genes are not typically essential (Bush et al., 2014) and may have minor effects on plant fitness (Swanson-Wagner et al., 2010). However, genes present in only one accession could contribute to the adaptation to specific selective constraints (Bush et al., 2014) and variation in quantitative traits (Swanson-Wagner et al., 2010). The observed phenotypic variation in growth rate and metabolism could be due to PAV, although no identified PAVs have been linked to trait differences in *Eutrema*. Further study utilizing molecular genetics to address the causes and consequences of natural variation in this species is needed to link these data types.

## DEGs and constitutive response to abiotic and biotic stresses

The low number of candidate DEGs between the two accessions of *E. salsugineum* used in this study was similar to other studies. 55% of our these had no reads in one of the two accessions, which could be due to a lack of sufficient depth to detect very low expression. Champigny et al. (2013) identified 381 DEGs that were present in our transcriptome but not identified as DEGs in our data. RNA-seq experiments are often underpowered, and the lack of overlap with the previous study may be primarily an issue of low read depth and replicate numbers. Consistent with this expectation, 58% of the 381 DEGs identified previously have low expression (<4 reads; Supplementary Table 1). In addition, 41% of the DEGs from Champigny et al. (2013) exhibit lower expression during the dark period in *A. thaliana* (Mockler et al., 2007), which may explain some of the lack of overlap in DEGs in the two studies. We identified 17 DEGs that were not identified by Champigny et al. (2013), possibly because they are only differentially expressed during the dark period or under our experimental conditions.

It appears that the relative abundance of most transcripts are similar in SH and YK, since only 0.2 and 1.9% of all genes were DEGs in our study and Champigny et al. (2013), respectively. Fourteen of the 31 DEGs identified in our study were also identified by Champigny et al. (2013) and the expression patterns were the same in both studies. This suggests that differential expression for these genes between SH and YK is consistent over light and dark periods and the two growth conditions.

Without exposure to abiotic or biotic stresses, YK expressed several stress-responsive genes (Table 4). This was also noted by Champigny et al. (2013). Two plant defensin genes within the same family were highly expressed in YK (Table 4). In *A. thaliana*, the plant defensin type 1 family (*PDF1*) is comprised of seven genes (Shahzad et al., 2013) with highly conserved sequences and identical mature peptides (Thomma et al., 2002). *AtPDF1* genes are induced by pathogens, non-host pathogens, methyl jasmonate (MeJA), and ethylene (ET) (De Coninck et al., 2010; Manners et al., 1998; Penninckx et al., 1996; Zimmerli et al., 2004). Also, expression of *AtPDF1s* in yeast results in zinc tolerance (Shahzad et al., 2013). Eight gene models in *E. salsugineum* are homologous to the *A. thaliana* plant defensin family. Of these, one gene is annotated as the likely ortholog of *AtPDF1.4*, whereas five are annotated as most closely related to *AtPDF1.2A*, and two are annotated as *AtPDF1.2C*. Higher expression of *PDF1.2* and *PDF1.4* was observed in the YK accession, however, paralogous *PDF1* genes are expressed in SH



(Supplementary Table 1).

The transcript abundance of a gene encoding *E. salsugineum* peptide methionine sulfoxide reductase 3 (*PMSR3*) was higher in YK than SH (Table 4). There are five orthologous *PMSR* genes in *A. thaliana*: *PMSR1* to 5 (Rouhier et al., 2006) that are also found in *E. salsugineum*. The expression of *PMSR3* is induced by arsenate (Paulose et al., 2010), suggesting the ability to respond to plant abiotic stresses. No function in tolerance or resistance has been established for this paralog in *A. thaliana*. However, knock-out of either *PMSR2* (Bechtold et al., 2004) or *PMSR4* (Romero et al., 2004) results in decreased oxidative stress tolerance and overexpression of either gene increases stress tolerance in *A. thaliana*. Expression of *PMSR4* (but not *PMSR1*, *PMSR2*, or *PMSR3*) was induced in response to UV and AgNO<sub>3</sub> in *E. salsuginea* SH (Mucha et al., 2015). It is plausible that the overexpression of *PMSR3* by YK could provide greater oxidative stress tolerance in this accession.

### ***E. salsugineum* accessions SH and YK differ in carbon metabolism**

In two metabolic profiling experiments, 125 and 144 metabolites were detected. Differences were identified between SH and YK for the 85 metabolites detected in both experiments (Table 5 and Supplementary Table 3). The derivatization method we utilized has been widely used to detect sugars (Gullberg et al., 2004), but is less accurate for identifying and quantifying amino acids (Kaspar et al., 2009). As a result, despite the fact that *Eutrema* accumulates higher concentrations of some amino acids than *Arabidopsis* (Eshel et al., 2017), many amino acids were not detected in our analyses.

Higher concentrations of fatty acids and fatty acid derivatives were measured in YK, including several previously identified as structural components of membrane lipids, cuticle components, and wall-resident suberin (Table 5 and Supplementary Table 3). Fatty acids contain more energy than carbohydrates when used as storage compounds and can act as an efficient storage form of reduced carbon (Taiz and Zeiger, 2010). High production of fatty acids is typical of rapidly growing tissues (Ohlrogge and Jaworski, 1997; Qin et al., 2007). Very-long-chain fatty acids (VLCFAs; C20:0 to C30:0) play an important role in cell elongation and expansion (Qin et al., 2007). Several VLCFAs, including docosanoic, hexacosanoic, pentacosanoic, tetracosanoic, and tricosanoic acids were more abundant in YK than SH (Table 5), consistent with our measurements of higher growth rates of YK as compared to SH in our growth conditions (manuscript in preparation). The VLCFA tetracosanoic acid, which plays an important role in root cell growth and expansion (Qin et al., 2007), was accumulated at a higher concentration in YK (Table 5 and Supplementary Table 3). In addition to carbon storage, lipids are important components of membranes and the leaf cuticle (Lynch and Dunn, 2004; Tresch et al., 2012; Zäuner et al., 2010). Our results are in agreement with Xu et al. (Xu et al., 2014), who measured a greater accumulation of C22 and C24 fatty acids in the epicuticular wax of YK over SH.

Overall, SH tissues had higher concentrations of sugars than YK. These measurements of nighttime sugar concentrations were similar to previous results obtained for fructose, glucose, and raffinose in leaves of *E. salsugineum* harvested during the day (Lee et al., 2012; Eshel et al., 2017). The products of starch breakdown, including maltose and glucose, were more abundant in SH (Table 5 and Supplementary Table 3), suggesting a higher rate of starch metabolism in the



lower-biomass SH accession This was consistent with the strong negative correlation between starch content and biomass observed in *A. thaliana* accessions (Sulpice et al., 2009).

A study of the correlation between specific metabolites and biomass accumulation in *A. thaliana* revealed twenty-three metabolites that were correlated with biomass (Meyer et al., 2007). We detected fourteen of these twenty-three metabolites (Table 5 and Supplementary Table 3). Of these, five differed between SH and YK. The concentrations of ascorbic acid, glycerol-3-phosphate, and raffinose were negatively correlated and putrescine was positively correlated with biomass in *Arabidopsis* (Meyer et al., 2007). The levels of these metabolites also corresponded to the differences in biomass between SH and YK, indicating that the relationships between metabolites and biomass found in *A. thaliana* were consistent in *E. salsugineum* (Table 5 and Supplementary Table 3). This suggests that, although stress tolerance is vastly different between these two species (Amtmann, 2009; Griffith et al., 2007; Lee et al., 2012; Volkov et al., 2003), the metabolic markers for biomass accumulation may be similar.

### Increased utilization rate as a hypothesis for metabolome heterosis

More than 58% of the metabolites in  $F_1$  plants were different from the predicted mid-parent concentration (Supplementary Table 3), indicating a non-additive effect of hybridity on the majority of the metabolome. The lower concentration of fructose and glucose in  $F_1$  hybrids suggests high rates of starch and sugar depletion to support rapid growth (Lisec et al., 2011). Transgressive heterosis was more commonly observed for metabolites that were not different between the two parents (Supplementary Table 5). This suggests that allelic variation affecting differential metabolite accumulation in the parents is not responsible for the observed heterosis in the  $F_1$  metabolome. Although it is surprising that differences between the parents were not predictive of a metabolite association with heterosis, it may be that the metabolomic consequences of heterosis derived from secondary effects of an increased growth rate in  $F_1$  hybrids, rather than a causative relationship between growth rate and specific metabolites or metabolite diversity. Differences in biomass polymers and metabolites involved in anabolic growth exhibited reduced pool sizes in the more rapidly growing YK as compared to SH (Table 5 and Supplementary Table 3), as well as in the very rapidly growing  $F_1$  plants as compared to the parents (Figure 1 and Supplementary Table 3). Consistent with the hypothesis that utilization rate drives the heterotic effects on metabolite pool sizes, the transgressive effect overwhelmingly resulted in lower concentrations of metabolites in the hybrids (Figure 1 and Supplementary Table 3 and 5). This hypothesis regarding the cause of metabolic heterosis may be a general phenomenon in plants. Indeed, the same associations have been previously observed in maize, in which largely negative overdominance for metabolites was found in the heterotic B73 x Mo17 hybrids (Lisec et al., 2011).

### Conclusions

Our study contributes to the annotation of the *E. salsugineum* genome and provides evidence of transcriptional and metabolic differences between the SH and YK accessions. Very few differences in gene expression were detected in the middle of the dark period between these two accessions, but YK has constitutively higher expression of several plant systematic defense genes. The high-quality SNPs identified in this study can be used with previously identified

SNPs to map traits that differ in these accessions, such as tolerance to various stresses. There is evidence for contrasting carbon metabolism in these two accessions, which correlates with observed growth differences. Furthermore, metabolite profiling of the accessions and F<sub>1</sub> hybrids supports the notion that the concentrations of key metabolites are correlated with growth rate, including the increased growth rate caused by heterosis.

Our hypothesis was that combined transcriptome and metabolome profiling of two contrasting *E. salsugineum* accessions might elucidate the pathway(s) related to the phenotypic differences between these contrasting accessions. The difference carbon metabolism identified in the metabolome profiling provides insights for growth differences between SH and YK. However, none of the 19 DEGs that have been annotated in the reference genome are related to the observed metabolic differences. There are two plausible explanations: 1) the additional 11 DEGs that are currently unannotated in the reference genome could provide additional evidence for the link between metabolome and transcriptome, or 2) by increasing the number of replicates in transcriptome study, more DEGs will be identified to support further pathway identification.

### Availability of supporting data

All the raw data supporting the results of this article have been deposited at (Edgar et al., 2002) are accessible through GEO Series accession GSE GSE67745 at (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE67745>).

### Abbreviations

ABA: Absciscic acid; DEGs: Differentially expressed genes; ESTs: Expressed sequence tags; ET: Ethylene; FDR: False discovery rate; GC/MS: Gas chromatography mass spectrometry; JA: Jasmonic acid; MeJA: Methyl jasmonate; PAV: Presence-absence variation; PDF: Plant defensin gene; PMSR: Peptide methionine sulfoxide reductase; QTL: Quantitative trait loci; SH: *E. salsugineum* Shandong accession; SNP: Single nucleotide polymorphism; VLCFA: Very long-chain fatty acids; YK: *E. salsugineum* Yukon accession.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

JY and MJG performed the experiments; JY, BPD, and MVM designed the experiments; JY, and BPD analyzed the data; JY, BPD, and MVM wrote the manuscript.

### Acknowledgements

Partial support to JY was provided by the U.S. Department of Agriculture, National Institute of Food and Agriculture – Agriculture and Food Research Initiative Grant no. 2010-85117-20607, the China Scholarship Council, and the Purdue Department of Horticulture and Landscape

Architecture. We thank Dr. Scott J. Emrich and Dr. Allison Regier (current address: Genome Institute, Washington University) in the Department of Computer Science and Engineering at University of Notre Dame and Dr. Michael Gribskov in the Department of Biological Sciences for assistance and discussion of RNA-seq data analysis. We thank Dr. Chan yul Yoo (current address: Biology Department, Duke University) for the assistance of RNA-seq sample collection and preparation. We thank Rob Eddy and Dan Hahn for assistance with plant growth, Dr. Phillip San Miguel and the Purdue University Genomics core for RNAseq data production and transcriptome analysis, Dr. Alexander Ulanov and Dr. Zhong (Lucas) Li at the Metabolomics Center at the University of Illinois for assistance with the metabolomic data, and Prof. David Rhodes for discussions regarding the metabolite data.

## Figures

**Figure 1.** Heterosis for metabolite concentrations in F<sub>1</sub> hybrids. Metabolites in F<sub>1</sub> hybrids that were higher than the high parent or lower than the low parent at  $P\text{-value} \leq 0.05$  based on two-tailed  $t$ -test are shown. The ratio is calculated as F<sub>1</sub>/high parent when F<sub>1</sub> had a higher concentration than the high parent and low parent/F<sub>1</sub> when F<sub>1</sub> had a lower concentration than the low parent.

## Supplementary data

**Supplementary Table 1. All genes identified in *E. salsugineum* Shandong (SH) and Yukon (YK) accessions.** Information for all identified genes includes a gene identification number (Gene\_id) as annotated in the JGI reference genome (Thhalv) or newly annotated (XLOC), the number of raw reads in Shandong (SH) or Yukon (YK) genotypes, the estimated fold change (foldChange), the  $P$ -value without (pval) and with (padj) FDR correction, gene position on the JGI reference genome (JGI scaffold, start, and end), homologous gene identification number for *Arabidopsis thaliana* (*A. thaliana* ID), *Arabidopsis lyrata* (*A. lyrata* ID), and *Schrenkiella parvula* (*S. parvula* ID), and gene name (Gene name) based on *A. thaliana* ID. Genes identified in both this study and 1) all plants (Shared with Champigny et al. 2013), 2) chamber-grown plants (Shared with Champigny et al. 2013 cabinet-grown plants), and 3) field-grown plants (Shared with Champigny et al. 2013 field-grown plants) are indicated. Genes identified as “True PAVs” were identified as PAVs in both Champigny et al. 2013 and our study with at least 1 read in SH but no read in YK (Yes; SH only) or at least 1 read in YK but no read in SH (Yes; YK only). Genes not identified as PAVs in either study are indicated (No).

**Supplementary Table 2.** Primer sequences used for qRT-PCR. Gene name (Name) and primer sequence (Sequence) for all genes used to test transcriptome expression differences by qRT-PCR.

**Supplementary Table 3.** All metabolites detected in *E. salsugineum* accessions Shandong (SH), Yukon (YK), and YK x SH F<sub>1</sub> hybrids. Data are means (n=4). Empty indicates absence or undetected metabolites. Genotypes are significantly different at  $P\text{-value} \leq 0.05$  (\*), 0.01 (\*\*), and 0.001 (\*\*\*) based on two-tailed  $t$  tests in Experiment 1 (E1), Experiment 2 (E2), or nested (N) analysis of the two experiments. In Experiment 2 (E2), S v Y indicates comparison of SH versus YK; F1 v S indicates comparison of YK x SH F<sub>1</sub> hybrids versus SH; F1 v Y indicates

comparison of YK x F1 hybrids versus YK; F1 v M indicates comparison of YK x SH F1 hybrids versus mid-parent; Nested indicates comparison of SH versus YK for two experiments combined together. Genotypes are significantly different at P-value  $\leq 0.05$  (\*), 0.01 (\*\*), and 0.001 (\*\*\*) based on two-tailed t test. Different letters within the same line indicate significant difference between genotypes at P-value  $\leq 0.05$  based on Tukey's studentized range test. Carbohydrates indicate sugar-related compounds and derivatives in citric acid cycle. Fatty acids indicate fatty acids and their derivatives. Others indicate all other compounds not belong to carbohydrates or fatty acids, or amino acids.

**Supplementary Table 4. Biallelic SNPs between *E. salsugineum* Shandong (SH) and Yukon (YK) accessions.** SNPs between *E. salsugineum* Shandong (SH) and Yukon (YK) identified by our transcriptome sequencing. SNP location is indicated according to the JGI *E. salsugineum* genome by scaffold and position within the scaffold, as well as the gene locus in which the SNP was identified. Reference and alternative alleles are indicated. High quality SNPs, based on our criteria described in the methods section are indicated. SNPs that were identified in this experiment and Champigny et al. 2013 (Shared with Champigny et al. 2013) and/or existing Sanger data (Present in Sanger data) are indicated.

**Supplementary Table 5. Heterosis effect of metabolites in YK x SH F1 hybrids of *E. salsugineum*.** Metabolites that were different between F1 and mid-parent showed no enrichment in those were different between the two parents based on a  $\chi^2$  test. Metabolites in F1 hybrids that showed transgressive heterosis were enriched in the metabolites with no variations between the two parents at P-value  $\leq 0.05$  (\*) based on a  $\chi^2$  test. The proportion of negative heterosis metabolites (23/24) is significantly different from the expected probability of 0.5 based on a Binomial exact test at P-value  $\leq 0.0001$  (\*\*\*)

## References

- Al-shehbaz, I. A., O’Kane, S. L., Jr., and Price, R. A. (1999). Generic placement of species excluded from *Arabidopsis* (Brassicaceae). *Novon* 9, 296–307.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Amtmann, A. (2009). Learning from evolution: *Thellungiella* generates new knowledge on essential and critical components of abiotic stress tolerance in plants. *Mol. Plant* 2, 3–12. doi: 10.1093/mp/ssn094
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106.
- Bazakos, C., Manioudaki, M. E., Therios, I., Voyiatzis, D., Kafetzopoulos, D., Awada, T., et al. (2012). Comparative transcriptome analysis of two olive cultivars in response to NaCl-stress. *PLoS One* 7, e42931. doi:10.1371/journal.pone.0042931
- Bechtold, U., Murphy, D. J., and Mullineaux, P. M. (2004). Arabidopsis peptide methionine sulfoxide reductase2 prevents cellular oxidative damage in long nights. *Plant Cell* 16, 908–919. doi: <http://dx.doi.org/10.1105/tpc.015818>
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57, 289–300.
- Bielecka, M., Watanabe, M., Morcuende, R., Scheible, W.-R., Hawkesford, M. J., Hesse, H., et al. (2014). Transcriptome and metabolome analysis of plant sulfate starvation and resupply provides novel information on transcriptional regulation of metabolism associated with sulfur, nitrogen and phosphorus nutritional responses in Arabidopsis. *Front. Plant Sci.* 5, 805. doi: 10.1007/s11105-014-0772-7
- Birnbaum, K., Shasha, D. E., Wang, J. Y., Jung, J. W., Lambert, G. M., Galbraith, D. W., et al. (2003). A gene expression map of the *Arabidopsis* root. *Science* 302, 1956–1960. doi: 10.1126/science.1090022
- Boke, H., Ozhuner, E., Turktas, M., Parmaksiz, I., Ozcan, S., and Unver, T. (2015). Regulation of the alkaloid biosynthesis by miRNA in opium poppy. *Plant Biotechnol. J.* 13, 409–420. doi: 10.1111/pbi.12346
- Bressan, R. A., Zhang, C., Zhang, H., Hasegawa, P. M., Bohnert, H. J., and Zhu, J.-K. (2001). Learning from the Arabidopsis experience. The next gene search paradigm. *Plant Physiol.* 127, 1354–1360. doi: <http://dx.doi.org/10.1104/pp.010752>
- Bush, S. J., Castillo-Morales, A., Tovar-Corona, J. M., Chen, L., Kover, P. X., and Urrutia, A. O. (2014). Presence-absence variation in *A. thaliana* is primarily associated with genomic signatures consistent with relaxed selective constraints. *Mol. Biol. Evol.* 31, 59–69. doi:10.1093/molbev/mst166



757 Champigny, M. J., Sung, W. W., Catana, V., Salwan, R., Summers, P. S., Dudley, S. A., et al.  
758 (2013). RNA-Seq effectively monitors gene expression in *Eutrema salsugineum* plants  
759 growing in an extreme natural habitat and in controlled growth cabinet conditions. *BMC*  
760 *Genomics* 14, 578. doi:10.1186/1471-2164-14-578

761 Chen, X., Hedley, P. E., Morris, J., Liu, H., Niks, R. E., and Waugh, R. (2011). Combining  
762 genetical genomics and bulked segregant analysis-based differential expression: an  
763 approach to gene localization. *Theor. Appl. Genet.* 122, 1375–1383. doi: 10.1007/s00122-  
764 011-1538-3

765 Czechowski, T., Bari, R. P., Stitt, M., Scheible, W.-R., and Udvardi, M. K. (2004). Real-time  
766 RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity  
767 reveals novel root- and shoot-specific genes. *Plant J.* 38, 366–379. Doi:10.1111/j.1365-  
768 313X.2004.02051.x

769 Dassanayake, M., Oh, D.-H., Haas, J. S., Hernandez, A., Hong, H., Ali, S., et al. (2011). The  
770 genome of the extremophile crucifer *Thellungiella parvula*. *Nat. Genet.* 43, 913–918.  
771 doi:10.1038/ng.889

772 De Coninck, B. M. A., Sels, J., Venmans, E., Thys, W., Goderis, I. J. W. M., Carron, D., et al.  
773 (2010). *Arabidopsis thaliana* plant defensin AtPDF1.1 is involved in the plant response to  
774 biotic stress. *New Phytol.* 187, 1075–1088. doi: 10.1111/j.1469-8137.2010.03326.x

775 Doherty, C. J., and Kay, S. A. (2010). Circadian control of global gene expression patterns. *Annu.*  
776 *Rev. Genet.* 44, 419–444. doi:10.1146/annurev-genet-102209-163432

777 Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene  
778 expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207–210.

779 Eshel, G., Shaked, R., Kazachovka, Y., Khan, A., Eppel, A., Cisneros, A., et al. (2017).  
780 *Anastatica hierochuntica*, an *Arabidopsis* desert relative, is tolerant to multiple abiotic  
781 stresses and exhibits species-specific and common stress tolerance strategies with its  
782 halophytic relative, *Eutrema (Thellungiella) salsugineum*. *Front. Plant Sci.* 7, 1992.

783 Etalo, D. W., Stulemeijer, I. J. E., van Esse, H. P., de Vos, R. C. H., Bouwmeester, H. J., and  
784 Joosten, M. H. A. J. (2013). System-wide hypersensitive response-associated transcriptome  
785 and metabolome reprogramming in tomato. *Plant Physiol.* 162, 1599–1617. doi:  
786 <http://dx.doi.org/10.1104/pp.113.217471>

787 Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R. N., and Willmitzer, L. (2000).  
788 Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* 18, 1157–1161.

789 Gordon, A., and Hannon, G. J. (2010). Fastx-toolkit. *FASTQ/A short-reads pre-processing tools*  
790 (unpublished) [http://hannonlab.cshl.edu/fastx\\_toolkit](http://hannonlab.cshl.edu/fastx_toolkit)

791 Griffith, M., Timonin, M., Wong, A. C. E., Gray, G. R., Akhter, S. R., Saldanha, M., et al.  
792 (2007). *Thellungiella*: an *Arabidopsis*-related model plant adapted to cold temperatures.  
793 *Plant Cell Environ.* 30, 529–538. doi: 10.1111/j.1365-3040.2007.01653.x

794 Gullberg, J., Jonsson, P., Nordström, A., Sjöström, M., and Moritz, T. (2004). Design of  
795 experiments: an efficient strategy to identify factors influencing extraction and  
796 derivatization of *Arabidopsis thaliana* samples in metabolomic studies with gas  
797 chromatography/mass spectrometry. *Anal. Biochem.* 331, 283–295.  
798 doi:10.1016/j.ab.2004.04.037

799 Gurkok, T., Turktas, M., Parmaksiz, I., and Unver, T. (2015). Transcriptome profiling of alkaloid  
800 biosynthesis in elicitor induced opium poppy. *Plant Mol. Biol. Rep.* 33, 673–688.doi: 1  
801 10.1007/s11105-014-0772-7

802 Hamanishi, E. T., Barchet, G. L. H., Dauwe, R., Mansfield, S. D., and Campbell, M. M. (2015).  
803 Poplar trees reconfigure the transcriptome and metabolome in response to drought in a  
804 genotype- and time-of-day-dependent manner. *BMC Genomics* 16, 329. doi:  
805 10.1186/s12864-015-1535-z

806 Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., et al.  
807 (2011). Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334, 83–86.  
808 doi: 10.1126/science.1209244  
809

810 Haun, W. J., Hyten, D. L., Xu, W. W., Gerhardt, D. J., Albert, T. J., Richmond, T., et al. (2011).  
811 The composition and origins of genomic variation among individuals of the soybean  
812 reference cultivar Williams 82. *Plant Physiol.* 155, 645–655. doi: [http://dx.doi.org/10.1104/](http://dx.doi.org/10.1104/pp.110.166736)  
813 [pp.110.166736](http://dx.doi.org/10.1104/pp.110.166736)

814 Huang, J., Lu, X., Yan, H., Chen, S., Zhang, W., Huang, R., et al. (2012a). Transcriptome  
815 characterization and sequencing-based identification of salt-responsive genes in *Millettia*  
816 *pinnata*, a semi-mangrove plant. *DNA Res.* 19, 195–207. doi:10.1093/dnares/dss004

817 Huang, W., Pérez-García, P., Pokhilko, A., Millar, a. J., Antoshechkin, I., Riechmann, J. L., et al.  
818 (2012b). Mapping the core of the *Arabidopsis* circadian clock defines the network structure  
819 of the oscillator. *Science* 336, 75–79. doi: 10.1126/science.1219075

820 Huber-Keener, K. J., Liu, X., Wang, Z., Wang, Y., Freeman, W., Wu, S., et al. (2012).  
821 Differential gene expression in tamoxifen-resistant breast cancer cells revealed by a new  
822 analytical model of RNA-Seq data. *PLoS One* 7, e41333. doi:10.1371/journal.pone.0041333  
823

824 Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., et al. (2011). The  
825 *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.*  
826 43, 476–481. doi:10.1038/ng.807

827 Hyten, D. L., Smith, J. R., Frederick, R. D., Tucker, M. L., Song, Q., and Cregan, P. B. (2009).  
828 Bulk segregant analysis using the GoldenGate assay to locate the *Rpp3* locus that confers  
829 resistance to soybean rust in soybean. *Crop Sci.* 49, 265. doi: 10.2135/cropsci2008.08.0511

830 Inan, G., Zhang, Q., Li, P., Wang, Z., Cao, Z., Zhang, H., et al. (2004). Salt cress. A halophyte  
831 and cryophyte *Arabidopsis* relative model system and its applicability to molecular genetic

analyses of growth and development of extremophile. *Plant Physiol.* 135, 1718–1737. doi: <http://dx.doi.org/10.1104/pp.104.041723>

Kaspar, H., Dettmer, K., Gronwald, W., and Oefner, P. J. (2009). Advances in amino acid analysis. *Anal. Bioanal. Chem.* 393, 445–452. doi: 10.1007/s00216-008-2421-1

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–10. doi:10.1093/nar/gkr1090

Lee, Y. P., Babakov, A., de Boer, B., Zuther, E., and Hincha, D. K. (2012). Comparison of freezing tolerance, compatible solutes and polyamines in geographically diverse collections of *Thellungiella sp.* and *Arabidopsis thaliana* accessions. *BMC Plant Biol.* 12, 131.

Lee, Y. P., Giorgi, F. M., Lohse, M., Kvederaviciute, K., Klages, S., Usadel, B., et al. (2013). Transcriptome sequencing and microarray design for functional genomics in the extremophile Arabidopsis relative *Thellungiella salsuginea* (*Eutrema salsugineum*). *BMC Genomics* 14, 793. doi:10.1186/1471-2164-14-793

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Lisec, J., Meyer, R. C., Steinfath, M., Redestig, H., Becher, M., Witucka-Wall, H., et al. (2008). Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant J.* 53, 960–972. doi: 10.1111/j.1365-313X.2007.03383.x

Lisec, J., Römisch-Margl, L., Nikoloski, Z., Piepho, H.-P., Giavalisco, P., Selbig, J., et al. (2011). Corn hybrids display lower metabolite variability and complex metabolite inheritance patterns. *Plant J.* 68, 326–336. doi: 10.1111/j.1365-313X.2011.04689.x

Liu, Q., Wang, X., Tzin, V., Romeis, J., Peng, Y., and Li, Y. (2016). Combined transcriptome and metabolome analyses to understand the dynamic responses of rice plants to attack by the rice stem borer *Chilo suppressalis* (Lepidoptera: Crambidae). *BMC Plant Biol.* 16, 259. doi: 10.1186/s12870-016-0946-6

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 25, 402–408. doi:10.1006/meth.2001.1262

Lynch, D. V., and Dunn, T. M. (2004). An introduction to plant sphingolipids and a review of recent advances in understanding their metabolism and function. *New Phytol.* 161, 677–702. doi: 10.1111/j.1469-8137.2003.00992.x

MacLeod, M. J. R., Dedrick, J., Ashton, C., Sung, W. W. L., Champigny, M. J., and Weretilnyk, E. A. (2014). Exposure of two *Eutrema salsugineum* (*Thellungiella salsuginea*) accessions to water deficits reveals different coping strategies in response to drought. *Physiol. Plant.*, doi: 10.1111/ppl.12316.

- 869 Manners, J. M., Penninckx, I. A., Vermaere, K., Kazan, K., Brown, R. L., Morgan, A., et al.  
870 (1998). The promoter of the plant defensin gene *PDF1.2* from *Arabidopsis* is systemically  
871 activated by fungal pathogens and responds to methyl jasmonate but not to salicylic acid.  
872 *Plant Mol. Biol.* 38, 1071–1080.
- 873 Masclaux-Daubresse, C., Clément, G., Anne, P., Routaboul, J.-M., Guiboileau, A., Soulay, F., et  
874 al. (2014). Stitching together the multiple dimensions of autophagy using metabolomics and  
875 transcriptomics reveals impacts on metabolism, development, and plant responses to the  
876 environment in *Arabidopsis*. *Plant Cell* 26, 1857–1877. doi:  
877 <http://dx.doi.org/10.1105/tpc.114.124677>
- 878 Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005).  
879 Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–  
880 380. doi:10.1038/nature03959
- 881 Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J.-I., Ebana, K., et al. (2015).  
882 Metabolome-genome-wide association study dissects genetic architecture for generating  
883 natural variation in rice secondary metabolism. *Plant J.* 81, 13–23. doi: 10.1111/tpj.12681
- 884 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010).  
885 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation  
886 DNA sequencing data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110
- 887 Meyer, R. C., Steinfath, M., Lisec, J., Becher, M., Witucka-Wall, H., Törjék, O., et al. (2007).  
888 The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *Proc.*  
889 *Natl. Acad. Sci. U. S. A.* 104, 4759–4764. doi: 10.1073/pnas.0609709104
- 890 Mockler, T. C., Michael, T. P., Priest, H. D., Shen, R., Sullivan, C. M., Givan, S. A., et al. (2007).  
891 The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern  
892 matching, and promoter analysis. *Cold Spring Harb. Symp. Quant. Biol.* 72, 353–363.  
893 doi:10.1101/sqb.2007.72.006
- 894 Mucha, S., Walther, D., Muller, T. M., Hinch, D. K., and Glaswisch, E. (2015). Substantial  
895 reprogramming of the *Eutrema salsugineum* (*Thellungiella salsuginea*) transcriptome in  
896 response to UV and silver nitrate challenge. *BMC Plant Biol.* 15, 137. doi: 10.1186/s12870-  
897 015-0506-5
- 898 Niederhuth, C. E., Patharkar, O. R., and Walker, J. C. (2013). Transcriptional profiling of the  
899 *Arabidopsis* abscission mutant *hae hsl2* by RNA-Seq. *BMC Genomics* 14, 37. doi:  
900 10.1186/1471-2164-14-37
- 901 Nielsen, R., Paul, J. S., Albrechtsen, A., and Song, Y. S. (2011). Genotype and SNP calling from  
902 next-generation sequencing data. *Nat. Rev. Genet.* 12, 443–451. doi: 10.1038/nrg2986
- 903 Ning, Z., Cox, A. J., and Mullikin, J. C. (2001). SSAHA: a fast search method for large DNA  
904 databases. *Genome Res.* 11, 1725–1729. doi: 10.1101/gr.194201
- 905 Oh, D.-H., Dassanayake, M., Haas, J. S., Kropornika, A., Wright, C., d’Urzo, M. P., et al. (2010).

906 Genome structures and halophyte-specific gene expression of the extremophile  
907 *Thellungiella parvula* in comparison with *Thellungiella salsuginea* (*Thellungiella halophila*)  
908 and *Arabidopsis*. *Plant Physiol.* 154, 1040–1052. doi: 10.1104/pp.110.163923

909 Ohlrogge, J. B., and Jaworski, J. G. (1997). Regulation of fatty acid synthesis. *Annu. Rev. Plant*  
910 *Physiol. Plant Mol. Biol.* 48, 109–136. doi: 10.1146/annurev.arplant.48.1.109

911 Orsini, F., D’Urzo, M. P., Inan, G., Serra, S., Oh, D.-H., Mickelbart, M. V., et al. (2010). A  
912 comparative study of salt tolerance parameters in 11 wild relatives of *Arabidopsis thaliana*.  
913 *J. Exp. Bot.* 61, 3787–3798. doi: 10.1093/jxb/erq188

914 Oshlack, A., Robinson, M. D., and Young, M. D. (2010). From RNA-seq reads to differential  
915 expression results. *Genome Biol.* 11, 220. doi: 10.1186/gb-2010-11-12-220

916 Page, A. F., Cseke, L. J., Minocha, R., Turlapati, S. A., Podila, G. K., Ulanov, A., et al. (2016).  
917 Genetic manipulation of putrescine biosynthesis reprograms the cellular transcriptome and  
918 the metabolome. *BMC Plant Biol.* 16, 113. doi: 10.1186/s12870-016-0796-2

919 Part C – GS De Novo Assembler, GS Reference Mapper, SFF Tools (2010). *454 Sequencing*  
920 *System Software Manual*, v 2.5.3 (454 Life Sciences Corp. A Roche Company), 1–241.

921 Paulose, B., Kandasamy, S., and Dhankher, O. P. (2010). Expression profiling of *Crambe*  
922 *abyssinica* under arsenate stress identifies genes and gene networks involved in arsenic  
923 metabolism and detoxification. *BMC Plant Biol.* 10, 108. doi: 10.1186/1471-2229-10-108

924 Penninckx, I. A., Eggermont, K., Terras, F. R., Thomma, B. P., De Samblanx, G. W., Buchala,  
925 A., et al. (1996). Pathogen-induced systemic activation of a plant defensin gene in  
926 *Arabidopsis* follows a salicylic acid-independent pathway. *Plant Cell* 8, 2309–2323. doi:  
927 10.1105/tpc.8.12.2309

928 Pickrell, J. K., Marioni, J. C., Pai, A. A., Degner, J. F., Engelhardt, B. E., Nkadori, E., et al.  
929 (2010). Understanding mechanisms underlying human gene expression variation with RNA  
930 sequencing. *Nature* 464, 768–772. doi: 10.1038/nature08872

931 Pilarska, M., Wiciarz, M., Jajic, I., Kozieradzka-Kiszkurno, M., Dobrev, P., Vankova, R., and  
932 Niewiadomska, E. (2016). A different pattern of production and scavenging of reactive  
933 oxygen species in halophytic *Eutrema salsugineum* (*Thellungiella salsuginea*) plants in  
934 comparison to *Arabidopsis thaliana* and its relation to salt stress signaling. *Front. Plant Sci.*  
935 7, 1179. doi: 10.3389/fpls.2016.01179

936 Qin, Y.-M., Hu, C.-Y., Pang, Y., Kastaniotis, A. J., Hiltunen, J. K., and Zhu, Y.-X. (2007).  
937 Saturated very-long-chain fatty acids promote cotton fiber and *Arabidopsis* cell elongation  
938 by activating ethylene biosynthesis. *Plant Cell* 19, 3692–3704. doi: 10.1105/tpc.107.054437

939 Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011). Identification of novel transcripts  
940 in annotated genomes using RNA-Seq. *Bioinformatics* 27, 2325–2329. doi:  
941 10.1093/bioinformatics/btr355



942 Romero, M., Berlett, B. S., Jensen, P. J., Pell, E. J., and Tien, M. (2004). Investigations into the  
943 role of the plastidial peptide methionine sulfoxide reductase in response to oxidative stress  
944 in Arabidopsis. *Plant Physiology* 136, 3784–3794. doi: 10.1104/pp.104.046656

945 Rouhier, N., Vieira Dos Santos, C., Tarrago, L., and Rey, P. (2006). Plant methionine sulfoxide  
946 reductase A and B multigenic families. *Photosynth. Res.* 89, 247–262. doi: 10.1007/s11120-  
947 006-9097-1

948 Satou, M., Enoki, H., Oikawa, A., Ohta, D., Saito, K., Hachiya, T., et al. (2014). Integrated  
949 analysis of transcriptome and metabolome of Arabidopsis albino or pale green mutants with  
950 disrupted nuclear-encoded chloroplast proteins. *Plant Mol. Biol.* 85, 411–428 doi:  
951 10.1007/s11103-014-0194-9

952 Shahzad, Z., Ranwez, V., Fizames, C., Marquès, L., Le Martret, B., Alassimone, J., et al. (2013).  
953 *Plant Defensin type 1 (PDF1)*: protein promiscuity and expression variation within the  
954 *Arabidopsis* genus shed light on zinc tolerance acquisition in *Arabidopsis halleri*. *New*  
955 *Phytol.* 200, 820–833. doi: 10.1111/nph.12396

956 Singh, A. K., Ulanov, A. V., Li, Z., Jayaswal, R. K., and Wilkinson, B. J. (2011). Metabolomes  
957 of the psychrotolerant bacterium *Listeria monocytogenes* 10403S grown at 37 °C and 8 °C.  
958 *Int. J. Food Microbiol.* 148, 107–114. doi: 10.1016/j.ijfoodmicro.2011.05.008

959 Springer, N. M., Ying, K., Fu, Y., Ji, T., Yeh, C.-T., Jia, Y., et al. (2009). Maize inbreds exhibit  
960 high levels of copy number variation (CNV) and presence/absence variation (PAV) in  
961 genome content. *PLoS Genet.* 5, e1000734. doi: 10.1371/journal.pgen.1000734

962 Stein, R. J., and Waters, B. M. (2012). Use of natural variation reveals core genes in the  
963 transcriptome of iron-deficient *Arabidopsis thaliana* roots. *J. Exp. Bot.* 63, 1039–1055. doi:  
964 10.1093/jxb/err343

965 Sulpice, R., Pyl, E.-T., Ishihara, H., Trenkamp, S., Steinfath, M., Witucka-Wall, H., et al. (2009).  
966 Starch as a major integrator in the regulation of plant growth. *Proc. Natl. Acad. Sci. U. S. A.*  
967 106, 10348–10353. doi: 10.1073/pnas.0903478106

968 Sumner, L. W., Lei, Z., Nikolau, B. J., and Saito, K. (2015). Modern plant metabolomics:  
969 advanced natural product gene discoveries, improved technologies, and future prospects.  
970 *Nat. Prod. Rep.* 32, 212–229. doi: 10.1039/c4np00072b

971 Swanson-Wagner, R. A., Eichten, S. R., Kumari, S., Tiffin, P., Stein, J. C., Ware, D., et al.  
972 (2010). Pervasive gene content variation and copy number variation in maize and its  
973 undomesticated progenitor. *Genome Res.* 20, 1689–1699. doi: 10.1101/gr.109165.110

974 Taiz, L., and Zeiger, E. (2010). *Plant Physiology 5th Ed.* Sunderland, MA: Sinauer Associates,  
975 Inc.

976 Taji, T., Seki, M., Satou, M., Sakurai, T., Kobayashi, M., Ishiyama, K., et al. (2004).  
977 Comparative genomics in salt tolerance between Arabidopsis and Arabidopsis-related  
978 halophyte salt cress using Arabidopsis microarray. *Plant Physiol.* 135, 1697–1709. doi:

979 10.1104/pp.104.039909

980 Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., et al. (2004). Mapman: a  
981 user-driven tool to display genomics data sets onto diagrams of metabolic pathways and  
982 other biological processes. *Plant J.* 37, 914–939. doi: 10.1111/j.1365-313X.2004.02016.x

983 Thomma, B. P. H. J., Cammue, B. P. A., and Thevissen, K. (2002). Plant defensins. *Planta* 216,  
984 193–202. doi: 10.1007/s00425-002-0902-6

985 Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010).  
986 Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and  
987 isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi:  
988 10.1038/nbt.1621

989 Tresch, S., Heilmann, M., Christiansen, N., Looser, R., and Grossmann, K. (2012). Inhibition of  
990 saturated very-long-chain fatty acid biosynthesis by mefluidide and perfluidone, selective  
991 inhibitors of 3-ketoacyl-CoA synthases. *Phytochemistry* 76, 162–171. doi:  
992 10.1016/j.phytochem.2011.12.023

993 Trick, M., Adamski, N. M., Mugford, S. G., Jiang, C.-C., Febrer, M., and Uauy, C. (2012).  
994 Combining SNP discovery from next-generation sequencing data with bulked segregant  
995 analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol.* 12, 14. doi:  
996 10.1186/1471-2229-12-14

997 Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics* 5, 99–  
998 114. doi: 10.2307/3001913

999 Volkov, V., Wang, B., Dominy, P. J., Fricke, W., and Amtmann, A. (2003). *Thellungiella*  
1000 *halophila*, a salt-tolerant relative of *Arabidopsis thaliana*, possesses effective mechanisms  
1001 to discriminate between potassium and sodium. *Plant Cell Environ.* 27, 1–14. doi:  
1002 10.1046/j.0016-8025.2003.01116.x

1003 Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for  
1004 transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484

1005 Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin* 1, 80–83.

1006 Wong, C. E., Li, Y., Whitty, B. R., Díaz-Camino, C., Akhter, S. R., Brandle, J. E., et al. (2005).  
1007 Expressed sequence tags from the Yukon ecotype of *Thellungiella* reveal that gene  
1008 expression in response to cold, drought and salinity shows little overlap. *Plant Mol. Biol.* 58,  
1009 561–574. doi: 10.1007/s11103-005-6163-6

1010 Wu, H.-J., Zhang, Z., Wang, J.-Y., Oh, D.-H., Dassanayake, M., Liu, B., et al. (2012). Insights  
1011 into salt tolerance from the genome of *Thellungiella salsuginea*. *Proc. Natl. Acad. Sci. U. S. A.*  
1012 109, 12219–12224. doi: 10.1073/pnas.1209954109

1013 Wu, S., Alseekh, S., Cuadros-Inostroza, Á., Fusari, C. M., Mutwil, M., Kooke, R., et al. (2016).  
1014 Combined use of genome-wide association data and correlation networks unravels key

1015 regulators of primary metabolism in *Arabidopsis thaliana*. *PLoS Genet.* 12, e1006363. doi:  
1016 10.1371/journal.pgen.1006363

1017 Wu, T. D., and Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for  
1018 mRNA and EST sequences. *Bioinformatics* 21, 1859–1875. doi:  
1019 10.1093/bioinformatics/bti310

1020 Xu, X., Feng, J., Lü, S., Lohrey, G. T., An, H., Zhou, Y., et al. (2014). Leaf cuticular lipids on  
1021 the Shandong and Yukon ecotypes of saltwater cress, *Eutrema salsugineum*, and their  
1022 response to water deficiency and impact on cuticle permeability. *Physiol. Plant.* 151, 446–  
1023 458. doi: 10.1111/ppl.12127

1024 Yang, R., Jarvis, D. E., Chen, H., Beilstein, M. A., Grimwood, J., Jenkins, J., et al. (2013). The  
1025 reference genome of the halophytic plant *Eutrema salsugineum*. *Front. Plant Sci.* 4, 46. doi:  
1026 10.3389/fpls.2013.00046

1027 Yeo, M. T. S., Carella, P., Fletcher, J., Champigny, M. J., Weretilnyk, E. a., and Cameron, R. K.  
1028 (2014). Development of a *Pseudomonas syringae*-*Eutrema salsugineum* pathosystem to  
1029 investigate disease resistance in a stress tolerant extremophile model plant. *Plant Pathol.*  
1030 doi: 10.1111/ppa.12271

1031 Zäuner, S., Ternes, P., and Warnecke, D. (2010). “Biosynthesis of sphingolipids in plants (and  
1032 some of their functions),” in *Madame Curie Bioscience Database*, eds. C. Chalfant and M.  
1033 D. Poeta (Landes Bioscience and Springer Science+Business Media), 249–263.

1034 Zimmerli, L., Stein, M., Lipka, V., Schulze-Lefert, P., and Somerville, S. (2004). Host and non-  
1035 host pathogens elicit different jasmonate/ethylene responses in *Arabidopsis*. *Plant J.* 40,  
1036 633–646. doi: 10.1111/j.1365-313X.2004.02236.x

**Table 1 Number of reads, percentage of reads uniquely aligned to reference genome (%), mean read length (bp) of *E. salsugineum* Shandong (SH) and Yukon (YK) accessions**

Accession	Total reads	Uniquely aligned (%)	Mean read length (bp)
SH	541,494	95.6	365.8
YK	491,926	94.4	361.8
Total	1,033,420	95.0	363.8

1037

1038

**Table 2 Summary of transcriptome assembly of *E. salsugineum* Shandong (SH) and Yukon (YK) accessions**

Number of assembled genes	17,888
Shared	14,577
SH only	1,643
YK only	1,668
Newly annotated genes	157
Shared	122
SH only	26
YK only	9
Annotated by best blast hit from <i>A. thaliana</i> , <i>A. lyrata</i> and/or <i>E. parvulum</i>	54
Newly annotated genes unique from genes annotated by Champigny et al., 2013	65
Newly annotated genes present in genes annotated by Champigny et al., 2013	92

1039



**Table 3 Summary of single nucleotide polymorphisms (SNPs) that differentiate *E. salsugineum* Shandong and Yukon accessions**

Number of genes with SNPs	6,182
Number of SNPs in genes	14,808
Number of high quality SNPs <sup>a</sup>	4,873
Number of SNPs identified by Champigny et al., 2013	47,317
Number of SNPs present in Champigny et al., 2013	10,883
Number of SNPs unique from Champigny et al., 2013	3,925

<sup>a</sup> Quality SNP satisfies technical and biologically relevant filtering criteria as discussed in the materials and methods.

**Table 4 Differently expressed genes (DEGs) of *E. salsugineum* Shandong (SH) and Yukon (YK) accessions**

<i>Eutrema</i> gene locus	SH	YK	Fold difference	Fold difference in qRT	<i>A. thaliana</i> best blast hit	Gene description
Thhalv10021382m.g	4	87	21.8	12.5	AT1G76810	Eukaryotic translation initiation factor 2 (eif-2) family protein
Thhalv10000662m.g	46	429	9.3	4.4	AT5G44420	Plant defensin 1.2A
Thhalv10015083m.g	34265	2494	-13.7	-5661.5		
Thhalv10011087m.g	1445	327	-4.4	n.s.	ATCG00480	ATP synthase subunit beta
XLOC_024729	237	25	-9.5	-2.0		
Thhalv10015718m.g	0	48	Absent in SH	Not detected in SH	AT5G56920	Cystatin/monellin superfamily protein
Thhalv10029390m.g	0	83	Absent in SH	Not detected in SH	AT4G11000	Ankyrin repeat family protein
Thhalv10021522m.g	1	244	244.0	144.9	AT3G16660	Pollen Ole e 1 allergen and extensin family protein
Thhalv10014718m.g	1	86	86.0	11.8	AT5G07470	Peptide methionine sulfoxide reductase 3
Thhalv10002969m.g	61	0	Absent in YK	-58.3	AT3G62210	Putative endonuclease or glycosyl hydrolase
Thhalv10009345m.g	58	0	Absent in YK	Not detected in YK		
Thhalv10014264m.g	53	0	Absent in YK	Not detected in YK	AT4G01560	Ribosomal RNA processing Brix domain protein
Thhalv10000285m.g	179	0	Absent in YK	Not detected in YK	AT4G20095	Protein of unknown function (DUF626)
Thhalv10022943m.g	148	0	Absent in YK	Not detected in YK		
Thhalv10023491m.g	47	0	Absent in YK	Not detected in YK	AT3G61100	Putative endonuclease or glycosyl hydrolase
Thhalv10018393m.g	70	1	-70.0	1.7	AT1G70160	Unknown protein
XLOC_005768	0	99	Absent in SH	Not detected in SH		
XLOC_017575	50	0	Absent in YK	Not detected in YK		
XLOC_003055	55	0	Absent in YK	Not detected in YK		
XLOC_024727	53	0	Absent in YK	Not detected in YK		
XLOC_003052	318	1	-318.0	-257.7		
XLOC_008740	135	3	-45.0	Not detected in YK		

XLOC_020731	76	0	Absent in YK	Not detected in YK
-------------	----	---	--------------	--------------------

**Table 4 Differently expressed genes (DEGs) of *E. salsugineum* Shandong (SH) and Yukon (YK) accessions**

<i>Eutrema</i> gene locus	SH	YK	Fold difference	<i>A. thaliana</i> best blast hit	Gene description
XLOC_017573	246	0	Absent in YK	Not detected in YK	
XLOC_015175	75	0	Absent in YK	Not detected in YK	
Genes with no specific primers available					
Thhalv10022994m.g	0	46	Absent in SH	Not tested	AT2G15220 Plant basic secretory protein (BSP) family protein
Thhalv10022932m.g	74	1	-74.0	Not tested	
Thhalv10014933m.g	295	32	-9.2	Not tested	AT5G59870 Histone H2A 6
Thhalv10019398m.g	6	217	36.2	Not tested	AT5G44430 Plant defensin 1.2C
Thhalv10029246m.g	0	47	Absent in SH	Not tested	AT4G11000 Ankyrin repeat family protein
XLOC_004723	187	16	-11.7	Not tested	ATCG01020 Ribosomal protein L32

**Table 5 Metabolite profiling differences between *E. salsugineum* Shandong (SH) and Yukon (YK) accessions. Metabolites for which tissue concentrations were different between *E. salsugineum* Shandong (SH) and Yukon (YK) accessions in one or two experiments and/or different based on a nested analysis of the two experiments.**

Metabolite	Experiment 1		Experiment 2		P-value		
	SH	YK	SH	YK	E1	E2	N
Carbohydrates							
2-keto-gluconic acid	1012.0	526.4	352.9	470.8	n.s.	*	n.s.
2PGA	2.1	0.0			**		
$\alpha$ -ketoglutaric acid	83.2	41.3	66.5	61.2	*	n.s.	*
Arabitol	45.4	60.9	12.9	19.5	n.s.	*	n.s.
Ascorbic acid	19.6	11.8	25.0	10.4	n.s.	**	**
Citric acid	6722.6	17762.7	14347.0	30526.0	n.s.	**	**
Digalactosylglycerol	2007.8	618.8	344.5	82.9	n.s.	***	n.s.
Disaccharide	9941.3	8651.1	10713.7	3862.7	n.s.	***	**
Erythritol	27.6	39.1	22.1	34.3	n.s.	*	*
Erythrose			976.8	591.3		*	
Fructose	65604.0	9500.3	14987.4	1284.4	***	***	***
Fumaric acid	232.3	429.0	111.5	263.1	n.s.	***	**
Galactaric acid			3.4	8.5		*	
Galactose	7730.3	4432.0	32.0	88.7	*	***	**
Gentiobiose			27.3	20.4		*	
Glucaric acid	3.7	3.0	16.1	10.4	n.s.	*	n.s.
Glucose	41843.5	28552.0	16547.4	6251.4	*	***	***
Glycerol-3-p	82.4	30.8	674.0	362.7	*	***	***
Isomaltose	75.1	0.0			**		
Itaconic acid	142.0	32.2			*		
Maltose	1119.8	678.6	1081.4	846.2	*	*	***
Melibiose	0.0	5.4			***		
Methylmaleic acid			16.7	7.1		*	
Pyruvic acid	42.4	38.6	59.8	44.0	n.s.	**	n.s.
Raffinose	477.5	322.5	218.9	152.7	*	n.s.	**

**Table 5 Metabolite profiling differences between *E. salsugineum* Shandong (SH) and Yukon (YK) accessions. Metabolites for which tissue concentrations were different between *E. salsugineum* Shandong (SH) and Yukon (YK) accessions in one or two experiments and/or different based on a nested analysis of the two experiments.**

Metabolite	Experiment 1		Experiment 2		P-value		
	SH	YK	SH	YK	E1	E2	N
Sedoheptulose	853.9	782.4	452.2	312.6	n.s.	**	n.s.
Sorbitol	17.0	39.5			*		
Turanose			13688.2	7035.2		**	
Xylitol	18.6	11.5	13.7	20.4	n.s.	*	n.s.
Xylose	165.8	81.1	64.6	49.3	*	n.s.	*
Fatty acids							
13-eicosenoic acid			5.7	7.6		*	
7-hexadecenoic acid			132.2	27.0		*	
9,12-octadecadienoic acid			6032.7	7434.6		*	
Docosanoic acid	50.5	241.3	176.5	1224.1	**	***	***
Hexacosanoic acid	13.2	28.9	0.0	20.9	*	***	***
Hexadecanol	0.0	12.2			***		
Octacosanol	9.5	1.9			**		
Tetracosanoic acid	1206.3	5001.7	8103.8	16232.9	*	*	**
Tetracosanol	2.2	0.0	2.5	4.5	***	**	n.s.
Tricosanoic acid	0.0	12.6	133.6	150.7	**	n.s.	n.s.
Tritriacontanol			47.1	37.6		*	
Amino acids							
Alanine			2.7	6.8		**	
Aspartic acid			928.9	2095.8		*	
Glycine			13.2	57.0		*	
N-acetylglutamic acid			158.0	1221.7		*	
Serine			216.4	818.2		*	
Threonine	0.0	3.8	311.1	383.3	***	*	*



**Table 5 Metabolite profiling differences between *E. salsugineum* Shandong (SH) and Yukon (YK) accessions. Metabolites for which tissue concentrations were different between *E. salsugineum* Shandong (SH) and Yukon (YK) accessions in one or two experiments and/or different based on a nested analysis of the two experiments.**

Metabolite	Experiment 1		Experiment 2		P-value		
	SH	YK	SH	YK	E1	E2	N
Valine	0.0	10.3	104.9	92.8	**	n.s.	n.s.
Others							
1,2,3-hydroxybutane	4.1	0.0			*		
2,3-dihydroxybutanedioic acid	11.8	27.0	5.8	9.6	*	*	**
2,4,6-tri-tert.-butylbenzenethiol	42.5	44.3	157.3	185.3	n.s.	*	n.s.
2-aminoethylphosphate			0.0	44.5		*	
2-desoxy-pentos-3-ulose	33.1	15.3			***		
2-hydroxybutanoic acid			11.3	0.0		*	
2-o-glycerol-galactopyranoside	2362.0	398.0	181.6	178.0	***	n.s.	*
2-oxo-3-hydroxypropanoic acid			17.3	75.7		**	
3-methylthiopropylamine			0.0	5.4		*	
5-o-coumaroyl-d-quinic acid			38.5	15.6		*	
Aconitic acid	202.7	103.5	12.7	37.5	*	*	n.s.
Tocopherol	7.8	4.0			*		
Chlorogenic acid			710.6	332.2		*	
Erythronic acid-1,4-lactone			24.2	38.0		*	
Ferulic acid	11.3	30.4	41.8	61.6	n.s.	**	**
Hydroxylamine			39.2	27.1		**	
Ketomalonic acid			0.0	3.1		*	
Malonic acid	59.7	89.5	0.0	2.6	n.s.	**	n.s.
P-hydroxyacetophenon	0.0	2.8			***		
Protocatechuic acid			5.9	2.1		*	

**Table 5 Metabolite profiling differences between *E. salsugineum* Shandong (SH) and Yukon (YK) accessions. Metabolites for which tissue concentrations were different between *E. salsugineum* Shandong (SH) and Yukon (YK) accessions in one or two experiments and/or different based on a nested analysis of the two experiments.**

Metabolite	Experiment 1		Experiment 2		P-value		
	SH	YK	SH	YK	E1	E2	N
Putrescine			0.0	22.4		**	
Stigmasterol	40.6	52.0	29.5	9.1	n.s.	**	n.s.
Vanillin	0.0	1.8			***		

Data are means (n=4). Empty indicates absence or undetected metabolites. Genotypes are significant different at P- value < 0.05 (\*), 0.01 (\*\*), and 0.001 (\*\*\*) based on two-tailed t test in Experiment 1 (E1), Experiment 2 (E2), or nested (N) analysis of two experiments combined together. Carbohydrates indicate sugar-related compounds and derivatives in citric acid cycle. Fatty acids indicate fatty acids and their derivatives. Others indicate all other compounds not belong to carbohydrates or fatty acids, or amino acids.



