Brigham Young University
BYU ScholarsArchive

All Theses and Dissertations

2012-12-13

Transcriptome Analysis of Drought Induced Stress in Chenopodium Quinoa

Joshua Arthur Raney
Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Animal Sciences Commons

BYU ScholarsArchive Citation

Raney, Joshua Arthur, "Transcriptome Analysis of Drought Induced Stress in Chenopodium Quinoa" (2012). *All Theses and Dissertations*. 3915.
https://scholarsarchive.byu.edu/etd/3915

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Transcriptome Analysis of Drought Induced Stress in

Chenopodium quinoa

Joshua A. Raney

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

P. Jeff Maughan, Chair
Joshua A. Udall
Sam B. St. Clair
Von D. Jolley

Department of Plant and Wildlife Science

Brigham Young University

December 2012

Copyright © 2012 Joshua A. Raney

All Rights Reserved

ABSTRACT

Transcriptome Analysis of Drought Induced Stress in
Chenopodium quinoa

Joshua A. Raney
Department of Plant and Wildlife Science, BYU
Master of Science

RNA-seq transcriptome analysis of *Chenopodium quinoa* at different water treatment levels was conducted in a greenhouse study using four water treatments (field capacity to drought) on a valley ecotype quinoa (variety Ingapirca) and an Altiplano Salares ecotype quinoa (variety Ollague). Physiological results support the earlier findings that the Salares ecotypes display greater tolerance to drought-like stress conditions than the valley ecotypes (as determined by growth rate, photosynthetic rate, stomatal conductance, and stem water potential). cDNA libraries from root tissue sample for each treatment x variety combination were sequenced using Illumina Hi-Seq technology in an RNA-seq experiment. *De novo* assembly of the transcriptome generated 20,337 unique transcripts. Gene expression analysis of the RNA-seq data identified 462 putative gene products that showed differential expression based on treatment and 27 putative gene products differential expressed based on variety x treatment, including significant expression differences in root tissue in response to increasing water stress. BLAST searches and gene ontology analysis show an overlap with drought tolerance stress and other abiotic stress mechanisms.

Keywords: *Chenopodium quinoa*, drought, Illumina sequencing, RNA-seq, transcriptome assembly

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to Dr. Jeff Maughan, my committee chair, for his continual support and help throughout the duration of this project. Without his encouragement and guidance this thesis would not have been possible. I would also like to thank the other members of my committee: Dr. Joshua A. Udall, Dr. Sam B. St. Clair, and Dr. Von D. Jolley. They have taught me many valuable lessons. Their expertise and advice has allowed this project be what it is. I would like to thank Angel Mujica and Alejandro Bonifacio for providing seed for this project and advice in the early planning stages of this project. J.C. Raney, Liz Soliai, Hassan Munir and others merit thanks for all their hard work and help with this project. Special thanks go to the Microarray and Genomic Analysis Core Facility at the University of Utah for sequencing my cDNA libraries. I am extremely grateful to Dave Elzinga and Justin Page who helped me for months with the bioinformatics (assembly and differential expression) portion of my project. I am grateful for the constant support and love I have received from my family and friends. Lastly, I would like to express my gratitude to the McKnight Foundation, the Benson Institute, Fundación Proinpa, and Brigham Young University for funding this research.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGEMENTS iii

TABLE OF CONTENTS iv

LIST OF TABLES vi

LIST OF FIGURES vii

LIST OF SUPPLEMENTAL FIGURES viii

Chapter 1. Transcriptome Analysis of Drought Induced Stress in *Chenopodium quinoa* 1

INTRODUCTION 2
MATERIALS AND METHODS 5
Greenhouse experiment 5
Phenotypic measurements 5
Drought treatments 6
cDNA library construction 7
Sequencing and assembly 7
Differential gene expression analysis 8
GO annotation and pathway analysis 9
RESULTS 10
Phenotypic response 10
Sequencing and assembly 11
Differential expression analysis 11
Homology and functional analysis 12
DISCUSSION 14
Phenotypic response 14
Transcriptome assembly and expression analysis 15
Biological insight 16
CONCLUSIONS 18
REFERENCES 19
TABLES 24
FIGURES 27
APPENDIX 30
SUPPLEMENTAL FIGURES 37

Chapter 2. Literature Review 39

INTRODUCTION 40
Taxonomy 40
History 41
Nutrition 42
Classification 43
Tolerance 44
Drought 45
Major photosynthetic pathways 46
Drought mechanisms 47
Protein classifications 50
Sequencing 52
Transcriptome analysis 53
RNA-Seq 54
Mapping 55
Differential expression analysis 56
Biological insights 56
REFERENCES 58
TABLES 65

LIST OF TABLES

TABLE 1.1: Treatment soil moisture tensions used in study……..……………………………...24

TABLE 1.2: ANOVA results..24

TABLE 1.3: Fischer exact test results………………………………………………………...25

TABLE 2.1: Crude protein content of quinoa...65

TABLE 2.2: Quinoa nutrition...65

TABLE 2.3: Major software packages for differential expression analysis.................................66

LIST OF FIGURES

FIGURE 1.1: Whisker plots of phenotypic measurements.. 27

FIGURE 1.2: Histogram of contig lengths... 28

FIGURE 1.3: Distribution of gene ontology of 462 DE treatment genes……………………… 29

FIGURE 1.4: Distribution of gene ontology of 27 DE variety x treatment interaction genes 30

LIST OF SUPPLEMENTAL FIGURES

SUPPLEMENTAL FIGURE 1.1: Flowchart of RNA-Seq analysis..37

SUPPLEMENTAL FIGURE 1.2: Experimental design and sequencing layout...........................38

Chapter 1. Transcriptome Analysis of Drought Induced Stress in *Chenopodium quinoa*

INTRODUCTION

Quinoa (*Chenopodium quinoa* Willd.) is an allotetraploid (2n=4x=36) species belonging to the goosefoot (*Chenopodium*) genus that consists of over 120 species of perennial or annual herbaceous flowering plants. Native to the Andean region of South American countries this pseudo-cereal grain crop has been cultivated for thousands of years. The Incan people treated the quinoa grain as sacred and referred to it as the "mother grain" (Koyro and Eisa 2008). It has also been referred to as the "super grain of the future" because of its high protein content – 7.5-22.1% (Tapia et al. 1979) and ideal balance of all the essential amino acids (Ruas et al. 1999). Quinoa has relatively high amounts of vitamins and minerals (Ruales and Nair 1993) and is gluten-free. Recent international attention is due to its unusually high nutritional value (Jacobsen 2003). Indeed, the United Nations has declared 2013 as the 'International Year of Quinoa' (FAO 2011).

More than three thousand accessions of cultivated and land race types of quinoa are held in germplasm banks throughout South America (Tapia 1980). Valencia-Chamorro (2003) categorized these varieties into five groups according to their altitudinal gradient and environment: sea level, valleys, the Yungas (subtropics), the salt flats (Salares) and the Altiplano (high plains). While valley quinoa types are cultivated around 2000-4000m they receive similar amounts of annual rainfall as the sea level quinoa, ranging from 800-1500mm. The Salares and Altiplano varieties are adapted to extreme environmental conditions that typify the Andean Altiplano: high altitude, drought, saline soils as well as frequent frost (Vacher 1998; Prado et al. 2000; Jacobsen et al. 2003). Both Salares and Altiplano quinoas are cultivated at altitudes around 4000m and receive as little as 200mm of annual rainfall (Risi and Galwey 1984; Tapia 1997). These regions have been known to have below freezing temperature and frost for up to 200 days a year (Jacobsen et al. 2007). The Salares types grow in highly saline soils with pH > 8.0.

Among all the abiotic and biotic stresses endured by crop plants, drought is considered the most important factor limiting crop productivity worldwide (Boyer, 1982; Bartels and Nelson, 1994). Current climate models indicate an increase in drought episodes and severity due to the effects of global warming (Cook et al. 2007). Such changes will have devastating economic and sociological impact on human population. For example, this year (2012), the United States experienced drought in almost 80% of its agricultural land – making 2012 the most extensive drought experienced by the United States since the 1950s. Yields for corn (122 bushels per acre) and soybean (37.8 bushels per acre) are the lowest yields since 1995 for corn and 2003 for soybean (USDA 2012). The decrease in corn and soybean production in 2012 drastically affected the supply, demand and price conditions of other crops, meat, poultry, dairy as well as other packaged and processed foods that use these crops. The 2012 drought has increased the 2013 price index for beef (up 4 - 5%), dairy products (up 3.5 - 4.5%), poultry (up 3 - 4%), and cereals and bakery products (up 3 – 4%; Lowrey and Nixon 2012). Similar outcomes related to drought stress are seen regularly across the globe in all crops. Understanding how plants respond to drought and developing adaptive agricultural strategies is an urgent and essential goal of plant and crop scientists.

While quinoa is commonly referred to as a drought and salt tolerant crop (Jacobsen 2003; Trognitz 2003), no research has been reported that investigates the genomic basis for drought tolerance differences among the quinoa ecotypes. Transcriptome research conducted in *Arabidopsis thaliana* (Kreps et al. 2002; Kawaguchi et al. 2004), *Oryza sativa* (Dubouzet et al. 2003; Gorantla et al. 2005), and *Sorghum bicolor* (Buchanan et al. 2005) has been used to successfully identify differentially expressed genes, implicating several biosynthetic pathways that assist in the overall tolerance to drought stress. Recent genomic research in quinoa has

successfully identified genes (e.g., SOS1, NHX1, etc.) associated with salt tolerance (Maughan et al. 2009; Morales et al. 2011), further suggesting the possibility of unraveling the genetic basis of drought resistance in quinoa through transcriptome analysis.

In this study we compared the response to different water treatments approximating drought-like conditions in two quinoa varieties – Ingapirca, a valley ecotype quinoa native to Ecuador, and Ollague, a Salares ecotype quinoa native to Chile (Mason et al. 2005). We used RNA sequencing (RNA-Seq) technology to identify putative gene products that are differentially expressed during progressive water stress. We assembled the first root transcriptome of quinoa, using Illumina reads (created by this study) and 454-pyrosequencing and Sanger reads (obtained from previous studies). Differentially expressed genes were identified and bioinformatic methods were employed to implicate specific pathways putatively associated with water stress in quinoa.

MATERIALS AND METHODS

Greenhouse experiment

Ingapirca and Ollague plants were propagated in greenhouse facilities at Brigham Young University (40.245N, -111.641W, 1410 m above sea level) in September 2011. Ingapirca is a valley ecotype quinoa native to Ecuador and Ollague is a Salares ecotype quinoa native to Chile. Seed for both varieties were kindly provided by Angel Mujica (Universidad del Altiplano, Puno, Peru). Daily average greenhouse conditions were maintained at a maximum temperature of 28°C, minimum temperature of 19°C, with a day length of 13 hours. Seedlings were thinned to three seeds per four-inch pot using Sunshine Basic Mix 1 soil (Sun Gro, Vancouver, British Columbia, Canada) supplemented with Osmocote fertilizer (Scotts, Marysville, OH). Three replicates, including the two quinoa varieties and four watering treatments, were set-up using a randomized complete block design. All plants were watered and treated equally for an initial period of three weeks.

Phenotypic measurements

Photosynthetic rate, stomatal conductance, transpiration rate, vapor pressure, and temperature of leaf were assayed using the LI-6400XT Portable Photosynthesis System (LI-COR, Lincoln, NE). Plant height was measured and averaged every third day. A pressure bomb (PMS Instrument Company, Albany, OR) with compressed nitrogen was used to measure the stem water potential, while chlorophyll a, b, and total chlorophyll contents was determined using a method described by Nagata and Yamashta (1992) for one plant per sample randomly chosen *a priori*. Quantitative data were analyzed statistically using the computer program R 2.15.1 (R Development Core Team 2012).

Drought treatments

The four drought treatments were induced and maintained by differential watering using the method that was described by Earl (2003). To determine the different drought treatments, 130 g of soil was water saturated and defined as the soil water holding capacity or "wet weight" (435.08 g). Oven-drying the saturated soil for four days resulted in a total "dry weight" of 72.58 g. The difference between the wet weight and dry weight (362.50 g) represents the soil water holding capacity of the pot. Field capacity was defined as 40% of the soil water holding capacity (145 g). The other estimated field capacities were calculated based on the 100% field capacity water weight of 145 g. The low, medium, and high water stress treatments were applied at 50%, 30% and 10% field capacity, respectively. Thus, the 100% field capacity treatment (hereafter referred to as the control) was maintained at 293 g (145 g water, 130 g soil, 18 g dry pot), while the low, medium, and high stress treatments were maintained at 220.5 g, 191.5 g, and 162.5 g respectively. Each pot was weighed independently several times a day, with water being added when the pot weight dropped below the weight thresholds.

These different drought treatments were initiated at day 23 when all plants reached the eight-leaf stage of development. The treatments were applied in a progressive manner. This was accomplished by all samples being maintained at a particular treatment for 1 day prior to progressing to higher drought treatments. Soil moisture tension and temperature of the soil was measured in control pots using a Davis 6440 watermark soil moisture sensors (Davis Instruments, Hayward, CA). This particular soil moisture sensor is an electrical resistance type sensor that converts electrical resistance to a calibrated reading of centibars of soil water suction, ranging from 10-200 centibars. Two sensors measured each drought treatment, which were averaged at the conclusion of the study to ensure distinct treatments (Table 1.1).

cDNA library construction

A flow chart that depicts cDNA library construction is provided as supplemental figure 1.1. Roots were collected from a representative sample from each treatment and quinoa variety. The tissue was immediately frozen in liquid nitrogen. RNA was extracted from root tissue using an RNeasy® Plant Mini Kit according to the manufacture's recommendation (Qiagen, Valencia, CA). RNA concentration and quality was assessed using a ND-1000 Spectrophotometer (NanoDrop, Wilmington, DE) and an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Conversion of mRNA into cDNA libraries was accomplished using the reagents and protocol provided in the Illumina® TruSeq™ RNA Sample Preparation Kit (Illumina, San Diego, CA). This procedure includes fragmenting the mRNA, synthesis of first and second strand cDNA, followed by end-repair, adenylating the 3' ends, and ligation of sequence adaptors. Adding unique sequence adaptors, also known as indexing or barcoding, allows for multiple libraries to be sequenced in a single lane. The quality of the synthesized cDNA libraries was verified using a DNA 7500 LabChip® Kit (Agilent Technologies, Santa Clara, CA) on a 2100 Bioanalyzer.

Sequencing and assembly

Sequencing was performed at the Microarray and Genomic Analysis Core Facility at the University of Utah (Salt Lake City, UT). The libraries were quantified and checked for quantity before being sequenced using an Illumina HiSeq 2000 instrument. The twenty-four cDNA libraries were divided evenly between two lanes of the flow cell (1 x 50 bases). The cDNA library numbering and sequencing layout is depicted in supplemental figure 1.2. The resulting Illumina fastq reads were trimmed using the computer program Sickle with a phred quality score of 20 (https://github.com/najoshi/sickle). In addition, 454-pyrosequencing and Sanger

sequence reads (obtained from previous studies) were also converted to fastq format. All reads were assembled *de novo* using the Trinity assembler (v. 2012-04-22; Grabherr et al. 2011). We then mapped the Illumina reads to the predicted Trinity contigs using GSNAP (v. 2012-05-23; Wu and Nacu 2010) with a max number of paths to print (-n) of 1, so each read could only map to one contig. The expression counts for each sample were calculated with a samtool script and the read counts for each gene were normalized by length and library size and rounded to the nearest whole number (Mortazavi et al. 2008).

Differential gene expression analysis

EdgeR is a Bioconductor package that is based on a negative binomial (NB) distribution. Robinson et al. (2010) modeled the counts for each gene (g) and sample (i) using the following algorithm:

$$Y_{gi} \sim NB(M_i p_{gj}, \phi_g)$$

where M_i represents the library size, p_{gj} represents the proportion of gene g in the experimental group j to which sample i belongs, and ϕ_g is the over-dispersion parameter that accounts for biological or sample-to-sample variation. Different models were also created that identified i) genes that were differentially expressed between the two varieties (Ingapirca versus Ollague across all treatments), ii) genes that were differentially expressed among the four drought treatments, and iii) gene that were differentially expressed based on the variety x treatment interaction, where the different variables represent the average expression of Ingapirca control, Ingapirca low stress, Ingapirca medium stress, Ingapirca high stress, Ollague control, Ollague low stress, Ollague medium stress, and Ollague high stress (see Appendix 1.1). Transcripts with false discovery rate (FDR) < 0.05 were extracted for further analysis and determined by an implementation of the Benjamini and Hochberg method, (Benjamini and Hochberg, 1995).

GO annotation and pathway analysis

Differentially expressed transcripts were compared to the RefSeq Protein public database using BLASTX (Altschul et al. 1997) with an e-value cut-off of $10e^{-10}$ to provide a functional annotation for each contig. We imported all differentially expressed transcripts into Blast2GO (version 2.6.1; Götz et al. 2008) where gene ontology (GO) terms were assigned and clustered based on biological process, cellular component, or molecular function. Mapping and annotation was performed using default parameters (*E-value* hit filter 1.0E-6, annotation cutoff 55, GO weight 5, HSP-hit coverage cut-off 20). GO enrichment was tested using the Fisher exact test, which uses the Gossip software integrated in the Blast2GO package (Blüthgen et al. 2005). This test assessed significant over-representation of GO terms in the list of the regulated genes (462 DE genes – test group) against the non-differentially expressed transcripts (20,337 contigs – reference group), using a cut-off threshold of FDR < 0.05. Specific gene products were identified as well as associated biological pathways as determined by the KEGG pathway (Ogata et al. 1999) mapping functionality offered in the Blast2GO program.

RESULTS

Phenotypic response

An initial greenhouse study compared the phenotypic responses associated with different water treatments between Ingapirca (valley ecotype) and Ollague (Salares ecotype). Measurements included heights, photosynthetic rate, stomatal conductance, and stem water potential. ANOVA analysis identified significant differences between the varieties and among all treatment levels (p = 3.01e-05 and p = 0.00212; Table 1.2) for plant height. The affects of water stress on plant height was evident in both varieties but more pronounced in the valley ecotype, Ingapirca, which had a more dramatic decrease in average height between the control and high water stress treatments than the respective treatments for Ollague. The Ollague plants were also negatively affected by the water stress but had a consistent decrease in average height among all treatments (Figure 1.1A).

Similar to height, photosynthetic rate and stomatal conductance measurements exhibited reduced metabolism due to stress treatments (Figures 1.1B and 1.1C). Ingapirca showed phenotypic effects typical to drought stress including loss of turgor pressure, causing the plant stem and leaves to droop, and curling of the leaves. The phenotypic effects were symptomatic of the underlying physiology. Under the high water stressed treatment Ingapirca had a lower average photosynthetic rate (5.5 µmol CO_2 m^{-2} s^{-1}) and stomatal conductance (0.031 µmol H_2O m^{-2} s^{-1}) when compared to the Ollague plants (9.4 µmol CO_2 m^{-2} s^{-1} and 0.081 µmol H_2O m^{-2} s^{-1}; Figures 1.1B and 1.1C). Significant differences for photosynthetic rate and stomatal conductance measurements were identified among the different water treatments in the ANOVA analysis (p = 0.00892 and p = 0.00265; Table 1.2).

Stem water potential is defined as the capacity of the plant to conduct water from the soil to the atmosphere. A more negative water potential indicates a greater tolerance to water stress conditions. MacCutchan and Shackel (1992) and Choné et al. (2001) used stem water potential as an indicator of water deficiencies in plum orchards and grapevines. ANOVA analysis revealed significant differences between varieties ($p = 0.00476$), among water treatments ($p = 1.44e-07$), and among the variety by treatment interactions ($p < 0.02551$; Table 1.2). Ollague exhibited lower average stem water potential (-2.97 MPa) when compared to Ingapirca (-1.95 MPa; Figure 1.1D).

Sequencing and assembly

Sequencing of twenty-four cDNA libraries were prepared from root tissue of each treatment and each variety. Sequencing resulted in 385,311,064 total reads (~19.27 Gb). A total of 373,835,465 reads (after assessing quality control and trimming) were combined with sources of quinoa EST sequences that were derived from various tissue sources (root, leaf, seedling, and seed tissue) and sequencing methodologies (454-pyrosequencing and Sanger sequencing; Reynolds et al. 2009; Morales et al. 2011). All reads were assembled *de novo* using the Trinity assembler (Grabherr et al. 2011). The combined assembly produced 20,337 unique consensus sequences (contigs), with a contig read length ranging from 201 to 18,777 bp with an average and median length of 525 bp and 435 bp, respectively (Figure 1.2).

Differential expression analysis

Implementing different models, EdgeR identified 6,170 differentially expressed (DE) contigs based on variety alone and 462 DE contigs based on the different water stress treatments. The variety x treatment DE contigs identified not only showed differential expression based on the water stress treatment, but the manner in which they were differentially expressed was also

influenced by their variety type. There were a total of 27 contigs (Appendix 1.2) that showed differential expression based on interaction terms.

Homology and functional analysis

Of the 462 DE contigs by Blast2GO, there were 251 sequences that resulted in functional gene ontology (GO) terms that were assigned to biological processes (BP), cellular components (CC), or molecular function (MF). It should be noted that the different GO categories are non-exclusive, resulting in some contigs being annotated in more than one category. Of the sequences that expressed GO terms, 175 sequences were assigned to the biological process category, 166 sequences to the cellular component category, and 177 to the molecular function category. The most dominant terms seen in the biological process category were cellular and metabolic process. Other major biological process identified were associated to various response stimuli: abiotic, abscisic acid, chemical, cold, endogenous, hormone, organic substance, osmotic, oxidative and salt stress (Figure 1.3A). Considering the molecular function the most dominant terms were catalytic activity and binding. Other major components were ATP-, DNA-, cation-, metal ion-, adenyl ribonucleotide-, and protein binding as well as enzymatic activity such as hydrolase, oxidoreductase, transferase, peroxidase and kinase activity (Figure 1.3B). The GO enrichment test identified 58 GO categories in our study, using a FDR < 0.05 (Table 1.3). All of the GO terms identified were over represented when comparing the regulated genes (462 DE treatment genes – test group) against the non-differentially expressed transcripts (20,337 contigs – reference group).

BLASTX searches of the RefSeq Protein database (e-value $< 10e^{-10}$) identified homology of 21 of the 27 identified DE variety x treatment interaction contigs. GO was performed on the 21 DE variety x treatment genes using Blast2GO. Of the total 21 DE interaction genes, 17

expressed GO term(s). Transferase activity, metal ion binding, cation binding, protein binding, and oxidoreductase activity are the major activities of interaction gene products as indicated by molecular function; while cellular metabolic process, primary metabolic process, response to stress, and response to abiotic stimulus are the major biological processes identified (Figure 1.4). While this study focused primarily on water deficiency effects, response to oxidative stress, oxidation-reduction process, and response to osmotic stress each contributed 5% and response to abscisic acid stimulus contributed 3% to the overall biological process ontology. The GO enrichment test failed to identify GO terms that were over or under represented among the 27 DE variety x treatment interaction genes.

DISCUSSION

Phenotypic response

The Salares ecotypes (variety Ollague) was more tolerant to the water stress, as measured by growth rate, photosynthetic rate, stomatal conductance, and stem water potential, when compared to the valley ecotype (variety Ingapirca; Morales et al. 2011). Root sensing of water stress is an important mechanism in Salares varieties (Jacobsen et al. 2009). Sensing water stress, roots initiates an increase in ABA synthesis leading to a decrease in turgor of stomata guard cells as well as inhibiting shoot growth (Jacobsen et al. 2009). Stomata closure decreases the CO_2 levels within the leaf reducing the light-independent reactions causing the photosynthetic rate to also decrease. This relationship was seen by Jensen et al. (2000) when the net photosynthesis of drought quinoa plants dropped from 14-24 μmol m^{-2} s^{-1} to 5-10 μmol m^{-2} s^{-1}. Certain quinoa varieties that have inherently higher leaf stomatal conductance to water vapor are capable of keeping higher photosynthesis rate (González et al. 2011). While we did not measure ABA concentrations in the root or xylem it appears that both varieties showed common drought responses with a decrease in leaf stomatal conductance and net photosynthetic rate. Interestingly, Ollague appeared to tolerate the high water stress treatment better than Ingapirca which may have been a result of higher ABA production in Ollague which was manifested by higher leaf stomatal conductance and a higher net photosynthesis rate . Lowering water potentials is another mechanism by which Altiplano Salares varieties may avoid water stress. Vacher (1998) showed that Altiplano Salares varieties under water stress had leaf water potential below -4.0 MPa accompanied by decreased stomatal conductance and photosynthesis rate. Our results indicated that Salares ecotype (variety Ollague,) exhibited significantly reduced stem water potential level (-2.97 MPa) under drought conditions when compared to the valley ecotype Ingapirca (-1.95

MPa). Tapia (1980) reported that varieties from the Altiplano, specifically Salares ecotype of Southern Bolivia, are usually 1.0-1.8 m tall while the varieties from the valley regions grow to be 2-4 m. In our experiments, Ollague plants were significantly smaller at the eight-leaf stage when compared to the Ingapirca. A slower growth rate may represent a coping mechanism that allows Salares ecotypes to tolerate water-limited environments (< 200 mm yearly rainfall) of the Altiplano (Pillco et al. 2007).

Transcriptome assembly and expression analysis

The development of next-generation sequencing (NGS) technology has made possible the cost effect characterization of non-model plant transcriptomes (Vera et al. 2008; Ekblom and Galindo 2010). We combined the Hi-Seq reads generated in this study with those developed through 454-pyrosequencing and Sanger sequencing (Reynolds et al. 2009; Morales et al. 2011) to produce a reference transcriptome for quinoa. Incorporation of reads from different NGS technologies with different read length and read depths created a long backbone of the transcriptional units while shorter read sequences improve accuracy of the assembly by providing sequencing depth coverage. Digital gene expression analysis, also termed RNA-seq, is one of many applications that can be performed using deep transcriptome sequencing (Metzker 2010). Using edgeR we were able to identify differentially expressed genes based on variety, treatment, and the variety x treatment interaction –ultimately making it a useful statistical method for detecting DE contigs in this study.

Biological insight

Plants in general respond to water-deficit conditions through a series of complex cellular, molecular, and physiological processes. Shinozaki and Yamaguchi-Shinozaki (2007) explained that the role genes products play in assisting drought tolerance can be classified as functional or regulatory. While regulatory proteins are responsible for signal transduction and response signaling, functional proteins involve cell maintenance, cell protection, and damage repair through proteins such as chaperones, late embryogenesis abundant (LEA) proteins, water channel proteins, etc. (Shinozaki et al. 2000).

At least two of the 27 interaction DE genes identified in this study can be classified as having regulatory functions, including one which showed high sequence homology with Naringenin, 2-oxogluturate 3-dioxygenase (comp42593_c0_seq1) and one which showed high sequence homology with a receptor-like cytosolic serine threoine-protein kinase rbk1-like protein (comp1839_c0_seq1). The function of Naringenin, 2-oxogluturate 3-dioxygenase is to catalyze the 3-beta-hydroxylation of 2S-flavanones to 2R, 3R-dihydroflavenols which are intermediates in the biosynthesis of flavonoids in plants. Hernández et al. (2004) showed that in *Cistus clusil* and other Mediterranean plants, tolerance to drought-like conditions is in part due to flavonoids contributing to the antioxidant defense. Similarly, SnRK2.6/OST1 (OPEN STOMATA 1) is a specific serine threoine-protein kinase that has implicated in drought-tolerance in *Arabidopsis thaliana* (Yunta et al. 2011). Contig comp56807_c0_seq1, which was overly expressed in Ollague, is a potential functional protein that has high homology with a putative chaperone-1-like protein that is known to function in response to high light intensity and heat in *Zea mays* (Lopez-Frias et al. 2011). Other gene products that showed high sequence homology in our BLAST search were comp44207_c0_seq1 and comp38166_c0_seq1; both of

which are associated with secologanin synthase activity which belongs to the family of oxidoreductases (Yamamoto et al. 2000). We also identified signal transducers, such as heat-shock proteins (contig comp42870_c0_seq1) which appeared in greatest abundance in the high water stress treatment Ollague samples. These heat shock proteins may be related indirect stress brought about by water deficiencies such as oxidative stress due to accumulation of reactive oxygen species (ROS; Martindale and Holbrook 2002). Homologs of both contig comp31939_c0_seq1, which was more expressed in Ollague, and contig comp56526_c0_seq1, which was more expressed in Ingapirca, respond to oxidation stress and showed less expression in the high water stress treatment suggesting a potential role in protecting the plant from the effects of ROS. Finally, a pathogenesis related gene protein (contig comp1469_c0_seq1) was overly expressed in Ollague under the severe drought-like condition. Kitajima and Sato (1999) reported that pathogenesis related proteins are induced by hormones or ROS in response to biotic stress; while more recently Przymusinski et al. (2004) showed that an increase of these proteins can be induced as a response to various abiotic stresses.

CONCLUSIONS

We report the first large transcriptome analysis of water stress in *C. quinoa*. Our greenhouse results support previous observations that Salares ecotypes display greater tolerance to drought-like stress conditions than the valley ecotypes, as determined by several phenotypic measurements: growth rates, photosynthetic rates, stomatal conductance, and stem water potential. A *de novo* assembly of the root transcriptome generated across two varieties and four water treatments generated 20,337 unique transcripts. Gene expression analysis of the RNA-seq data identified 462 putative gene products that showed differentially expression based on treatment and 27 differentially expressed putative gene products based on variety x treatment interaction, including significant increases in expression in the root tissue in response to increasing water stress. BLAST searches and gene ontology analysis show an overlap with drought tolerance stress and other abiotic stress mechanisms, suggesting an overlap in response to abiotic and biotic stresses.

REFERENCES

Altschul SF, Madden TL, *et al*. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389-3402

Bartels D, Nelson D (1994) Approaches to improve stress tolerance using molecular genetics. Plant Cell Environ 17:659-667

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B 57:289-300

Blüthgen N, Brand K, *et al*. (2005) Biological profiling of gene groups utilizing Gene Ontology. Genome Inform 16:106-115

Boyer, JS (1982) Plant productivity and environment. Science 218:443–448

Buchanan CD, Lim S, *et al*. (2005) *Sorghum bicolor's* transcriptome response to dehydration, high salinity and ABA. Plant Mol Biol 58:699–720

Choné X, Van Leeuwen C, *et al.* (2001) Stem water potential is a sensitive indicator of grapevine water status. Annals of Botany 87:477-483

Cook ER, Seager R, *et al*. (2007) North American droughts: Reconstructions, causes and consequences. Earth Sci Rev 81:93-134

Dubouzet JG, Sakuma Y, *et al.* (2003) *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. The Plant Journal 33:751-763

Earl HJ (2003) A precise gravimetric method for simulating drought stress in pot experiments. Crop Sci 43:1868-1873

Ekblom R, Galindo J (2010) Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity

FAO (2011) Quinoa: ancient culture to contribute to global food security towards the declaration of 2013 as "International Year of Quinoa". Plurinational State of Bolivia Ministry of Foreign Affairs: Annex 1 www.fao.org/docrep/meeting/022/mb442e.pdf

González JA, Bruno M, *et al.* (2011) Genotypic variation of gas exchange parameters and leaf stable carbon and nitrogen isotopes in ten quinoa cultivars grown under drought. Journal of Agronomy and Crop Science 197:81-93

Gorantla M, Babu PR, *et al.* (2005) Functional genomics of drought stress response in rice: Transcript mapping of annotated unigenes of an indica rice (*Oryza sativa* L. cv. Nagina 22). Current Science 89:496-514

Götz S, Garcia-Gomez JM (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research 36:3420-3435

Grabherr MG, Haas BJ, *et al.* (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol 29:644-652. doi: 10.1038/nbt.1883. PubMed PMID: 21572440

Hernández I, Alegre L, Munné-Bosch S (2004) Drought-induced changes in flavonoids and other low molecular weight antioxidants in *Cistus clusii* grown under Mediterranean field conditions. Tree Physiol 24:1303-1311

Jacobsen SE (2003) The Worldwide potential for quinoa (*Chenopodium quinoa* Willd.) potential. Food Rev Int 19:167–177

Jacobsen SE, Mujica A, Jensen CR (2003) The resistance of quinoa (Chenopodium quinoa Willd.) to adverse abiotic factors. Food Rev Int 19:99–109

Jacobsen SE, Monteros C, *et al.* (2007) Frost resistance mechanisms in quinoa (*Chenopodium quinoa* Willd.). European Journal of Agronomy 26:471-475

Jacobsen SE, Liu F, Jensen CR (2009) Does root-sourced ABA play a role for regulation of stomata under drought in quinoa (*Chenopodium quinoa* Willd.). Scientia Horticulturae 122:281-287

Kawaguchi R, Girke T, *et al.* (2004) Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in *Arabidopsis thaliana*. Plant J 38:823-839

Kitajima S, Sato F (1999) Plant pathogenesis-related proteins: Molecular mechanisms of gene expression and protein function. Journal of Biochemistry 125:1-8

Koyro HW, Eisa SS (2008) Effect of salinity on composition, viability and germination of seeds of *Chenopodium quinoa* Willd. Plant Soil 302:79-90

Kreps JA, Wu Y, *et al*. (2002) Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. Plant Physiol 130:2129-2141

Lopez-Frias G, Martinez LM, *et al.* (2011) Role of HSP101 in the stimulation of nodal root development from the coleoptilar node by light and temperature in maize (*Zea mays L.*) seedlings. J Exp Bot 62: 4661-4673

Lowry A, Nixon R (July 25, 2012) Severe drought seen as driving cost of food up. New York Times http://www.nytimes.com/2012/07/26/business/food-prices-to-rise-in-wake-of-severe-drought.html?_r=0

MacCutchan H, Shackel KA (1992) Stem water potential as a sensitive indicator of water stress in prune trees. Journal of American Society of Horticulture and Science 117:607-611

Martindale JL, Holbrook NJ (2002) Cellular response to oxidative stress: Signaling for suicide and survival. Journal of Cellular Physiology 192:1-15

Mason SL, Stevens MR, *et al*. (2005) Development and use of microsatellite markers for germplasm characterization in quinoa (Chenopodium quinoa Willd.). Crop Sci 45:1618-1630

Maughan PJ, Turner TB, *et al*. (2009) Characterization of Salt Overly Sensitive 1 (SOS1) gene homoeologs in quinoa (Chenopodium quinoa Willd.). Genome 52:647-657

Metzker ML (2010) Sequencing technologies – the next generation. Nature Reviews Genetics 11:31-46

Morales AJ, Bajgain P, *et al*. (2011) Physiological responses of Chenopodium quinoa to salt stress. International Journal of Plant Physiology and Biochemistry 3:219-232

Mortazavi A, Williams BA, *et al*. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods 5:621-628 doi:10.1038/nmeth.1226

Nagata M, Yamashita I (1992) Simple method for simultaneous determination of chlorophyll and carotenoids in tomato fruit. Nippon Shokuhin Kogyo Gakkaish 39:925–928

Ogata H, Goto S, *et al*. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucl Acids Res 27:29-34

Pillco R, Uvo CB, *et al.* (2007) Precipitation variability and regionalization over the Southern Altiplano, Bolivia. International Journal of Climatology pp. 149-164

Prado RE, Boero C, *et al*. (2000) Effect of NaCl on germination, growth, and soluble sugar content in Chenopodium quinoa Willd. seeds. Bot Bull Acad Sci 41:27–34

Przymusinski R, Rucinska R, Gwozdz EA (2004) Increased accumulation of pathogenesis-related proteins in response of lupine roots to various abiotic stresses. Environmental and Experimental Botany 52:53-61

R Development Core Team (2012) R: A language and environment for statistical computing, version 2.15.1. website: http://www.R-project.org

Risi C J, Galwey NW (1984) The Chenopodium grains of the Andes: Inca crops for modern agriculture. Adv Appl Biol 10:145-215

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140

Ruales J, Nair BM (1993) Content of fat, vitamins and minerals in quinoa (Chenopodium quinoa, Willd.) seed. Food Chem 48:131-136

Ruas, PM, Bonifacio A, *et al*. (1999) Genetic relationship among 19 accessions of six species of Chenopodium L. by random amplified polymorphic DNA fragments(RAPD). Euphytica 105:25–32

Shinozaki K, Yamaguchi-Shinozaki, *et al.* (2003) Regulatory network of gene expression in the drought and cold stress responses. Current Opinion in Plant Biology 6:410-417

Shinozaki K, Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. Journal of Experimental Botany 58:221-227

Tapia M, Gandarillas H, *et al*. (1979) Quinua y kañiwa: Cultivos andinos. CIID-IICA. Bogotá, Colombia

Tapia ME (1980) Origen, distribución geográfica y sistemas de producción de la quinua. En: Reunión sobre genética y Fitomejoramiento de la quinua. PISCA-UNTA-IBTA-IICA-CIID. Puno, Perú

Tapia M (1997) Cultivos andinos sub explotados y su aporte a la alimentación. Segunda Edición. F.A.O. Santiago-Chile

Trognitz BR (2003) Prospects of breeding quinoa for tolerance to abiotic stress. Food Reviews International 19:129-137

USDA (2012) U.S. Drought 2012: Farm and food impacts. Retrieved September 28, 2012, from http://www.ers.usda.gov/newsroom/us-drought-2012-farm-and-food-impacts.aspx

Vacher JJ (1998) Responses of two main Andean crops, quinoa (*Chenopodium quinoa* Willd.) and papa amarga (Solanum juzepezukii Buk.) to drought on the Bolivian Altiplano: Significance of local adaptation. Agric Ecosyst Environ 68:99–108.

Valencia-Chamorro SA (2003) Quinoa. In: Caballero B.: Encyclopedia of Food Science and Nutrition. Vol. 8. Academic Press, Amsterdam: 4895–4902

Vera JC, Wheat CW, *et al*. (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Molecular Ecology 17:1636-1647

Wilhite DA, Hayes MJ, *et al.* (2000) Planning for drought: Moving from crisis to risk management. Journal of American Water Resources Association 36:697-710

Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26:873-881

Yamamoto H, Katano N, *et al.* (2000) Secologanin synthase which catalyzes the oxidative cleavage of loganin into secologanin is a cytochrome P450. Phytochemistry 53:7-12

Yunta C, Martinez-Ripoll M, Albert A (2011) SnRK2.6/OST1 from *Arabidopsis thaliana*: Cloning, expression, purification, crystallization and preliminary X-ray analysis of K50N and D160A mutants. Acta Cryst 67:364-368

TABLES

Table 1.1 The different water treatments used in this study with reported weights maintained (g) throughout the experiment and average soil moisture tension (bars) that were measured once treatments reached their final maintained weights.

Treatment	Weights maintained (g)	Average Soil Moisture Tension (centibars)
1	293	-10.33
2	220	-19.00
3	190	-34.17
4	160	-176.17

Table 1.2 ANOVA results of various measurements conducted during the greenhouse study. ANOVA analysis was performed using Program R 2.15.1 (R Development Core Team 2012).

		Df	Mean Sq	Pr(>F)
Height				
	cult	1	107.32	3.01e-05 ***
	treat	3	24.96	0.00212**
	cult:treat	3	1.48	0.71704
	Residuals	16	3.25	
Photosynthetic Rate				
	cult	1	19.03	0.17591
	treat	3	51.76	0.00892**
	cult:treat	3	4.21	0.72475
	Residuals	16	9.49	
Stomatal Conductance				
	cult	1	0.00003	0.92036
	treat	3	0.023674	0.00265**
	cult:treat	3	0.00194	0.62502
	Residuals	16	0.00324	
Stem Water Potential				
	cult	1	0.776	0.00476**
	treat	3	2.8115	1.44e-07***
	cult:treat	3	0.2937	0.02551*
	Residuals	16	0.0725	

--

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 1.3 Overrepresented GO terms (58) in drought regulated genes. The test group is represented by the regulated genes with at least one GO assigned (462 sequences). The reference group is constituted by the non-regulated transcripts with at least one GO assigned (20,337 sequences). The cut-off threshold was FDR < 0.05.

GO-ID	Go description	FDR	Test (%)	Reference (%)
GO:0016798	hydrolase activity, acting on glycosyl bonds	8.35E-08	12.0	2.2
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	1.63E-07	11.1	2.0
GO:0005975	carbohydrate metabolic process	5.19E-06	21.7	8.3
GO:0042221	response to chemical stimulus	1.08E-04	33.6	18.2
GO:0050896	response to stimulus	1.15E-04	48.4	30.9
GO:0009719	response to endogenous stimulus	1.71E-04	18.4	7.5
GO:0044262	cellular carbohydrate metabolic process	1.71E-04	17.5	6.9
GO:0000272	polysaccharide catabolic process	1.71E-04	5.5	0.7
GO:0005976	polysaccharide metabolic process	2.42E-04	11.5	3.5
GO:0006979	response to oxidative stress	2.83E-04	10.1	2.8
GO:0016161	beta-amylase activity	3.58E-04	2.8	0.1
GO:0009505	plant-type cell wall	3.77E-04	8.3	2.0
GO:0006950	response to stress	5.27E-04	30.4	16.9
GO:0016684	oxidoreductase activity, acting on peroxide as acceptor	5.88E-04	4.6	0.6
GO:0006804	peroxidase reaction	5.88E-04	4.6	0.6
GO:0004601	peroxidase activity	5.88E-04	4.6	0.6
GO:0009725	response to hormone stimulus	0.0013	16.1	6.9
GO:0016491	oxidoreductase activity	0.0015	17.5	7.9
GO:0016052	carbohydrate catabolic process	0.0026	9.2	2.8
GO:0010033	response to organic substance	0.0035	18.9	9.2
GO:0016209	antioxidant activity	0.0043	4.6	0.8
GO:0016160	amylase activity	0.0044	2.8	0.2
GO:0005618	cell wall	0.0055	12.4	5.0
GO:0009409	response to cold	0.0055	9.2	3.1
GO:0044264	cellular polysaccharide metabolic process	0.0055	9.2	3.1
GO:0030312	external encapsulating structure	0.0061	12.4	5.1
GO:0009628	response to abiotic stimulus	0.0064	24.0	13.5
GO:0010480	microsporocyte differentiation	0.0064	1.8	0.0
GO:0003824	catalytic activity	0.0098	60.4	46.8
GO:0006970	response to osmotic stress	0.0098	11.1	4.4
GO:0009251	glucan catabolic process	0.0098	2.8	0.3
GO:0005983	starch catabolic process	0.0098	2.8	0.3
GO:0044247	cellular polysaccharide catabolic process	0.0098	2.8	0.3
GO:0009311	oligosaccharide metabolic process	0.0119	7.8	2.5
GO:0005985	sucrose metabolic process	0.0119	7.4	2.3
GO:0005984	disaccharide metabolic process	0.0133	7.4	2.3
GO:0006073	cellular glucan metabolic process	0.0154	8.3	2.9
GO:0044042	glucan metabolic process	0.0159	8.3	2.9
GO:0009737	response to abscisic acid stimulus	0.0162	9.2	3.4
GO:0051740	ethylene binding	0.0167	1.4	0.0
GO:0072328	alkene binding	0.0167	1.4	0.0

GO:0080167	response to karrikin	0.0213	4.6	1.0
GO:0009753	response to jasmonic acid stimulus	0.0242	5.1	1.3
GO:0009266	response to temperature stimulus	0.0268	10.6	4.5
GO:0016830	carbon-carbon lyase activity	0.0336	3.7	0.7
GO:0071369	cellular response to ethylene stimulus	0.0336	2.8	0.4
GO:0006026	aminoglycan catabolic process	0.0336	2.8	0.4
GO:0055114	oxidation-reduction process	0.0346	15.2	7.9
GO:0008295	spermidine biosynthetic process	0.0346	1.4	0.0
GO:0016137	glycoside metabolic process	0.0355	7.8	2.9
GO:0009651	response to salt stress	0.0376	9.7	4.1
GO:0016829	lyase activity	0.0376	6.5	2.1
GO:0009814	defense response, incompatible interaction	0.0397	4.1	0.9
GO:0006595	polyamine metabolic process	0.0404	1.8	0.1
GO:0005982	starch metabolic process	0.0407	6.9	2.4
GO:0046351	disaccharide biosynthetic process	0.0409	2.8	0.4
GO:0009873	ethylene mediated signaling pathway	0.0491	2.3	0.3
GO:0005996	monosaccharide metabolic process	0.0498	7.8	3.0

FIGURES

Figure 1.1 Whisker plots of heights, photosynthetic rate, stomatal conductance, and steam water potentials measured. Drought treatments were classified by total weight maintained throughout the experiment of a given sample where treatment 1 = 293 g, treatment 2 = 220 g, treatment 3 = 190 g, and treatment 4 = 160 g total weight. Inga represents the Ingapirca variety while Olla represents the Ollague variety. A: Heights; B: Photosynthetic rate; C: Stomatal conductance; D: Stem water potential.

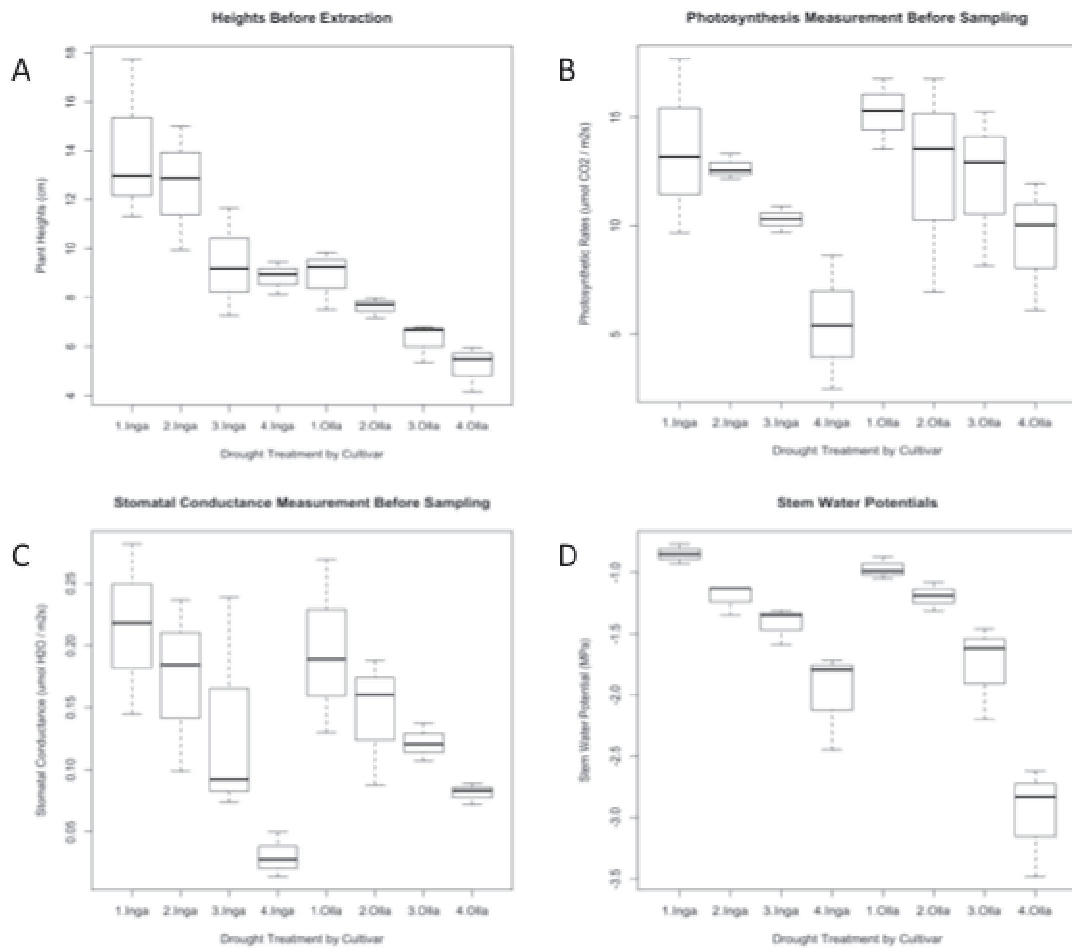


Figure 1.2 Histogram depicting the unique contig lengths of 454, Illumina, and Sanger sequencing reads.

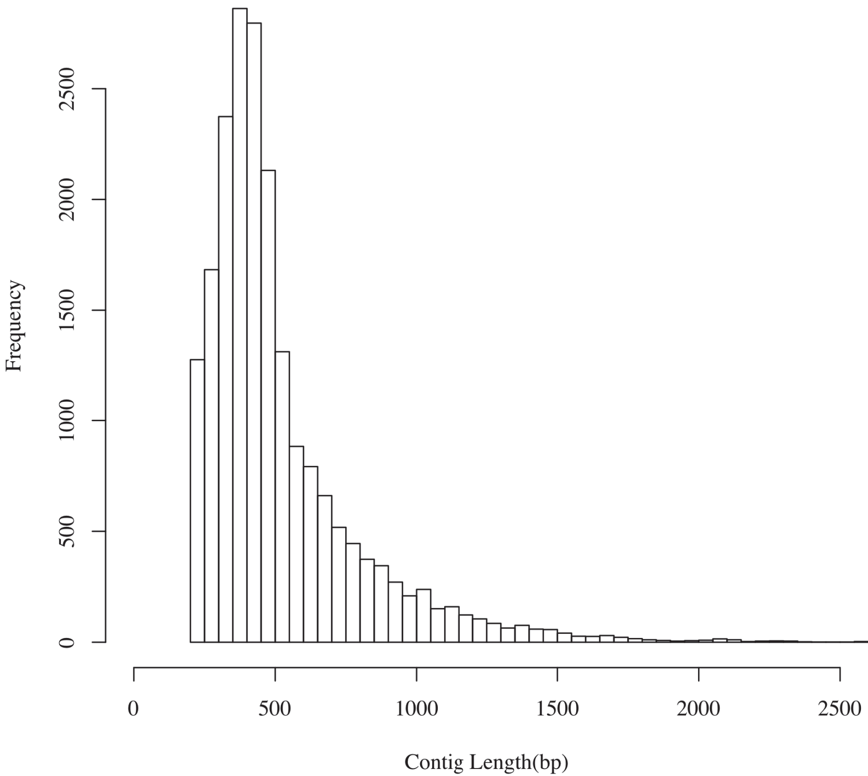


Figure 1.3 Distribution of 462 DE genes that were differentially expressed based on treatment according to GO annotation. A: Biological Process (BP); B: Molecular Function (MF).

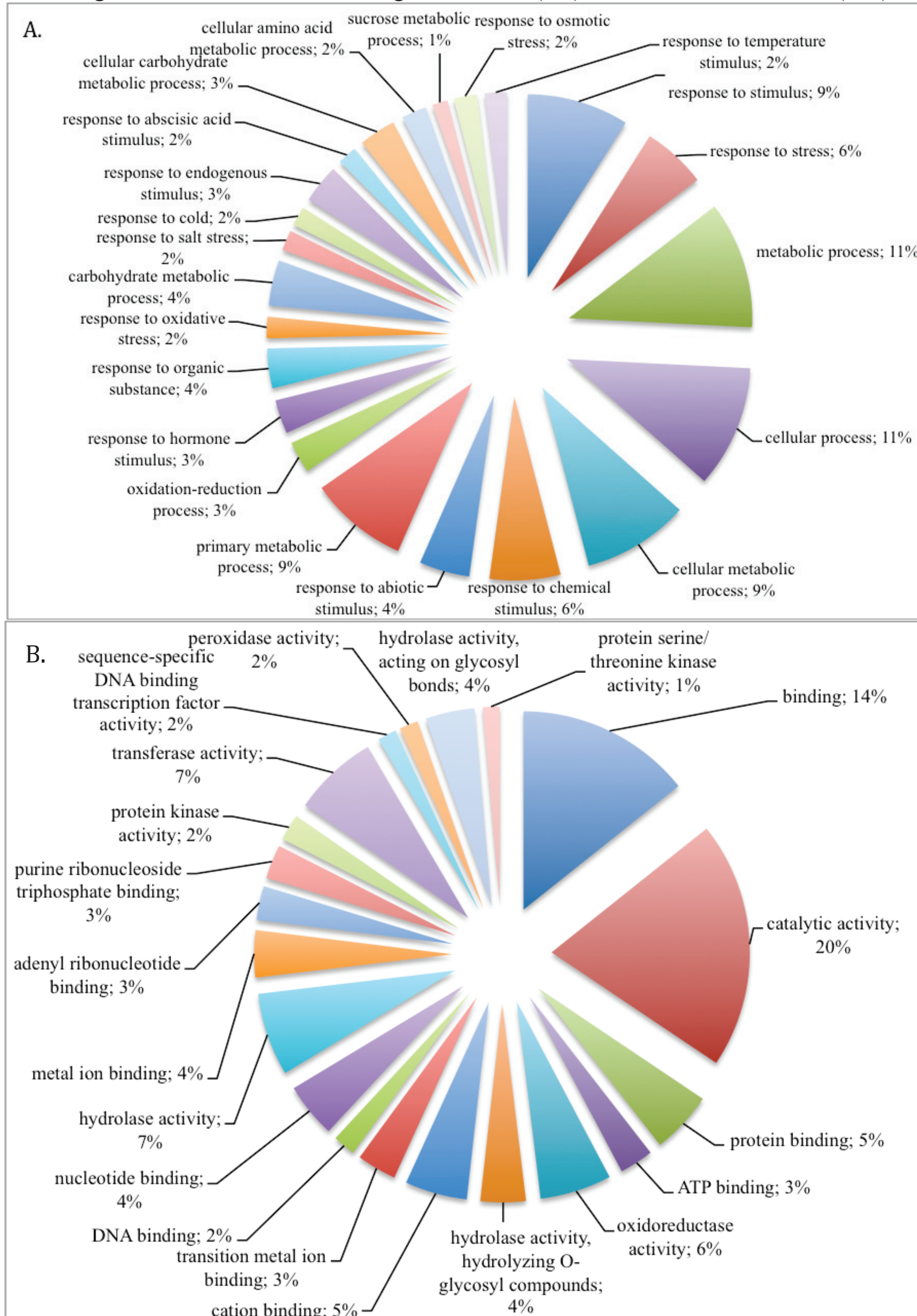
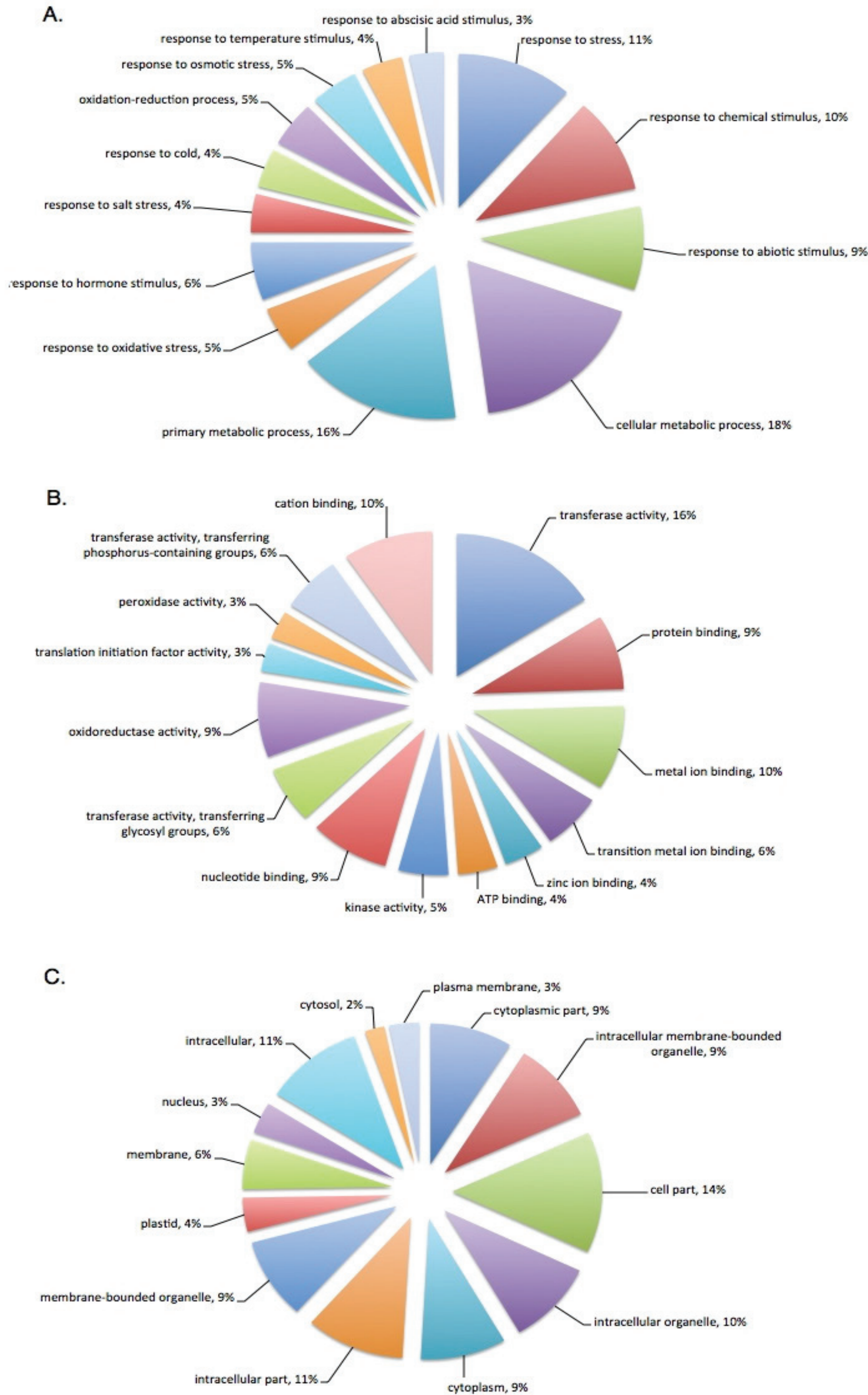


Figure 1.4 Distribution of the 27 DE genes that were differential expressed based on the variety x treatment interaction in the three main GO categories. A: Biological process (BP); B: Molecular function (MF); C: Cellular component (CC).



APPENDIX

Appendix 1.1 Script created and used to identify differentially expressed genes using edgeR.

```
#!/fslhome/jtpage/bin/Rscript
#PBS -l nodes=1:ppn=4,pmem=500mb,walltime=12:00:00
#PBS -N july9
#PSSS -l nodes=1:ppn=32:bigmem,pmem=1gb,walltime=12:00:00
#PBS -m bea
#PBS -M jaraney@gmail.com
library("edgeR")
INFILE="/fslhome/bcboy4/compute/scripts/edgeR/collapsedFilteredCountsJuly9.txt"

counts <- read.table(INFILE, header=T, row.names=1)
no_1_3O_counts <- cbind(counts[1:14],counts[16:24])
groups <- factor(c(1,1,1,2,2,2,3,3,3,4,4,4,5,5,6,6,6,7,7,7,8,8,8))

design <- model.matrix(~0+groups)

dge <- DGEList(counts=no_1_3O_counts, group=groups)
dge <- calcNormFactors(dge)
dge <- estimateGLMCommonDisp(dge, design)
dge <- estimateGLMTrendedDisp(dge, design)
dge <- estimateGLMTagwiseDisp(dge, design)

fit <- glmFit(dge, design)

lrt.m1 <- glmLRT(dge, fit, contrast=c(1,1,1,1,-1,-1,-1,-1))
lrt.m2 <- glmLRT(dge, fit, contrast=c(1,-1,0,0,1,-1,0,0))
lrt.m3 <- glmLRT(dge, fit, contrast=c(1,1,-2,0,1,1,-2,0))
lrt.m4 <- glmLRT(dge, fit, contrast=c(1,1,1,-3,1,1,1,-3))
lrt.m5 <- glmLRT(dge, fit, contrast=c(1,-1,0,0,-1,1,0,0))
lrt.m6 <- glmLRT(dge, fit, contrast=c(1,1,-2,0,-1,-1,2,0))
lrt.m7 <- glmLRT(dge, fit, contrast=c(1,1,1,-3,-1,-1,-1,3))

write.table(topTags(lrt.m1 , n=10000),
file="/fslhome/bcboy4/compute/scripts/edgeR/allModels/redoNov5/m1.results", sep="\t", quote=F)
write.table(topTags(lrt.m2 , n=10000),
file="/fslhome/bcboy4/compute/scripts/edgeR/allModels/redoNov5/m2.results", sep="\t", quote=F)
write.table(topTags(lrt.m3 , n=10000),
file="/fslhome/bcboy4/compute/scripts/edgeR/allModels/redoNov5/m3.results", sep="\t", quote=F)
write.table(topTags(lrt.m4 , n=10000),
file="/fslhome/bcboy4/compute/scripts/edgeR/allModels/redoNov5/m4.results", sep="\t", quote=F)
write.table(topTags(lrt.m5 , n=10000),
file="/fslhome/bcboy4/compute/scripts/edgeR/allModels/redoNov5/m5.results", sep="\t", quote=F)
write.table(topTags(lrt.m6 , n=10000),
file="/fslhome/bcboy4/compute/scripts/edgeR/allModels/redoNov5/m6.results", sep="\t", quote=F)
write.table(topTags(lrt.m7 , n=10000),
file="/fslhome/bcboy4/compute/scripts/edgeR/allModels/redoNov5/m7.results", sep="\t", quote=F)
```

Appendix 1.2 Sequences of the 27 differentially expressed genes based on variety x treatment interaction that were identified via edgeR.

>comp1018_c0_seq1
TGACAACTTACCTTCAATTGTTTTTGTGTTGCTTCCTTTGTAAGTGTTGGCTGATTTAGGATACTCTCA
GTAACTGAGTGTACTACCTTACTAATATCAGCATCAACACCCTAGTTTTAGCATCTTTCTCTCCCTGAT
TATCAGCATCAACATTCTGATTATCAGCATCTTTCTCTCCTTTCTCTTCATTCTTCTCTTCTTTCTTTTTAT
TATTCTTTTTTTGCTCCACCTTCTTTAGTTTTTTTTATAATACAAAGTAATATTATTTATTCATCAACCC
GGAACAAATATTGTACACAGTACACATGCTCACACCCTCAAGGCCTCAATTATTCTCTATAGAAAAAG
TGTGAAACCCTCACAAATTAAAAAATTAAATAATTAAACAAATAAAAGAAAGGAATGAAGCTAAAAT
TAGACTTTTGATACAAACAAACTTAAAAATTACACAAGAATTTATTCATGTCTACGAGTACAATACAA
GTATGCAACACAGTATTTTAATTATCTTTGATTTTGAATTAATTCAATAAATGCTTATCAGATCAGCTT
ATCGCGTCATGCCTCCTGCCATTGATGGAGGTACCATCCCCTGGTTTTTCCTCTCCTCATTACCTTGCCA
TCCGAGGGCAAGATCATAGCCACGATGTTGAAGTTCACGGTACTCTTGGCACAAGGCACAAGCTTCAC
ACCAAAAGTGAGTACAACAATCTCCACAACTATTCTCTGGCATTCCAAGTTGAGCCTTCAACTTTGAAC
GATAAGTGCATGAATATATCCATTGACATCCAGTCAATACCAAAAGTAGTGAATACAATGCTCCACTT
ACACCACATGATGATGAACCTTTGTCAACAATCTCTGCTATACGTCCAAAAGTGATGCATGGGCACCA
ACATGTCATACAACAAAGAGAACAATCAGAGAAACAATCACAAAGACCACTAGAGAACTCTCCAGCA
ATTGCAACTGATTGATTTGGTGGAGCTGAAATTGGAACACCAATTGCATGTTGATGATCTTGCTTCATG
TTTTTATTATCTAGAAATAATAATTGTACACGAAAAATAAATGTAACGTTTATTAAGTTTCTACTTTGT
AAAAAAATG
>comp1469_c0_seq1
GGAGTACCACATATATAAAATTAAGTGAATTTCACTAAAAAAAGATCTAGTGAGCTTATTCCAATTCA
CTCTTTATTATAACAACATGAAGTTAAAGGAAAGGTTTAAAAGAGATACATATATATCATATTGATCA
TATATTATTATTATTATTATTCCAAGGTTGTTCTCAGTTTATTGAGAATTAGTCGTTAATTTTATCAAAA
CATTTAGTAAGGTTTCTGGCCAACCCAATTGCCTGGAGGATCATAGTTACATGTAACGAAATAACCTC
CATTGTTACATTGAACCCTAGCACATCCTAGACGAACCGAGTTACGCCAAACCACTTGAGTATAA
>comp1792_c0_seq1
TTTAGTTTAGGGTAAGGGGAAAGTTACAAAAAAAAAGGAAAAGGGTAAAAGTTAAAAAGCCAAAAA
GAAGAAAGGAACAACGCTTCTGACCTTGCTGTCATCGCGAAATCTTTCTTCTGAAATGAAGCAGGCCG
AGGTTGAGAAAATAAAAACCACACAAATTTGATTCAACAATGGATATAAAAGGCTTCTTCAATCCAAT
CATATGATTGATTCGACAATGGATATGTGAGGGTTTAATCAAATCATATTGTGACAATGTGTAAAAAA
TGTACACTTGCCGGTGAGCTGTAGATAAGTGGTTCAATCAGGTAAAGGTGACTTAATGGATCAATGCT
GCTATAATTGTAGGTAAATCTAATGGAAGGAAAACCAAAAATAGGACAATAAAAAAATAAAATACAA
AAAAAGGGAAACTTGGACTCTTTTCTTATGAACAGAGTTGTAACTAAGGGCTAGCATTGGCGACCCCT
GCGGCAGGAGAGAGGGCCTGCACTTGCTGTGATGGTTCCGACCATGGAAGAGGACTTAGCTCCCAAG
CTTGAGGCTCTTGCGTAACTAGCTGAGGCTCCAGCTCCGGATGGGGTGAAGTATGCACCGTTCAGCAG
TAGATCTCCCTCAGACCTCCAGTTCCAACCCTTCCATTTG
>comp1839_c0_seq1
ATCCATTGCAGATAAGAGAGCTCTTAGAATTTCGATACCGGAACCCGTAACTCTTCCCCATAATGACC
GTCTCAGCGGATGAAATTCATAAAACCCAACAAAACCCAGAAACAGAATTAGAAGATGCATTCAAAA
ATCTGGGAATCAAAACAAGCGAAGCATCTCCAAGTGGAGTGCTGGATATTCCAATTTCAGATGGTTCC
TCTGTTTCTGACAGCCCAAAAAGCAGCAGTAGCAGCAGCAGCAGCAGTCCCTCAGAAAAACCTTATAC
CCCTCAAAAGGACAGACGAAGGCGAAAGCGAAAGCGAAAGCAGAGGCGGAGGAGGAGGTGGAGGTG
GAGGCAAAGGAGGAGGAGCAACAATATTCAATCCATTGACTTATGGGATTCAATTCAAGGGATTTTTT
GACATAATGAAGAGGAAATCAGGGGTTAGAAGATTGTCCACTACTACTATAAATTTGCTTAGTAACTA
TGATCTCTCTGCGAAGAGCCTGCGCAAGAAATTGGCTAGGATTCGCAGCGCCGAGGATAGTTTCGATT
GCGGTCTTCTATCGGCAGCTAAGCCTTCTTGGAGGAACTTTAAGTATTCAGAGCTTGCTGTTGCTACAG
ACCATTTTAGTCCTGGGAGGTTGCTAGGGAAAGGGG
>comp2075_c0_seq1
GCTTCGTATAAAATTAACTCAAAGTTGATTTTTGCCCAATTTATCTTGAAATTTCATTACATCATCCACT
CTAGATTCAATTACCAATAGTTGAAATCCGAATAACGTACATGATCAAATTATTCCACATTCAAAGATT
AATTAAAGTTTGCAAACACTTGAGTAAAACAAAATTTTAACTAACAACAATCAAACTAAAAAACAAA
GGACTAATTAGTGTTGAACATAAACATCATATTCAACAACACTAGCACCACTTTTATCCAACATTGTAA

TAGTTTTAGCTTGTGCATACCCTGTTGCAAACCTAAACACACCACTTCCACCAACAACACCCATTTCTC
GTACTTTATTCATCACTTGATTTCTTCCTAACACACTTAGAGTACTTCCATTAAATTCACCCTCTATAAA
CACCAAGTTCATTACCATCAATAGCCCCGGTTGCTTTTGGTCCGCAAACCCGTACATTCCTTGCGCGTA
CCCCACGATTTTCGAGTCTTTTTCGGGCCCTTCGGTCAAAGGATCATCGATCATAACAAGCGCGCCAAA
ACCACTTGAGGTTTTGTCCGTGTTAGGGGCTTGGGCAATACGTAGCGCAGTCTGGTTAGACCCTTGGAC
GTAGTCATGGAAGTAAAAGTGTAGATGGGTTAAGTTTGTTGTTGTTTAAGACCCTATTTCTTCTTTTGTT
ATTGGATTTTGAGAATGCATGATTTGTGGATGAAGAACTCAAGATTAGTAAGAAAGAAAGAAGGAAA
>comp3355_c0_seq1
AAAAACACAAATCAAAGTAAAGTGTATTTTTCGCGATATCTAGCTTAACTAACTAGCAAAGGCTGAAG
AAATGGCTAAGTCTACTTACTTCGTTAGCCTCCTTGCACTTTCTCTCATTTTCATGCTTGTAGTGAACAC
TGAAATTCAAATGGCGGAAGCAAAAACTTGTGGATGGTTGTACAAAGCTTGGTGGACAGATAACTGCT
CCTACAAGTGCAAGAACGTCTGGGATGCCTACTATGGAAAGAAAGAGTGGTACCAAGGAAAATGGGC
TTGCTTCTGTTACTATGGACCTTGTTAAGATAC
>comp4060_c0_seq1
GTCGATTCGTAAGAACGGAGGAATCATCTCTTTAAAAACACGAAATTACAAGTGCATACATTGCAATT
TTGCTAGCAAAGTTACAAGGAGTATGTCGTGGAGTGCTATCCACATGGCCACAAATGCACGAGTCTTT
TTTTAACGTACCACAACCACATTTTGCAATAACAAGAACGTACAACTAACAACATATATTAAGACGAG
ATCTTAACAATTTCCCTTGAAATGGGATAAATTTAAAAATGATCAATCACTTGAAGAGACCTTTAGCCT
TGTCCATCAAACTCTCCTCTGGTGGTGGAGGAGCGGGTTTAACAAAATCCCCGGCCTTGTTAAGAGCA
TCTCCAACGGTGTTCTTGATGTTTTCAGCCATTTCTCCGGCAGATTTGGAAATTTGAGCCGTGTCAATCT
TGTCAGTCTGTTGGGTGACGGTTGTTTTCACTGTCTCCCCAGCATTCTTGGCACCATCTACTATGTTAGT
CATGGTTGTTTTCTCAGCCATTGCTTAGTATGAGAATATTTGGCTAAGGTTTTACGGGTGTG
>comp4773_c0_seq1
GACCAAAAACATGAAAGTTGGTCATGTGCTATTATTCATGGTTGCCATTGTTGTGGTAGGCACCGTGA
CACAGCCAACTTTGGCAGTGCGTCGAACTCCGCTCACCGCAAACACCATACAACAGAATCCCCTCTCA
AACATCATGCTCGAAACGGTCAACAATGGAGCTGCTGGTTGGTGCCAGCAAAGGAACGAAATCTGCA
ATGCTTACTTTAAGTGTTGTCCAGACCTTGTATGCTACACTCGTGGAACTTGTGGAGATCTGGTCTAAC
AACTTTGCACTGCGGTCGGCATCGAGATACTTAAGTAATTATAGAATGTATGCTGTTTAAGAATAATAT
ACGATGTATAAGTTAATATGAAGCTTTGGGTATACCATGCTTGTCAAGTTGTCAATAATGCATGTAACG
AAACGAATAAGG
>comp4911_c0_seq1
GCATTTTATAGCTTTTCTTACCATAATTAGCCTTACTTCGCAAATTAGCTCAATTTTAGCAATCCATCGT
CCAATATACACAGATTTTCTTAAAAGATCATGCAAGTCAACACTTTATCCTGAACTTTGCTACAACTCC
CTTTCAACCTATGCAAGTGACATCCAAAAAAATCCTCACATGCTTGCAAATGCCGCCCTCTTGGTGACC
TTAGCCACCACCAGGACCACAACCGCGATGATATCTAGGATGTCGAATGGCCCGGGAATACGGCCAA
GGGAGGCAGGAGCTATGAAGGACTGTTTGGAAGAACTAAATGATGCAGTAAGTGAGCTTACAAAAAT
CCATCAAGGAGATGAATCATATTAAGGATAGTAAGAGAGTAGGGCTTATGATGAATGATATACAAAC
TTGGGTGAGTGCTGCATTGACCGATGCGGACACGTGTACGGCTGGATTTGATGGAAAGGTCATGGATG
GAAAGTTGAAGAGTGTTGTGAGAGGGAAAATTGTGAC
>comp5517_c0_seq1
AAAAAAGCTTCGCACTTTTCCATAAAAATGGCGAGTTGGGTACTCTCAGAATGTGGTTTAAGGCCTCTT
CCTAAAACCTACCCTAAACCCAGAACAGGGTTTTCCTCAACTTTCAATAACAATCCCTTAAAAATCAA
AACTTCCCCTGTTTTTCCCCCTGTTTTATCAGATCTGAAATCAATTTCTTGTTGTTCTAATTCTAGAGAA
AAAAAGTGGGAGGTTAATGTTAGTGCCCCATTAAGGGTATCTAACCCACCTCCAATTGAAGAAGATAA
TGATGAAAATGATAATAAAGTTATTGATGAATTTAATCCTGGTGCACCACCTCCATTCAAATTGGCTGA
TATTAAGAATGCAATTCCTAAACATTGCTGGGTTAAAAATCCTTGGAAATCAATGAGTTATGTTTTTAG
GGATATTGCTGTTGTTTTTGGATTAGCTGCTACTGCTGCTTATTTCAACAAATGGATTGTTTGGCCTCTT
TATTGGATTGCTCAGGGTACCATGTTCTGGGCTCTCTTTGTTCTTGGTCATGACTGTGGTCATGGAAGC
TTTTCTAATAATCCAAAGCTTAATAGTGTTGTGGGCCACTTGCTTCACTCTTCAATCCTAGTTCCTTATC
ATGGATGGAGAATTAGTCATAGGACTCATCATCAAAACCATGGGCATGTTGAGAATGATGAATCATGG
CATCCAATGTCTGAAAAACTAAACAATGGTATGGACTATTTGTCAAAGAAGTTAAGGTTCACCTTTCC
ATTTCCCTTGCTTGCATACCCTATATACCTGTGGGTACGGAGTCCTGGAAAATCAGGTTCTCATTATCA
TCCGGATAGTGATTTGTTCACTCCAA

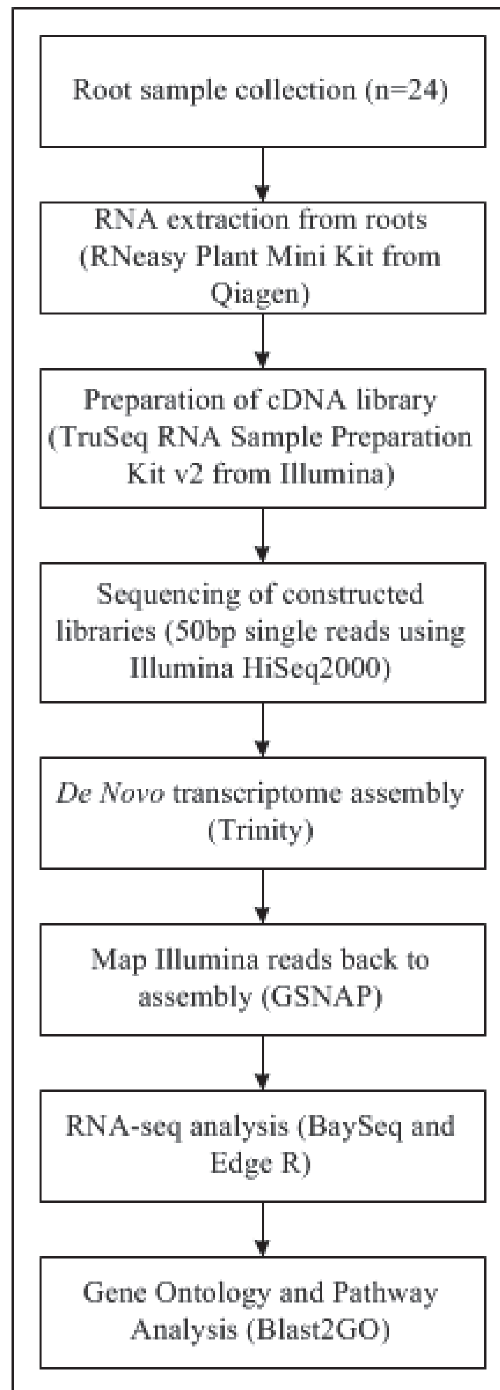
>comp25853_c0_seq1
GTTGAAATTGATAGTGATATTGCCAGCCATTGGATCAACAGTACAAGTGTGTTCGACGAATCTGGAAG
TGTAATAATTGATACCAATATAATCAAAGGCGCCTTTGATTAGTTGCTTTTCAGCTTTTGTGAATTCTG
GTAGCCCCTTCACATTCTTTTTCATTGTTTCTGGGTAGTCCCCACGAATAAGTGGGTCCATGTACCATCC
CAAGGTGAAATCCAACTGCATTTTTGAAGCAGCAAAGTCTTTAGGATCAGTTGAATGTGGCATAAACC
ATTTACATGGGAGCGAAATTCCGATTTTCCCTTTTGTGATTTCTGGTATTTATCTTTGTAGAGTTTCGCG
GCAGAAGCATGAGCTAAGATGATATGGTGGGTGGCAATGAATGAACCTCTTGGATCTTCTGAAGATCG
CTAGGCAATCCATCTGTACCCATACGCCCAAATACATAA
>comp31939_c0_seq1
GTCTGTTGAGTTCTGGTCAAGTAGAAGTTGTCTTGTTCCCATCTACCATTGACTGGGCGACAAGCTCTG
CAGCTGTGATACTAGGCTAGTGTCCATGGTGCTGTCTGGAGCTCCGGTATTAGAAAAATTGAACAACC
TTGAACTAAAAGTAGCACATCTTGCCAATCCTATTGTATGAGCACCTTGTAAGGCCACAACGTCAGTT
AAGTTGAGGCCTTGATCACCAAACTTAGAAATGACGCTGTCTAGGGGATCAAACGGAGATGGAAGGT
TTGAGTTTGCTCCGGTCTGGTTGGCCACCATTCCATCTCTCCTTCCTAGCAATACTCTATATGAATGCCC
TCCACTCTGCAACATTAAATTATTAGTCAAGTACATACGTAACCAAGACTAATGAATTGCGGGCACCT
AGTCACCTACTGCTATTTTGTCCTTTTCATCTGTGATTTTTCTAGATGCACGACAATATGACATACCAAT
GCTATTGAATCGCGGGCAGCTATTGCTAGTATATCAGCACAGGACACAACTCCACTGCATGAACTCTC
CACTGCACTTTTGATCTTGTCCACAACGTCAAACCCTCTTAGAGAATTCTGGTTAGGAAGAGCGAGCTT
TTCACTGTCGCTTCCATCCAGAAGAACAGACCCATCACAACCCTATTCACAAAACCATAAGATATTGTT
GTTAGAGACGAAAATCTG
>comp35446_c0_seq1
ACGTTTTCATCTTCCTTAGACTCTTTGTTCGAGTTCTCCAAACTCGAATACTTGATCGAATCAGTAGCAC
TAGCCGTAGAGTTATGCTGAATCTTACGATAATCTAGACCTAGCTCTAACCCTCCGGCTTCTACCTTTC
TCTTTTATTCAAAGAATCATCACTTTCTTGTTTCTTATTGTTGTTATCATGACTATGTTTCTTGTTAGAGC
TCCAACTACCAAGAGTTAGAGAAACCATCTCAGGCTCTTGGTTTTGATCATCATCTTCATTGTTGTTGT
TGCTCGATGACTTCGTTTCTATTGATTGTTTAGATTGATCCATCGACGAAATTTGCTTCTCTAGAGATTG
ATATTCCTTCACAATTTTCG
>comp38166_c0_seq1
GAGAACAAAGCCTTGGTATACATACGTGTGTGGTATGTACAATATGAAAATAGACCAAAATAATCCTA
TTGAAATGAAATAATAAAATACCTGGATCCAGGAATATAAACTGACTGCATGGCTTGTACAGTAATAG
AAGTATGCTCTCTTAAAAGTTCAAACACTCTCCTTCCTTCTTCAAAACTACTTCCAAATGCAGCTCGAG
AGATTACATCTGCTGACAGATTGTGTAAAAATGACCACATTTCTACCTCTTGAGATCCCGTTTCGGGTG
TCGTCTTCTCCCACTCTTGGATCATCTCATTGCAACTCGCATGAAATGCTGGTAGCATAAGCTTAATTTT
CTCCATATGGAAGGCAGGATTAAGCAATTTCCTTTCTTGGGCCCACTTATGGCCTTCGAGTCTAACCAA
ACCATTAGTAAGCCTCTTCAAAAGTGGAGTCGATATTTTGGAAAATCATAAATC
>comp38447_c0_seq1
GGTTTACAAAATTTCCTACAGTATAAACACAGGGATCTGTTCATAACAGGAGAAAGTTATGCAGGCCA
CTATATTCCTCAACTTGCAAATCTGATGGTCCAATTCAACAAAAAGGAGAAAGTGTTTAACTTGAAAG
GAGTTGCGCTTGGAAACCCGGTTTTGGAGTTTGCTACAGATTTCAACTCGAGGGCAGAGTATTTCTGGT
CACATGGATTAATATCGGATTCAACCTACAAAATTTTTACCAAAGCCTGCAATTATTCAAGATATGTGA
GTGAGTATTATCGAGGATATGTTTCACCAGTATGTTCCAAGGTGATGAGGCAAGTTGGTATGGAAACA
AGCAAATTTGTGGACAAGTATGATGTTACCCTCGACGTTTGTATATCTTCTGTGCTATCTCAATCTAAG
ATCATTAGTCCTCAGGAGTCAAACGGTGAGAGCATAGATGTCTGTATTGAGGATGAGACCGTGAGCTA
CTTGAACCGGAAAGATGTTCAAAAGGCTCTTCATGCCCGACTAGTTAATGTTCGTCAATGGACTGTGT
GCAGCAACATATTGGATTATCAGCTCCTTGATGTAGAGATACCCACAATAACCATTGTAGGATCACTT
GTCAAGAATGGAATCCCTGTCTTAGTTTACAGCGGGATCAAGATTCAGTTATTCCACTAACTGGAAGC
CGCACACTTGTGCATCAACTAGCTAGGGAAACTACGGCTGAACACAACCACT
>comp42488_c0_seq1
ATTTTGTTACCCTCCTCTCTCTTTATTTTTCTTGGTAATTAATCTACGGAGTAGTAGGTTTTCATGCAGA
TCCATAAGGATCAGATAGCTTTTATGCTGCCTGCCATGAAATGTCCAAGGAAGAAGAAAAGGAGGAG
TTTTATGCTCTAGCATATATCATATTGCATCTGCAATTCACCAAAGTGGCCTATGTCGAAGAGAAGGAT
TGGATTTGTCGATGCCATCAGCATCATGGCCATGGACTCAAGAACAAGTATTGATTGGGCATTAAGAG
ACAAAATTAGTACCAGATGTTGTAATCTTTATTACAACACAAACTGGGAAAATGATGCCTTGACATGT
TTGTCTGATTCTCATAGTCTTCACGAGACGGTTGGTGCAGGCTCATGTCATTGGTGGTGGTATTGATAT
GGCTGGGCTTTGGACCGACCAATTCGGAAGGCTTGGGGATGTGTCCTCATGTTTTGGGTGGTATCGATC

TCCTTTCGCTATCAATACGATTCACGGTAATAGGCCTGATCTTTTTGACTTGGCCATGTCAGGCATTTGT
CGAAGAAATCATCATCTTGAATGCCCTACAATTTTACCTTGTATCTCAAGTAGTCG
>comp42593_c0_seq1
GGCGAGACTACTCAAGATGGCCTAGCAAACCTGAGCAATGGAAGGCCATGGTTGAAGCCTACGGCGA
AAATATGATGAACTTGTCCTGCAAGATGTTCGAGCTCTTGTCCGAGGGTTTGGGGTTAGACAAAGATG
CCTATGCAAAGTCAATGGGTGATATTAGCCAGCATTTGGTGGTTAACCATTACCCTAAGTGCCCACAA
CCCGAGCTAGCTGTCGGTCTTAAGCGACACACCGATCCAAGCTCCATTACCTTGCTCCTCCAAGACCAT
GTTGGTGGGTTACAAGCTACAAAAGATGGTAGGCAAAACTTGGGTCACCGTTCGACCTATCAAAAATG
CTTTGGTCCTTAACCTTGGTGATCATGCCTATTTCATGAGCAATGGAAGATTCAAGAGTGCCGATCATC
AAGCAATCGTGAACTCGATCGAGACACGAACGTCGATAGCAGTATTCATGAATCCAGCACAAGATGC
AAAGGTTTATCCGCTCAACGTAAAAGAGGGAGAGAAGGCAGTGTTGGAAGAGCCTATCACGTATGCT
GAGATGAATAACAGGCACAAGAATAGGTATCTCGAAATCGTTCAACTAAAGAAATTGGCTAAGGAAG
AAAATTGGTCACAAGAAGAGCTTGATCTCAAGCTTAGCCAGCTTAATTAAATCGATTTAGCCAGCATC
TAAT
>comp42870_c0_seq1
GAAGGATGTGGCTAAACGTCTTGCTGAGAGGGGTATTGCCTTGGCTGTTACAGATGGTGCTCTAGATT
ACGTCTTAGCAGAGAGTTACGATCCAGTATATGGTGCAAGGCCTATTAGGAGGTGGTTAGAAAAGAA
AGTTGTCACAGAATTATCGAAAATGCTCATCAAGGAGGAAATTGATGAGAATTCAACCGTCTTCATTG
ATGCTGGGATGGATGGACTTGCGTACCGCGTGGAGCAAACTGGTGGGATGGTTAATGCTGCAACTGGG
CAGAGGGCTGAGGTCCTGATTCAAGTCCCTAATGGACCACGAAGTAACGCCACACAAGCTGCAAAGA
AGTTAAAGATCGAGGAACTTGATGATGAGGATGATGAGATGATGGACTGAAGAGCTCAAGTATATGA
TGATGAAAGCTTTCCTTTTTATGGGCTTTTTACCTAATTTCATGGATTAGCAAATTTTCTTTATAAAAAG
GCTTATGTAATGTTGTTCAATCTGTTAATGAGTGTTGTTCAAATTTTC
>comp43304_c0_seq1
CTATTCGAGGTGGTGGTGGTTCGAGTTTTTGTGTCATCCTTTCCTATAAAATAAAATTGGTTAAAGTAC
CTAAAAAGGTCACTGTTTTAGGGTTTCAAGGACACTAGACCAAACCTAACTGATATTGTTGATCAATA
CCAACATGTTGCTCCAAATTTGGACCGTGATGCCTTCATTAGGCTTACCCTAGGTGTTACAAACTCGAC
CAAAACTGGCCTTTTGACAAACCAAGCAACATATAGATGCTTGTTTTTAGGTGACACCCAATCCCTCCT
ATCCCTAGTACACGAGAATTTTCCCTTGTTAGGCCTAGAAAAGTCGGATTGCCTTGAGATGAGTTGGAT
TGAATCTGTCCTTTTCTATTATGACTTTGCTATGGAAACTCCACTTTCGACTTTGTTAGATAGGCACCCT
TCAAGTTTAACTTCCCTCAAAAGAAAGTCCGATTATTTGAAGAAG
>comp43968_c0_seq1
GAAAGCATACTTCTTTTCTGAATCTCTCTGCAGCGAGAAACTGATCAGACATCAGCTTGAAGACCAAA
AAGATGAAGTTACTAACTTGGATGCATCGCAAGTTTCGGCAAGGTAATGGCGAAACACCGAAGGAAT
TTCCCCTAGGGCAGCAATTCCTGGACGACCCTCACTTCTATCAGAAGCAAAACCAAGGGAAACAACGA
AGAGCATCATTTACAAGTGTGGAGACAAGAATTTTGCTGAATACGAAAGAAAGCATCGTCAGCTGATG
AGTTTCCTGAGTACTTCGAAGGATTTTCTAGCAATCGGCACTTTTGGGGTAGCAGATGCAGTCATCCCA
GAGCCAGAGACACCAACCTTTGAGTGTCATTTGATAAGCTTGCAGAAGGAGAGGCTGAAGTGACAGA
AAATGAATTGAGGCTCATCAATGATGAGTTAGAGAGGGTGCTTGCTTCTGAAGCTAAAGATGATGGTT
GGACTGATTCTTCAGGGAGAAACAGTCATGTTAGCAATGGAAGAATCAGTCATGCAAGCACCATCACT
CTGAGTGGGAAGCCACTGGAAATTCCAGAAAGCACAGGAGGCACAGGAAGTAATTTGATGTACTGTC
CACTACAAGAATATCTGTTAGGCTCAGCTGTAGAAGTGCCAGAGAAAGCAACAGTGCCGAAAAAGGA
GAACAGGACATCATTGGGAGAACTATTCGAGAGGCACAAGGCAGAAGAAACCTCGGATGTCAAGTGC
AAACGTGGG
>comp44207_c0_seq1
GTGGCCTTCAAAATTCCTTGTGAAAATCTCTCTGGGTTAAATTCCTTTGCATCAGTACCCCATAATTTTT
CACTATGATGAATAAAAATAGTTGGCAAATTAACATGTACACCCGCCGGTAGAACCATATCGCCTACT
CTCATTTCGTGGCTAACCTTTCGATTTAGGGAAAACACCGGCGAATATAGTCGTAGAACCTCATGAAG
TATCATGTTCACCTTTTTGAGTTGGATTAAGCCATCAACATGAGGTGTATTGTGACCGAACATGGCGAA
AATCTCTTCTCGAGCTTGTTCTTGCCAATTTGTATGTTTGCTTAGCAAAACAAGAGTCCATACCAACAA
CGACGAAGTGGTGTCTTGTCCGGCAACATAGAACATTTTGCACTCTTCAATAACCTCTTTCATACTCAT
CTTAAATTGTTGTTGTTTGTTATCATTTCCATGTAATAATTGATGATCTTGTACTCCTTCTAAGTTCGAT
TTTAGCAATAACCCTAACAAGTCATCTTTTAGCTGCTTCACCATTCTCCATGTTTTTTGCCTCTTTTTGA
ATCATATCTTTTAACAATGATTCCATTTCATTGACGATTTTCGCCTCCCTCTTGTTTGTTTTAGTAGGAA
GAAACCTGGATCCAGGAATATAAACTG

>comp49229_c0_seq1
GTTGAATTGCAGGTAACACAGAAGCGCCATAGCATCAGTTTTGCTGCTGTTCAGAAGCTGATGCTATG
GGAGATCTGTAGGAGAATATTTTTTAGGAAAATTACAAACGGAATCGTTTCCACCGTTAAGAGAATTT
TTTTTAAGTAGCAGTGATAACTTATGTTGAGTATGCTATCATATGATTATTTTTCTACGGAATATTATCC
TATGAAAATGGAGACATGTACGGAAAATGTCTTCCACAAAAATAAACGGAGCCGCCTCCACTGTAAA
GAGAGCTTTTTAAGCAGCAGTGGTAATTTATGTGATTATATGCCAATTATGCCACCAGCCCACCACCCA
CATTAGCCATGTGGATGTACCATAATTATAAAATGCTCCAGGATAGAGACCCGACTTCTTAATGAGCA
AGTAGTTTGAGGTGGAACCTACATGACTGACCTGGCCACAAAATCAGTTTAAGGTGGGCTCACCGGGA
TACAGAATCAGTTTGAGGTGGAACCAATATGGCTGACCAACCTGGCCACAGAATCAGTTTAAG
>comp55875_c0_seq1
ATGTTCTTATCATTCCATCTTTTTGAATCTACAAACATCATATCCATGATAGATTACATACATTTTATAC
CAAGCTACCAAAGCATGAACAATATAACAAATTACACTTTCGTTTAGGACACAAATAATTGACATCAG
ACTATCCCGGTACATTTAATCATCGGATCTAGACCCTGGTGCACGAGGGCATCTCAGGTACGTTTTGCA
CCTATCAATTGACCCGCCTCCACATAAGCAGTTCCCATAACTATTATTTGCAGTCTTACAGTATGGGCT
CAGGCCATTAGTTTTATCGATGCAATTATCGGCAACGACAAGGTTGGTGCTCAAACTGAAAACTAGTA
GGATCAACAGAAGAGTTGAAGGGGAGAAACAAGGACGACGAACGAACAGAATGTTACTCTTACTCAT
GGTGTTTACCTCTCTCGTATGATCGATTTAATTTCTTATAAAAATGGAAAAAAT
>comp56526_c0_seq1
TAGAGTTTGATCAACACTTATGGCTACAAAAATGCACCTAGTGTTTTTGGCTTCCTTTCATTGATACCG
CCTTTTATGGTTGCTCAAGCAGACCCGGGTTATAGCTATGGGTACGGCTTTGGATGGGGTAGTGGTGGT
GGGAACGACTATAGCTCGTTGTCTCCGTACTTCTATGACTTCTCTTGCCCTCAAGCTAACAACATTATT
GCTTCGGTCTTAGAGAATGCTGTTGCACAAGACCCTCGGATGGCTGCTTCTTTGCTTAGGCTTCATTTT
CATGACTGCTTTGTTCAGGGGTGTGATGCTTCTATACTGCTAGATGATAGCGCGACATTTGTAAGTGAG
AAAAGGCTAAACCAAATCTTAATTCCTTGAGAGGTTTACTGTCATTGATGAAATTAAAGCCAGGTTAG
AAGAAGCATGCCCTGAAACTGTCTCTTGTGCTGATATCTTGGCCTTGGTAGCTCGCGCCGCCAGTGTCT
TGAGTGGAGGACCTAATTGGGAAGTCCCATTAGGAAGAAGGGACTCAAAAGCAGCAAATATGCTACT
TGCAAATG
>comp56807_c0_seq1
TGAGGCATGTGCAAACGTACAGAGTTCAGCTAGACAGCCAGCCAGAGGAGATAGATAATTTGGAGAG
GAAGAGAATGCAGTTAGAAATTGAGCTACATGCTCTTGAGAAGGAGAAAGATAAAGCTAGCAAAGCA
CGGTTAGTGGAGGTTACTAAGGAGCTTGAGGACTTGAGGGATAAACTTCAGCCTCTAAAGATGAAATA
TATGAAGGAGAAGGAGAGGATTGATGAGCTCCGTAGACTGAAGCAAAAAGAGAAGAGCTCATTATTT
CTGCACAAGAGGCTGAAAGAAGATATGATCTTGCAAGGGCTGCTGATTTGAGGTATGGAGCAATCCA
GGAAGTCGAGGCAGCTATTGCTAGATTGGAATCTAATGCTACCGATGTTGAATTGTTGACTGAAACTG
TTGGGCCAGAGCAGATTGCAGAAGTGGTGAGCCGATGGACTGGTATTCCAGTCACCAGACTGGG
>comp79264_c0_seq1
GGGATGTTAAATTTCAGAATGAAGATGATTTACTCTTTAGCAATAACACAGAAGCTATAGACAGAGGG
GATGAAAAGAAAATTTATACACTGAATCTCAACAAGAACAAGAAAATCAAAAGATTGAACTTTCAAA
AGAAGAGAGCTTTCCCAAAGATCAAGCCTCTGAAGTAAATAAAACCAAAGAATTCACACATTCCGAA
GTGAACTCTGATGAAGACATGAAAAAAATGGTGCTGACTTTACAACAAGTGTAACTAAAGAAGAAAT
GATGCAAGAGTTTATGGAGTCTGATCCAGAAAAAAGACTAGAGGTGACAAAATGAGAAGGAATCAAA
CAG
>comp82670_c0_seq1
TAAAATTCCAGATTCTATCAAACCCAGAATTTTTAGCTGAAGGAACTGCAATCCAAGACTTTTTTAACC
CAGACAGAGTTCTCATAGGAGGAAGAGAAACCCCTGGAGGCCAAAAGGCTGTACAAGAACTCAAAAA
CGTGTACGCGCAATGGGTGCCTGAGGACCGGATCCTAACCACTAATCTGTGGTCAGCCGAGCTGTCTA
AGCTCGCGGCTAATGCATTTTTAGCCCAAAGAATTTCATCTGTAAATGCTATGTCCGCCCTGTGTGAGG
CCACGGGTGCTAATGTGTCACAAGTCGCCCATGCTGTTGTAGGACACAAGAATTGGACCAAAATTTTT
GAATGCTAGTGTTGGTTTGGAGGGTCATGTTTCCAAAATGATATACTTAACTTAGTGTACATCTGTGAT
GTAATGGCTTACATGAGGTCGCGAATTATTGGAAGCAAGTGATTAAGATTAATGATTACCAAAAG

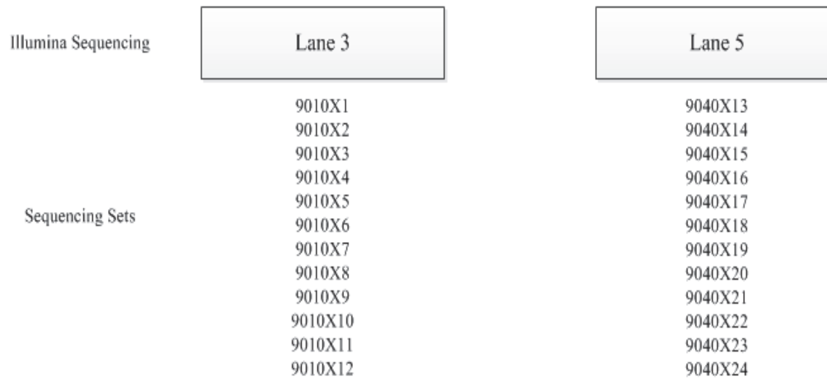
SUPPLEMENTAL FIGURES

Supplemental Figure 1.1 Flowchart of cDNA preparation, sequencing and RNA-seq analysis to identify drought tolerant genes in the root tissue of quinoa.



Supplemental Figure 1.2 Experimental design, cDNA preparation and sequence layout. Each replicate consisted of both Ingapirca and Ollague varieties that were subjected to one of four drought treatments. RNA was extracted from root tissue and synthesized into cDNA libraries (resulting in 24 total cDNA library samples). A unique adaptor was attached to each cDNA allowing 12 libraries to sequenced on the Illumina HiSeq 2000.

Replicate	Replicate 1								Replicate 2								Replicate 3							
Variety	Ingapirca				Ollague				Ingapirca				Ollague				Ingapirca				Ollague			
Treatment	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
cDNA construction sample number	9040 X21	9040 X19	9040 X23	9040 X24	9040 X17	9040 X20	9040 X18	9040 X22	9040 X16	9040 X10	9040 X13	9040 X8	9040 X14	9040 X4	9040 X12	9040 X3	9040 X15	9040 X6	9040 X5	9040 X1	9040 X7	9040 X11	9040 X9	9040 X2



Chapter 2. Literature Review

INTRODUCTION

Chenopodium quinoa Willd., commonly referred to as quinoa, is a hardy, nutritional grain native to the Andean region of South America. Known for its ability to tolerate many abiotic stresses such as drought, salinity, frost, extreme temperature, and poor edaphic conditions, quinoa is a valuable crop to the people of the developing Andean countries in which it natively grows. Lack of breeding in quinoa has allowed for high genetic diversity in different varieties allowing for many comparative studies. Even though quinoa is accepted as a salt and drought tolerant crop it has been understudies when compared to other crops: maize, soybean, wheat, or rice. By identifying gene products associated with quinoa, we will be able to understand better the mechanism and overall biology that makes some quinoa varieties tolerate drought stress more than other varieties and crops.

Taxonomy

Commonly referred to as the goosefoots, the genus *Chenopodium* is made up of over 120 species that grow all over the world (Giusti 1970). The many species contained in the *Chenopodium* genus are separated into 16 sections based on morphology (Aellen 1960). Species pertaining to this genus are diploid, tetraploid or hexaploid and have a basic chromosome number of x=8 or x=9. In addition, *Chenopodium* belongs to the sub-family Chenopodioidea of the family Amaranthaceae. About 368 species belonging to the Chenopodioidea sub-family have been classified as being tolerant to abiotic stress, approximately 28% of the total sub-family (Orcutt and Nilsen 2000). Wilson indicated that species of the *Chenopodium* genus may be cultivated, weedy, wild or may include both cultivated and weedy types. Three main cultivated species of *Chenopodium* are canihua (*C. pallidicaule*), huauzontle (*C. berlandieri* spp. *nuttalliae*), and quinoa (*C. quinoa* Willd.). While huauzontle is native to Mexico the other two

species are mostly native to the Andean region of South America. Evidence shows partial domestication of other *Chenopodium* species in Denmark, Greece (Risi and Galwey 1984), the Himalayas (Partap and Kapoor 1985) and the United States (Wilson 1981).

History

Quinoa has a distinguished and long history in South America having been cultivated for many centuries. Archeological evidence suggests that quinoa was cultivated since 5000 BC in Peru (Tapia 1979) and since 3000 BC in Chile (Tapia 1982). by the Incas. Able to tolerate high altitudes (up to 4000m), cold temperatures, drought, poor soil conditions and high salinity, quinoa was the staple crop cultivated in the highland regions of the Incan Empire (Cusack 1984). Being called *chisiya mama* or "mother grain", the South American Indians considered it a sacred food source. The cultivation, harvest, and consumption of quinoa were implemented regularly in Incan ceremonies and traditions. One such example was a yearly offering that was performed at the beginning of each year's planting in which quinoa was offered in a fountain of gold to Inti, the Incan sun god.

By the 1500s quinoa was grown along the Andes from Colombia to Argentina and Chile. When the Spanish conquistadors arrived in the New World during the 16th they introduced several new crops such as barley (*Hordeum vulgare L.)* and wheat (*Triticum aestivum L.*) displacing quinoa due to the ceremonial traditions that accompanied its cultivation. As a result, the production of this staple crop diminished greatly during this, the Colonial Period of South America, until recently (Cardoza and Tapia 1979). According to Risi and Galwey, quinoa production continued in harsher climate areas where other crops failed to grow as well as areas where Incans persisted regardless of regulations during Spanish rule.

Even in the past century quinoa production declined from 1947 to 1975 due to growth of

urban populations of the Andean region. Total area in cultivation went from 47,000 ha to 15,000 ha during this time period (Tapia 1979). Since the mid-1970s quinoa growth has increased significantly and continues to gain popularity. Currently quinoa is grown in most regions it was previously grown in the 1500s before the conquistadors coming to the New World. In high elevation terrain it grows form southern Colombia (5° N) to southern Argentina/Chile (30° S); while at sea level it has been cultivated on the southern, coastal regions of Chile (40° S) (Risi and Galwey 1984). At the household level, it is normally intercropped by the subsistence farmer (Fleming and Galwey 1995). It is typically at the higher, harsher elevations where quinoa is found in monoculture. Recently, the worldwide health food market has acknowledged the superior nutritional value of this pseudocereal, which has promoted exportation of quinoa to Europe and North America.

Nutrition

Because of its high nutritional value, quinoa has great potential as a commercialized crop around the world. Besides for being gluten free it is well known for being high in protein, seeds containing 12-18% protein (Risi and Galwey 1984; Oshodi et al. 1999; Mastebroek et al. 2000). Risi and Galwey (1984) showed that other valued crops such as barley, rice, and wheat are inferior in comparison with wheat having 10-12% protein with barley and rice having even less. In addition, quinoa is commonly referred to as a complete protein, containing all essential amino acids required by the human body. In comparison, it has a better amino acid balance than most crops (Risi and Galwey 1984; Table 2.1); and is particular high in lysine, an essential amino acid that is deficient in other crops (Coulter and Lorenz 1990).

Furthermore, the quinoa seed is rich in many vitamins and minerals satisfying necessary nutritional requirements. Ruales and Nair (1993) stated it as being a good source in many of

vitamin C and vitamin E as well as many of the B vitamins: including thiamine (B1), riboflavin (B2), niacin (B3), panothenic acid (B5), biotin (B7), and folic acid (B9). In addition, its mineral content of calcium, copper, iron, magnesium, manganese, phosphorus, potassium, and zinc exceeds that of other grains in comparison (Ruales and Nair 1993; Fleming and Galwey 1995; Table 2.2). This pseudocereal grain is also a better carbohydrates source than most cereal grains like wheat, rice, or rye; while at the same time being naturally low in fat – the seed fat content is about 5-7%. Seeds have been found to have significant amounts of antioxidants like coumaric, ferulic, hydroxybenzoic and vanillic acids. Antioxidant flavonoids kaempferol and quercetin are sometimes in greater concentration in quinoa than in high-flavonoid berries like the cranberry (Hirose et al. 2010; Repo-Carrasco et al. 2010).

Even though quinoa is a high-nutritive grain, there are certain antinutritional compounds affecting the overall nutritional value. Saponins are amphipathic glycosides that are found in the seed coat not the seed itself. Used as a defense mechanism against predators, this bitter tasting compound can found in most quinoa varieties. Removal of the seed coat allows a saponin-free grain without seriously affecting the nutritional value of quinoa (Ruales and Nair 1993).

Classification

Quinoa is known for its diversity and many different environment in which the different varieties grow. Growing from sea level to the high Altiplano, the different varieties are known to have different length of growth periods – time from emergence to anthesis (Risi and Galwey 1989). Plant heights, seed color, leaf shape, and saponin content also varies among different plant ecotypes.

The five major environmental quinoa types are: sea level, subtropical, valley, Salares (salt flats) and Altiplano (high plains). Tapia (1980; 1997) described the following environmental and

plant characteristics that are native to each ecotype. The sea level type is grown in southern Chile mostly around the latitude of 40°S and receives about 800-1,500 mm of rain annually. The plants are around 2m tall, mostly not branched, and will flower in long days. The seed are small, yellow, translucent, and contain high amounts of saponins. Popular sea level varieties are Baer, Quechuco de Cautin, and Pichará de Maule. Subtropical types are found in the Yungas region of Bolivia, which receive anywhere from 1,000-2,000 mm of rain each year. The subtropical quinoa plant has an intense green coloring that turns orange at maturation. It produces small, orange seeds. The valley type grows in the Andean valley between 2,000-4,000 m in central and northern Peru where the annual precipitation ranges from 700-1,500 mm. Plants are 2-4 m tall, mostly branched, and have a growth period of 7 months. Popular valley varieties are Blanca de Junín, Rosada de Junín, Amarilla de Marangani, Dulce de Quitopampa, and Dulce de Lazo. Salares types come from the Bolivian salt flats, at an altitude of 4,000 m, and receive about 250-400 mm of rainfall annually. The plants grow in soil with a pH often over 8. Most have black seeds with high saponin content and sharp edges. The other characteristics resemble the Altiplano quinoa. 'Real' and the ancestral 'Sajama' are well-known Salares quinoas. The Altiplano types are from surrounding areas of Lake Titicaca, cultivated at 4,000 m, and yearly receive 400-800 mm of precipitation. The plants are 1-1.8 m tall, mainly unbranched, and with a growth period of 4-7 months. Popular varieties include 'Chewecca', 'Kanccolla', and 'Blanca de Juli'.

Tolerance

One of the most desirable traits of quinoa is its ability to tolerate many abiotic stresses such as salt, frost, extreme temperatures, and drought. Prado et al. (2000) demonstrated that quinoa, a halophytic grain crop, germinates in 0.4M NaCl with reduced germination rates. More recently,

Maughan et al. (2009) characterized the Salt Overly Sensitive 1 (SOS1) gene, which encodes a plasma membrane Na+/H+ antiporter that facilitates germination and growth under saline conditions, in quinoa.

Native to the Altiplano, quinoa is cultivated in temperatures ranging from 7-10° C where nocturnal frost is common. According to Risi and Galwey (1984), quinoa has been known to resist temperatures as low as -4° C for extended periods. In 1998, Vacher performed a field trial in Bolivia where he supported the idea that quinoa is a drought tolerant crop. He found that quinoa had a rapid stomatal closure and great decrease in leaf water potentials reached as low as -4 MPa while maintaining high leaf water use efficiency. The overall tolerance of quinoa allows it to grow in the Altiplano and Salares areas making it an ideal crop to be grown in other harsh environmental areas.

Drought

Drought causes both biological and physiological effects on the development of a plant. Drought stress has been known to have an effect on growth, photosynthesis, respiration as well as water and nutrient relations. It has a direct effect on reducing leaf size, stem extension and root proliferation (Chaves et al. 2003). Drought stress disturbs plant water relations and reduces water-use efficiency (Farooq et al. 2009). Water-use efficiency, which refers to the ratio of CO_2 fixed to water lost, varies among different types of plants where mature leaves of C_3 plants average 0.5-1.5mmol(mol H20)$^{-1}$ while C4 and CAM plants average 1-2 and 4-10mmol(mol H20)$^{-1}$ respectively (LeHouerou 1984).

Drought has a direct effect on the plant's ability to undergo proper photosynthesis. A reduction in photosynthesis has many secondary consequences, one of which is the formation of high levels of reactive oxygen species (Logan 2007). Reactive oxygen species are dangerous to

the overall welfare of the plant because they cause oxidative damage to membrane lipids and proteins. In addition, drought stress affects the photosynthetic machinery and induces stomata closure. Stomatal closure leads to reduced CO_2 assimilation by leaves and increased photorespiration. Closing of the stomata also disturbs the activity of various enzymes including those of CO_2 fixation and adenosine triphosphate (ATP) synthesis (Farooq et al. 2009).

Major photosynthetic pathways

There are three main photosynthetic pathways seen among plants that enable them to fix carbon. The most simple photosynthetic pathway is that of C_3 plants. During the daytime, CO_2 diffuses through the stomatal pores to the chloroplast. In the mesophyll cell chloroplasts CO_2 is combined with a 5-carbon compound (ribulose 1,5-biphosphate) by Rubisco (ribulose 1,5 bisphosphate carboxylase/ oxygenase) resulting in the formation of two 3-carbon molecules (3-phosphoglycerate) in the chloroplasts. The formation of 3-phosphoglycerate leads to regeneration of ribulose 1,5-biphosphate thus continuing the cycle.

In C_4 plants CO_2 diffuses through the stomatal pores to the cyotsol of the mesophyll cells during the daytime. CO_2 is incorporated into phosphoenolpyruvate (PEP) by PEP carboxylase resulting in oxaloacetate. In the chloroplasts, addition of two hydrogens using NADPH allows oxaloacetate to be reduced to malate (Hatch and Slack 1966). Malate, the reduction of oxaloacetate, is able to diffuse into the bundle sheath cells. In these cells decarboxylation of malate can occur by NAD-malic enzyme producing CO_2 and pyruvate. CO_2 facilitates its assimilation through the Calvin cycle while pyruvate is able to move back to the chloroplast where it is able to be converted to PEP by pyruvate phosphate dikinase. In addition, aspartate can be formed through transamination of oxaloacetate by alanine. It should be noted however that this conversion is formed in the cytosol and also forms pyruvate as a product. Aspartate, like

oxaloacetate, has the ability to diffuse into the bundle sheath cells where it is able to be transaminated back oxaloacetate and then reduced by NAD-dependent malate dehydrogenase to malate. Once again this malate compound can be decarboxylated forming both CO_2 and pyruvate. Finally, in some C_4 species aspartate forms oxaloacetate which gets decarboxylated by PEP carboxykinase (requiring ATP) to form PEP. The CO_2 that is released by oxaloacetate or malate diffuses into the chloroplasts and is fixed by Rubisco as it was previously explained for C_3 plants.

Similar to C_4 plants CAM plants use PEP carboxylase for the fixing of CO_2 into oxaloacetate, but entry of atmospheric CO_2 occurs at night when stoma of CAM plants open (Cushman 2001). Cystolic NAD^+ malic dehydrogenase alters oxaloacetate and forms malate. Malate is then converted into malic acid and transported into the vacuole of the chlorenchyma cell for overnight storage. During the daytime when the stoma are closed, preventing water loss and CO_2 intake, malic acid is converted by to malate and diffuses from the vacuole to the cytosol. At this point malate is able to be decarboxylated by $NADP^+$ malic enzyme and converted to CO_2 and pyruvate which diffuses into the chloroplast. The CO_2 is then fixed by Rubisco as it was previously explained for C_3 plants.

Drought mechanisms

Because of a plant’s sessile nature, they must be able to resist drought to be successful. The different mechanisms by which plants resist drought are either by escaping the drought or tolerating it. Drought escape is the ability of a plant to complete its life cycle before deficiencies in soil or plant water develop. Plants can escape drought by having a rapid phenological development as well as developmental plasticity (Turner 1979). A rapid phenological development includes rapid growth, maturation and flowering/fruiting allowing for reproduction

before the environment alters their fitness level. Developmental plasticity refers to the plant's ability to grow in different times or in different climates as to escape the period of drought that can be too severe of a stress for the survival of the plant (Blum 2009).

Drought tolerance is the degree to which a plant can adapt to water deficient conditions. Plants have adapted several ways to increase tolerance to drought as one of the major abiotic stressors. Stomata adaptations such as reduced number of stomata or production of waxy leaf surfaces results in reduced water loss and ability to tolerate water deficiencies. Another main adaptation seen in drought tolerant plants is root proliferation. This adaptation allows for increased water absorption and ability to survive compared to neighboring plants. Plants can also tolerate drought by maintaining turgor by lowering the osmotic potential that arises from the accumulation of solutes as a response to water deficiency while at the same time reducing the amount of water loss (Mitra 2001). Plants can reduce the loss of water in several ways, including folding or rolling of the leaf to reduce absorption of radiation (O'Toole and Moya 1978) as well as altering the leaf surface area in order to reduce the evaporation surface (Turner 1986).

Plants must be able to respond to endogenous or external stimuli to survive. To respond effectively, plants must perceive drought stress signals, relay those signals through various intermediate compounds with the ultimate goal of targeting a response to defend against or avoid the stressor (drought). Often these response are consist of alterations in gene expression.

Drought lowers the external water potential causing an immediate reduction of water delivery to plant tissues and growth (Munns 2002). Even though no plant-produced molecule has been identified as an osmosensor, it is believed that membrane receptors sense the altered environment (drought or lack of water) and initiate the cascade of events that allow the plant to respond to the drought stress. In yeast, osmotic stress response is initiated by histidine kinases

SLN1 or SHO1 that feed into the high-osmolarity glycerol (HOG) pathway. SLN1 is a two-component histidine kinase and is capable of sensing cellular turgor in prokaryotes (Reiser et al. 2003). Activation of SLN1 results in MAPK that transcribes genes for glycerol synthesis. Cytokinin response 1 (Cre1) is an Arabidopsis hybrid histidine kinase. CRE1 can substitute the SLN1 osmosensing function in that it also is regulated by changes in turgor, in the presence of cytokinin (Reiser et al. 2003). Another Arabidopsis histidine kinase, AtHK1, was identified as a homologue of SLN1. It functions as an osmosensor in yeast and complements the yeast double mutant, *sln1*Δ *sho1*Δ, to transmit the stress signal to a downstream MAPK cascade (Urao et al. 1999).

Even though the initial perception of drought stress in plants is still unclear, a rapid synthesis of the hormone abscisic acid (ABA) has been observed as a response to drought. ABA has a central role in drought response in maintaining water homeostasis and dehydration response (Li et al. 2006). Biosynthesis of ABA requires rapid regulation of zeaxanthin epoxidase (*ZEP*), 9-*cis*-epoxycarotenoid dioxygenase (*NCED3*) (Qin and Zeevaart 1999), aldehyde oxidase 3 (*AAO3*) and molybdenum cofactor sulfurase (*MCSU*) (Xiong and Zhu 2003). Once synthesized ABA is transported to the guard cells from the roots or vascular tissue where it binds to one of the several guard cell surface receptors. Once bound to the guard cell surface receptors ABA signals secondary messengers including cytoplasmic calcium (Ca^{2+}_{cyt}) and phospholipid based compounds such as inositol phosphate (IP_3).

Signaling of these secondary messengers has several consequences with the biggest being the closure of the stomata. The binding of ABA to one of the cell surface receptors increases the cytoplasmic pH or activates OST1 kinase (Mustilli et al. 2002); either of which will produce ROS by way of NADPH oxidases. The production of ROS activates Ca^{2+} influx of the Ca^{2+} ion

channel that allows Ca^{2+} ions to enter the guard cell freely. Another pathway induced by ABA is the accumulation of nitric oxide (NO) via NO synthase (Fujita et al. 2004) or nitrate reductase (Conrad et al. 1968). NO is the beginning signaling intermediate that leads to increased cytoplasmic Ca^{2+} from the vacuole (Li et al. 2006). A third signaling pathway increases cytoplasmic Ca^{2+} from the vacuole by means of phospholipid based compounds secondary messengers such as inositol phosphate (IP_3) (Alexandra et al. 1990). ABA activates phospholipase C (PLC) which leads to the formation of (IP_3). In addition, ABA activates phospholipase D (PLD) which leads to the formation of phosphatidic acid (PA). As a result, PA binds to ABI1 which inhibits stomatal closure (Mishra et al. 2006). The increased cytoplasmic Ca^{2+} from these different pathways leads to an increased cytosolic Ca^{2+} concentration. The increased cytolic Ca^{2+} concentration in turn promotes stomatal closure by means of activating GORK1, the gene responsible for outward K^+ channel from the guard cell (Hosy et al. 2003), as well as inhibiting the inward K^+ channel.

Protein classifications

Products of drought-induced genes can be classified into functional proteins and regulatory proteins (Shinozaki et al. 2003). Functional proteins include proteins such as chaperones, late embryogenesis abundant (LEA) proteins, water channel proteins, detoxification enzymes, key enzymes for osmolyte biosynthesis and various proteases. Regulatory proteins are involved in the regulation of signal transduction and stress-responsive gene expression(Shinozaki and Yamaguchi-Shinozaki 2000; Seki et al. 2002). Various transcription factors, protein kinases, protein phosphatases, enzymes involved in phospholipid metabolism and other signaling molecules fall into this group of regulatory proteins (Shinozaki and Yamaguchi-Shinozaki 2007).

ABA also plays a vital role in the expression of drought stress responsive genes. Yamaguchi-Shinozaki and Shinozaki (2005) showed that there are at least five regulatory systems for gene expression; three of which are ABA-dependent while the remaining two are ABA-independent. In the ABA-dependent pathways there are three main groups of transcription factors that play important roles in the response of the plant to drought and other environmental stresses. Transcription factors are simply sequence-specific DNA-binding proteins that control transcription of genetic information from DNA to mRNA. One of the transcription factors in the ABA-dependent pathways is AREB/ABF, which is made up of two basic leucine zipper (bZIP) transcription factors. AREB/ABF binds to the ABA responsive element (ABRE) which in turn activates gene expression of *RD29B* and *RD20A* whose function entails responding to desiccation stress in Arabidopsis by encoding drought-inducible hydrophilic proteins (Yamaguchi-Shinozaki et al. 1992). The second ABA-dependent pathway uses MYC and MYB transcriptional factors to bind to *cis*-elements in the RD22 promoter and co-operatively activate *RD22* (Abe et al. 2003). *RD22* is a desiccation and dehydration-responsive gene found in Arabidopsis (Abe et al. 1997). Lastly, Fujita et al. (2004) identified a drought-inducible *RD26* gene encoding a NAC transcription factor.

The ABA-independent stress induced pathways include the Apetala2/ethylene responsive element binding factor (AP2/ERF), dehydration response element binding protein (DREB2) and the dehydration response element binding protein /C-repeat binding factor (DREB1/CBF). DREB2 and DREB1/CBR are subfamilies of AP2/ERF. DREB1/CBF interacts with the dehydration responsive element C-repeat (DRE/CRT) sequence of the RD29A promoter to activate the drought stress genes. Most of the CBF/DREB1 target genes contain the DRE motif with a conserved (G/A)CCGAC sequence in their promoter. In addition, CBF/DREB1 genes are

activated by cold stress in addition to other abiotic stresses (Kasuga et al. 1999). Homologs of CBF/DREB1 and DREB2 have been identified in rice and overexpression of these genes have shown stress-responsive gene expression and stress tolerance (Dubouzet et al. 2003). The other ABA-independent pathway incorporates NAC and HD-ZIP transcription factors that bind to the *cis*-acting element of early response to dehydration (ERD1). The ERD1 gene is not only induced by dehydration but is also up regulated during natural senescence and dark-induced senescence (Simpson et al. 2003). Other transcription factor families that have been implicated in drought stress are bHLH, Cys2Cys2 zinc-finger, Cys2His2 zinc-finger as well as others.

Sequencing

Sanger sequencing has been the manner by which sequencing was performed for many years. Recent sequencing developments have significantly reduced the cost of sequencing as well as the speed at which millions of DNA fragment can be sequenced (Metzker 2010). These new sequencing techniques (termed "next generation sequencing" or "massive parallel sequencing") are not only less expensive than Sanger sequencing, but also eliminate the need of bacterial cloning prior to sequencing. There are three main NGS platforms available, including Roche GS-FLX 454 Genome Sequencer, the Illumina Genome Analyzer (formerly Solexa), and the ABI SOLiD analyzer.

The Roche 454 Genome Sequencer uses sequencing technology known as pyrosequencing. Its amplification protocol uses an emulsion PCR step in which the target DNA strand comes bound to a beaded PicoTiter plate. Currently, read lengths are about 400-600 base pairs (bp) with a throughput of 500Mbp/run. The base accuracy has been reported to be 95-99% accurate with higher accuracy rates when corrected (Margulies et al. 2005; Blow 2007). The Illumina Genome Analyzer is based on sequencing-by-synthesis (SBS) technology and amplifies the target DNA

via bridge amplification, also known as cluster PCR. Presently, the Genome Analyzer can produce reads of 100 bp with a throughput of ~200 Gbp/run (Zhang et al. 2011) and a base accuracy > 99.5%. The ABI SOLiD platform uses a ligation-based sequencing approach and an emulsion PCR amplification procedure. This platform has the capability to produce reads of 50-100 bp in length with a throughput greater than 100Gbp/run and a base accuracy reported to reach 99.94% (Zhang et al. 2011). While all NGS platforms create a lot of sequence data, the Illumina allows for the greatest volume of reads per run and the preferred sequencer for transcriptomics (since read numbers of specific cDNA targets are utilized to quantify gene expression changes).

As previously mentioned Illumina uses polymerase-based sequence-by-synthesis (SBS) to sequence. Like the cDNA library preparation, the workflow for the Illumina sequencing is a multi-step procedure. The process begins with a random annealing of the library to the surface of the flow cell. This hybridization is done by an active heating and cooling step followed by an incubation with reactants (Mardis 2008). Surface adaptors are attached to the bound DNA molecules which allow for bridge amplification and clusters formation on the flow cell. After amplification, this unique approach allows for each cluster to be supplied with polymerase and four differentially labeled nucleotide bases. The addition of specific bases are then monitored via imagining of the flow cell after each cycle of the sequencing process (Mardis 2008). Repetition of this process allows for the determination of the complete sequence. A significant advantages to the Illumina SBS process is that all four nucleotides are in one reaction and that problems with homopolymer repeats are minimized – a problem for 454 pyrosequencing.

Transcriptome analysis

The transcriptome is the set of all RNA molecules: including mRNA, rRNA, tRNA, and

other non-coding RNA, produced in cells. It reflects all expressed genes at a particular time (developmental stage) or physical condition (under stress conditions). Analyzing the gene expression is known as functional genomics and is essential in understanding the function of genes and how they are altered under stress conditions. Recently, there have been several methodologies to analyze the transcriptome of an organism, including both hybridization- or sequenced-based approaches. Hybridization-based approaches used fluorescently labeled cDNA and hybridizes to a microarray chip. Once the microarray has been designed it is able to be compared to controls to make qualitative measurements of gene expression. Limitations with microarrays consist of reliance of existing genome sequence as well as backgrounds of cross-hybridization (Okoniewski and Miller 2006). Sequenced based approach determines the cDNA sequence directly and doesn't deal with hybridizing the cDNA at all. Some of these sequence-based approaches are serial analysis of gene expression (SAGE) (Velculescu et al. 1995), cap analysis of gene expression (CAGE) (Kodzius et al. 2006) or massively parallel signature sequencing (MPSS) (Brenner et al. 2000). While all these implement different techniques they sequenced the cDNA using the Sanger sequencing method which is slower and more expensive than the next-generation sequencing.

RNA-Seq

RNA sequencing (RNA-Seq), also known as "Whole Transcriptome Shotgun Sequencing" (Ryan et al. 2008), is a newer sequencing-based approach to analyze the transcriptome. RNA-Seq uses next generation sequencing (NGS) technology to quantitatively measure accurately the expression level of thousands of genes by sequencing RNA in a particular cell or tissue. Because transcriptomes provide information regarding gene regulation and protein production, the RNA-Seq technology has become widely used in the studying diseases, such as

cancer (Maher et al. 2009). RNA-Seq technology has been applied to different research experiments: detecting alternative splicing (Gan et al. 2010), detecting gene fusions (Maher et al. 2009) identifying novel transcripts (Robertson et al. 2010), identifying splice junctions, and quantifying expression of genes and transcripts (Mortazavi et al 2008). Since RNA-Seq is applicable to various research studies it has been referred to as “a revolutionary tool for transcriptomics” (Wang et al. 2009). Wang et al. explained that the RNA-Seq technology is capable with the major next-generation sequencing platforms.

A standard RNA-Seq pipeline varies depending the aims of the experiment. The general pipeline for differential expression testing includes: making millions of short reads, mapping and assembling, normalizing, testing for differential expression, and making conclusion based on results. Regardless of the platform implemented, RNA samples are typically fragmented, converted to cDNA, and made into a nucleotide library that is sequenced by a next-generation sequencer producing reads. The number of reads and the length of the reads will vary depending on the sequencing platform. Regardless of the platform, the reads (sequences) generated are then assembled: mapped to a reference or assembled as a *de novo* transcriptome.

Mapping

There are many short read aligners available for public use. The aligner software one uses depends on several factors such as sequencing platform and purpose. There are two classic classification widely used for reference read (mapping-first) aligners: Hash Look-up Table Algorithms (Hash-table) and Burrows-Wheeler Transform (BWT)-based methods. Hash-table aligners, such as GSNAP (Wu and Nacu 2010) and SOAP (Li et al. 2008), are created to detect complicated differences between the reference and read sequence by increasing computational requirements. BWT-based aligners, such as Bowtie (Langmead et al. 2009), BWA (Li and

Durban 2009), and SOAP2 (Li et al. 2009), are extremely slow when aligning complex alignments arise but are efficient in mapping reads that closely match the reference (Oshlack 2010). When there is not an assembly, a *de novo* (assembly-first) assembly must be performed. *De novo* assemblers, such as ABySS (Birol et al. 2009), SOAPdenovo (Li 2009), Trinity (Grabherr et al. 2011), and Velvet (Zerbino and Birney 2008), all directly assemble the sequence reads into transcripts. The assembly software uses de Bruijn graphs to model the overlapping sequences of k-mers (usually 25-50 base pairs) from the reads. Applying algorithms that parse the de Bruijn graphs based k-1 sequence conservation allow for the creation of contigs or scaffolds.

Differential expression analysis

One aim of RNA-Seq technology is analyzing expression changes of gene products between different samples or samples at different states. RNA-Seq analysis can quantify the differential expression by the number of reads mapped to the gene or transcripts; whereas, microarray analysis only identifies more or less expression by fluorescence levels. There are several main parametric approaches based on known probability distributions, such as Binomial, Poisson, and Negative Binomial, which are implemented in the tools used for differential expression analysis. The usage of publicly available software packages largely depends on the aims and design of the experiment. Table 2.3 shows the major software packages that are used in differential expression analysis.

Biological insights

The primary purpose of a RNA-Seq is to understand the biology occurring that is quantitatively represented in the analysis. Genes or gene products that are differentially expressed can be annotated by using BLAST to find significant matches to genes previously

annotated. Blast2GO is another tool used to provide insight on the biological process, cellular components, and molecular function of putative gene products involved. The biological process ontology puts the various molecular functions in a biological context; the cellular components ontology describes the location of the gene products in the cell; and the molecular function ontology describes the activities of the gene products (Ashburner et al. 2000). It functionally annotates the putative genes by blasting them against the non-redundant (nr) protein database (at NCBI) and annotates based on sequence similarity (Conesa et al. 2005). This annotation can be combined with other information, such as metabolic pathways. This can be performed in Blast2GO using KEGG (Kyoto Encyclopedia of Genes and Genomes), which enables visualization of the metabolic pathways within the transcriptome (Götz et al. 2008).

REFERENCES

Abe H, Yamaguchi-Shinozaki K, *et al*. (1997) Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. The Plant Cell Online 9:1859-1868

Abe H, Urao T, *et al*. (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) Function as transcriptional activators in abscisic acid signaling. The Plant Cell Online 15:63-78

Aellen P (1960) *Chenopodium*. In Hegi G (ed) Illustrierte flora von mitteleuropa, 2nd edn, vol 3C Hanser, Munich, pp 533-657

Alexandra J, Lassalles JP, *et al*. (1990) Opening of Ca2+ channels in isolated red beet root vacuole membrane by inositol 1,4,5-trisphosphate. Nature 343:567-570

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106

Ashburner M, Ball CA, *et al.* (2000) Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25-29

Birol I, *et al*. (2009) De novo transcriptome assembly with ABySS. Bioinformatics 25:2872-2877

Blow N (2007) Genomics: The personal side of genomics. Nature 449:627-630

Blum A (2009) Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress. Field Crop Res 112:119-123

Brenner S, Johnson M, *et al*. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotech 18:630-634

Cardozo A, Tapia ME (1979) Valor nutritivo. In ME Tapia (ed) Quinua and kaniwa. Instituto Interamericano de Ciencias Agricolas, Bogota, Colombia, pp 149-192

Chaves MM, Maroco JS, *et al.* (2003) Understanding plant responses to drought – from genes to the whole plant. Functional Plant Biology 30:239-264

Chen G, Wang C, Shi T (2011) Overview of available methods for diverse RNA-Seq data analyses. Sci China Life Sci 54(12):1121-1128 doi:10.1007/s11427-011-4255-x

Conesa A, Götz S, *et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21 (18): 3674–3676. doi:10.1093/bioinformatics/bti610. PMID 16081474.

Conrad G, Mehl E, *et al.* (1968). Région de la Gara Azzel Matti, Sahara central. Paris, Éd. du Centre national de la recherche schientifique

Coulter L, Lorenz K (1990) Quinoa – Composition, nutritional value, food applications. Lebensm-Wiss u Technol 23:203-207

Cusack DF (1984) Quinua: grain of the Incas. The Ecologist 14:21-31

Cushman JC (2001) Crassulacean Acid Metabolism. A plastic photosynthetic adaptation to arid environments. Plant Physiology 127:1439-1448

Dubouzet JG, Sakuma Y, *et al*. (2003) OsDREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. The Plant Journal 33:751-763

Farooq M, Wahid A, *et al.* (2009) Plant drought stress: Effects, mechanisms and management. Sustainable Agriculture. Springer Netherlands:153-188

Fleming JE, Galwey NW (1995) Quinoa (*Chenopodium quinoa)*. In: Williams JT (ed) Cereals and pseudocereals. Chapman & Hall, London, pp 3-83

Fujita M, Fujita Y, *et al*. (2004) A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. The Plant Journal 39:863-876

Gan Q, Chepelev I, *et al*. (2010) Dynamic regulation of alternative splicing and chromatin structure in Drosophila gonads revealed by RNA-seq. Cell Res 20:763-783

Giusti L (1970) El género *Chenopodium* en Argentina. Darwiniana 16:98-105

Götz S, García-Gomez JM, et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research 36:3420-3435

Grabherr MG, Haas BJ, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29:644-652

Hardcastle TJ, Kelly KA (2010) baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11:422

Hatch MD, Slack CR (1966) Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. Biochem J 101:103-111

Hirose Y, Fujita T, Ishii T (2010) Antioxidative properties and flavonoid composition of *Chenopodium quinoa* seed cultivated in Japan. Food Chemistry 119:1300-1306

Hosy E, Vavasseur A, *et al.* (2003) The Arabidopsis outward K+ channel GORK is involved in regulation of stomatal movements and plant transpiration. Proceedings of the National Academy of Science 100:5549-5554

Kasuga M, Liu Q, *et al.* (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nat Biotech 17:287-291

Kodzius R, Kojima M, *et al.* (2006) CAGE: cap analysis of gene expression. Nat Meth 3:211-222

Landmead B, Trapnell C, *et al*. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

Landmead B, Hansen KD, Leek JT (2010) Cloud-scale RNA-sequencing differential expression analysis with Myrna. Genome Biol 11:R83

LeHouerou HN (1984) Rain use efficiency: A unifying concept in arid-land ecology. Journal of Arid Environments 7:213-247

Li H, Durbin R (2009) Fast and accurate short read alignments with Burrows-Wheeler transform. Bioinformatics 25:1754-1760

Li R, Li Y, *et al*. (2008) SOAP: short oligonucleotide alignment program. Bioinformatics 24:713-714

Li R, Yu C, *et al*. (2009) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25:1966-1967

Li S, Assmann SM, *et al.* (2006) Predicting essential components of signal transduction networks: A dynamic model of guard cell abscisic acid signaling. PLoS Biol 4:e312

Logan BA (2007) Reactive oxygen species and photosynthesis. Blackwell Publishing Ltd.
Maher CA, Kumar-Sinha C, *et al*. (2009) Transcriptome Sequencing to Detect Gene Fusions in Cancer. Nature 458: 97–101 doi:10.1038/nature07638. PMC 2725402. PMID 19136943.
Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends in Genetics 24:133-141

Margulies M, Egholm M, *et al*. (2005) Genome sequencing in microfabricated high-density picrolitre reactors. Nature 437:376-380

Maughan PJ, Turner TB, *et al.* (2009) Characterization of Salt Overly Sensitive 1 (SOS1) gene homoeologs in quinoa (*Chenopodium quinoa* Willd.). Genome 52:647-657

Mastebroek HD, Limburg H, *et al.* (2000) Occurrence of sapogenins in leaves and seeds of quinoa (*Chenopodium quinoa* Willd.). J Sci Food Agric 80:152-156

Metzker ML (2010) Sequencing technologies [mdash] the next generation. Nat Rev Genet 11:31-46

Mishra G, Zhang W, *et al.* (2006) A bifurcating pathway directs Abscisic Acid effects on stomatal closure and opening in Arabidopsis. Science 312:264-266

Mitra J (2001) Genetics and genetic improvement of drought resistance in crop plants. Curr Sci 80:758-763

Mortazavi A, Williams BA *et al*. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621-628

Munns R (2002) Comparative physiology of salt and water stress. Plant, cell & environment 25:239-250

Mustilli AC, Merlot S, *et al.* (2002) Arabidopsis OST1 protein kinase mediates the regulation of stomatal aperture by Abscisic Acid and acts upstream of reactive oxygen species production. The Plant Cell Online 14:3089-3099

Okoniewski M, Miller C (2006) Hybridization interactions between probe sets in short oligo microarrays lead to spurious correlations. BMC Bioinformatics 7:276

Orcutt DM, Nilsen ET (2000) The Physiology of Plants Under Stress. John Wiley & Sons, Inc., New York

Oshlack A, Robinson MD, Young MD (2010) From RNA-seq read to differential expression results. Genome Biology 11:220

Oshodi AA, Ogungbenle HN, Oladimeji MO (1999) Chemical composition, nutritionally valuable minerals and functional properties of benniseed (*Sesamum radiatum*), pearl millet (*Pennisetum typhoides*) and quinoa (*Chenopodium quinoa*) flours. Int J Food Sci Nutr 50:325-331

O'Toole JC, Moya TB (1978) Genotypic variation in maintenance of leaf water potential in rice. Crop Sci 18:873-876

Partap T, Kapoor P (1985) The Himalayan India grain chenopods II. Comparative morphology. Agr Ecosyt Environ 14:201-220

Prado FE, Boero C, *et al.* (2000) Effect of NaCl on germination, growth, and soluble sugar content in *Chenopodium quinoa* Willd. Seeds. Bot Bull Acad Sin 41:27:34

Qin X, Zeevaart JAD (1999) The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory step of abscisic acid biosynthesis in water-stress bean. Proceedings of the National Academy of Sciences 96:15354-15361

Reiser V, Raitt DC, *et al*. (2003) Yeast osmosensor Sln1 and plant cytokinin receptor Cre1 respond to changes in turgor pressure. The Journal of Cell Biology 161:1035-1040

Repo-Carrasco-Valencia R, Hellstrom JK, Pihlava JM (2010) Flavonoids and other phenolic compounds in Andean indigenous grains: Quinoa (*Chenopodium quinoa),* kaniwa (*Chenopodium pallidicaule*) and kiwichi (*Amarathus caudatus*). Food Chemistry 120:128-133

Risi J, Galwey NW (1984) The *Chenopodium* grains of the Andes: Inca crops for modern agriculture. Adv Appl Biol 10:145-216

Risi J, Galwey NW (1989) The pattern of genetic diversity in the Andean grain crop quinoa (*Chenopodium quinoa* Willd.). I. Associations between characteristics. Euphytica 41:147-162

Robertson G, Schein J, *et al.* (2010) De novo assembly and analysis of RNA-seq data. Nat Methods 7:909-912

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140

Ruales J, Nair BM (1993) Content of fat, vitamins and minerals in quinoa (*Chenopodium quinoa* Willd.) seeds. Food Chem 48:131-136

Ryan DM, Bainbridge M, *et al*. (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. BioTechniques 45(1):81-94

Seki M, Narusaka M, *et al*. (2002) Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. The Plant Journal 31:279-292

Shinozaki K, Yamaguchi-Shinozaki K (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. Current Opinion in Plant Biology 3:217-223

Shinozaki K, Yamaguchi-Shinozaki, *et al.* (2003) Regulatory network of gene expression in the drought and cold stress responses. Current Opinion in Plant Biology 6:410-417

Shinozaki K, Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. Journal of Experimental Botany 58:221-227

Simpson SD, Nakashima K, *et al*. (2003) Two different novel cis-acting elements of erd1, a clpA homologous Arabidopsis gene function in induction by dehydration stress and dark-induced senescence. The Plant Journal 33:259-270

Tapia ME (1979) Historia y distribucion geografica. In ME Tapia (ed) Quinua y kaniwa: Cultivos andinos. Instituto Interamericano de Ciencias Agricolas, Bogota, Colombia, pp 11-19
Tapia ME (1980) Origen, distribución geográfica y sistemas de producción de la quinua. Reunión sobre genética y Fitomejoramiento de la quinua. PISCA-UNTA-IBTA-IICA-CIID. Puno, Perú

Tapia M (1982) The Environment, crops and agricultural systems in the Andes of Southern Peru IICA

Tarazona S, Garcia-Alcalde F, *et al.* (2011) Differential expression in RNA-seq: A matter of depth. Genome Res 21:2213-2223

Turner NC (1979) Drought resistance and adaptation to water deficient in crop plants. In H Mussell and RC Staples, eds. Stress physiology in crop plants. Wiley Interscience, New York, pp 343-372

Turner NC (1986) Crop water deficits a decade of progress. Adv in Agr 39:1-51
Urao T. Yakubov B, *et al*. (1999) A transmembrane hybrid-type histidine kinase in Arabidopsis functions as an osmosensor. The Plant Cell Online 11:1743-1754

Vacher JJ (1998) Responses of two main Andean crops, quinoa (*Chenopodium quinoa* Willd.) and papa amarga (*Solanum juzepczukii* Buk.) to drought on the Bolivian Altiplano: Significance of local adaptation. Agric Ecosys and Environ 68:99-108

Velculescu VE, Zhang L (1995) Serial analysis of gene expression. Science 270:484-487
Wang L, Feng Z, *et al*. (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 26:136-138

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10(1):57-63 doi:10.1038/nrg2484. PMC 2949280. PMID 19015660.

Wilson HD (1981) Domesticated *Chenopodium* of the Ozark cliff dwellers. Econ Bot 35:233-239

Wilson HD (1990) *Quinua* and relatives (*Chenopodium* sect. *Chenopodium* subsect. *Cellulata*). Econ Bot 44:92-110

Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26:873-881

Xiong L, Zhu JK (2003) Regulation of abscisic acid biosynthesis. Plant Physiology 133:29-36

Yamaguchi-Shinozaki K, Koizumi M, *et al.* (1992) Molecular cloning and characterization of 9 cDNAs for genes that are responsive to desiccation in *Arabidopsis thaliana*: Sequence analysis of one cDNA clone that encodes a putative transmembrane channel protein. Plant and Cell Physiology 33:217-224

Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research 18:821-829

Zhang J, Chiodini R, *et al.* (2011) The impact of next-generation sequencing on genomics. Journal of Genetics and Genomics 38:95-109

TABLES

Table 2.1 Crude protein content (percentage of dry matter) and essential amino acid composition (percentage of protein) of quinoa and some common crops. Adapted from Risi and Galwey (1984).

Amino Acid	Source Quinoa	Wheat	Barley	Maize	Soybean
Isoleucine	6.4	3.8	3.8	3.8	4.9
Leucine	7.1	6.8	7.0	12.0	7.6
Lysine	6.6	2.9	3.6	3.0	6.4
Methionine	2.4	1.7	1.7	2.2	1.4
Phenylalanine	3.5	4.5	5.2	4.8	4.9
Threonine	4.8	3.1	3.5	3.7	4.2
Tryptophan	1.1	0.8	1.1	1.3	0.9
Valine	4.0	4.7	5.5	5.1	5.0
Arginine	7.4	4.8	5.2	4.9	7.2
Histidine	2.7	2.2	2.2	2.9	2.5
Crude protein	13.0	10.7	11.9	9.0	36.5

Table 2.2 Nutritional analysis of 100 g of cooked quinoa. Nutrient levels presented as percent daily recommended levels (DRL) based on FAO/WHO/UN nutritional recommendations. Adapted from Fleming and Galwey (1995).

		%DRL
Vitamins		
	Vitamin B6	10%
	Pantothenic acid (Vitamin B5)	9-15%
	Folic acid (Vitamin B9)	12%
	Biotin (Vitamin B7)	7-24%
	Riboflavin (B2) and niacin (B3)	6%
Minerals		
	Calcium	4-6%
	Phosphorus	11-16%
	Magnesium	23-76%
	Potassium	15-19%
	Iron	27-40%
	Zinc	10-15%
	Copper	47-200%
	Sodium	1-2%
Amino acids		
	Histidine	200%
	Isoleucine	377%
	Lysine	347%
	Methionine and cysteine	312%
	Phenylalanine and tyrosine	363%
	Threonine	411%
	Tryptophan	180%
	Valine	346%

Table 2.3 Major software packages for differential expression analysis. Adapted from Chen (2011).

Name	Website	Reference
baySeq	http://www.bioconductor.org/packages/2.8/bioc/html/baySeq.html	Hardcastle 2010
DEGseq	http://bioinfo.au.tsinghua.edu.cn/software/degseq/	Wang 2010
DESeq	http://www-huber.embl.de/users/anders/DESeq/	Anders 2010
EdgeR	http://bioconductor.org/packages/release/bioc/html/edgeR.html	Robinson 2010
Myrna	http://bowtie-bio.sourceforge.net/myrna/index.shtml	Langmead 2010
NOISeq	http://bioinfo.cipf.es/noiseq/doku.php?id=start	Tarazona 2011